Associated with document Ref. Ares(2015)2187417 - 26/05/2015

EUROPEAN COMMISSION
DIRECTORATE-GENERAL FOR RESEARCH & INNOVATION

Innovation Union and European Research Area
Research infrastructure

GRANT AGREEMENT

NUMBER — 654024 — SoBigData

This **Agreement** ('the Agreement') is **between** the following parties:

on the one part,

*the **European Union** ('the EU'), represented by the European Commission ('the Commission')*[1],

represented for the purposes of signature of this Agreement by Head of Unit, DIRECTORATE-GENERAL FOR RESEARCH & INNOVATION, Innovation Union and European Research Area , Administration and finance, Pascale CID,

and

on the other part,

1. 'the coordinator':

CONSIGLIO NAZIONALE DELLE RICERCHE (CNR), CF80054330586, established in PIAZZALE ALDO MORO 7, ROMA 00185, Italy, IT02118311006, represented for the purposes of signing the Agreement by CLAUDIO MONTANI

and the following other beneficiaries, if they sign their 'Accession Form' (see Annex 3 and Article 56):

2. **THE UNIVERSITY OF SHEFFIELD (USFD)**, RC000667, established in FIRTH COURT WESTERN BANK, SHEFFIELD S10 2TN, United Kingdom, GB648238808,

3. **UNIVERSITA DI PISA (UNIPI)**, 125096/CF80003670504, established in LUNGARNO PACINOTTI 43/44, PISA 56126, Italy, IT00286820501,

4. **FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV (Fraunhofer)** EV, VR4461, established in HANSASTRASSE 27C, MUENCHEN 80686, Germany, DE129515865,

5. **TARTU ULIKOOL (UT)**, 74001073, established in ULIKOOLI 18, TARTU 50090, Estonia, EE100030417,

6. **SCUOLA IMT (ISTITUZIONI, MERCATI, TECNOLOGIE) ALTI STUDI DI LUCCA (IMT LUCCA)**, CF92037570469, established in PIAZZA SAN PONZIANO 6, LUCCA 55100, Italy, IT02235840465,

7. **GOTTFRIED WILHELM LEIBNIZ UNIVERSITAET HANNOVER (LUH)**, established in Welfengarten 1, HANNOVER 30167, Germany, DE811245527,

8. **KING'S COLLEGE LONDON (KCL)**, RC000297, established in Strand, LONDON WC2R 2LS, United Kingdom, GB627403551,

9. **SCUOLA NORMALE SUPERIORE (SNS)**, CF80005050507, established in PIAZZA DEI CAVALIERI 7, PISA 56126, Italy, IT00420000507,

1 Text in *italics* shows the options of the Model Grant Agreement that are applicable to this Agreement.

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

10. **AALTO-KORKEAKOULUSAATIO (AALTO UNIVERSITY)** FI1, 22283574, established in OTAKAARI 1, ESPOO 02150, Finland, FI22283574,

*11. **EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH (ETH Zurich)**, GESETZSR414110, established in Raemistrasse 101, ZUERICH 8092, Switzerland, CHE115203630MWST, as 'beneficiary not receiving EU funding' (see Article 9),*

12. **TECHNISCHE UNIVERSITEIT DELFT (TU Delft)**, 27364265, established in STEVINWEG 1, DELFT 2628 CN, Netherlands, NL001569569B01,

Unless otherwise specified, references to 'beneficiary' or 'beneficiaries' include the coordinator.

The parties referred to above have agreed to enter into the Agreement under the terms and conditions below.

By signing the Agreement or the Accession Form, the beneficiaries accept the grant and agree to implement it under their own responsibility and in accordance with the Agreement, with all the obligations and conditions it sets out.

The Agreement is composed of:

Terms and Conditions

Annex 1 Description of the action

Annex 2 Estimated budget for the action

Annex 3 Accession Forms

Annex 4 Model for the financial statements

Annex 5 Model for the certificate on the financial statements

Annex 6 Model for the certificate on the methodology

TERMS AND CONDITIONS

TABLE OF CONTENTS

CHAPTER 1 GENERAL....11

ARTICLE 1 — SUBJECT OF THE AGREEMENT....11

CHAPTER 2 ACTION....11

ARTICLE 2 — ACTION TO BE IMPLEMENTED....11

ARTICLE 3 — DURATION AND STARTING DATE OF THE ACTION....11

ARTICLE 4 — ESTIMATED BUDGET AND BUDGET TRANSFERS....11

4.1 Estimated budget....11

4.2 Budget transfers....11

CHAPTER 3 GRANT....11

ARTICLE 5 — GRANT AMOUNT, FORM OF GRANT, REIMBURSEMENT RATES AND FORMS OF COSTS....11

5.1 Maximum grant amount....11

5.2 Form of grant, reimbursement rates and forms of costs....12

5.3 Final grant amount — Calculation....12

5.4 Revised final grant amount — Calculation....14

ARTICLE 6 — ELIGIBLE AND INELIGIBLE COSTS....14

6.1 General conditions for costs to be eligible....14

6.2 Specific conditions for direct costs to be eligible....15

6.3 Conditions for costs of linked third parties to be eligible....21

6.4 Conditions for in-kind contributions provided by third parties free of charge to be eligible....21

6.5 Ineligible costs....21

6.6 Consequences of declaration of ineligible costs....22

CHAPTER 4 RIGHTS AND OBLIGATIONS OF THE PARTIES....22

SECTION 1 RIGHTS AND OBLIGATIONS RELATED TO IMPLEMENTING THE ACTION....22

ARTICLE 7 — GENERAL OBLIGATION TO PROPERLY IMPLEMENT THE ACTION....22

7.1 General obligation to properly implement the action....22

7.2 Consequences of non-compliance....22

ARTICLE 8 — RESOURCES TO IMPLEMENT THE ACTION — THIRD PARTIES INVOLVED IN THE ACTION....22

ARTICLE 9 — IMPLEMENTATION OF ACTION TASKS BY BENEFICIARIES NOT RECEIVING EU FUNDING....23
9.1 Rules for the implementation of action tasks by beneficiaries not receiving EU funding....23
9.2 Consequences of non-compliance....23
ARTICLE 10 — PURCHASE OF GOODS, WORKS OR SERVICES....23
10.1 Rules for purchasing goods, works or services....23
10.2 Consequences of non-compliance....24
ARTICLE 11 — USE OF IN-KIND CONTRIBUTIONS PROVIDED BY THIRD PARTIES AGAINST PAYMENT....24
11.1 Rules for the use of in-kind contributions against payment....24
11.2 Consequences of non-compliance....24
ARTICLE 12 — USE OF IN-KIND CONTRIBUTIONS PROVIDED BY THIRD PARTIES FREE OF CHARGE....25
12.1 Rules for the use of in-kind contributions free of charge....25
12.2 Consequences of non-compliance....25
ARTICLE 13 — IMPLEMENTATION OF ACTION TASKS BY SUBCONTRACTORS....25
13.1 Rules for subcontracting action tasks....25
13.2 Consequences of non-compliance....26
ARTICLE 14 — IMPLEMENTATION OF ACTION TASKS BY LINKED THIRD PARTIES....26
ARTICLE 15 — FINANCIAL SUPPORT TO THIRD PARTIES....26
15.1 Rules for providing financial support to third parties....26
15.2 Financial support in the form of prizes....26
15.3 Consequences of non-compliance....26
ARTICLE 16 — PROVISION OF TRANS-NATIONAL OR VIRTUAL ACCESS TO RESEARCH INFRASTRUCTURE....26
16.1 Rules for providing trans-national access to research infrastructure....26
16.2 Rules for providing virtual access to research infrastructure....28
16.3 Consequences of non-compliance....28
SECTION 2 RIGHTS AND OBLIGATIONS RELATED TO THE GRANT ADMINISTRATION....29
ARTICLE 17 – GENERAL OBLIGATION TO INFORM....29
17.1 General obligation to provide information upon request....29
17.2 Obligation to keep information up to date and to inform about events and circumstances likely to affect the Agreement....29
17.3 Consequences of non-compliance....29
ARTICLE 18 — KEEPING RECORDS — SUPPORTING DOCUMENTATION....29
18.1 Obligation to keep records and other supporting documentation....29
18.2 Consequences of non-compliance....31

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

ARTICLE 19 — SUBMISSION OF DELIVERABLES 31
19.1 Obligation to submit deliverables 31
19.2 Consequences of non-compliance 31
ARTICLE 20 — REPORTING — PAYMENT REQUESTS 31
20.1 Obligation to submit reports 31
20.2 Reporting periods 31
20.3 Periodic reports — Requests for interim payments 31
20.4 Final report — Request for payment of the balance 33
20.5 Information on cumulative expenditure incurred 33
20.6 Currency for financial statements and conversion into euro 34
20.7 Language of reports 34
20.8 Consequences of non-compliance — Suspension of the payment deadline — Termination 34
ARTICLE 21 — PAYMENTS AND PAYMENT ARRANGEMENTS 34
21.1 Payments to be made 34
21.2 Pre-financing payment — Amount — Amount retained for the Guarantee Fund 34
21.3 Interim payments — Amount — Calculation 35
21.4 Payment of the balance — Amount — Calculation — Release of the amount retained for the Guarantee Fund 35
21.5 Notification of amounts due 36
21.6 Currency for payments 36
21.7 Payments to the coordinator — Distribution to the beneficiaries 36
21.8 Bank account for payments 37
21.9 Costs of payment transfers 37
21.10 Date of payment 37
21.11 Consequences of non-compliance 37
ARTICLE 22 — CHECKS, REVIEWS, AUDITS AND INVESTIGATIONS — EXTENSION OF FINDINGS 38
22.1 Checks, reviews and audits by the Commission 38
22.2 Investigations by the European Anti-Fraud Office (OLAF) 40
22.3 Checks and audits by the European Court of Auditors (ECA) 40
22.4 Checks, reviews, audits and investigations for international organisations 40
22.5 Consequences of findings in checks, reviews, audits and investigations —Extension of findings 40
22.6 Consequences of non-compliance 42
ARTICLE 23 — EVALUATION OF THE IMPACT OF THE ACTION 42
23.1 Right to evaluate the impact of the action 42

23.2 Consequences of non-compliance ... 42

SECTION 3 RIGHTS AND OBLIGATIONS RELATED TO BACKGROUND AND RESULTS ... 43

SUBSECTION 1 GENERAL ... 43

ARTICLE 23a — MANAGEMENT OF INTELLECTUAL PROPERTY ... 43

23a.1 Obligation to take measures to implement the Commission Recommendation on the management of intellectual property in knowledge transfer activities ... 43

23a.2 Consequences of non-compliance ... 43

SUBSECTION 2 RIGHTS AND OBLIGATIONS RELATED TO BACKGROUND ... 43

ARTICLE 24 — AGREEMENT ON BACKGROUND ... 43

24.1 Agreement on background ... 43

24.2 Consequences of non-compliance ... 43

ARTICLE 25 — ACCESS RIGHTS TO BACKGROUND ... 44

25.1 Exercise of access rights — Waiving of access rights — No sub-licensing ... 44

25.2 Access rights for other beneficiaries, for implementing their own tasks under the action ... 44

25.3 Access rights for other beneficiaries, for exploiting their own results ... 44

25.4 Access rights for affiliated entities ... 44

25.5 Access rights for third parties ... 45

25.6 Consequences of non-compliance ... 45

SUBSECTION 3 RIGHTS AND OBLIGATIONS RELATED TO RESULTS ... 45

ARTICLE 26 — OWNERSHIP OF RESULTS ... 45

26.1 Ownership by the beneficiary that generates the results ... 45

26.2 Joint ownership by several beneficiaries ... 46

26.3 Rights of third parties (including personnel) ... 46

26.4 *EU* ownership, to protect results ... 46

26.5 Consequences of non-compliance ... 47

ARTICLE 27 — PROTECTION OF RESULTS — VISIBILITY OF EU FUNDING ... 47

27.1 Obligation to protect the results ... 47

27.2 *EU* ownership, to protect the results ... 48

27.3 Information on EU funding ... 48

27.4 Consequences of non-compliance ... 48

ARTICLE 28 — EXPLOITATION OF RESULTS ... 48

28.1 Obligation to exploit the results ... 48

28.2 Results that could contribute to European or international standards — Information on EU funding ... 48

28.3 Consequences of non-compliance ... 49

ARTICLE 29 — DISSEMINATION OF RESULTS — OPEN ACCESS — VISIBILITY OF EU FUNDING....49
29.1 Obligation to disseminate results....49
29.2 Open access to scientific publications....49
29.3 Open access to research data....50
29.4 Information on EU funding — Obligation and right to use the EU emblem....50
29.5 Disclaimer excluding *Commission* responsibility....50
29.6 Consequences of non-compliance....50
ARTICLE 30 — TRANSFER AND LICENSING OF RESULTS....51
30.1 Transfer of ownership....51
30.2 Granting licenses....51
30.3 *Commission* right to object to transfers or licensing....51
30.4 Consequences of non-compliance....51
ARTICLE 31 — ACCESS RIGHTS TO RESULTS....52
31.1 Exercise of access rights — Waiving of access rights — No sub-licensing....52
31.2 Access rights for other beneficiaries, for implementing their own tasks under the action....52
31.3 Access rights for other beneficiaries, for exploiting their own results....52
31.4 Access rights of affiliated entities....52
31.5 Access rights for the EU institutions, bodies, offices or agencies and EU Member States....52
31.6 Access rights for third parties....52
31.7 Consequences of non-compliance....53
SECTION 4 OTHER RIGHTS AND OBLIGATIONS....53
ARTICLE 32 — RECRUITMENT AND WORKING CONDITIONS FOR RESEARCHERS....53
32.1 Obligation to take measures to implement the European Charter for Researchers and Code of Conduct for the Recruitment of Researchers....53
32.2 Consequences of non-compliance....53
ARTICLE 33 — GENDER EQUALITY....53
33.1 Obligation to aim for gender equality....53
33.2 Consequences of non-compliance....53
ARTICLE 34 — ETHICS....54
34.1 Obligation to comply with ethical principles....54
34.2 Activities raising ethical issues....54
34.3 Activities involving human embryos or human embryonic stem cells....55
34.4 Consequences of non-compliance....55

ARTICLE 35 — CONFLICT OF INTERESTS....55
35.1 Obligation to avoid a conflict of interests....55
35.2 Consequences of non-compliance....55
ARTICLE 36 — CONFIDENTIALITY....55
36.1 General obligation to maintain confidentiality....55
36.2 Consequences of non-compliance....56
ARTICLE 37 — SECURITY-RELATED OBLIGATIONS....57
37.1 Results with a security recommendation....57
37.2 Classified results....57
37.3 Activities involving dual-use goods or dangerous materials and substances....57
37.4 Consequences of non-compliance....57
ARTICLE 38 — PROMOTING THE ACTION — VISIBILITY OF EU FUNDING....57
38.1 Communication activities by beneficiaries....57
38.2 Communication activities by the *Commission*....58
38.3 Consequences of non-compliance....59
ARTICLE 39 — PROCESSING OF PERSONAL DATA....59
39.1 Processing of personal data by the Commission....59
39.2 Processing of personal data by the beneficiaries....60
39.3 Consequences of non-compliance....60
ARTICLE 40 — ASSIGNMENTS OF CLAIMS FOR PAYMENT AGAINST THE *COMMISSION*....60
CHAPTER 5 DIVISION OF BENEFICIARIES' ROLES AND RESPONSIBILITIES....60
ARTICLE 41 — DIVISION OF BENEFICIARIES' ROLES AND RESPONSIBILITIES....60
41.1 Roles and responsibilities towards the *Commission*....60
41.2 Internal division of roles and responsibilities....60
41.3 Internal arrangements between beneficiaries — Consortium agreement....61
41.4 Relationship with complementary beneficiaries — Collaboration agreement....62
41.5 Relationship with partners of a joint action — Coordination agreement....62
CHAPTER 6 REJECTION OF COSTS — REDUCTION OF THE GRANT — RECOVERY — PENALTIES — DAMAGES — SUSPENSION — TERMINATION — FORCE MAJEURE....62
SECTION 1 REJECTION OF COSTS — REDUCTION OF THE GRANT — RECOVERY — PENALTIES....62
ARTICLE 42 — REJECTION OF INELIGIBLE COSTS....62
42.1 Conditions....62
42.2 Ineligible costs to be rejected — Calculation — Procedure....62
42.3 Effects....63

ARTICLE 43 — REDUCTION OF THE GRANT 63
43.1 Conditions 63
43.2 Amount to be reduced — Calculation — Procedure 63
43.3 Effects 64
ARTICLE 44 — RECOVERY OF UNDUE AMOUNTS 64
44.1 Amount to be recovered — Calculation — Procedure 64
ARTICLE 45 — ADMINISTRATIVE AND FINANCIAL PENALTIES 68
45.1 Conditions 68
45.2 Duration — Amount of penalty — Calculation 68
45.3 Procedure 68
SECTION 2 LIABILITY FOR DAMAGES 69
ARTICLE 46 — LIABILITY FOR DAMAGES 69
46.1 Liability of the *Commission* 69
46.2 Liability of the beneficiaries 69
SECTION 3 SUSPENSION AND TERMINATION 70
ARTICLE 47 — SUSPENSION OF PAYMENT DEADLINE 70
47.1 Conditions 70
47.2 Procedure 71
ARTICLE 48 — SUSPENSION OF PAYMENTS 71
48.1 Conditions 71
48.2 Procedure 71
ARTICLE 49 — SUSPENSION OF THE ACTION IMPLEMENTATION 72
49.1 Suspension of the action implementation, by the beneficiaries 72
49.2 Suspension of the action implementation, by the *Commission* 72
ARTICLE 50 — TERMINATION OF THE AGREEMENT OR OF THE PARTICIPATION OF ONE OR MORE BENEFICIARIES 73
50.1 Termination of the Agreement by the beneficiaries 73
50.2 Termination of the participation of one or more beneficiaries, by the beneficiaries 74
50.3 Termination of the Agreement or the participation of one or more beneficiaries, by the *Commission* 76
SECTION 4 FORCE MAJEURE 80
ARTICLE 51 — FORCE MAJEURE 80
CHAPTER 7 FINAL PROVISIONS 81
ARTICLE 52 — COMMUNICATIONS BETWEEN THE PARTIES 81
52.1 Form and means of communication 81

52.2 Date of communication.....81
52.3 Addresses for communication.....82
ARTICLE 53 — INTERPRETATION OF THE AGREEMENT.....82
53.1 Precedence of the Terms and Conditions over the Annexes.....82
53.2 Privileges and immunities.....82
ARTICLE 54 — CALCULATION OF PERIODS, DATES AND DEADLINES.....82
ARTICLE 55 — AMENDMENTS TO THE AGREEMENT.....83
55.1 Conditions.....83
55.2 Procedure.....83
ARTICLE 56 — ACCESSION TO THE AGREEMENT.....83
56.1 Accession of the beneficiaries mentioned in the Preamble.....83
56.2 Addition of new beneficiaries.....84
ARTICLE 57 — APPLICABLE LAW AND SETTLEMENT OF DISPUTES.....84
57.1 Applicable law.....84
57.2 Dispute settlement.....84
ARTICLE 58 — ENTRY INTO FORCE OF THE AGREEMENT.....85

CHAPTER 1 GENERAL

ARTICLE 1 — SUBJECT OF THE AGREEMENT

This Agreement sets out the rights and obligations and the terms and conditions applicable to the grant awarded to the beneficiaries for implementing the action set out in Chapter 2.

CHAPTER 2 ACTION

ARTICLE 2 — ACTION TO BE IMPLEMENTED

The grant is awarded for the action entitled '***SoBigData Research Infrastructure — SoBigData***' **('action')**, as described in Annex 1.

ARTICLE 3 — DURATION AND STARTING DATE OF THE ACTION

The duration of the action will be ***48*** **months** as of *1 September 2015* ('**starting date of the action**').

ARTICLE 4 — ESTIMATED BUDGET AND BUDGET TRANSFERS

4.1 Estimated budget

The '**estimated budget**' for the action is set out in Annex 2.

It contains the estimated eligible costs and the forms of costs, broken down by beneficiary and budget category (see Articles 5, 6). *It also contains the estimated costs of the beneficiaries not receiving EU funding (see Article 9).*

4.2 Budget transfers

The estimated budget breakdown indicated in Annex 2 may be adjusted by transfers of amounts between beneficiaries or between budget categories (or both). This does not require an amendment according to Article 55, if the action is implemented as described in Annex 1.

However, the beneficiaries may not add costs relating to subcontracts not provided for in Annex 1, unless such additional subcontracts are approved by an amendment or in accordance with Article 13.

CHAPTER 3 GRANT

ARTICLE 5 — GRANT AMOUNT, FORM OF GRANT, REIMBURSEMENT RATES AND FORMS OF COSTS

5.1 Maximum grant amount

The '**maximum grant amount**' is **EUR 5,000,000.00** (five million EURO).

5.2 Form of grant, reimbursement rates and forms of costs

The grant reimburses ***100% of the action's eligible costs*** (see Article 6) ('**reimbursement of eligible costs grant**') (see Annex 2).

The estimated eligible costs of the action are EUR **5,917,500.00** (five million nine hundred and seventeen thousand five hundred EURO).

Eligible costs (see Article 6) must be declared under the following forms ('**forms of costs**'):

(a) for **direct personnel costs**:

- as actually incurred costs ('**actual costs**') or
- on the basis of an amount per unit calculated by the beneficiary in accordance with its usual cost accounting practices ('**unit costs**').

Personnel **costs for SME owners** or **beneficiaries that are natural persons** not receiving a salary (see Article 6.2, Points A.4 and A.5) must be declared on the basis of the amount per unit set out in Annex 2 (**unit costs**);

(b) for **direct costs for subcontracting**: as actually incurred costs (**actual costs**);

(c) for **direct costs of providing financial support to third parties**: *not applicable;*

(d) for **other direct costs**: as actually incurred costs (**actual costs**);

(e) for **indirect costs**: on the basis of a flat-rate applied as set out in Article 6.2, Point E ('**flat-rate costs**');

(f) ***specific cost category(ies):*** *not applicable.*

5.3 Final grant amount — Calculation

The '**final grant amount**' depends on the actual extent to which the action is implemented in accordance with the Agreement's terms and conditions.

This amount is calculated by the *Commission* — when the payment of the balance is made (see Article 21.4) — in the following steps:

Step 1 – Application of the reimbursement rates to the eligible costs

Step 2 – Limit to the maximum grant amount

Step 3 – Reduction due to the no-profit rule

Step 4 – Reduction due to improper implementation or breach of other obligations

5.3.1 Step 1 — Application of the reimbursement rates to the eligible costs

The reimbursement rate(s) (see Article 5.2) are applied to the eligible costs (actual costs, unit costs and flat-rate costs; see Article 6) declared by the beneficiaries (see Article 20) and approved by the *Commission* (see Article 21).

5.3.2 Step 2 — Limit to the maximum grant amount

If the amount obtained following Step 1 is higher than the maximum grant amount set out in Article 5.1, it will be limited to the latter.

5.3.3 Step 3 — Reduction due to the no-profit rule

The grant must not produce a profit.

'**Profit**' means the surplus of the amount obtained following Steps 1 and 2 plus the action's total receipts, over the action's total eligible costs.

The '**action's total eligible costs**' are the consolidated total eligible costs approved by the *Commission*.

The '**action's total receipts**' are the consolidated total receipts generated during its duration (see Article 3).

The following are considered **receipts**:

(a) income generated by the action; if the income is generated from selling equipment or other assets purchased under the Agreement, the receipt is up to the amount declared as eligible under the Agreement;

(b) financial contributions given by third parties to the beneficiary specifically to be used for the action, and

(c) in-kind contributions provided by third parties free of charge and specifically to be used for the action, if they have been declared as eligible costs.

The following are however not considered receipts:

(a) income generated by exploiting the action's results (see Article 28);

(b) financial contributions by third parties, if they may be used to cover costs other than the eligible costs (see Article 6);

(c) financial contributions by third parties with no obligation to repay any amount unused at the end of the period set out in Article 3.

If there is a profit, it will be deducted from the amount obtained following Steps 1 and 2.

5.3.4 Step 4 — Reduction due to improper implementation or breach of other obligations — Reduced grant amount — Calculation

If the grant is reduced (see Article 43), the *Commission* will calculate the reduced grant amount by deducting the amount of the reduction (calculated in proportion to the improper implementation of the action or to the seriousness of the breach of obligations in accordance with Article 43.2) from the maximum grant amount set out in Article 5.1.

The final grant amount will be the lower of the following two:

- the amount obtained following Steps 1 to 3 or

- the reduced grant amount following Step 4.

5.4 Revised final grant amount — Calculation

If — after the payment of the balance (in particular, after checks, reviews, audits or investigations; see Article 22) — the *Commission* rejects costs (see Article 42) or reduces the grant (see Article 43), it will calculate the '**revised final grant amount**' for the beneficiary concerned by the findings.

This amount is calculated by the *Commission* on the basis of the findings, as follows:

- in case of **rejection of costs**: by applying the reimbursement rate to the revised eligible costs approved by the *Commission* for the beneficiary concerned;

- in case of **reduction of the grant**: by calculating the concerned beneficiary's share in the grant amount reduced in proportion to its improper implementation of the action or to the seriousness of its breach of obligations (see Article 43.2).

In case of **rejection of costs and reduction of the grant**, the revised final grant amount for the beneficiary concerned will be the lower of the two amounts above.

ARTICLE 6 — ELIGIBLE AND INELIGIBLE COSTS

6.1 General conditions for costs to be eligible

'**Eligible costs**' are costs that meet the following criteria:

(a) for **actual costs**:

(i) they must be actually incurred by the beneficiary;

(ii) they must be incurred in the period set out in Article 3, with the exception of costs relating to the submission of the periodic report for the last reporting period and the final report (see Article 20);

(iii) they must be indicated in the estimated budget set out in Annex 2;

(iv) they must be incurred in connection with the action as described in Annex 1 and necessary for its implementation;

(v) they must be identifiable and verifiable, in particular recorded in the beneficiary's accounts in accordance with the accounting standards applicable in the country where the beneficiary is established and with the beneficiary's usual cost accounting practices;

(vi) they must comply with the applicable national law on taxes, labour and social security, and

(vii) they must be reasonable, justified and must comply with the principle of sound financial management, in particular regarding economy and efficiency;

(b) for **unit costs**:

(i) they must be calculated as follows:

{amounts per unit set out in Annex 2 or calculated by the beneficiary in accordance with its usual cost accounting practices (see Article 6.2, Point A)

multiplied by

the number of actual units};

(ii) the number of actual units must comply with the following conditions:

- the units must be actually used or produced in the period set out in Article 3;
- the units must be necessary for implementing the action or produced by it, and
- the number of units must be identifiable and verifiable, in particular supported by records and documentation (see Article 18);

(c) for **flat-rate costs**:

(i) they must be calculated by applying the flat-rate set out in Annex 2, and

(ii) the costs (actual costs or unit costs) to which the flat-rate is applied must comply with the conditions for eligibility set out in this Article.

6.2 Specific conditions for costs to be eligible

Costs are eligible if they comply with the general conditions (see above) and the specific conditions set out below for each of the following budget categories:

A. direct personnel costs;
B. direct costs of subcontracting;
C. *not applicable;*
D. other direct costs;
E. indirect costs;
F. *not applicable.*

'Direct costs' are costs that are directly linked to the action implementation and can therefore be attributed to it directly. They must not include any indirect costs (see Point E below).

'Indirect costs' are costs that are not directly linked to the action implementation and therefore cannot be attributed directly to it.

A. Direct personnel costs

Types of eligible personnel costs

A.1 **Personnel costs** are eligible, if they are related to personnel working for the beneficiary under an employment contract (or equivalent appointing act) and assigned to the action ('**costs for employees (or equivalent)**'). They must be limited to salaries (including during parental leave), social security contributions, taxes and other costs included in the **remuneration**, if they arise from national law or the employment contract (or equivalent appointing act).

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

Beneficiaries that are non-profit legal entities[2] may also declare as personnel costs **additional remuneration** for personnel assigned to the action (including payments on the basis of supplementary contracts regardless of their nature), if:

(a) it is part of the beneficiary's usual remuneration practices and is paid in a consistent manner whenever the same kind of work or expertise is required;

(b) the criteria used to calculate the supplementary payments are objective and generally applied by the beneficiary, regardless of the source of funding used.

Additional remuneration for personnel assigned to the action is eligible up to the following amount:

(a) if the person works full time and exclusively on the action during the full year: up to EUR 8 000;

(b) if the person works exclusively on the action but not full-time or not for the full year: up to the corresponding pro-rata amount of EUR 8 000, or

(c) if the person does not work exclusively on the action: up to a pro-rata amount calculated as follows:

{{EUR 8 000

divided by

the number of annual productive hours (see below)},

multiplied by

the number of hours that the person has worked on the action during the year}.

A.2 The **costs for natural persons working under a direct contract** with the beneficiary other than an employment contract are eligible personnel costs, if:

(a) the person works under the beneficiary's instructions and, unless otherwise agreed with the beneficiary, on the beneficiary's premises;

(b) the result of the work carried out belongs to the beneficiary, and

(c) the costs are not significantly different from those for personnel performing similar tasks under an employment contract with the beneficiary.

A.3 The **costs of personnel seconded by a third party against payment** are eligible personnel costs, if the conditions in Article 11.1 are met.

[2] For the definition, see Article 2.1(14) of the Rules for Participation Regulation No 1290/2013: '**non-profit legal entity**' means a legal entity which by its legal form is non-profit-making or which has a legal or statutory obligation not to distribute profits to its shareholders or individual members.

A.4 **Costs of owners** of beneficiaries that are small and medium-sized enterprises ('**SME owners**') who are working on the action and who do not receive a salary are eligible personnel costs, if they correspond to the amount per unit set out in Annex 2 multiplied by the number of actual hours worked on the action.

A.5 **Costs of 'beneficiaries that are natural persons'** not receiving a salary are eligible personnel costs, if they correspond to the amount per unit set out in Annex 2 multiplied by the number of actual hours worked on the action.

A.6 ***Personnel costs for providing trans-national access to research infrastructure*** *are eligible only if also the conditions set out in Article 16.1.1 are met.* ***Personnel costs for providing virtual access to research infrastructure*** *are eligible only if also the conditions set out in Article 16.2 are met.*

Calculation

Personnel costs must be calculated by the beneficiaries as follows:

{{hourly rate

multiplied by

the number of actual hours worked on the action},

plus

for non-profit legal entities: additional remuneration to personnel assigned to the action under the conditions set out above (Point A.1)}.

The number of actual hours declared for a person must be identifiable and verifiable (see Article 18).

The total number of hours declared in EU or Euratom grants, for a person for a year, cannot be higher than the annual productive hours used for the calculations of the hourly rate. Therefore, the maximum number of hours that can be declared for the grant is:

{the number of annual productive hours for the year (see below)

minus

total number of hours declared by the beneficiary for that person in that year for other EU or Euratom grants}.

The '**hourly rate**' is one of the following:

(a) for personnel costs declared as **actual costs:** the hourly rate is the amount calculated as follows:

{actual annual personnel costs (excluding additional remuneration) for the person

divided by

number of annual productive hours}.

The beneficiaries must use the annual personnel costs and the number of annual productive hours for each financial year covered by the reporting period. If a financial year is not closed at the end of the reporting period, the beneficiaries must use the hourly rate of the last closed financial year available.

For the 'number of annual productive hours', the beneficiaries may choose one of the following:

(i) 'fixed number of hours': 1 720 hours for persons working full time (or corresponding pro-rata for persons not working full time);

(ii) 'individual annual productive hours': the total number of hours worked by the person in the year for the beneficiary, calculated as follows:

{annual workable hours of the person (according to the employment contract, applicable collective labour agreement or national law)

plus

overtime worked

minus

absences (such as sick leave and special leave)}.

'Annual workable hours' means the period during which the personnel must be working, at the employer's disposal and carrying out his/her activity or duties under the employment contract, applicable collective labour agreement or national working time legislation.

If the contract (or applicable collective labour agreement or national working time legislation) does not allow to determine the annual workable hours, this option cannot be used;

(iii) 'standard annual productive hours': the 'standard number of annual hours' generally applied by the beneficiary for its personnel in accordance with its usual cost accounting practices. This number must be at least 90% of the 'standard annual workable hours'.

If there is no applicable reference for the standard annual workable hours, this option cannot be used.

For all options, the actual time spent on **parental leave** by a person assigned to the action may be deducted from the number of annual productive hours;

(b) for personnel costs declared on the basis of **unit costs**: the hourly rate is one of the following:

(i) for SME owners or beneficiaries that are natural persons: the hourly rate set out in Annex 2 (see Points A.4 and A.5 above), or

(ii) for personnel costs declared on the basis of the beneficiary's usual cost accounting practices: the hourly rate calculated by the beneficiary in accordance with its usual cost accounting practices, if:

- the cost accounting practices used are applied in a consistent manner, based on objective criteria, regardless of the source of funding;

- the hourly rate is calculated using the actual personnel costs recorded in the beneficiary's accounts, excluding any ineligible cost or costs included in other budget categories.

  The actual personnel costs may be adjusted by the beneficiary on the basis of budgeted or estimated elements. Those elements must be relevant for calculating the personnel costs, reasonable and correspond to objective and verifiable information;

and

- the hourly rate is calculated using the number of annual productive hours (see above).

B. Direct costs of subcontracting (including related duties, taxes and charges such as non-deductible value added tax (VAT) paid by the beneficiary) are eligible if the conditions in Article 13.1.1 are met.

*Subcontracting costs **for providing trans-national access to research infrastructure** are eligible only if also the conditions set out in Article 16.1.1 are met.*

*Subcontracting costs **for providing virtual access to research infrastructure** are eligible only if also the conditions set out in Article 16.2 are met.*

C. Direct costs of providing financial support to third parties *not applicable.*

D. Other direct costs

D.1 **Travel costs and related subsistence allowances** (including related duties, taxes and charges such as non-deductible value added tax (VAT) paid by the beneficiary) are eligible if they are in line with the beneficiary's usual practices on travel.

*Travel costs **for providing trans-national access to research infrastructure** are eligible only if also the conditions set out in Article 16.1.1 are met.*

D.2 *The **depreciation costs of equipment, infrastructure or other assets** (new or second-hand) as recorded in the beneficiary's accounts are eligible, if they were purchased in accordance with Article 10.1.1 and written off in accordance with international accounting standards and the beneficiary's usual accounting practices.*

*The **costs of renting or leasing** equipment, infrastructure or other assets (including related duties, taxes and charges such as non-deductible value added tax (VAT) paid by the beneficiary) are also eligible, if they do not exceed the depreciation costs of similar equipment, infrastructure or assets and do not include any financing fees.*

*The costs of equipment, infrastructure or other assets **contributed in-kind against payment** are eligible, if they do not exceed the depreciation costs of similar equipment, infrastructure or assets, do not include any financing fees and if the conditions in Article 11.1 are met.*

The only portion of the costs that will be taken into account is that which corresponds to the duration of the action and rate of actual use for the purposes of the action.

As an exception, the beneficiaries must not declare such costs (i.e. costs of renting, leasing, purchasing depreciable equipment, infrastructure and other assets) for providing trans-national or virtual access to research infrastructure (see Article 16).

D.3 **Costs of other goods and services** (including related duties, taxes and charges such as non-deductible value added tax (VAT) paid by the beneficiary) are eligible, if they are:

(a) purchased specifically for the action and in accordance with Article 10.1.1 or

(b) contributed in kind against payment and in accordance with Article 11.1.

Such goods and services include, for instance, consumables and supplies, dissemination (including open access), protection of results, certificates on the financial statements (if they are required by the Agreement), certificates on the methodology, translations and publications.

*Costs of other goods and services **for providing trans-national access to research infrastructure** are eligible only if also the conditions set out in Article 16.1.1 are met.*

*Costs of other goods and services **for providing virtual access to research infrastructure** are eligible only if also the conditions set out in Article 16.2 are met.*

D.4 **Capitalised and operating costs of 'large research infrastructure'**[3]*: Not applicable*

E. Indirect costs

Indirect costs are eligible if they are declared on the basis of the flat-rate of 25% of the eligible direct costs (see Article 5.2 and Points A to D above), from which are excluded:

(a) costs of subcontracting and

(b) costs of in-kind contributions provided by third parties which are not used on the beneficiary's premises;

(c) *not applicable;*

(d) *not applicable.*

[3] '**Large research infrastructure**' means research infrastructure of a total value of at least EUR 20 million, for a beneficiary, calculated as the sum of historical asset values of each individual research infrastructure of that beneficiary, as they appear in its last closed balance sheet before the date of the signature of the Agreement or as determined on the basis of the rental and leasing costs of the research infrastructure.

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

Beneficiaries receiving an operating grant[5] financed by the EU or Euratom budget cannot declare indirect costs for the period covered by the operating grant.

F. *Specific cost category(ies)*

Not applicable

6.3 Conditions for costs of linked third parties to be eligible

not applicable

6.4 Conditions for in-kind contributions provided by third parties free of charge to be eligible

In-kind contributions provided free of charge are eligible direct costs (for the beneficiary), if the costs incurred by the third party fulfil — *mutatis mutandis* — the general and specific conditions for eligibility set out in this Article (Article 6.1 and 6.2) and Article 12.1.

6.5 Ineligible costs

'**Ineligible costs**' are:

(a) costs that do not comply with the conditions set out above (Article 6.1 to 6.4), in particular:

(i) costs related to return on capital;

(ii) debt and debt service charges;

(iii) provisions for future losses or debts;

(iv) interest owed;

(v) doubtful debts;

(vi) currency exchange losses;

(vii) bank costs charged by the beneficiary's bank for transfers from the *Commission*;

(viii) excessive or reckless expenditure;

(ix) deductible VAT;

(x) costs incurred during suspension of the implementation of the action (see Article 49);

(b) costs declared under another EU or Euratom grant (including grants awarded by a Member State and financed by the EU or Euratom budget and grants awarded by bodies other than the

[5] For the definition, see Article 121(1)(b) of Regulation (EU, Euratom) No 966/2012 of the European Parliament and of the Council of 25 October 2012 on the financial rules applicable to the general budget of the Union and repealing Council Regulation (EC, Euratom) No 1605/2002 (OJ L 218, 26.10.2012, p.1) ('**Financial Regulation No 966/2012**'): '**operating grant**' means direct financial contribution, by way of donation, from the budget in order to finance the functioning of a body which pursues an aim of general EU interest or has an objective forming part of and supporting an EU policy.

Commission for the purpose of implementing the EU or Euratom budget); in particular, indirect costs if the beneficiary is already receiving an operating grant financed by the EU or Euratom budget in the same period.

6.6 Consequences of declaration of ineligible costs

Declared costs that are ineligible will be rejected (see Article 42).

This may also lead to any of the other measures described in Chapter 6.

CHAPTER 4 RIGHTS AND OBLIGATIONS OF THE PARTIES

SECTION 1 RIGHTS AND OBLIGATIONS RELATED TO IMPLEMENTING THE ACTION

ARTICLE 7 — GENERAL OBLIGATION TO PROPERLY IMPLEMENT THE ACTION

7.1 General obligation to properly implement the action

The beneficiaries must implement the action as described in Annex 1 and in compliance with the provisions of the Agreement and all legal obligations under applicable EU, international and national law.

7.2 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

ARTICLE 8 — RESOURCES TO IMPLEMENT THE ACTION — THIRD PARTIES INVOLVED IN THE ACTION

The beneficiaries must have the appropriate resources to implement the action.

If it is necessary to implement the action, the beneficiaries may:

- purchase goods, works and services (see Article 10);
- use in-kind contributions provided by third parties against payment (see Article 11);
- use in-kind contributions provided by third parties free of charge (see Article 12);
- call upon subcontractors to implement action tasks described in Annex 1 (see Article 13);
- call upon linked third parties to implement action tasks described in Annex 1 (see Article 14).

In these cases, the beneficiaries retain sole responsibility towards the *Commission* and the other beneficiaries for implementing the action.

ARTICLE 9 — IMPLEMENTATION OF ACTION TASKS BY BENEFICIARIES NOT RECEIVING EU FUNDING

9.1 Rules for the implementation of action tasks by beneficiaries not receiving EU funding

Beneficiaries not receiving EU funding must implement the action tasks attributed to them in Annex 1 according to Article 7.1.

Their costs are estimated in Annex 2 but:

- *will not be reimbursed and*
- *will not be taken into account for the calculation of the grant (see Articles 5.2, 5.3 and 5.4, and 21).*

Chapter 3, Articles 10 to 15, 18.1.2, 20.3(b), 20.4(b), 20.6, 21, 23a, 26.4, 27.2, 28.1, 28.2, 30.3, 31.5, 40, 42, 43, 44, 47 and 48 do not apply to these beneficiaries.

The beneficiary will not be subject to financial checks, reviews and audits under Article 22.

Beneficiaries not receiving EU funding may provide in-kind contributions to another beneficiary. In this case, they will be considered as a third party for the purpose of Articles 11 and 12.

9.2 Consequences of non-compliance

If a beneficiary not receiving EU funding breaches any of its obligations under this Article, its participation of the Agreement may be terminated (see Article 50).

Such breaches may also lead to any of the other measures described in Chapter 6 that are applicable to it.

ARTICLE 10 — PURCHASE OF GOODS, WORKS OR SERVICES

10.1 Rules for purchasing goods, works or services

10.1.1 If necessary to implement the action, the beneficiaries may purchase goods, works or services.

The beneficiaries must make such purchases ensuring the best value for money or, if appropriate, the lowest price. In doing so, they must avoid any conflict of interests (see Article 35).

The beneficiaries must ensure that the Commission, the European Court of Auditors (ECA) and the European Anti-Fraud Office (OLAF) can exercise their rights under Articles 22 and 23 also towards their contractors.

10.1.2 Beneficiaries that are 'contracting authorities' within the meaning of Directive 2004/18/EC[6] or 'contracting entities' within the meaning of Directive 2004/17/EC[7] must comply with the applicable national law on public procurement.

10.2 Consequences of non-compliance

If a beneficiary breaches any of its obligations under Article 10.1.1, the costs related to the contract concerned will be ineligible (see Article 6) and will be rejected (see Article 42).

If a beneficiary breaches any of its obligations under Article 10.1.2, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

ARTICLE 11 — USE OF IN-KIND CONTRIBUTIONS PROVIDED BY THIRD PARTIES AGAINST PAYMENT

11.1 Rules for the use of in-kind contributions against payment

If necessary to implement the action, the beneficiaries may use in-kind contributions provided by third parties against payment.

The beneficiaries may declare costs related to the payment of in-kind contributions as eligible (see Article 6.1 and 6.2), up to the third parties' costs for the seconded persons, contributed equipment, infrastructure or other assets or other contributed goods and services.

The third parties and their contributions must be set out in Annex 1. The *Commission* may however approve in-kind contributions not set out in Annex 1 without amendment (see Article 55), if:

- they are specifically justified in the periodic technical report and

- their use does not entail changes to the Agreement which would call into question the decision awarding the grant or breach the principle of equal treatment of applicants.

The beneficiaries must ensure that the Commission, the European Court of Auditors (ECA) and the European Anti-Fraud Office (OLAF) can exercise their rights under Articles 22 and 23 also towards the third parties.

11.2 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the costs related to the payment of the in-kind contribution will be ineligible (see Article 6) and will be rejected (see Article 42).

Such breaches may also lead to any of the other measures described in Chapter 6.

---

[6] Directive 2004/18/EC of the European Parliament and of the Council of 31 March 2004 on the coordination of procedures for the award of public work contracts, public supply contracts and public service contracts (OJ L 134, 30.04.2004, p. 114).

[7] Directive 2004/17/EC of the European Parliament and of the Council of 31 March 2004 coordinating the procurement procedures of entities operating in the water, energy, transport and postal services sectors (OJ L 134, 30.04.2004, p. 1).

ARTICLE 12 — USE OF IN-KIND CONTRIBUTIONS PROVIDED BY THIRD PARTIES FREE OF CHARGE

12.1 Rules for the use of in-kind contributions free of charge

If necessary to implement the action, the beneficiaries may use in-kind contributions provided by third parties free of charge.

The beneficiaries may declare costs incurred by the third parties for the seconded persons, contributed equipment, infrastructure or other assets or other contributed goods and services as eligible in accordance with Article 6.4.

The third parties and their contributions must be set out in Annex 1. The *Commission* may however approve in-kind contributions not set out in Annex 1 without amendment (see Article 55), if:

- they are specifically justified in the periodic technical report and
- their use does not entail changes to the Agreement which would call into question the decision awarding the grant or breach the principle of equal treatment of applicants.

The beneficiaries must ensure that the Commission, the European Court of Auditors (ECA) and the European Anti-Fraud Office (OLAF) can exercise their rights under Articles 22 and 23 also towards the third parties.

12.2 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the costs incurred by the third parties related to the in-kind contribution will be ineligible (see Article 6) and will be rejected (see Article 42).

Such breaches may also lead to any of the other measures described in Chapter 6.

ARTICLE 13 — IMPLEMENTATION OF ACTION TASKS BY SUBCONTRACTORS

13.1 Rules for subcontracting action tasks

13.1.1 If necessary to implement the action, the beneficiaries may award subcontracts covering the implementation of certain action tasks described in Annex 1.

Subcontracting may cover only a limited part of the action.

The beneficiaries must award the subcontracts ensuring the best value for money or, if appropriate, the lowest price. In doing so, they must avoid any conflict of interests (see Article 35).

The tasks to be implemented and the estimated cost for each subcontract must be set out in Annex 1 and the total estimated costs of subcontracting per beneficiary must be set out in Annex 2. The *Commission* may however approve subcontracts not set out in Annex 1 and 2 without amendment (see Article 55), if:

- they are specifically justified in the periodic technical report and
- they do not entail changes to the Agreement which would call into question the decision awarding the grant or breach the principle of equal treatment of applicants.

The beneficiaries must ensure that the Commission, the European Court of Auditors (ECA) and the European Anti-Fraud Office (OLAF) can exercise their rights under Articles 22 and 23 also towards their subcontractors.

13.1.2 The beneficiaries must ensure that their obligations under Articles 35, 36, 38 and 46 also apply to the subcontractors.

Beneficiaries that are 'contracting authorities' within the meaning of Directive 2004/18/EC or 'contracting entities' within the meaning of Directive 2004/17/EC must comply with the applicable national law on public procurement.

13.2 Consequences of non-compliance

If a beneficiary breaches any of its obligations under Article 13.1.1, the costs related to the subcontract concerned will be ineligible (see Article 6) and will be rejected (see Article 42).

If a beneficiary breaches any of its obligations under Article 13.1.2, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

ARTICLE 14 — IMPLEMENTATION OF ACTION TASKS BY LINKED THIRD PARTIES

Not applicable

ARTICLE 15 — FINANCIAL SUPPORT TO THIRD PARTIES

15.1 Rules for providing financial support to third parties

Not applicable

15.2 Financial support in the form of prizes

Not applicable

15.3 Consequences of non-compliance

Not applicable

ARTICLE 16 — PROVISION OF TRANS-NATIONAL OR VIRTUAL ACCESS TO RESEARCH INFRASTRUCTURE

16.1 Rules for providing trans-national access to research infrastructure

*16.1.1 '**Access providers**'[11] must provide access to research infrastructure or installations[12] in accordance with the following conditions:*

[11] '**Access provider**' means a beneficiary or third party (subcontractor or linked third party) that is in charge of providing access to one or more research infrastructure or installations, or part of them, as described in Annex 1.

(a) access which must be provided:

The access must be free of charge, trans-national access to research infrastructure or installations for selected user-groups.

This access must include the logistical, technological and scientific support and the specific training that is usually provided to external researchers using the infrastructure.

(b) categories of users that may have access:

*Trans-national access must be provided to selected '**user-groups**', i.e. teams of one or more researchers (users) led by a '**user group leader**'.*

The user group leader and the majority of the users must work in a country other than the country(ies) where the installation is located.

This rule does not apply:

- *if access is provided by an International organisation, the Joint Research Centre (JRC), an ERIC or similar legal entities;*

- *in case of remote access to a set of installations located in different countries offering the same type of service.*

Only user groups that are allowed to disseminate the results they have generated under the action may benefit from the access, unless the users are working for SMEs.

Access for user groups with a majority of users not working in a EU or associated country is limited to 20% of the total amount of units of access provided under the grant, unless a higher percentage is foreseen in Annex 1;

(c) procedure and criteria for selecting user groups:

The user groups must request access by submitting (in writing) a description of the work that they wish to carry out and the names, nationalities and home institutions of the users.

*The user groups must be selected by a **selection panel** set up by the access providers.*

The selection panel must be composed of international experts in the field, at least half of them independent from the beneficiaries, unless otherwise specified in Annex 1.

The selection panel must assess all proposals received and recommend a short-list of the user groups that should benefit from access.

The selection panel must base its selection on scientific merit, taking into account that priority should be given to user groups composed of users who:

[12] '**Installation**' means a part or a service of a research infrastructure that could be used independently from the rest. A research infrastructure consists of one or more installations.

- *have not previously used the installation and*

- *are working in countries where no equivalent research infrastructure exist.*

It will apply the principles of transparency, fairness and impartiality.

(d) other conditions:

The access provider must request written approval from the Commission (see Article 52) for the selection of user groups requiring visits to the installation(s) exceeding 3 months, unless such visits are foreseen in Annex 1.

16.1.2 In addition, the access provider must:

- *advertise widely, including on a dedicated website, the access offered under the Agreement;*

- *promote equal opportunities in advertising the access and take into account the gender dimension when defining the support provided to users;*

- *ensure that users comply with the terms and conditions of this Agreement;*

- *ensure that his obligations under Articles 35, 36, 38 and 46 also apply to the users.*

16.2 Rules for providing virtual access to research infrastructure

*'**Access providers**' must provide access to research infrastructure or installations in accordance with the following conditions:*

(a) access which must be provided:

The access must be free of charge, virtual access to research infrastructure or installations.

'Virtual access' means open and free access through communication networks to resources needed for research, without selecting the researchers to whom access is provided;

(b) other conditions:

The access provider must have the virtual access services assessed periodically by a board composed of international experts in the field, at least half of whom must be independent from the beneficiaries, unless otherwise specified in Annex 1.

16.3 Consequences of non-compliance

If a beneficiary breaches any of its obligations under Articles 16.1.1 and 16.2, the costs of access will be ineligible (see Article 6) and will be rejected (see Article 42).

If a beneficiary breaches any of its obligations under Articles 16.1.2, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

SECTION 2 RIGHTS AND OBLIGATIONS RELATED TO THE GRANT ADMINISTRATION

ARTICLE 17 — GENERAL OBLIGATION TO INFORM

17.1 General obligation to provide information upon request

The beneficiaries must provide — during implementation of the action or afterwards and in accordance with Article 41.2 — any information requested in order to verify eligibility of the costs, proper implementation of the action and compliance with any other obligation under the Agreement.

17.2 Obligation to keep information up to date and to inform about events and circumstances likely to affect the Agreement

Each beneficiary must keep information stored in the 'Beneficiary Register' (via the electronic exchange system; see Article 52) up to date, in particular, its name, address, legal representatives, legal form and organisation type.

Each beneficiary must immediately inform the coordinator — which must immediately inform the *Commission* and the other beneficiaries — of any of the following:

(a) **events** which are likely to affect significantly or delay the implementation of the action or the EU's financial interests, in particular:

(i) changes in its legal, financial, technical, organisational or ownership situation

(b) **circumstances** affecting:

(i) the decision to award the grant or

(ii) compliance with requirements under the Agreement.

17.3 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

ARTICLE 18 — KEEPING RECORDS — SUPPORTING DOCUMENTATION

18.1 Obligation to keep records and other supporting documentation

The beneficiaries must — for a period of *five* years after the payment of the balance — keep records and other supporting documentation in order to prove the proper implementation of the action and the costs they declare as eligible.

They must make them available upon request (see Article 17) or in the context of checks, reviews, audits or investigations (see Article 22).

If there are on-going checks, reviews, audits, investigations, litigation or other pursuits of claims under the Agreement (including the extension of findings; see Articles 22), the beneficiaries must keep the records and other supporting documentation until the end of these procedures.

The beneficiaries must keep the original documents. Digital and digitalised documents are considered originals if they are authorised by the applicable national law. The *Commission* may accept non-original documents if it considers that they offer a comparable level of assurance.

18.1.1 Records and other supporting documentation on the scientific and technical implementation

The beneficiaries must keep records and other supporting documentation on scientific and technical implementation of the action in line with the accepted standards in the respective field.

18.1.2 Records and other documentation to support the costs declared

The beneficiaries must keep the records and documentation supporting the costs declared, in particular the following:

(a) for **actual costs**: adequate records and other supporting documentation to prove the costs declared, such as contracts, subcontracts, invoices and accounting records. In addition, the beneficiaries' usual cost accounting practices and internal control procedures must enable direct reconciliation between the amounts declared, the amounts recorded in their accounts and the amounts stated in the supporting documentation;

(b) for **unit costs**: adequate records and other supporting documentation to prove the number of units declared. *This documentation must include records of the names, nationalities, and home institutions of users, as well as the nature and quantity of access provided to them.* Beneficiaries do not need to identify the actual eligible costs covered or to keep or provide supporting documentation (such as accounting statements) to prove the amount per unit.

In addition, **for direct personnel costs declared as unit costs calculated in accordance with the beneficiary's usual cost accounting practices**, the beneficiaries must keep adequate records and documentation to prove that the cost accounting practices used comply with the conditions set out in Article 6.2, Point A.

The beneficiaries may submit to the Commission, for approval, a certificate (drawn up in accordance with Annex 6) stating that their usual cost accounting practices comply with these conditions ('**certificate on the methodology**'). If the certificate is approved, costs declared in line with this methodology will not be challenged subsequently, unless the beneficiaries have concealed information for the purpose of the approval.

(c) for **flat-rate costs**: adequate records and other supporting documentation to prove the eligibility of the costs to which the flat-rate is applied. The beneficiaries do not need to identify the costs covered or provide supporting documentation (such as accounting statements) to prove the amount declared at a flat-rate.

In addition, for **personnel costs** (declared as actual costs or on the basis of unit costs), the beneficiaries must keep **time records** for the number of hours declared. The time records must be in writing and approved by the persons working on the action and their supervisors, at least monthly. In the absence of reliable time records of the hours worked on the action, the *Commission* may accept alternative

evidence supporting the number of hours declared, if it considers that it offers an adequate level of assurance.

As an exception, for **persons working exclusively on the action**, there is no need to keep time records, if the beneficiary signs a **declaration** confirming that the persons concerned have worked exclusively on the action.

18.2 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, costs insufficiently substantiated will be ineligible (see Article 6) and will be rejected (see Article 42), and the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

ARTICLE 19 — SUBMISSION OF DELIVERABLES

19.1 Obligation to submit deliverables

The coordinator must submit the '**deliverables**' identified in Annex 1, in accordance with the timing and conditions set out in it.

19.2 Consequences of non-compliance

If the coordinator breaches any of its obligations under this Article, the *Commission* may apply any of the measures described in Chapter 6.

ARTICLE 20 — REPORTING — PAYMENT REQUESTS

20.1 Obligation to submit reports

The coordinator must submit to the *Commission* (see Article 52) the technical and financial reports set out in this Article. These reports include the requests for payment and must be drawn up using the forms and templates provided in the electronic exchange system (see Article 52).

20.2 Reporting periods

The action is divided into the following '**reporting periods**':
- RP1: from month 1 to month 18
- *RP2: from month 19 to month 36*
- *RP3: from month 37 to month 48*

20.3 Periodic reports — Requests for interim payments

The coordinator must submit a periodic report within 60 days following the end of each reporting period.

The **periodic report** must include the following:

(a) a '**periodic technical report**' containing:

(i) an **explanation of the work carried out** by the beneficiaries;

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

(ii) an **overview of the progress** towards the objectives of the action, including milestones and deliverables identified in Annex 1.

This report must include explanations justifying the differences between work expected to be carried out in accordance with Annex 1 and that actually carried out.

The report must also detail the exploitation and dissemination of the results and — if required in Annex 1 — an updated '**plan for the exploitation and dissemination of the results**';

The report must detail the access activity, indicating the members of the selection panel, the selection procedure, the exact amount of access provided to the user groups, the description of their work, and information on the users (including names, nationality and home institutions).

The reports must detail the access activity, with statistics on the virtual access provided in the period, including quantity, geographical distribution of users and, whenever possible, information/statistics on scientific outcomes (publications, patents, etc) acknowledging the use of the infrastructure;

(iii) a **summary** for publication by the *Commission*;

(iv) the answers to the '**questionnaire**', covering issues related to the action implementation and the economic and societal impact, notably in the context of the Horizon 2020 key performance indicators and the Horizon 2020 monitoring requirements;

(b) a '**periodic financial report**' containing:

(i) an '**individual financial statement**' (see Annex 4) from each beneficiary, for the reporting period concerned.

The individual financial statement must detail the eligible costs (actual costs, unit costs and flat-rate costs; see Article 6) for each budget category (see Annex 2).

The beneficiaries must declare all eligible costs, even if — for actual costs, unit costs and flat-rate costs — they exceed the amounts indicated in the estimated budget (see Annex 2). Amounts which are not declared in the individual financial statement will not be taken into account by the *Commission*.

If an individual financial statement is not submitted for a reporting period, it may be included in the periodic financial report for the next reporting period.

The individual financial statements of the last reporting period must also detail the **receipts of the action** (see Article 5.3.3).

Each beneficiary must **certify** that:

- the information provided is full, reliable and true;

- the costs declared are eligible (see Article 6);

- the costs can be substantiated by adequate records and supporting documentation (see Article 18) that will be produced upon request (see Article 17) or in the context of checks, reviews, audits and investigations (see Article 22), and

- for the last reporting period: that all the receipts have been declared (see Article 5.3.3);

(ii) an **explanation of the use of resources** and the information on subcontracting (see Article 13) and in-kind contributions provided by third parties (see Articles 11 and 12) from each beneficiary, for the reporting period concerned;

(iii) *not applicable;*

(iv) a '**periodic summary financial statement**' (see Annex 4), created automatically by the electronic exchange system, consolidating the individual financial statements for the reporting period concerned and including — except for the last reporting period — the **request for interim payment**.

20.4 Final report — Request for payment of the balance

In addition to the periodic report for the last reporting period, the coordinator must submit the final report within 60 days following the end of the last reporting period.

The **final report** must include the following:

(a) a '**final technical report**' with a **summary** for publication containing:

(i) an overview of the results and their exploitation and dissemination;

(ii) the conclusions on the action, and

(iii) the socio-economic impact of the action;

(b) a '**final financial report**' containing:

(i) a '**final summary financial statement**' (see Annex 4), created automatically by the electronic exchange system, consolidating the individual financial statements for all reporting periods and including the **request for payment of the balance** and

(ii) a '**certificate on the financial statements**' (drawn up in accordance with Annex 5) for each beneficiary , if it requests a total contribution of EUR 325 000 or more, as reimbursement of actual costs and unit costs calculated on the basis of its usual cost accounting practices (see Article 5.2 and Article 6.2, Point A).

20.5 Information on cumulative expenditure incurred

Not applicable

20.6 Currency for financial statements and conversion into euro

Financial statements must be drafted in euro.

Beneficiaries with accounting established in a currency other than the euro must convert the costs recorded in their accounts into euro, at the average of the daily exchange rates published in the C series of the *Official Journal of the European Union*, calculated over the corresponding reporting period.

If no daily euro exchange rate is published in the *Official Journal of the European Union* for the currency in question, they must be converted at the average of the monthly accounting rates published on the Commission's website, calculated over the corresponding reporting period.

Beneficiaries with accounting established in euro must convert costs incurred in another currency into euro according to their usual accounting practices.

20.7 Language of reports

All reports (technical and financial reports, including financial statements) must be submitted in the language of the Agreement.

20.8 Consequences of non-compliance — Suspension of the payment deadline — Termination

If the reports submitted do not comply with this Article, the *Commission* may suspend the payment deadline (see Article 47) and apply any of the other measures described in Chapter 6.

If the coordinator breaches its obligation to submit the reports and if it fails to comply with this obligation within 30 days following a written reminder sent by the *Commission*, the Agreement may be terminated (see Article 50).

ARTICLE 21 — PAYMENTS AND PAYMENT ARRANGEMENTS

21.1 Payments to be made

The following payments will be made to the coordinator:

- one **pre-financing payment**;
- one or more **interim payments**, on the basis of the request(s) for interim payment (see Article 20), and
- one **payment of the balance**, on the basis of the request for payment of the balance (see Article 20).

21.2 Pre-financing payment — Amount — Amount retained for the Guarantee Fund

The aim of the pre-financing is to provide the beneficiaries with a float.

It remains the property of the *EU* until the payment of the balance.

The amount of the pre-financing payment will be EUR **1,875,000.00** (one million eight hundred and seventy five thousand EURO).

The *Commission* will — except if Article 48 applies — make the pre-financing payment to the coordinator within 30 days either from the entry into force of the Agreement (see Article 58) or from 10 days before the starting date of the action (see Article 3), whichever is the latest.

An amount of EUR **250,000.00** (two hundred and fifty thousand EURO), corresponding to 5% of the maximum grant amount (see Article 5.1), is retained by the *Commission* from the pre-financing payment and transferred into the '**Guarantee Fund**'.

21.3 Interim payments — Amount — Calculation

Interim payments reimburse the eligible costs incurred for the implementation of the action during the corresponding reporting periods.

The *Commission* will pay to the coordinator the amount due as interim payment within 90 days from receiving the periodic report (see Article 20.3), except if Articles 47 or 48 apply.

Payment is subject to the approval of the periodic report. Its approval does not imply recognition of the compliance, authenticity, completeness or correctness of its content.

The **amount due as interim payment** is calculated by the *Commission* in the following steps:

Step 1 – Application of the reimbursement rates

Step 2 – Limit to 90% of the maximum grant amount

21.3.1 Step 1 — Application of the reimbursement rates

The reimbursement rate(s) (see Article 5.2) are applied to the eligible costs (actual costs, unit costs and flat-rate costs ; see Article 6) declared by the beneficiaries (see Article 20) and approved by the *Commission* (see above) for the concerned reporting period.

21.3.2 Step 2 — Limit to 90% of the maximum grant amount

The total amount of pre-financing and interim payments must not exceed 90% of the maximum grant amount set out in Article 5.1. The maximum amount for the interim payment will be calculated as follows:

{90% of the maximum grant amount (see Article 5.1)

minus

{pre-financing and previous interim payments}}.

21.4 Payment of the balance — Amount — Calculation — Release of the amount retained for the Guarantee Fund

The payment of the balance reimburses the remaining part of the eligible costs incurred by the beneficiaries for the implementation of the action.

If the total amount of earlier payments is greater than the final grant amount (see Article 5.3), the payment of the balance takes the form of a recovery (see Article 44).

If the total amount of earlier payments is lower than the final grant amount, the *Commission* will pay the balance within 90 days from receiving the final report (see Article 20.4), except if Articles 47 or 48 apply.

Payment is subject to the approval of the final report. Its approval does not imply recognition of the compliance, authenticity, completeness or correctness of its content.

The **amount due as the balance** is calculated by the *Commission* by deducting the total amount of pre-financing and interim payments (if any) already made, from the final grant amount determined in accordance with Article 5.3:

{final grant amount (see Article 5.3)

minus

{pre-financing and interim payments (if any) made}}.

At the payment of the balance, the amount retained for the Guarantee Fund (see above) will be released and:

- if the balance is positive: the amount released will be paid in full to the coordinator together with the amount due as the balance;

- if the balance is negative (payment of the balance taking the form of recovery): it will be deducted from the amount released (see Article 44.1.2). If the resulting amount:

    - is positive, it will be paid to the coordinator

    - is negative, it will be recovered.

The amount to be paid may however be offset — without the beneficiary's consent — against any other amount owed by the beneficiary to the Commission or an executive agency (under the EU or Euratom budget), up to the maximum EU contribution indicated, for that beneficiary, in the estimated budget (see Annex 2).

21.5 Notification of amounts due

When making payments, the *Commission* will formally notify to the coordinator the amount due, specifying whether it concerns an interim payment or the payment of the balance.

For the payment of the balance, the notification will also specify the final grant amount.

In the case of reduction of the grant or recovery of undue amounts, the notification will be preceded by the contradictory procedure set out in Articles 43 and 44.

21.6 Currency for payments

The *Commission* will make all payments in euro.

21.7 Payments to the coordinator — Distribution to the beneficiaries

Payments will be made to the coordinator.

Payments to the coordinator will discharge the *Commission* from its payment obligation.

The coordinator must distribute the payments between the beneficiaries without unjustified delay.

Pre-financing may however be distributed only:

(a) if the minimum number of beneficiaries set out in the call for proposals has acceded to the Agreement (see Article 56) and

(b) to beneficiaries that have acceded to the Agreement (see Article 56).

21.8 Bank account for payments

All payments will be made to the following bank account:

Name of bank: BANCA NAZIONALE DEL LAVORO S.P.A.
Address of branch: PIAZZALE ALDO MORO 7 ROMA, Italy
Full name of the account holder: INCASSI GIORNALIERI DA ALTRE DIPENDENZE
Full account number (including bank codes):
IBAN code: IT57S0100503392000000218155

21.9 Costs of payment transfers

The cost of the payment transfers is borne as follows:

- the *Commission* bears the cost of transfers charged by its bank;
- the beneficiary bears the cost of transfers charged by its bank;
- the party causing a repetition of a transfer bears all costs of the repeated transfer.

21.10 Date of payment

Payments by the *Commission* are considered to have been carried out on the date when they are debited to its account.

21.11 Consequences of non-compliance

21.11.1 If the *Commission* does not pay within the payment deadlines (see above), the beneficiaries are entitled to **late-payment interest** at the rate applied by the European Central Bank (ECB) for its main refinancing operations in euros ('reference rate'), plus three and a half points. The reference rate is the rate in force on the first day of the month in which the payment deadline expires, as published in the C series of the *Official Journal of the European Union*.

If the late-payment interest is lower than or equal to EUR 200, it will be paid to the coordinator only upon request submitted within two months of receiving the late payment.

Late-payment interest is not due if all beneficiaries are EU Member States (including regional and local government authorities or other public bodies acting on behalf of a Member State for the purpose of this Agreement).

Suspension of the payment deadline or payments (see Articles 47 and 48) will not be considered as late payment.

Late-payment interest covers the period running from the day following the due date for payment (see above), up to and including the date of payment.

Late-payment interest is not considered for the purposes of calculating the final grant amount.

21.11.2 If the coordinator breaches any of its obligations under this Article, the grant may be reduced (see Article 43) and the Agreement or the participation of the coordinator may be terminated (see Article 50).

Such breaches may also lead to any of the other measures described in Chapter 6.

ARTICLE 22 — CHECKS, REVIEWS, AUDITS AND INVESTIGATIONS — EXTENSION OF FINDINGS

22.1 Checks, reviews and audits by the Commission

22.1.1 Right to carry out checks

The Commission will — during the implementation of the action or afterwards — check the proper implementation of the action and compliance with the obligations under the Agreement, including assessing deliverables and reports.

For this purpose the Commission may be assisted by external persons or bodies.

The Commission may also request additional information in accordance with Article 17. The Commission may request beneficiaries to provide such information to it directly.

Information provided must be accurate, precise and complete and in the format requested, including electronic format.

22.1.2 Right to carry out reviews

The Commission may — during the implementation of the action or afterwards — carry out reviews on the proper implementation of the action (including assessment of deliverables and reports), compliance with the obligations under the Agreement and continued scientific or technological relevance of the action.

Reviews may be started **up to two years after the payment of the balance**. They will be formally notified to the coordinator or beneficiary concerned and will be considered to have started on the date of the formal notification.

If the review is carried out on a third party (see Articles 10 to 16), the beneficiary concerned must inform the third party.

The Commission may carry out reviews directly (using its own staff) or indirectly (using external persons or bodies appointed to do so). It will inform the coordinator or beneficiary concerned of the identity of the external persons or bodies. They have the right to object to the appointment on grounds of commercial confidentiality.

The coordinator or beneficiary concerned must provide — within the deadline requested — any information and data in addition to deliverables and reports already submitted (including information

on the use of resources). The Commission may request beneficiaries to provide such information to it directly.

The coordinator or beneficiary concerned may be requested to participate in meetings, including with external experts.

For **on-the-spot** reviews, the beneficiaries must allow access to their sites and premises, including to external persons or bodies, and must ensure that information requested is readily available.

Information provided must be accurate, precise and complete and in the format requested, including electronic format.

On the basis of the review findings, a '**review report**' will be drawn up.

The Commission will formally notify the review report to the coordinator or beneficiary concerned, which has 30 days to formally notify observations ('**contradictory review procedure**').

Reviews (including review reports) are in the language of the Agreement.

22.1.3 Right to carry out audits

The Commission may — during the implementation of the action or afterwards — carry out audits on the proper implementation of the action and compliance with the obligations under the Agreement.

Audits may be started **up to two years after the payment of the balance**. They will be formally notified to the coordinator or beneficiary concerned and will be considered to have started on the date of the formal notification.

If the audit is carried out on a third party (see Articles 10 to 16), the beneficiary concerned must inform the third party.

The Commission may carry out audits directly (using its own staff) or indirectly (using external persons or bodies appointed to do so). It will inform the coordinator or beneficiary concerned of the identity of the external persons or bodies. They have the right to object to the appointment on grounds of commercial confidentiality.

The coordinator or beneficiary concerned must provide — within the deadline requested — any information (including complete accounts, individual salary statements or other personal data) to verify compliance with the Agreement. The Commission may request beneficiaries to provide such information to it directly.

For **on-the-spot** audits, the beneficiaries must allow access to their sites and premises, including to external persons or bodies, and must ensure that information requested is readily available.

Information provided must be accurate, precise and complete and in the format requested, including electronic format.

On the basis of the audit findings, a '**draft audit report**' will be drawn up.

The Commission will formally notify the draft audit report to the coordinator or beneficiary concerned, which has 30 days to formally notify observations ('**contradictory audit procedure**'). This period may be extended by the Commission in justified cases.

The '**final audit report**' will take into account observations by the coordinator or beneficiary concerned. The report will be formally notified to it.

Audits (including audit reports) are in the language of the Agreement.

The Commission may also access the beneficiaries' statutory records for the periodical assessment of unit costs or flat-rate amounts.

22.2 Investigations by the European Anti-Fraud Office (OLAF)

Under Regulations No 883/2013[15] and No 2185/96[16] (and in accordance with their provisions and procedures) the European Anti-Fraud Office (OLAF) may — at any moment during implementation of the action or afterwards — carry out investigations, including on-the-spot checks and inspections, to establish whether there has been fraud, corruption or any other illegal activity affecting the financial interests of the EU.

22.3 Checks and audits by the European Court of Auditors (ECA)

Under Article 287 of the Treaty on the Functioning of the European Union (TFEU) and Article 161 of the Financial Regulation No 966/2012[17], the European Court of Auditors (ECA) may — at any moment during implementation of the action or afterwards — carry out audits.

The ECA has the right of access for the purpose of checks and audits.

22.4 Checks, reviews, audits and investigations for international organisations

Not applicable

22.5 Consequences of findings in checks, reviews, audits and investigations — Extension of findings

22.5.1 Findings in this grant

Findings in checks, reviews, audits or investigations carried out in the context of this grant may lead to the rejection of ineligible costs (see Article 42), reduction of the grant (see Article 43), recovery of undue amounts (see Article 44) or to any of the other measures described in Chapter 6.

Rejection of costs or reduction of the grant after the payment of the balance will lead to a revised final grant amount (see Article 5.4).

Findings in checks, reviews, audits or investigations may lead to a request for amendment for the modification of Annex 1 (see Article 55).

---

[15] Regulation (EU, Euratom) No 883/2013 of the European Parliament and of the Council of 11 September 2013 concerning investigations conducted by the European Anti-Fraud Office (OLAF) and repealing Regulation (EC) No 1073/1999 of the European Parliament and of the Council and Council Regulation (Euratom) No 1074/1999 (OJ L 248, 18.09.2013, p. 1).

[16] Council Regulation (Euratom, EC) No 2185/1996 of 11 November 1996 concerning on-the-spot checks and inspections carried out by the Commission in order to protect the European Communities' financial interests against fraud and other irregularities (OJ L 292, 15.11.1996, p. 2).

[17] Regulation (EU, Euratom) No 966/2012 of the European Parliament and of the Council of 25 October 2012 on the financial rules applicable to the general budget of the Union and repealing Council Regulation (EC, Euratom) No 1605/2002 ('**Financial Regulation No 966/2012**') (OJ L 298, 26.10.2012, p. 1).

Checks, reviews, audits or investigations that find systemic or recurrent errors, irregularities, fraud or breach of obligations may also lead to consequences in other EU or Euratom grants awarded under similar conditions ('**extension of findings from this grant to other grants**').

Moreover, findings arising from an OLAF investigation may lead to criminal prosecution under national law.

22.5.2 Findings in other grants

The Commission may extend findings from other grants to this grant ('**extension of findings from other grants to this grant**'), if:

(a) the beneficiary concerned is found, in other EU or Euratom grants awarded under similar conditions, to have committed systemic or recurrent errors, irregularities, fraud or breach of obligations that have a material impact on this grant and

(b) those findings are formally notified to the beneficiary concerned — together with the list of grants affected by the findings — no later than two years after the payment of the balance of this grant.

The extension of findings may lead to the rejection of costs (see Article 42), reduction of the grant (see Article 43), recovery of undue amounts (see Article 44), suspension of payments (see Article 48), suspension of the action implementation (see Article 49) or termination (see Article 50).

22.5.3 Procedure

The Commission will formally notify the beneficiary concerned the systemic or recurrent errors and its intention to extend these audit findings, together with the list of grants affected.

22.5.3.1 If the findings concern **eligibility of costs**: the formal notification will include:

(a) an invitation to submit observations on the list of grants affected by the findings;

(b) the request to submit **revised financial statements** for all grants affected;

(c) the **correction rate for extrapolation** established by the Commission on the basis of the systemic or recurrent errors, to calculate the amounts to be rejected if the beneficiary concerned:

(i) considers that the submission of revised financial statements is not possible or practicable or

(ii) does not submit revised financial statements.

The beneficiary concerned has 90 days from receiving notification to submit observations, revised financial statements or to propose a duly substantiated **alternative correction method**. This period may be extended by the Commission in justified cases.

The amounts to be rejected will be determined on the basis of the revised financial statements, subject to their approval.

If the Commission does not receive any observations or revised financial statements, does not accept the observations or the proposed alternative correction method or does not approve the revised

financial statements, it will formally notify the beneficiary concerned the application of the initially notified correction rate for extrapolation.

If the Commission accepts the alternative correction method proposed by the beneficiary concerned, it will formally notify the application of the accepted alternative correction method.

22.5.3.2 If the findings concern **improper implementation** or a **breach of another obligation**: the formal notification will include:

(a) an invitation to submit observations on the list of grants affected by the findings and

(b) the flat-rate the Commission intends to apply according to the principle of proportionality.

The beneficiary concerned has 90 days from receiving notification to submit observations or to propose a duly substantiated alternative flat-rate.

If the Commission does not receive any observations or does not accept the observations or the proposed alternative flat-rate, it will formally notify the beneficiary concerned the application of the initially notified flat-rate.

If the Commission accepts the alternative flat-rate proposed by the beneficiary concerned, it will formally notify the application of the accepted alternative flat-rate.

22.6 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, any insufficiently substantiated costs will be ineligible (see Article 6) and will be rejected (see Article 42).

Such breaches may also lead to any of the other measures described in Chapter 6.

ARTICLE 23 — EVALUATION OF THE IMPACT OF THE ACTION

23.1 Right to evaluate the impact of the action

The Commission may carry out interim and final evaluations of the impact of the action measured against the objective of the *EU* programme.

Evaluations may be started during implementation of the action and up to *five* years after the payment of the balance. The evaluation is considered to start on the date of the formal notification to the coordinator or beneficiaries.

The Commission may make these evaluations directly (using its own staff) or indirectly (using external bodies or persons it has authorised to do so).

The coordinator or beneficiaries must provide any information relevant to evaluate the impact of the action, including information in electronic format.

23.2 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the *Commission* may apply the measures described in Chapter 6.

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

SECTION 3 RIGHTS AND OBLIGATIONS RELATED TO BACKGROUND AND RESULTS

SUBSECTION 1 GENERAL

ARTICLE 23a — MANAGEMENT OF INTELLECTUAL PROPERTY

23a.1 Obligation to take measures to implement the Commission Recommendation on the management of intellectual property in knowledge transfer activities

Beneficiaries that are universities or other public research organisations must take measures to implement the principles set out in Points 1 and 2 of the Code of Practice annexed to the Commission Recommendation on the management of intellectual property in knowledge transfer activities[18].

This does not change the obligations set out in Subsections 2 and 3 of this Section.

The beneficiaries must ensure that researchers and third parties involved in the action are aware of them.

23a.2 Consequences of non-compliance

If a beneficiary breaches its obligations under this Article, the *Commission* may apply any of the measures described in Chapter 6.

SUBSECTION 2 RIGHTS AND OBLIGATIONS RELATED TO BACKGROUND

ARTICLE 24 — AGREEMENT ON BACKGROUND

24.1 Agreement on background

The beneficiaries must identify and agree (in writing) on the background for the action ('**agreement on background**').

'**Background**' means any data, know-how or information — whatever its form or nature (tangible or intangible), including any rights such as intellectual property rights — that:

(a) is held by the beneficiaries before they acceded to the Agreement, and

(b) is needed to implement the action or exploit the results.

24.2 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

[18] Commission Recommendation C (2008) 1329 of 10.4.2008 on the management of intellectual property in knowledge transfer activities and the Code of Practice for universities and other public research institutions attached to this recommendation.

ARTICLE 25 — ACCESS RIGHTS TO BACKGROUND

25.1 Exercise of access rights — Waiving of access rights — No sub-licensing

To exercise access rights, this must first be requested in writing ('**request for access**').

'**Access rights**' means rights to use results or background under the terms and conditions laid down in this Agreement.

Waivers of access rights are not valid unless in writing.

Unless agreed otherwise, access rights do not include the right to sub-license.

25.2 Access rights for other beneficiaries, for implementing their own tasks under the action

The beneficiaries must give each other access — on a royalty-free basis — to background needed to implement their own tasks under the action, unless the beneficiary that holds the background has — before acceding to the Agreement —:

(a) informed the other beneficiaries that access to its background is subject to legal restrictions or limits, including those imposed by the rights of third parties (including personnel), or

(b) agreed with the other beneficiaries that access would not be on a royalty-free basis.

25.3 Access rights for other beneficiaries, for exploiting their own results

The beneficiaries must give each other access — under fair and reasonable conditions — to background needed for exploiting their own results, unless the beneficiary that holds the background has — before acceding to the Agreement — informed the other beneficiaries that access to its background is subject to legal restrictions or limits, including those imposed by the rights of third parties (including personnel).

'**Fair and reasonable conditions**' means appropriate conditions, including possible financial terms or royalty-free conditions, taking into account the specific circumstances of the request for access, for example the actual or potential value of the results or background to which access is requested and/or the scope, duration or other characteristics of the exploitation envisaged.

Requests for access may be made — unless agreed otherwise — up to one year after the period set out in Article 3.

25.4 Access rights for affiliated entities

Unless otherwise agreed in the consortium agreement, access to background must also be given — under fair and reasonable conditions (see above; Article 25.3) and unless it is subject to legal restrictions or limits, including those imposed by the rights of third parties (including personnel) —

to affiliated entities[19] established in an EU Member State or '**associated country**' [20], if this is needed to exploit the results generated by the beneficiaries to which they are affiliated.

Unless agreed otherwise (see above; Article 25.1), the affiliated entity concerned must make the request directly to the beneficiary that holds the background.

Requests for access may be made — unless agreed otherwise — up to one year after the period set out in Article 3.

25.5 Access rights for third parties

The access provider must — unless it is subject to legal restrictions or limits, including those imposed by the rights of third parties (including personnel) — give users royalty-free access to background needed to implement the action.

The access provider must inform the users as soon as possible of any restriction which might substantially affect the granting of access rights.

25.6 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

SUBSECTION 3 RIGHTS AND OBLIGATIONS RELATED TO RESULTS

ARTICLE 26 — OWNERSHIP OF RESULTS

26.1 Ownership by the beneficiary that generates the results

Results are owned by the beneficiary that generates them.

---

[19] For the definition, see Article 2.1(2) of the Rules for Participation Regulation No 1290/2013: '**affiliated entity**' means any legal entity that is under the direct or indirect control of a participant, or under the same direct or indirect control as the participant, or that is directly or indirectly controlling a participant.
'Control' may take any of the following forms:
(a) the direct or indirect holding of more than 50% of the nominal value of the issued share capital in the legal entity concerned, or of a majority of the voting rights of the shareholders or associates of that entity;
(b) the direct or indirect holding, in fact or in law, of decision-making powers in the legal entity concerned.
However the following relationships between legal entities shall not in themselves be deemed to constitute controlling relationships:
(a) the same public investment corporation, institutional investor or venture-capital company has a direct or indirect holding of more than 50% of the nominal value of the issued share capital or a majority of voting rights of the shareholders or associates;
(b) the legal entities concerned are owned or supervised by the same public body.

[20] For the definition, see Article 2.1(3) of the Rules for Participation Regulation No 1290/2013: '**associated country**' means a third country which is party to an international agreement with the Union, as identified in *Article 7 of Horizon 2020 Framework Programme Regulation No 1291/2013. Article 7 sets out the conditions for association of non-EU countries to Horizon 2020.*

'**Results**' means any (tangible or intangible) output of the action such as data, knowledge or information — whatever its form or nature, whether it can be protected or not — that is generated in the action, as well as any rights attached to it, including intellectual property rights.

26.2 Joint ownership by several beneficiaries

Two or more beneficiaries own results jointly if:

(a) they have jointly generated them and

(b) it is not possible to:

(i) establish the respective contribution of each beneficiary, or

(ii) separate them for the purpose of applying for, obtaining or maintaining their protection (see Article 27).

The joint owners must agree (in writing) on the allocation and terms of exercise of their joint ownership ('**joint ownership agreement**'), to ensure compliance with their obligations under this Agreement.

Unless otherwise agreed in the joint ownership agreement, each joint owner may grant non-exclusive licences to third parties to exploit jointly-owned results (without any right to sub-license), if the other joint owners are given:

(a) at least 45 days advance notice and

(b) fair and reasonable compensation.

Once the results have been generated, joint owners may agree (in writing) to apply another regime than joint ownership (such as, for instance, transfer to a single owner (see Article 30) with access rights for the others).

26.3 Rights of third parties (including personnel)

If third parties (including personnel) may claim rights to the results, the beneficiary concerned must ensure that it complies with its obligations under the Agreement.

If a third party generates results, the beneficiary concerned must obtain all necessary rights (transfer, licences or other) from the third party, in order to be able to respect its obligations as if those results were generated by the beneficiary itself.

If obtaining the rights is impossible, the beneficiary must refrain from using the third party to generate the results.

26.4 *EU* ownership, to protect results

26.4.1 *The EU* may — with the consent of the beneficiary concerned — assume ownership of results to protect them, if a beneficiary intends — up to four years after the period set out in Article 3 — to disseminate its results without protecting them, except in any of the following cases:

(a) the lack of protection is because protecting the results is not possible, reasonable or justified (given the circumstances);

(b) the lack of protection is because there is a lack of potential for commercial or industrial exploitation, or

(c) the beneficiary intends to transfer the results to another beneficiary or third party established in an EU Member State or associated country, which will protect them.

Before the results are disseminated and unless any of the cases above under Points (a), (b) or (c) applies, the beneficiary must formally notify the *Commission* and at the same time inform it of any reasons for refusing consent. The beneficiary may refuse consent only if it can show that its legitimate interests would suffer significant harm.

If the *Commission* decides to assume ownership, it will formally notify the beneficiary concerned within 45 days of receiving notification.

No dissemination relating to these results may before the end of this period or, if the *Commission* takes a positive decision, until it has taken the necessary steps to protect the results.

26.4.2 *The EU* may — with the consent of the beneficiary concerned — assume ownership of results to protect them, if a beneficiary intends — up to four years after the period set out in Article 3 — to stop protecting them or not to seek an extension of protection, except in any of the following cases:

(a) the protection is stopped because of a lack of potential for commercial or industrial exploitation;

(b) an extension would not be justified given the circumstances.

A beneficiary that intends to stop protecting results or not seek an extension must — unless any of the cases above under Points (a) or (b) applies — formally notify the *Commission* at least 60 days before the protection lapses or its extension is no longer possible and at the same time inform it of any reasons for refusing consent. The beneficiary may refuse consent only if it can show that its legitimate interests would suffer significant harm.

If the *Commission* decides to assume ownership, it will formally notify the beneficiary concerned within 45 days of receiving notification.

26.5 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such breaches may also lead to the any of the other measures described in Chapter 6.

ARTICLE 27 — PROTECTION OF RESULTS — VISIBILITY OF EU FUNDING

27.1 Obligation to protect the results

Each beneficiary must examine the possibility of protecting its results and must adequately protect them — for an appropriate period and with appropriate territorial coverage — if:

(a) the results can reasonably be expected to be commercially or industrially exploited and

(b) protecting them is possible, reasonable and justified (given the circumstances).

When deciding on protection, the beneficiary must consider its own legitimate interests and the legitimate interests (especially commercial) of the other beneficiaries.

27.2 *EU* ownership, to protect the results

If a beneficiary intends not to protect its results, to stop protecting them or not seek an extension of protection, *The EU* may — under certain conditions (see Article 26.4) — assume ownership to ensure their (continued) protection.

27.3 Information on EU funding

Applications for protection of results (including patent applications) filed by or on behalf of a beneficiary must — unless the *Commission* requests or agrees otherwise or unless it is impossible — include the following:

> "The project leading to this application has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 654024".

27.4 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such a breach may also lead to any of the other measures described in Chapter 6.

ARTICLE 28 — EXPLOITATION OF RESULTS

28.1 Obligation to exploit the results

Each beneficiary must — up to four years after the period set out in Article 3 — take measures aiming to ensure '**exploitation**' of its results (either directly or indirectly, in particular through transfer or licensing; see Article 30) by:

(a) using them in further research activities (outside the action);

(b) developing, creating or marketing a product or process;

(c) creating and providing a service, or

(d) using them in standardisation activities.

This does not change the security obligations in Article 37, which still apply.

28.2 Results that could contribute to European or international standards — Information on EU funding

If results are incorporated in a standard, the beneficiary concerned must — unless the *Commission* requests or agrees otherwise or unless it is impossible — ask the standardisation body to include the following statement in (information related to) the standard:

> "Results incorporated in this standard received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 654024".

28.3 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced in accordance with Article 43.

Such a breach may also lead to any of the other measures described in Chapter 6.

ARTICLE 29 — DISSEMINATION OF RESULTS — OPEN ACCESS — VISIBILITY OF EU FUNDING

29.1 Obligation to disseminate results

Unless it goes against their legitimate interests, each beneficiary must — as soon as possible — '**disseminate**' its results by disclosing them to the public by appropriate means (other than those resulting from protecting or exploiting the results), including in scientific publications (in any medium).

This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39, all of which still apply.

A beneficiary that intends to disseminate its results must give advance notice to the other beneficiaries of — unless agreed otherwise — at least 45 days, together with sufficient information on the results it will disseminate.

Any other beneficiary may object within — unless agreed otherwise — 30 days of receiving notification, if it can show that its legitimate interests in relation to the results or background would be significantly harmed. In such cases, the dissemination may not take place unless appropriate steps are taken to safeguard these legitimate interests.

If a beneficiary intends not to protect its results, it may — under certain conditions (see Article 26.4.1) — need to formally notify the *Commission* before dissemination takes place.

29.2 Open access to scientific publications

Each beneficiary must ensure open access (free of charge online access for any user) to all peer-reviewed scientific publications relating to its results.

In particular, it must:

(a) as soon as possible and at the latest on publication, deposit a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications;

Moreover, the beneficiary must aim to deposit at the same time the research data needed to validate the results presented in the deposited scientific publications.

(b) ensure open access to the deposited publication — via the repository — at the latest:

(i) on publication, if an electronic version is available for free via the publisher, or

(ii) within six months of publication (twelve months for publications in the social sciences and humanities) in any other case.

(c) ensure open access — via the repository — to the bibliographic metadata that identify the deposited publication.

The bibliographic metadata must be in a standard format and must include all of the following:

- the terms *"European Union (EU)" and "Horizon 2020"*;
- the name of the action, acronym and grant number;
- the publication date, and length of embargo period if applicable, and
- a persistent identifier.

29.3 Open access to research data

Not applicable

29.4 Information on EU funding — Obligation and right to use the EU emblem

Unless the *Commission* requests or agrees otherwise or unless it is impossible, any dissemination of results (in any form, including electronic) must:

(a) display the EU emblem and

(b) include the following text:

"This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 654024".

When displayed together with another logo, the EU emblem must have appropriate prominence.

For the purposes of their obligations under this Article, the beneficiaries may use the EU emblem without first obtaining approval from the *Commission*.

This does not however give them the right to exclusive use.

Moreover, they may not appropriate the EU emblem or any similar trademark or logo, either by registration or by any other means.

29.5 Disclaimer excluding *Commission* responsibility

Any dissemination of results must indicate that it reflects only the author's view and that the *Commission* is not responsible for any use that may be made of the information it contains.

29.6 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such a breach may also lead to any of the other measures described in Chapter 6.

ARTICLE 30 — TRANSFER AND LICENSING OF RESULTS

30.1 Transfer of ownership

Each beneficiary may transfer ownership of its results.

It must however ensure that its obligations under Articles 26.2, 26.4, 27, 28, 29, 30 and 31 also apply to the new owner and that this owner has the obligation to pass them on in any subsequent transfer.

This does not change the security obligations in Article 37, which still apply.

Unless agreed otherwise (in writing) for specifically-identified third parties or unless impossible under applicable EU and national laws on mergers and acquisitions, a beneficiary that intends to transfer ownership of results must give at least 45 days advance notice (or less if agreed in writing) to the other beneficiaries that still have (or still may request) access rights to the results. This notification must include sufficient information on the new owner to enable any beneficiary concerned to assess the effects on its access rights.

Unless agreed otherwise (in writing) for specifically-identified third parties, any other beneficiary may object within 30 days of receiving notification (or less if agreed in writing), if it can show that the transfer would adversely affect its access rights. In this case, the transfer may not take place until agreement has been reached between the beneficiaries concerned.

30.2 Granting licenses

Each beneficiary may grant licences to its results (or otherwise give the right to exploit them), if:

(a) this does not impede the rights under Article 31 and

(b) *not applicable.*

In addition to Points (a) and (b), exclusive licences for results may be granted only if all the other beneficiaries concerned have waived their access rights (see Article 31.1).

This does not change the dissemination obligations in Article 29 or security obligations in Article 37, which still apply.

30.3 *Commission* right to object to transfers or licensing

Not applicable

30.4 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such a breach may also lead to any of the other measures described in Chapter 6.

ARTICLE 31 — ACCESS RIGHTS TO RESULTS

31.1 Exercise of access rights — Waiving of access rights — No sub-licensing

The conditions set out in Article 25.1 apply.

The obligations set out in this Article do not change the security obligations in Article 37, which still apply.

31.2 Access rights for other beneficiaries, for implementing their own tasks under the action

The beneficiaries must give each other access — on a royalty-free basis — to results needed for implementing their own tasks under the action.

31.3 Access rights for other beneficiaries, for exploiting their own results

The beneficiaries must give each other — under fair and reasonable conditions (see Article 25.3) — access to results needed for exploiting their own results.

Requests for access may be made — unless agreed otherwise — up to one year after the period set out in Article 3.

31.4 Access rights of affiliated entities

Unless agreed otherwise in the consortium agreement, access to results must also be given — under fair and reasonable conditions (Article 25.3) — to affiliated entities established in an EU Member State or associated country, if this is needed for those entities to exploit the results generated by the beneficiaries to which they are affiliated.

Unless agreed otherwise (see above; Article 31.1), the affiliated entity concerned must make any such request directly to the beneficiary that owns the results.

Requests for access may be made — unless agreed otherwise — up to one year after the period set out in Article 3.

31.5 Access rights for the EU institutions, bodies, offices or agencies and EU Member States

The beneficiaries must give access to their results — on a royalty-free basis — to EU institutions, bodies, offices or agencies, for developing, implementing or monitoring EU policies or programmes.

Such access rights are limited to non-commercial and non-competitive use.

This does not change the right to use any material, document or information received from the beneficiaries for communication and publicising activities (see Article 38.2).

31.6 Access rights for third parties

The access provider must give the users royalty-free access to the results needed to implement the action.

31.7 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

SECTION 4 OTHER RIGHTS AND OBLIGATIONS

ARTICLE 32 — RECRUITMENT AND WORKING CONDITIONS FOR RESEARCHERS

32.1 Obligation to take measures to implement the European Charter for Researchers and Code of Conduct for the Recruitment of Researchers

The beneficiaries must take all measures to implement the principles set out in the Commission Recommendation on the European Charter for Researchers and the Code of Conduct for the Recruitment of Researchers[22], in particular regarding:

- working conditions;
- transparent recruitment processes based on merit, and
- career development.

The beneficiaries must ensure that researchers and third parties involved in the action are aware of them.

32.2 Consequences of non-compliance

If a beneficiary breaches its obligations under this Article, the *Commission* may apply any of the measures described in Chapter 6.

ARTICLE 33 — GENDER EQUALITY

33.1 Obligation to aim for gender equality

The beneficiaries must take all measures to promote equal opportunities between men and women in the implementation of the action. They must aim, to the extent possible, for a gender balance at all levels of personnel assigned to the action, including at supervisory and managerial level.

33.2 Consequences of non-compliance

If a beneficiary breaches its obligations under this Article, the *Commission* may apply any of the measures described in Chapter 6.

[22] Commission Recommendation 2005/251/EC of 11 March 2005 on the European Charter for Researchers and on a Code of Conduct for the Recruitment of Researchers (OJ L 75, 22.3.2005, p. 67).

ARTICLE 34 — ETHICS

34.1 Obligation to comply with ethical principles

The beneficiaries must carry out the action in compliance with:

(a) ethical principles (including the highest standards of research integrity — as set out, for instance, in the European Code of Conduct for Research Integrity[23] — and including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct) and

(b) applicable international, EU and national law.

Funding will not be granted for activities carried out outside the EU if they are prohibited in all Member States.

The beneficiaries must ensure that the activities under the action have an exclusive focus on civil applications.

The beneficiaries must ensure that the activities under the action do not:

(a) aim at human cloning for reproductive purposes;

(b) intend to modify the genetic heritage of human beings which could make such changes heritable (with the exception of research relating to cancer treatment of the gonads, which may be financed), or

(c) intend to create human embryos solely for the purpose of research or for the purpose of stem cell procurement, including by means of somatic cell nuclear transfer.

34.2 Activities raising ethical issues

Activities raising ethical issues must comply with the '**ethics requirements**' set out in Annex 1.

Before the beginning of an activity raising an ethical issue, the coordinator must submit (see Article 52) to the *Commission* copy of:

(a) any ethics committee opinion required under national law and

(b) any notification or authorisation for activities raising ethical issues required under national law.

If these documents are not in English, the coordinator must also submit an English summary of the submitted opinions, notifications and authorisations (containing, if available, the conclusions of the committee or authority concerned).

If these documents are specifically requested for the action, the request must contain an explicit reference to the action title. The coordinator must submit a declaration by each beneficiary concerned that all the submitted documents cover the action tasks.

[23] The European Code of Conduct for Research Integrity of ALLEA (All European Academies) and ESF (European Science Foundation) of March 2011.
http://www.esf.org/fileadmin/Public_documents/Publications/Code_Conduct_ResearchIntegrity.pdf

34.3 Activities involving human embryos or human embryonic stem cells

Activities involving research on human embryos or human embryonic stem cells may be carried out only if:

- they are set out in Annex 1 or
- the coordinator has obtained explicit approval (in writing) from the *Commission* (see Article 52).

34.4 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43) and the Agreement or participation of the beneficiary may be terminated (see Article 50).

Such breaches may also lead to any of the other measures described in Chapter 6.

ARTICLE 35 — CONFLICT OF INTERESTS

35.1 Obligation to avoid a conflict of interests

The beneficiaries must take all measures to prevent any situation where the impartial and objective implementation of the action is compromised for reasons involving economic interest, political or national affinity, family or emotional ties or any other shared interest (‘**conflict of interests**’).

They must formally notify to the *Commission* without delay any situation constituting or likely to lead to a conflict of interests and immediately take all the necessary steps to rectify this situation.

The *Commission* may verify that the measures taken are appropriate and may require additional measures to be taken by a specified deadline.

35.2 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43) and the Agreement or participation of the beneficiary may be terminated (see Article 50).

Such breaches may also lead to any of the other measures described in Chapter 6.

ARTICLE 36 — CONFIDENTIALITY

36.1 General obligation to maintain confidentiality

During implementation of the action and for four years after the period set out in Article 3, the parties must keep confidential any data, documents or other material (in any form) that is identified as confidential at the time it is disclosed (‘**confidential information**’).

If a beneficiary requests, the *Commission* may agree to keep such information confidential for an additional period beyond the initial four years.

If information has been identified as confidential only orally, it will be considered to be confidential only if this is confirmed in writing within 15 days of the oral disclosure.

Unless otherwise agreed between the parties, they may use confidential information only to implement the Agreement.

The beneficiaries may disclose confidential information to their personnel or third parties involved in the action only if they:

(a) need to know to implement the Agreement and

(b) are bound by an obligation of confidentiality.

This does not change the security obligations in Article 37, which still apply.

The *Commission* may disclose confidential information to its staff, other EU institutions and bodies or third parties, if:

(a) this is necessary to implement the Agreement or safeguard the EU's financial interests and

(b) the recipients of the information are bound by an obligation of confidentiality.

Under the conditions set out in Article 4 of the Rules for Participation Regulation No 1290/2013[24], the Commission must moreover make available information on the results to other EU institutions, bodies, offices or agencies as well as Member States or associated countries.

The confidentiality obligations no longer apply if:

(a) the disclosing party agrees to release the other party;

(b) the information was already known by the recipient or is given to him without obligation of confidentiality by a third party that was not bound by any obligation of confidentiality;

(c) the recipient proves that the information was developed without the use of confidential information;

(d) the information becomes generally and publicly available, without breaching any confidentiality obligation, or

(e) the disclosure of the information is required by EU or national law.

36.2 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

[24] Regulation (EU) No 1290/2013 of the European Parliament and of the Council of 11 December 2013 laying down the rules for participation and dissemination in "Horizon 2020 - the Framework Programme for Research and Innovation (2014-2020)" (OJ L 347, 20.12.2013 p.81).

ARTICLE 37 — SECURITY-RELATED OBLIGATIONS

37.1 Results with a security recommendation

Not applicable

37.2 Classified results

Not applicable

37.3 Activities involving dual-use goods or dangerous materials and substances

Not applicable

37.4 Consequences of non-compliance

Not applicable

ARTICLE 38 — PROMOTING THE ACTION — VISIBILITY OF EU FUNDING

38.1 Communication activities by beneficiaries

38.1.1 Obligation to promote the action and its results

The beneficiaries must promote the action and its results, by providing targeted information to multiple audiences (including the media and the public) in a strategic and effective manner.

This does not change the dissemination obligations in Article 29, the confidentiality obligations in Article 36 or the security obligations in Article 37, all of which still apply.

Before engaging in a communication activity expected to have a major media impact, the beneficiaries must inform the *Commission* (see Article 52).

38.1.2 Information on EU funding — Obligation and right to use the EU emblem

Unless the *Commission* requests or agrees otherwise or unless it is impossible, any communication activity related to the action (including in electronic form, via social media, etc.) and any infrastructure, equipment and major results funded by the grant must:

(a) display the EU emblem and

(b) include the following text:

> For communication activities: "This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 654024".
>
> For infrastructure, equipment and major results: "This *[infrastructure][equipment][insert type of result]* is part of a project that has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 654024".

When displayed together with another logo, the EU emblem must have appropriate prominence.

For the purposes of their obligations under this Article, the beneficiaries may use the EU emblem without first obtaining approval from the *Commission*.

This does not, however, give them the right to exclusive use.

Moreover, they may not appropriate the EU emblem or any similar trademark or logo, either by registration or by any other means.

38.1.3 Disclaimer excluding the *Commission* responsibility

Any communication activity related to the action must indicate that it reflects only the author's view and that the *Commission* is not responsible for any use that may be made of the information it contains.

38.2 Communication activities by the *Commission*

38.2.1 Right to use beneficiaries' materials, documents or information

The *Commission* may use, for its communication and publicising activities, information relating to the action, documents notably summaries for publication and public deliverables as well as any other material, such as pictures or audio-visual material that it receives from any beneficiary (including in electronic form).

This does not change the confidentiality obligations in Article 36 and the security obligations in Article 37, all of which still apply.

However, if the *Commission's* use of these materials, documents or information would risk compromising legitimate interests, the beneficiary concerned may request the *Commission* not to use it (see Article 52).

The right to use a beneficiary's materials, documents and information includes:

(a) **use for its own purposes** (in particular, making them available to persons working for the *Commission* or any other EU institution, body, office or agency or body or institutions in EU Member States; and copying or reproducing them in whole or in part, in unlimited numbers);

(b) **distribution to the public** (in particular, publication as hard copies and in electronic or digital format, publication on the internet, as a downloadable or non-downloadable file, broadcasting by any channel, public display or presentation, communicating through press information services, or inclusion in widely accessible databases or indexes);

(c) **editing or redrafting** for communication and publicising activities (including shortening, summarising, inserting other elements (such as meta-data, legends, other graphic, visual, audio or text elements), extracting parts (e.g. audio or video files), dividing into parts, use in a compilation);

(d) **translation**;

(e) giving **access in response to individual requests** under Regulation No 1049/2001[25], without the right to reproduce or exploit;

[25] Regulation (EC) No 1049/2001 of the European Parliament and of the Council of 30 May 2001 regarding public access to European Parliament, Council and Commission documents, OJ L 145, 31.5.2001, p. 43.

(f) **storage** in paper, electronic or other form;

(g) **archiving**, in line with applicable document-management rules, and

(h) the right to authorise **third parties** to act on its behalf or sub-license the modes of use set out in Points (b),(c),(d) and (f) to third parties if needed for the communication and publicising activities of the *Commission*.

If the right of use is subject to rights of a third party (including personnel of the beneficiary), the beneficiary must ensure that it complies with its obligations under this Agreement (in particular, by obtaining the necessary approval from the third parties concerned).

Where applicable (and if provided by the beneficiaries), the *Commission* will insert the following information:

> "© – [year] – [name of the copyright owner]. All rights reserved. Licensed to the *European Union (EU)* under conditions."

38.3 Consequences of non-compliance

If a beneficiary breaches any of its obligations under this Article, the grant may be reduced (see Article 43).

Such breaches may also lead to any of the other measures described in Chapter 6.

ARTICLE 39 — PROCESSING OF PERSONAL DATA

39.1 Processing of personal data by the Commission

Any personal data under the Agreement will be processed by the Commission under Regulation No 45/2001[26] and according to the 'notifications of the processing operations' to the Data Protection Officer (DPO) of the Commission (publicly accessible in the DPO register).

Such data will be processed by the '**data controller**' of the Commission for the purposes of implementing, managing and monitoring the Agreement or protecting the financial interests of the EU or Euratom (including checks, reviews, audits and investigations; see Article 22).

The persons whose personal data are processed have the right to access and correct their own personal data. For this purpose, they must send any queries about the processing of their personal data to the data controller, via the contact point indicated in the 'service specific privacy statement(s) (SSPS)' that are published on the Commission websites.

They also have the right to have recourse at any time to the European Data Protection Supervisor (EDPS).

[26] Regulation (EC) No 45/2001 of the European Parliament and of the Council of 18 December 2000 on the protection of individuals with regard to the processing of personal data by the Community institutions and bodies and on the free movement of such data (OJ L 8, 12.01.2001, p. 1).

39.2 Processing of personal data by the beneficiaries

The beneficiaries must process personal data under the Agreement in compliance with applicable EU and national law on data protection (including authorisations or notification requirements).

The beneficiaries may grant their personnel access only to data that is strictly necessary for implementing, managing and monitoring the Agreement.

The beneficiaries must inform the personnel whose personal data are collected and processed by the Commission . For this purpose, they must provide them with the service specific privacy statement (SSPS) (see above), before transmitting their data to the Commission .

39.3 Consequences of non-compliance

If a beneficiary breaches any of its obligations under Article 39.2, the *Commission* may apply any of the measures described in Chapter 6.

ARTICLE 40 — ASSIGNMENTS OF CLAIMS FOR PAYMENT AGAINST THE *COMMISSION*

The beneficiaries may not assign any of their claims for payment against the *Commission* to any third party, except if approved by the *Commission* on the basis of a reasoned, written request by the coordinator (on behalf of the beneficiary concerned).

If the *Commission* has not accepted the assignment or the terms of it are not observed, the assignment will have no effect on it.

In no circumstances will an assignment release the beneficiaries from their obligations towards the *Commission*.

CHAPTER 5 DIVISION OF BENEFICIARIES' ROLES AND RESPONSIBILITIES

ARTICLE 41 — DIVISION OF BENEFICIARIES' ROLES AND RESPONSIBILITIES

41.1 Roles and responsibilities towards the *Commission*

The beneficiaries have full responsibility for implementing the action and complying with the Agreement.

The beneficiaries are jointly and severally liable for the **technical implementation** of the action as described in Annex 1. If a beneficiary fails to implement its part of the action, the other beneficiaries become responsible for implementing this part (without being entitled to any additional EU funding for doing so), unless the *Commission* expressly relieves them of this obligation.

The **financial responsibility** of each beneficiary is governed by Articles 44, 45 and 46.

41.2 Internal division of roles and responsibilities

The internal roles and responsibilities of the beneficiaries are divided as follows:

(a) Each **beneficiary** must:

(i) keep information stored in the 'Beneficiary Register' (via the electronic exchange system) up to date (see Article 17);

(ii) inform the coordinator immediately of any events or circumstances likely to affect significantly or delay the implementation of the action (see Article 17);

(iii) submit to the coordinator in good time:

- individual financial statements for itself and, if required, certificates on the financial statements (see Article 20);
- the data needed to draw up the technical reports (see Article 20);
- ethics committee opinions and notifications or authorisations for activities raising ethical issues (see Article 34);
- any other documents or information required by the Commission under the Agreement, unless the Agreement requires the beneficiary to submit this information directly to the Commission.

(b) The **coordinator** must:

(i) monitor that the action is implemented properly (see Article 7);

(ii) act as the intermediary for all communications between the beneficiaries and the *Commission* (in particular, providing the *Commission* with the information described in Article 17), unless the Agreement specifies otherwise;

(iii) request and review any documents or information required by the *Commission* and verify their completeness and correctness before passing them on to the *Commission*;

(iv) submit the deliverables and reports to the *Commission* (see Articles 19 and 20);

(v) ensure that all payments are made to the other beneficiaries without unjustified delay (see Article 21);

(vi) inform the *Commission* of the amounts paid to each beneficiary, when required under the Agreement (see Articles 44 and 50) or requested by the *Commission*.

The coordinator may not delegate the above-mentioned tasks to any other beneficiary or subcontract them to any third party.

41.3 Internal arrangements between beneficiaries — Consortium agreement

*The beneficiaries must have internal arrangements regarding their operation and co-ordination to ensure that the action is implemented properly. These internal arrangements must be set out in a written '**consortium agreement**' between the beneficiaries, which may cover:*

- *internal organisation of the consortium;*

- *management of access to the electronic exchange system;*

- *distribution of EU funding;*

- *additional rules on rights and obligations related to background and results (including whether access rights remain or not, if a beneficiary is in breach of its obligations) (see Section 3 of Chapter 4);*

- *settlement of internal disputes;*

- *liability, indemnification and confidentiality arrangements between the beneficiaries.*

The consortium agreement must not contain any provision contrary to the Agreement.

41.4 Relationship with complementary beneficiaries — Collaboration agreement

Not applicable

41.5 Relationship with partners of a joint action — Coordination agreement

Not applicable

CHAPTER 6 REJECTION OF COSTS — REDUCTION OF THE GRANT — RECOVERY — PENALTIES — DAMAGES — SUSPENSION — TERMINATION — FORCE MAJEURE

SECTION 1 REJECTION OF COSTS — REDUCTION OF THE GRANT — RECOVERY — PENALTIES

ARTICLE 42 — REJECTION OF INELIGIBLE COSTS

42.1 Conditions

42.1.1 The *Commission* will — at the time of an **interim payment, at the payment of the balance** or **afterwards** — reject any costs which are ineligible (see Article 6), in particular following checks, reviews, audits or investigations (see Article 22).

42.1.2 The rejection may also be based on the **extension of findings from other grants to this grant**, under the conditions set out in Article 22.5.2.

42.2 Ineligible costs to be rejected — Calculation — Procedure

Ineligible costs will be rejected in full.

If the *Commission* rejects costs **without reduction of the grant** (see Article 43) or **recovery of undue amounts** (see Article 44), it will formally notify the coordinator or beneficiary concerned the rejection of costs, the amounts and the reasons why (if applicable, together with the notification of amounts due; see Article 21.5). The coordinator or beneficiary concerned may — within 30 days of receiving notification — formally notify the *Commission* of its disagreement and the reasons why.

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

If the *Commission* rejects costs **with reduction of the grant** or **recovery of undue amounts** , it will formally notify the rejection in the '**pre-information letter**' on reduction or recovery set out in Articles 43 and 44.

42.3 Effects

If the *Commission* rejects costs at the time of an **interim payment** or **the payment of the balance**, it will deduct them from the total eligible costs declared, for the action, in the periodic or final summary financial statement (see Articles 20.3 and 20.4). It will then calculate the interim payment or payment of the balance as set out in Articles 21.3 or 21.4.

If the *Commission* — **after an interim payment but before the payment of the balance** — rejects costs declared in a periodic summary financial statement, it will deduct them from the total eligible costs declared, for the action, in the next periodic summary financial statement or in the final summary financial statement. It will then calculate the interim payment or payment of the balance as set out in Articles 21.3 or 21.4.

If the *Commission* rejects costs **after the payment of the balance**, it will deduct the amount rejected from the total eligible costs declared, by the beneficiary, in the final summary financial statement. It will then calculate the revised final grant amount as set out in Article 5.4.

ARTICLE 43 — REDUCTION OF THE GRANT

43.1 Conditions

43.1.1 The *Commission* may — **at the payment of the balance** or **afterwards** — reduce the maximum grant amount (see Article 5.1), if the action has not been implemented properly as described in Annex 1 or another obligation under the Agreement has been breached.

43.1.2 The *Commission* may also reduce the maximum grant amount on the basis of the **extension of findings from other grants to this grant**, under the conditions set out in Article 22.5.2.

43.2 Amount to be reduced — Calculation — Procedure

The amount of the reduction will be proportionate to the improper implementation of the action or to the seriousness of the breach.

Before reduction of the grant, the *Commission* will formally notify a '**pre-information letter**' to the coordinator or beneficiary concerned:

- informing it of its intention to reduce the grant, the amount it intends to reduce and the reasons why and

- inviting it to submit observations within 30 days of receiving notification

If the *Commission* does not receive any observations or decides to pursue reduction despite the observations it has received, it will formally notify **confirmation** of the reduction (if applicable, together with the notification of amounts due; see Article 21).

43.3 Effects

If the *Commission* reduces the grant at the time of **the payment of the balance**, it will calculate the reduced grant amount for the action and then determine the amount due as payment of the balance (see Articles 5.3.4 and 21.4).

If the *Commission* reduces the grant **after the payment of the balance**, it will calculate the revised final grant amount for the beneficiary concerned (see Article 5.4). If the revised final grant amount for the beneficiary concerned is lower than its share of the final grant amount, the *Commission* will recover the difference (see Article 44).

ARTICLE 44 — RECOVERY OF UNDUE AMOUNTS

44.1 Amount to be recovered — Calculation — Procedure

The *Commission* will — after **termination of the participation of a beneficiary, at the payment of the balance** or **afterwards** — claim back any amount that was paid but is not due under the Agreement.

Each beneficiary's financial responsibility in case of recovery is limited to its own debt, except for the amount retained for the Guarantee Fund (see Article 21.4).

44.1.1 Recovery after termination of a beneficiary's participation

If recovery takes place after termination of a beneficiary's participation (including the coordinator), the *Commission* will claim back the undue amount from the beneficiary concerned, by formally notifying it a debit note (see Article 50.2 and 50.3). This note will specify the amount to be recovered, the terms and the date for payment.

If payment is not made by the date specified in the debit note, the Commission will **recover** the amount:

(a) by '**offsetting**' it — without the beneficiary's consent — against any amounts owed to the beneficiary concerned by the Commission or an executive agency (from the EU or Euratom budget).

In exceptional circumstances, to safeguard the EU's financial interests, the *Commission* may offset before the payment date specified in the debit note;

(b) *not applicable;*

(c) by **taking legal action** (see Article 57) or by **adopting an enforceable decision** under Article 299 of the Treaty on the Functioning of the EU (TFEU) and Article 79(2) of the Financial regulation No 966/2012.

If payment is not made by the date specified in the debit note, the amount to be recovered (see above) will be increased by **late-payment interest** at the rate set out in Article 21.11, from the day following the payment date in the debit note, up to and including the date the Commission receives full payment of the amount.

Partial payments will be first credited against expenses, charges and late-payment interest and then against the principal.

Bank charges incurred in the recovery process will be borne by the beneficiary, unless Directive 2007/64/EC[27] applies.

44.1.2 Recovery at payment of the balance

If the payment of the balance takes the form of a recovery (see Article 21.4), the *Commission* will formally notify a '**pre-information letter**' to the coordinator:

- informing it of its intention to recover, the amount due as the balance and the reasons why;
- specifying that it intends to deduct the amount to be recovered from the amount retained for the Guarantee Fund;
- requesting the coordinator to submit a report on the distribution of payments to the beneficiaries within 30 days of receiving notification, and
- inviting the coordinator to submit observations within 30 days of receiving notification.

If no observations are submitted or the *Commission* decides to pursue recovery despite the observations it has received, it will **confirm recovery** (together with the notification of amounts due; see Article 21.5) and:

- pay the difference between the amount to be recovered and the amount retained for the Guarantee Fund, **if the difference is positive** or
- formally notify to the coordinator a **debit note** for the difference between the amount to be recovered and the amount retained for the Guarantee Fund, **if the difference is negative**. This note will also specify the terms and the date for payment.

If the coordinator does not repay the *Commission* by the date in the debit note and has not submitted the report on the distribution of payments: the Commission will **recover** the amount set out in the debit note from the coordinator (see below).

If the coordinator does not repay the *Commission* by the date in the debit note, but has submitted the report on the distribution of payments: the *Commission* will:

(a) identify the beneficiaries for which the amount calculated as follows is negative:

{{{{beneficiary's costs declared in the final summary financial statement and approved by the *Commission* multiplied by the reimbursement rate set out in Article 5.2 for the beneficiary concerned}

divided by

the EU contribution for the action calculated according to Article 5.3.1}

[27] Directive 2007/64/EC of the European Parliament and of the Council of 13 November 2007 on payment services in the internal market amending Directives 97/7/EC, 2002/65/EC, 2005/60/EC and 2006/48/EC and repealing Directive 97/5/EC (OJ L 319, 05.12.2007, p. 1).

multiplied by

the final grant amount (see Article 5.3)}**}**,

minus

{pre-financing and interim payments received by the beneficiary}**}**.

(b) formally notify to each beneficiary identified according to point (a) a **debit note** specifying the terms and date for payment. The amount of the debit note is calculated as follows:

{{amount calculated according to point (a) for the beneficiary concerned

divided by

the sum of the amounts calculated according to point (a) for all the beneficiaries identified according to point (a)}

multiplied by

the amount set out in the debit note formally notified to the coordinator}.

If payment is not made by the date specified in the debit note, the *Commission* will **recover** the amount:

(a) by '**offsetting**' it — without the beneficiary's consent — against any amounts owed to the beneficiary concerned by the Commission or an executive agency (from the EU or Euratom budget).

In exceptional circumstances, to safeguard the EU's financial interests, the *Commission* may offset before the payment date specified in the debit note;

(b) by **drawing on the Guarantee Fund**. The Commission will formally notify the beneficiary concerned the debit note on behalf of the Guarantee Fund and recover the amount:

(i) *not applicable;*

(ii) by **taking legal action** (see Article 57) or by **adopting an enforceable decision** under Article 299 of the Treaty on the Functioning of the EU (TFEU) and Article 79(2) of the Financial Regulation No 966/2012.

If payment is not made by the date in the debit note, the amount to be recovered (see above) will be increased by **late-payment interest** at the rate set out in Article 21.11, from the day following the payment date in the debit note, up to and including the date the Commission receives full payment of the amount.

Partial payments will be first credited against expenses, charges and late-payment interest and then against the principal.

Bank charges incurred in the recovery process will be borne by the beneficiary, unless Directive 2007/64/EC applies.

44.1.3 Recovery of amounts after payment of the balance

If, for a beneficiary, the revised final grant amount (see Article 5.4) is lower than its share of the final grant amount, it must repay the difference to the *Commission*.

The beneficiary's share of the final grant amount is calculated as follows:

{{{beneficiary's costs declared in the final summary financial statement and approved by the *Commission* multiplied by the reimbursement rate set out in Article 5.2 for the beneficiary concerned}

divided by

the EU contribution for the action calculated according to Article 5.3.1}

multiplied by

the final grant amount (see Article 5.3)}.

If the coordinator has not distributed amounts received (see Article 21.7), the *Commission* will also recover these amounts.

The *Commission* will formally notify a **pre-information letter** to the beneficiary concerned:

- informing it of its intention to recover, the due amount and the reasons why and
- inviting it to submit observations within 30 days of receiving notification.

If no observations are submitted or the *Commission* decides to pursue recovery despite the observations it has received, it will **confirm** the amount to be recovered and formally notify to the beneficiary concerned a **debit note**. This note will also specify the terms and the date for payment.

If payment is not made by the date specified in the debit note, the *Commission* will **recover** the amount:

(a) by '**offsetting**' it — without the beneficiary's consent — against any amounts owed to the beneficiary concerned by the Commission or an executive agency (from the EU or Euratom budget).

In exceptional circumstances, to safeguard the EU's financial interests, the *Commission* may offset before the payment date specified in the debit note;

(b) by **drawing on the Guarantee Fund**. The Commission will formally notify the beneficiary concerned the debit note on behalf of the Guarantee Fund and recover the amount:

(i) *not applicable;*

(ii) by **taking legal action** (see Article 57) or by **adopting an enforceable decision** under Article 299 of the Treaty on the Functioning of the EU (TFEU) and Article 79(2) of the Financial Regulation No 966/2012.

If payment is not made by the date in the debit note, the amount to be recovered (see above) will be increased by **late-payment interest** at the rate set out in Article 21.11, from the day following the

date for payment in the debit note, up to and including the date the Commission receives full payment of the amount.

Partial payments will be first credited against expenses, charges and late-payment interest and then against the principal.

Bank charges incurred in the recovery process will be borne by the beneficiary, unless Directive 2007/64/EC applies.

ARTICLE 45 — ADMINISTRATIVE AND FINANCIAL PENALTIES

45.1 Conditions

Under Articles 109 and 131(4) of the Financial Regulation No 966/2012, the *Commission* may impose **administrative** and **financial penalties** if a beneficiary:

(a) has committed substantial errors, irregularities or fraud or is in serious breach of its obligations under the Agreement or

(b) has made false declarations about information required under the Agreement or for the submission of the proposal (or has not supplied such information).

Each beneficiary is responsible for paying the financial penalties imposed on it.

Under Article 109(3) of the Financial Regulation No 966/2012, the Commission may — under certain conditions and limits — publish decisions imposing administrative or financial penalties.

45.2 Duration — Amount of penalty — Calculation

Administrative penalties exclude the beneficiary from all contracts and grants financed from the EU or Euratom budget for a maximum of five years from the date the infringement is established by the *Commission*.

If the beneficiary commits another infringement within five years of the date the first infringement is established, the *Commission* may extend the exclusion period up to 10 years.

Financial penalties will be between 2% and 10% of the maximum EU contribution indicated, for the beneficiary concerned, in the estimated budget (see Annex 2).

If the beneficiary commits another infringement within five years of the date the first infringement is established, the *Commission* may increase the rate of financial penalties to between 4% and 20%.

45.3 Procedure

Before applying a penalty, the *Commission* will formally notify the beneficiary concerned:

- informing it of its intention to impose a penalty, its duration or amount and the reasons why and
- inviting it to submit observations within 30 days.

If the *Commission* does not receive any observations or decides to impose the penalty despite of observations it has received, it will formally notify **confirmation** of the penalty to the beneficiary

concerned and — in case of financial penalties — deduct the penalty from the payment of the balance or formally notify a **debit note**, specifying the amount to be recovered, the terms and the date for payment.

If payment is not made by the date specified in the debit note, the Commission may **recover** the amount:

(a) by '**offsetting**' it — without the beneficiary's consent — against any amounts owed to the beneficiary concerned by the Commission or an executive agency (from the EU or Euratom budget).

In exceptional circumstances, to safeguard the EU's financial interests, the *Commission* may offset before the payment date specified in the debit note;

(b) by **taking legal action** (see Article 57) or by **adopting an enforceable decision** under Article 299 of the Treaty on the Functioning of the EU (TFEU) and Article 79(2) of the Financial Regulation No 966/2012.

If payment is not made by the date in the debit note, the amount to be recovered (see above) will be increased by **late-payment interest** at the rate set out in Article 21.11, from the day following the payment date in the debit note, up to and including the date the Commission receives full payment of the amount.

Partial payments will be first credited against expenses, charges and late-payment interest and then against the principal.

Bank charges incurred in the recovery process will be borne by the beneficiary, unless Directive 2007/64/EC applies.

SECTION 2 LIABILITY FOR DAMAGES

ARTICLE 46 — LIABILITY FOR DAMAGES

46.1 Liability of the *Commission*

The *Commission* cannot be held liable for any damage caused to the beneficiaries or to third parties as a consequence of implementing the Agreement, including for gross negligence.

The *Commission* cannot be held liable for any damage caused by any of the beneficiaries or third parties involved in the action, as a consequence of implementing the Agreement.

46.2 Liability of the beneficiaries

46.2.1 Conditions

Except in case of force majeure (see Article 51), the beneficiaries must compensate the *Commission* for any damage it sustains as a result of the implementation of the action or because the action was not implemented in full compliance with the Agreement.

Each beneficiary is responsible for paying the damages claimed from it.

46.2.2 Amount of damages - Calculation

The amount the *Commission* can claim from a beneficiary will correspond to the damage caused by that beneficiary.

46.2.3 Procedure

Before claiming damages, the *Commission* will formally notify the beneficiary concerned:

- informing it of its intention to claim damages, the amount and the reasons why and
- inviting it to submit observations within 30 days.

If the *Commission* does not receive any observations or decides to claim damages despite the observations it has received, it will formally notify **confirmation** of the claim for damages and a **debit note**, specifying the amount to be recovered, the terms and the date for payment.

If payment is not made by the date specified in the debit note, the Commission may **recover** the amount:

(a) by '**offsetting**' it — without the beneficiary's consent — against any amounts owed to the beneficiary concerned by the Commission or an executive agency (from the EU or Euratom budget).

In exceptional circumstances, to safeguard the EU's financial interests, the *Commission* may offset before the payment date specified in the debit note;

(b) by **taking legal action** (see Article 57) or by **adopting an enforceable decision** under Article 299 of the Treaty on the Functioning of the EU (TFEU) and Article 79(2) of the Financial Regulation No 966/2012.

If payment is not made by the date in the debit note, the amount to be recovered (see above) will be increased by **late-payment interest** at the rate set out in Article 21.11, from the day following the payment date in the debit note, up to and including the date the Commission receives full payment of the amount.

Partial payments will be first credited against expenses, charges and late-payment interest and then against the principal.

Bank charges incurred in the recovery process will be borne by the beneficiary, unless Directive 2007/64/EC applies.

SECTION 3 SUSPENSION AND TERMINATION

ARTICLE 47 — SUSPENSION OF PAYMENT DEADLINE

47.1 Conditions

The *Commission* may — at any moment — suspend the payment deadline (see Article 21.2 to 21.4) if a request for payment (see Article 20) cannot be approved because:

(a) it does not comply with the provisions of the Agreement (see Article 20);

(b) the technical reports or financial reports have not been submitted or are not complete or additional information is needed, or

(c) there is doubt about the eligibility of the costs declared in the financial statements and additional checks, reviews, audits or investigations are necessary.

47.2 Procedure

The *Commission* will formally notify the coordinator of the suspension and the reasons why.

The suspension will **take effect** the day notification is sent by the *Commission* (see Article 52).

If the conditions for suspending the payment deadline are no longer met, the suspension will be **lifted** — and the remaining period will resume.

If the suspension exceeds two months, the coordinator may request the *Commission* if the suspension will continue.

If the payment deadline has been suspended due to the non-compliance of the technical or financial reports (see Article 20) and the revised report or statement is not submitted or was submitted but is also rejected, the *Commission* may also terminate the Agreement or the participation of the beneficiary (see Article 50.3.1(l)).

ARTICLE 48 — SUSPENSION OF PAYMENTS

48.1 Conditions

The *Commission* may — at any moment — suspend, in whole or in part, the pre-financing payment and interim payments for one or more beneficiaries or the payment of the balance for all beneficiaries, if a beneficiary:

(a) has committed or is suspected of having committed substantial errors, irregularities, fraud or serious breach of obligations in the award procedure or under this Agreement or

(b) has committed — in other EU or Euratom grants awarded to it under similar conditions — systemic or recurrent errors, irregularities, fraud or serious breach of obligations that have a material impact on this grant (**extension of findings from other grants to this grant**; see Article 22.5.2).

48.2 Procedure

Before suspending payments, the *Commission* will formally notify the coordinator:

- informing it of its intention to suspend payments and the reasons why and

- inviting it to submit observations within 30 days of receiving notification.

If the *Commission* does not receive observations or decides to pursue the procedure despite the observations it has received, it will formally notify **confirmation** of the suspension. Otherwise, it will formally notify that the suspension procedure is not continued.

The suspension will **take effect** the day the confirmation notification is sent by the *Commission*.

If the conditions for resuming payments are met, the suspension will be **lifted**. The *Commission* will formally notify the coordinator.

During the suspension, the periodic report(s) (see Article 20.3) must not contain any individual financial statements from the beneficiary concerned. When the *Commission* resumes payments, the coordinator may include them in the next periodic report.

The beneficiaries may suspend implementation of the action (see Article 49.1) or terminate the Agreement or the participation of the beneficiary concerned (see Article 50.1 and 50.2).

ARTICLE 49 — SUSPENSION OF THE ACTION IMPLEMENTATION

49.1 Suspension of the action implementation, by the beneficiaries

49.1.1 Conditions

The beneficiaries may suspend implementation of the action or any part of it, if exceptional circumstances — in particular *force majeure* (see Article 51) — make implementation impossible or excessively difficult.

49.1.2 Procedure

The coordinator must immediately formally notify to the *Commission* the suspension (see Article 52), stating:

- the reasons why and
- the expected date of resumption.

The suspension will **take effect** the day this notification is received by the *Commission*.

Once circumstances allow for implementation to resume, the coordinator must immediately formally notify the *Commission* and request an **amendment** of the Agreement to set the date on which the action will be resumed, extend the duration of the action and make other changes necessary to adapt the action to the new situation (see Article 55) — unless the Agreement or the participation of a beneficiary has been terminated (see Article 50).

The suspension will be **lifted** with effect from the resumption date set out in the amendment. This date may be before the date on which the amendment enters into force.

Costs incurred during suspension of the action implementation are not eligible (see Article 6).

49.2 Suspension of the action implementation, by the *Commission*

49.2.1 Conditions

The *Commission* may suspend implementation of the action or any part of it:

(a) if a beneficiary has committed or is suspected of having committed substantial errors, irregularities, fraud or serious breach of obligations in the award procedure or under this Agreement;

(b) if a beneficiary has committed — in other EU or Euratom grants awarded to it under similar conditions — systemic or recurrent errors, irregularities, fraud or serious breach of obligations that have a material impact on this grant (**extension of findings from other grants to this grant**; see Article 22.5.2), or

(c) if the action is suspected of having lost its scientific or technological relevance.

49.2.2 Procedure

Before suspending implementation of the action, the *Commission* will formally notify the coordinator:

- informing it of its intention to suspend the implementation and the reasons why and

- inviting it to submit observations within 30 days of receiving notification.

If the *Commission* does not receive observations or decides to pursue the procedure despite the observations it has received, it will formally notify **confirmation** of the suspension. Otherwise, it will formally notify that the procedure is not continued.

The suspension will **take effect** five days after confirmation notification is received by the coordinator (or on a later date specified in the notification).

It will be **lifted** if the conditions for resuming implementation of the action are met.

The coordinator will be formally notified of the lifting and the Agreement will be **amended** to set the date on which the action will be resumed, extend the duration of the action and make other changes necessary to adapt the action to the new situation (see Article 55) — unless the Agreement has already been terminated (see Article 50).

The suspension will be lifted with effect from the resumption date set out in the amendment. This date may be before the date on which the amendment enters into force.

Costs incurred during suspension are not eligible (see Article 6).

The beneficiaries may not claim damages due to suspension by the *Commission* (see Article 46).

Suspension of the action implementation does not affect the *Commission's* right to terminate the Agreement or participation of a beneficiary (see Article 50), reduce the grant or recover amounts unduly paid (see Articles 43 and 44).

ARTICLE 50 — TERMINATION OF THE AGREEMENT OR OF THE PARTICIPATION OF ONE OR MORE BENEFICIARIES

50.1 Termination of the Agreement by the beneficiaries

50.1.1 Conditions and procedure

The beneficiaries may terminate the Agreement.

The coordinator must formally notify termination to the *Commission* (see Article 52), stating:

- the reasons why and

- the date the termination will take effect. This date must be after the notification.

If no reasons are given or if the *Commission* considers the reasons do not justify termination, the Agreement will be considered to have been '**terminated improperly**'.

The termination will **take effect** on the day specified in the notification.

50.1.2 Effects

The coordinator must — within 60 days from when termination takes effect — submit:

(i) a periodic report (for the open reporting period until termination; see Article 20.3) and

(ii) the final report (see Article 20.4).

If the *Commission* does not receive the reports within the deadline (see above), only costs which are included in an approved periodic report will be taken into account.

The *Commission* will **calculate** the final grant amount (see Article 5.3) and the balance (see Article 21.4) on the basis of the reports submitted. Only costs incurred until termination are eligible (see Article 6). Costs relating to contracts due for execution only after termination are not eligible.

Improper termination may lead to a reduction of the grant (see Article 43).

After termination, the beneficiaries' obligations (in particular Articles 20, 22, 23, Section 3 of Chapter 4, 36, 37, 38 and 40) continue to apply.

50.2 Termination of the participation of one or more beneficiaries, by the beneficiaries

50.2.1 Conditions and procedure

The participation of one or more beneficiaries may be terminated by the coordinator, on request of the beneficiary concerned or on behalf of the other beneficiaries.

The coordinator must formally notify termination to the *Commission* (see Article 52) and inform the beneficiary concerned.

If the coordinator's participation is terminated without its agreement, the formal notification must be done by another beneficiary (acting on behalf of the other beneficiaries).

The notification must include:

- the reasons why;

- the opinion of the beneficiary concerned (or proof that this opinion has been requested in writing);

- the date the termination takes effect. This date must be after the notification, and

- a request for amendment (see Article 55), with a proposal for reallocation of the tasks and the estimated budget of the beneficiary concerned (see Annexes 1 and 2) and, if necessary, the addition of one or more new beneficiaries (see Article 56). If termination takes effect after the period set out in Article 3, no request for amendment must be included unless the beneficiary concerned is the coordinator. In this case, the request for amendment must propose a new coordinator.

If this information is not given or if the *Commission* considers that the reasons do not justify termination, the participation will be considered to have been **terminated improperly**.

The termination will **take effect** on the day specified in the notification.

50.2.2 Effects

The coordinator must — within 30 days from when termination takes effect — submit:

(i) a report on the distribution of payments to the beneficiary concerned and

(ii) if termination takes effect during the period set out in Article 3, a '**termination report**' from the beneficiary concerned, for the open reporting period until termination, containing an overview of the progress of the work, an overview of the use of resources, the individual financial statement and, if applicable, the certificate on the financial statement (see Articles 20.3 and 20.4).

The information in the termination report must also be included in the periodic report for the next reporting period (see Article 20.3).

If the request for amendment is rejected by the *Commission*, (because it calls into question the decision awarding the grant or breaches the principle of equal treatment of applicants), the Agreement may be terminated according to Article 50.3.1(c).

If the request for amendment is accepted by the *Commission*, the Agreement is **amended** to introduce the necessary changes (see Article 55).

The *Commission* will **calculate** — on the basis of the periodic reports, the termination report and the report on the distribution of payments — if the (pre-financing and interim) payments received by the beneficiary concerned exceed the beneficiary's EU contribution (calculated by applying the reimbursement rate(s) to the eligible costs declared by the beneficiary and approved by the *Commission*). Only costs incurred by the beneficiary concerned until termination takes effect are eligible (see Article 6). Costs relating to contracts due for execution only after termination are not eligible.

- If the payments received **exceed the amounts due**:

  - if termination takes effect during the period set out in Article 3 and the request for amendment is accepted, the beneficiary concerned must repay to the coordinator the amount unduly received. The *Commission* will formally notify the amount unduly received and request the beneficiary concerned to repay it to the coordinator within 30 days of receiving notification. If it does not repay the coordinator, the *Commission* will draw upon the Guarantee Fund to pay the coordinator and then notify a **debit note** on behalf of the Guarantee Fund to the beneficiary concerned (see Article 44);

- in all other cases (in particular if termination takes effect after the period set out in Article 3), the *Commission* will formally notify a **debit note** to the beneficiary concerned. If payment is not made by the date in the debit note, the Guarantee Fund will pay to the *Commission* the amount due and the *Commission* will notify a debit note on behalf of the Guarantee Fund to the beneficiary concerned (see Article 44);

- if the beneficiary concerned is the former coordinator, it must repay the new coordinator according to the procedure above, unless:

    - termination is after an interim payment and

    - the former coordinator has not distributed amounts received as pre-financing or interim payments (see Article 21.7).

    In this case, the *Commission* will formally notify a **debit note** to the former coordinator. If payment is not made by the date in the debit note, the Guarantee Fund will pay to the *Commission* the amount due. The *Commission* will then pay the new coordinator and notify a debit note on behalf of the Guarantee Fund to the former coordinator (see Article 44).

• If the payments received **do not exceed the amounts due**: amounts owed to the beneficiary concerned will be included in the next interim or final payment.

If the *Commission* does not receive the termination report within the deadline (see above), only costs included in an approved periodic report will be taken into account.

If the *Commission* does not receive the report on the distribution of payments within the deadline (see above), it will consider that:

- the coordinator did not distribute any payment to the beneficiary concerned and that

- the beneficiary concerned must not repay any amount to the coordinator.

Improper termination may lead to a reduction of the grant (see Article 43) or termination of the Agreement (see Article 50).

After termination, the concerned beneficiary's obligations (in particular Articles 20, 22, 23, Section 3 of Chapter 4, 36, 37, 38 and 40) continue to apply.

50.3 Termination of the Agreement or the participation of one or more beneficiaries, by the *Commission*

50.3.1 Conditions

The *Commission* may terminate the Agreement or the participation of one or more beneficiaries, if:

(a) one or more beneficiaries do not accede to the Agreement (see Article 56);

(b) a change to their legal, financial, technical, organisational or ownership situation is likely to substantially affect or delay the implementation of the action or calls into question the decision to award the grant;

(c) following termination of participation for one or more beneficiaries (see above), the necessary changes to the Agreement would call into question the decision awarding the grant or breach the principle of equal treatment of applicants (see Article 55);

(d) implementation of the action is prevented by force majeure (see Article 51) or suspended by the coordinator (see Article 49.1) and either:

(i) resumption is impossible, or

(ii) the necessary changes to the Agreement would call into question the decision awarding the grant or breach the principle of equal treatment of applicants;

(e) a beneficiary is declared bankrupt, being wound up, having its affairs administered by the courts, has entered into an arrangement with creditors, has suspended business activities, or is subject to any other similar proceedings or procedures under national law;

(f) a beneficiary (or a natural person who has the power to represent or take decisions on its behalf) has been found guilty of professional misconduct, proven by any means;

(g) a beneficiary does not comply with the applicable national law on taxes and social security;

(h) the action has lost scientific or technological relevance;

(i) *not applicable;*

(j) *not applicable;*

(k) a beneficiary (or a natural person who has the power to represent or take decisions on its behalf) has committed fraud, corruption, or is involved in a criminal organisation, money laundering or any other illegal activity affecting the EU's financial interests;

(l) a beneficiary (or a natural person who has the power to represent or take decisions on its behalf) has — in the award procedure or under the Agreement — committed:

(i) substantial errors, irregularities, fraud or

(ii) serious breach of obligations, including improper implementation of the action, submission of false information, failure to provide required information, breach of ethical principles;

(m) a beneficiary has committed — in other EU or Euratom grants awarded to it under similar conditions — systemic or recurrent errors, irregularities, fraud or serious breach of obligations that have a material impact on this grant ('**extension of findings from other grants to this grant**').

50.3.2 Procedure

Before terminating the Agreement or participation of one or more beneficiaries, the *Commission* will formally notify the coordinator:

- informing it of its intention to terminate and the reasons why and

- inviting it, within 30 days of receiving notification, to submit observations and — in case of Point (l.ii) above — to inform the *Commission* of the measures to ensure compliance with the obligations under the Agreement.

If the *Commission* does not receive observations or decides to pursue the procedure despite the observations it has received, it will formally notify to the coordinator **confirmation** of the termination and the date it will take effect. Otherwise, it will formally notify that the procedure is not continued.

The termination will **take effect**:

- for terminations under Points (b), (c), (e), (g), (h), (j), and (l.ii) above: on the day specified in the notification of the confirmation (see above);

- for terminations under Points (a), (d), (f), (i), (k), (l.i) and (m) above: on the day after the notification of the confirmation is received by the coordinator.

50.3.3 Effects

(a) for **termination of the Agreement**:

The coordinator must — within 60 days from when termination takes effect — submit:

(i) a periodic report (for the last open reporting period until termination; see Article 20.3) and

(ii) a final report (see Article 20.4).

If the Agreement is terminated for breach of the obligation to submit the reports (see Articles 20.8 and 50.3.1(l)), the coordinator may not submit any reports after termination.

If the *Commission* does not receive the reports within the deadline (see above), only costs which are included in an approved periodic report will be taken into account.

The *Commission* will **calculate** the final grant amount (see Article 5.3) and the balance (see Article 21.4) on the basis of the reports submitted. Only costs incurred until termination takes effect are eligible (see Article 6). Costs relating to contracts due for execution only after termination are not eligible.

This does not affect the *Commission's* right to reduce the grant (see Article 43) or to impose administrative and financial penalties (Article 45).

The beneficiaries may not claim damages due to termination by the *Commission* (see Article 46).

After termination, the beneficiaries' obligations (in particular Articles 20, 22, 23, Section 3 of Chapter 4, 36, 37, 38 and 40) continue to apply.

(b) for **termination of the participation of one or more beneficiaries**:

The coordinator must — within 60 days from when termination takes effect — submit:

(i) a report on the distribution of payments to the beneficiary concerned;

(ii) a request for amendment (see Article 55), with a proposal for reallocation of the tasks and estimated budget of the beneficiary concerned (see Annexes 1 and 2) and, if necessary, the addition of one or more new beneficiaries (see Article 56). If termination is notified after the period set out in Article 3, no request for amendment must be submitted unless the beneficiary concerned is the coordinator. In this case the request for amendment must propose a new coordinator, and

(iii) if termination takes effect during the period set out in Article 3, a **termination report** from the beneficiary concerned, for the open reporting period until termination, containing an overview of the progress of the work, an overview of the use of resources, the individual financial statement and, if applicable, the certificate on the financial statement (see Article 20).

The information in the termination report must also be included in the periodic report for the next reporting period (see Article 20.3).

If the request for amendment is rejected by the *Commission* (because it calls into question the decision awarding the grant or breaches the principle of equal treatment of applicants), the Agreement may be terminated according to Article 50.3.1(c).

If the request for amendment is accepted by the *Commission*, the Agreement is **amended** to introduce the necessary changes (see Article 55).

The *Commission* will **calculate** — on the basis of the periodic reports, the termination report and the report on the distribution of payments — if the (pre-financing and interim) payments received by the beneficiary concerned exceed the beneficiary's EU contribution (calculated by applying the reimbursement rate(s) to the eligible costs declared by the beneficiary and approved by the *Commission*). Only costs incurred by the beneficiary concerned until termination takes effect are eligible (see Article 6). Costs relating to contracts due for execution only after termination are not eligible.

- If the payments received **exceed the amounts due**:

  - if termination takes effect during the period set out in Article 3 and the request for amendment is accepted, the beneficiary concerned must repay to the coordinator the amount unduly received. The *Commission* will formally notify the amount unduly received and request the beneficiary concerned to repay it to the coordinator within 30 days of receiving notification. If it does not repay the coordinator, the *Commission* will draw upon the Guarantee Fund to pay the coordinator and then notify a debit note on behalf of the Guarantee Fund to the beneficiary concerned (see Article 44);

  - in all other cases, in particular if termination takes effect after the period set out in Article 3, the *Commission* will formally notify a **debit note** to the beneficiary concerned. If payment is not made by the date in the debit note, the Guarantee Fund will pay to the *Commission* the amount due and the *Commission* will notify a debit note on behalf of the Guarantee Fund to the beneficiary concerned (see Article 44);

- if the beneficiary concerned is the former coordinator, it must repay the new coordinator the amount unduly received, unless:

  - termination takes effect after an interim payment and

  - the former coordinator has not distributed amounts received as pre-financing or interim payments (see Article 21.7)

  In this case, the *Commission* will formally notify a **debit note** to the former coordinator. If payment is not made by the date in the debit note, the Guarantee Fund will pay to the *Commission* the amount due. The *Commission* will then pay the new coordinator and notify a debit note on behalf of the Guarantee Fund to the former coordinator (see Article 44).

• If the payments received **do not exceed the amounts due**: amounts owed to the beneficiary concerned will be included in the next interim or final payment.

If the *Commission* does not receive the termination report within the deadline (see above), only costs included in an approved periodic report will be taken into account.

If the *Commission* does not receive the report on the distribution of payments within the deadline (see above), it will consider that:

- the coordinator did not distribute any payment to the beneficiary concerned, and that

- the beneficiary concerned must not repay any amount to the coordinator.

After termination, the concerned beneficiary's obligations (in particular Articles 20, 22, 23, Section 3 of Chapter 4, 36, 37, 38 and 40) continue to apply.

SECTION 4 FORCE MAJEURE

ARTICLE 51 — FORCE MAJEURE

'Force majeure' means any situation or event that:

- prevents either party from fulfilling their obligations under the Agreement,

- was unforeseeable, exceptional situation and beyond the parties' control,

- was not due to error or negligence on their part (or on the part of third parties involved in the action), and

- proves to be inevitable in spite of exercising all due diligence.

The following cannot be invoked as force majeure:

- any default of a service, defect in equipment or material or delays in making them available, unless they stem directly from a relevant case of force majeure,

- labour disputes or strikes, or

- financial difficulties.

Any situation constituting force majeure must be formally notified to the other party without delay, stating the nature, likely duration and foreseeable effects.

The parties must immediately take all the necessary steps to limit any damage due to force majeure and do their best to resume implementation of the action as soon as possible.

The party prevented by force majeure from fulfilling its obligations under the Agreement cannot be considered in breach of them.

CHAPTER 7 FINAL PROVISIONS

ARTICLE 52 — COMMUNICATION BETWEEN THE PARTIES

52.1 Form and means of communication

Communication under the Agreement (information, requests, submissions, 'formal notifications', etc.) must:

- be made in writing and

- bear the number of the Agreement.

Until the payment of the balance: all communication must be made through the electronic exchange system and using the forms and templates provided there.

After the payment of the balance: formal notifications must be made by registered post with proof of delivery ('formal notification on paper').

Communications in the electronic exchange system must be made by persons authorised according to the 'Terms and Conditions of Use of the electronic exchange system'. For naming the authorised persons, each beneficiary must have designated — before the signature of this Agreement — a 'Legal Entity Appointed Representative (LEAR)'. The role and tasks of the LEAR are stipulated in his/her appointment letter (see Terms and Conditions of Use of the electronic exchange system).

If the electronic exchange system is temporarily unavailable, instructions will be given on the Commission websites.

52.2 Date of communication

Communications are considered to have been made when they are sent by the sending party (i.e. on the date and time they are sent through the electronic exchange system).

Formal notifications through the **electronic** exchange system are considered to have been made when they are received by the receiving party (i.e. on the date and time of acceptance by the receiving party, as indicated by the time stamp). A formal notification that has not been accepted within 10 days after sending is considered to have been accepted.

Formal notifications **on paper** sent by **registered post** with proof of delivery (only after the payment of the balance) are considered to have been made on either:

- the delivery date registered by the postal service or
- the deadline for collection at the post office.

If the electronic exchange system is temporarily unavailable, the sending party cannot be considered in breach of its obligation to send a communication within a specified deadline.

52.3 Addresses for communication

The **electronic** exchange system must be accessed via the following URL:

https://ec.europa.eu/research/participants/portal/desktop/en/projects/

The *Commission* will formally notify the coordinator and beneficiaries in advance any changes to this URL.

Formal notifications on paper (only after the payment of the balance) addressed **to the *Commission*** must be sent to the following address:

European Commission
DIRECTORATE-GENERAL FOR RESEARCH & INNOVATION
Research infrastructure
B-1049 Brussels Belgium

Formal notifications on paper (only after the payment of the balance) addressed **to the beneficiaries** must be sent to their legal address as specified in the 'Beneficiary Register'.

ARTICLE 53 — INTERPRETATION OF THE AGREEMENT

53.1 Precedence of the Terms and Conditions over the Annexes

The provisions in the Terms and Conditions of the Agreement take precedence over its Annexes.

Annex 2 takes precedence over Annex 1.

53.2 Privileges and immunities

Not applicable

ARTICLE 54 — CALCULATION OF PERIODS, DATES AND DEADLINES

In accordance with Regulation No 1182/71[28], periods expressed in days, months or years are calculated from the moment the triggering event occurs.

The day during which that event occurs is not considered as falling within the period.

[28] Regulation (EEC, Euratom) No 1182/71 of the Council of 3 June 1971 determining the rules applicable to periods, dates and time-limits (OJ L 124, 8.6.1971, p. 1).

ARTICLE 55 — AMENDMENTS TO THE AGREEMENT

55.1 Conditions

The Agreement may be amended, unless the amendment entails changes to the Agreement which would call into question the decision awarding the grant or breach the principle of equal treatment of applicants.

Amendments may be requested by any of the parties.

55.2 Procedure

The party requesting an amendment must submit a request for amendment signed in the electronic exchange system (see Article 52).

The coordinator submits and receives requests for amendment on behalf of the beneficiaries (see Annex 3).

If a change of coordinator is requested without its agreement, the submission must be done by another beneficiary (acting on behalf of the other beneficiaries).

The request for amendment must include:

- the reasons why;
- the appropriate supporting documents;
- for a change of coordinator without its agreement: the opinion of the coordinator (or proof that this opinion has been requested in writing).

The *Commission* may request additional information.

If the party receiving the request agrees, it must sign the amendment in the electronic exchange system within 45 days of receiving notification (or any additional information the *Commission* has requested). If it does not agree, it must formally notify its disagreement within the same deadline. The deadline may be extended, if necessary for the assessment of the request. If no notification is received within the deadline, the request is considered to have been rejected

An amendment **enters into force** on the day of the signature of the receiving party.

An amendment **takes effect** on the date agreed by the parties or, in the absence of such an agreement, on the date on which the amendment enters into force.

ARTICLE 56 — ACCESSION TO THE AGREEMENT

56.1 Accession of the beneficiaries mentioned in the Preamble

The other beneficiaries must accede to the Agreement by signing the Accession Form (see Annex 3) in the electronic exchange system (see Article 52) within 30 days after its entry into force (see Article 58).

They will assume the rights and obligations under the Agreement with effect from the date of its entry into force (see Article 58).

If a beneficiary does not accede to the Agreement within the above deadline, the coordinator must — within 30 days — request an amendment to make any changes necessary to ensure proper implementation of the action. This does not affect the *Commission's* right to terminate the Agreement (see Article 50).

56.2 Addition of new beneficiaries

In justified cases, the beneficiaries may request the addition of a new beneficiary.

For this purpose, the coordinator must submit a request for amendment in accordance with Article 55. It must include an Accession Form (see Annex 3) signed by the new beneficiary in the electronic exchange system (see Article 52).

New beneficiaries must assume the rights and obligations under the Agreement with effect from the date of their accession specified in the Accession Form (see Annex 3).

ARTICLE 57 — APPLICABLE LAW AND SETTLEMENT OF DISPUTES

57.1 Applicable law

The Agreement is governed by the applicable EU law, supplemented if necessary by the law of Belgium.

57.2 Dispute settlement

If a dispute concerning the interpretation, application or validity of the Agreement cannot be settled amicably, the General Court — or, on appeal, the Court of Justice of the European Union — has sole jurisdiction. Such actions must be brought under Article 272 of the Treaty on the Functioning of the EU (TFEU).

As an exception, if such a dispute is between the Commission and EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH, the competent Belgian courts have sole jurisdiction.

If a dispute concerns administrative or financial penalties, offsetting or an enforceable decision under Article 299 TFEU (see Articles 44, 45 and 46), the beneficiaries must bring action before the General Court — or, on appeal, the Court of Justice of the European Union — under Article 263 TFEU.

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

ARTICLE 58 — ENTRY INTO FORCE OF THE AGREEMENT

The Agreement will enter into force on the day of signature by the *Commission* or the coordinator, depending on which is later.

SIGNATURES

For the coordinator

CLAUDIO MONTANI with ECAS id cmontani signed in the Participant Portal on 27/05/2015 at 08:36:48 (transaction id SigId-130-iDynBYAJdkhsHy90o4zmxzzK2BVHJWM4Li7B=9IrO4nuv8r6zXBYspxrw18O58WV4FMQuLcLbZ2JYIPYTp4-PHsUMVSXYCkKfUjX5YDKy-F8FzR96b=AA=hNDxGknBTQEcx8l8NFznO2TcurdF2LzO). Timestamp by third party at Wed May 27 09:36:55 CEST 2015

For the *Commission*

Pascale CID with ECAS id cidlupa signed in the Participant Portal on 25/05/2015 at 08:50:05 (transaction id SigId-113-xt5paDpzOzdevashFaJCRPzIIhx3nzPTsBG0WsZGDxmKGNh1CD-LPyAoDGuYd643HsKKS3Ad8bRHnE01w8-PHsUMVSXYC=Hy5EW6BzsR-9dwv0tf7tslrb93tlFtgTrsWczaZK0Y3X6XyEKV5NX0). Timestamp by third party at Thu Jun 25 09:50:10 CEST 2015

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

EUROPEAN COMMISSION

DIRECTORATE-GENERAL FOR RESEARCH & INNOVATION

Research infrastructure



ANNEX 1 (part A)

Research and Innovation action

NUMBER — 654024 — SoBigData

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

Table of Contents

1.1. The project summary....3
1.2. The list of beneficiaries....4
1.3. Workplan Tables - Detailed implementation....5
1.3.1. WT1 List of work packages....5
1.3.2. WT2 List of deliverables....6
1.3.3. WT3 Work package descriptions....10
Work package 1....10
Work package 2....12
Work package 3....16
Work package 4....20
Work package 5....25
Work package 6....28
Work package 7....35
Work package 8....41
Work package 9....45
Work package 10....50
Work package 11....55
1.3.4. WT4 List of milestones....59
1.3.5. WT5 Critical Implementation risks and mitigation actions....60
1.3.6 WT6 Summary of project effort in person-months....64
1.3.7. WT7 Tentative schedule of project reviews....65
1.3.8. WT8 Summary of transnational / virtual access provision per installation....66
1.4. Ethics Requirements....67

1.1. The project summary

Project Number [1]	654024	Project Acronym [2]	SoBigData

One form per project

General information

Project title [3]	SoBigData Research Infrastructure
Starting date [4]	01/09/2015
Duration in months [5]	48
Call (part) identifier [6]	H2020-INFRAIA-2014-2015
Topic	INFRAIA-1-2014-2015 Integrating and opening existing national and regional research infrastructures of European interest
Fixed EC Keywords	Knowledge infrastructure, ENGINEERING AND TECHNOLOGY
Free keywords	Big Data, Multidisciplinary European Infrastructure, Social Data Mining, Starting Community

Abstract [7]

One of the most pressing and fascinating challenges scientists face today, is understanding the complexity of our globally interconnected society. The big data arising from the digital breadcrumbs of human activities promise to let us scrutinize the ground truth of individual and collective behaviour at an unprecedented detail and scale. There is an urgent need to harness these opportunities for scientific advancement and for the social good. The main obstacle to this accomplishment, besides the scarcity of data scientists, is the lack of a large-scale open infrastructure, where big data and social mining research can be carried out.To this end, SoBigData proposes to create the Social Mining & Big Data Ecosystem: a research infrastructure (RI) providing an integrated ecosystem for ethic-sensitive scientific discoveries and advanced applications of social data mining on the various dimensions of social life, as recorded by “big data”. Building on several established national infrastructures, SoBigData will open up new research avenues in multiple research fields, including mathematics, ICT, and human, social and economic sciences, by enabling easy comparison, re-use and integration of state-of-the-art big social data, methods, and services, into new research. It will not only strengthen the existing clusters of excellence in social data mining research, but also create a pan-European, inter-disciplinary community of social data scientists, fostered by extensive training, networking, and innovation activities. In addition, as an open research infrastucture, SoBigData will promote repeatable and open science.Although SoBigData is primarily aimed at serving the needs of researchers, the openly available datasets and open source methods and services provided by the new research infrastructure will also impact industrial and other stakeholders (e.g. government bodies, non-profit organisations, funders, policy makers).

1.2. List of Beneficiaries

Project Number [1]	654024	Project Acronym [2]	SoBigData

List of Beneficiaries

No	Name	Short name	Country	Project entry month[8]	Project exit month
1	CONSIGLIO NAZIONALE DELLE RICERCHE	CNR	Italy	1	48
2	THE UNIVERSITY OF SHEFFIELD	USFD	United Kingdom	1	48
3	UNIVERSITA DI PISA	UNIPI	Italy	1	48
4	FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV	Fraunhofer	Germany	1	48
5	TARTU ULIKOOL	UT	Estonia	1	48
6	SCUOLA IMT (ISTITUZIONI, MERCATI, TECNOLOGIE) ALTI STUDI DI LUCCA	IMT LUCCA	Italy	1	48
7	GOTTFRIED WILHELM LEIBNIZ UNIVERSITAET HANNOVER	LUH	Germany	1	48
8	KING'S COLLEGE LONDON	KCL	United Kingdom	1	48
9	SCUOLA NORMALE SUPERIORE	SNS	Italy	1	48
10	AALTO-KORKEAKOULUSAATIO	AALTO UNIVERSITY	Finland	1	48
11	EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH	ETH Zurich	Switzerland	1	48
12	TECHNISCHE UNIVERSITEIT DELFT	TU Delft	Netherlands	1	48

1.3. Workplan Tables - Detailed implementation

1.3.1. WT1 List of work packages

WP Number [9]	WP Title	Lead beneficiary [10]	Person-months [11]	Start month [12]	End month [13]
WP1	Management	1 - CNR	38.00	1	48
WP2	NA1_Legal and Ethical Framework	12 - TU Delft	24.00	1	48
WP3	NA2_Dissemination, Impact, and Sustainability	2 - USFD	35.00	1	48
WP4	NA3_Training	8 - KCL	48.00	1	48
WP5	NA4_Accelerating Innovation	11 - ETH Zurich	32.00	1	48
WP6	TA1_Transnational Access	2 - USFD	49.00	13	48
WP7	VA1_Virtual Access	1 - CNR	28.00	13	48
WP8	JRA1_Big Data Ecosystem	7 - LUH	75.00	1	39
WP9	JRA2_Integrating Big Data Analytics Methods and Techniques	3 - UNIPI	119.00	1	39
WP10	JRA3_SoBigData e-Infrastructure	1 - CNR	77.00	1	39
WP11	NA5_Evaluation	4 - Fraunhofer	61.00	13	45
		Total	586.00		

1.3.2. WT2 list of deliverables

Deliverable Number [14]	Deliverable Title	WP number [9]	Lead beneficiary	Type [15]	Dissemination level [16]	Due Date (in months) [17]
D1.1	All Relevant Boards established	WP1	1 - CNR	Report	Public	3
D2.1	Ethics Board Establishment	WP2	12 - TU Delft	Report	Public	6
D2.2	Legal and Ethical Framework for SoBigData 1	WP2	7 - LUH	Report	Public	12
D2.3	Legal and Ethical Framework for SoBigData 2	WP2	7 - LUH	Report	Public	24
D2.4	Legal and Ethical Framework for SoBigData 3	WP2	7 - LUH	Report	Public	36
D2.5	Value-Sensitive Design & Privacy-by-Design technologies for big data analytics 1	WP2	12 - TU Delft	Report	Public	12
D2.6	Value-Sensitive Design & Privacy-by-Design technologies for big data analytics 2	WP2	12 - TU Delft	Report	Public	24
D2.7	IP principles and business models 1	WP2	12 - TU Delft	Report	Public	12
D2.8	IP principles and business models 2	WP2	12 - TU Delft	Report	Public	24
D2.9	IP principles and business models 3	WP2	12 - TU Delft	Report	Public	36
D3.1	Project web site, project presentation, and social media presence	WP3	1 - CNR	Demonstrator	Public	3
D3.2	Fact sheets aimed at different stakeholders	WP3	1 - CNR	Report	Public	3
D3.3	Initial dissemination and impact plan	WP3	2 - USFD	Report	Public	3
D3.4	Periodic dissemination and impact report and plan for following year 1	WP3	1 - CNR	Report	Public	18
D3.5	Periodic dissemination and impact report and plan for following year 2	WP3	1 - CNR	Report	Public	36

Deliverable Number [14]	Deliverable Title	WP number [9]	Lead beneficiary	Type [15]	Dissemination level [16]	Due Date (in months) [17]
D3.6	Final dissemination and community building events	WP3	10 - AALTO UNIVERSITY	Report	Public	48
D3.7	Sustainability Plan	WP3	1 - CNR	Report	Public	47
D4.1	Periodic Training Planning Reports 1	WP4	2 - USFD	Report	Public	18
D4.2	Periodic Training Planning Reports 2	WP4	2 - USFD	Report	Public	36
D4.3	Periodic Training Planning Reports 3	WP4	2 - USFD	Report	Public	48
D4.4	Data Scientist Training Materials 1	WP4	8 - KCL	Report	Public	20
D4.5	Data Scientist Training Materials 2	WP4	8 - KCL	Report	Public	34
D4.6	Data Scientist Training Materials 3	WP4	8 - KCL	Report	Public	46
D4.7	Training Programme 1	WP4	8 - KCL	Report	Public	6
D4.8	Training Programme 2	WP4	8 - KCL	Report	Public	24
D4.9	Training Programme 3	WP4	8 - KCL	Report	Public	36
D5.1	Preliminary Innovation and Action Plan	WP5	2 - USFD	Report	Public	9
D5.2	Final Innovation Report and Action Plan	WP5	2 - USFD	Report	Public	46
D5.3	Entrepreneurial skills materials and activities undertaken	WP5	3 - UNIPI	Report	Public	47
D5.4	Innovation Accelerator 1	WP5	11 - ETH Zurich	Demonstrator	Public	24
D5.5	Innovation Accelerator 2	WP5	11 - ETH Zurich	Demonstrator	Public	42
D6.1	Periodic reports on TA activities 1	WP6	2 - USFD	Report	Public	18
D6.2	Periodic reports on TA activities 2	WP6	2 - USFD	Report	Public	36
D6.3	Periodic reports on TA activities 3	WP6	2 - USFD	Report	Public	48
D7.1	VA e-Infrastructure Service Provision and Operation Report 1	WP7	1 - CNR	Report	Public	18

Deliverable Number [14]	Deliverable Title	WP number [9]	Lead beneficiary	Type [15]	Dissemination level [16]	Due Date (in months) [17]
D7.2	VA e-Infrastructure Service Provision and Operation Report 2	WP7	1 - CNR	Report	Public	36
D7.3	VA e-Infrastructure Service Provision and Operation Report 3	WP7	1 - CNR	Report	Public	48
D8.1	Data Management report	WP8	1 - CNR	Websites, patents filling, etc.	Public	3
D8.2	Crowd sensing platform	WP8	1 - CNR	Other	Public	12
D8.3	Analytical Crawling Platform	WP8	7 - LUH	Other	Public	24
D8.4	Living Archive Platform	WP8	11 - ETH Zurich	Other	Public	12
D9.1	Social mining method and service integration 1	WP9	10 - AALTO UNIVERSITY	Other	Public	12
D9.2	Social mining method and service integration 2	WP9	1 - CNR	Other	Public	24
D9.3	Social mining method and service integration 3	WP9	1 - CNR	Other	Public	36
D10.1	Best practices and guidelines towards interoperability	WP10	1 - CNR	Websites, patents filling, etc.	Public	6
D10.2	SoBigData e-Infrastructure release plan 1	WP10	1 - CNR	Report	Public	6
D10.3	SoBigData e-Infrastructure release plan 2	WP10	1 - CNR	Report	Public	20
D10.4	SoBigData e-Infrastructure release plan 3	WP10	1 - CNR	Report	Public	38
D10.5	SoBigData e-Infrastructure software release 1	WP10	1 - CNR	Demonstrator	Public	11
D10.6	SoBigData e-Infrastructure software release 2	WP10	1 - CNR	Demonstrator	Public	23

Deliverable Number [14]	Deliverable Title	WP number [9]	Lead beneficiary	Type [15]	Dissemination level [16]	Due Date (in months) [17]
D10.7	SoBigData e-Infrastructure software release 3	WP10	1 - CNR	Demonstrator	Public	39
D10.8	Resource adaptation to register to the e-infrastructure 1	WP10	1 - CNR	Other	Public	18
D10.9	Resource adaptation to register to the e-infrastructure 2	WP10	1 - CNR	Other	Public	36
D10.10	Resource adaptation to register to the e-infrastructure 3	WP10	1 - CNR	Other	Public	39
D10.11	Data processing workflow specification language	WP10	1 - CNR	Websites, patents filling, etc.	Public	12
D10.12	VRE specification and software 1	WP10	1 - CNR	Other	Public	12
D10.13	VRE specification and software 2	WP10	1 - CNR	Other	Public	36
D10.14	VRE specification and software 3	WP10	1 - CNR	Other	Public	39
D11.1	Evaluation Framework Design and Positioning	WP11	8 - KCL	Report	Public	24
D11.2	Evaluation Framework Toolkit and Datasets 1	WP11	2 - USFD	Other	Public	30
D11.3	Evaluation Framework Toolkit and Datasets 2	WP11	2 - USFD	Other	Public	45
D11.4	SoBigData Evaluation Framework: Engagement and Sustainability Report	WP11	11 - ETH Zurich	Report	Public	45

1.3.3. WT3 Work package descriptions

Work package number [9]	WP1	**Lead beneficiary** [10]	1 - CNR
Work package title	Management		
Start month	1	**End month**	48

Objectives

- Ensure administrative and financial management of the project;
- Develop a spirit of co-operation between the partners;
- Enable smooth progress of the work
- Ensure project reporting
- Enable an interface with the European Commission
- Manage the contract and assure compliance with the Commission's reporting requirements

Description of work and role of partners

WP1 - Management [Months: 1-48]
CNR, USFD, UNIPI, Fraunhofer, UT, IMT LUCCA, LUH, KCL, SNS, AALTO UNIVERSITY, ETH Zurich
WP1 globally aims at ensuring a strong and efficient day-to-day management so that the project meets its objectives on time and within budget constraints

T1.1 Project administration and financial management.
This task encompasses the essential aspects of leadership and project management required for the successful planning, undertaking and finalising of any project. These aspects are especially important given the project will deliver eleven (11) work packages, including many inter-dependent ones. The project involves seven (7) national infrastructures and a total of twelve (12) partners. CNR will manage the project. It will establish and maintain the central project management team and undertake the full range of managerial, administrative, supporting and co-ordinating activities necessary and appropriate to a project of this size.
The SoBigData coordinator Fosca Giannotti is an experienced infrastructure organiser and project coordinator and is directly supported by deputies i) Kalina Bontcheva from Sheffield who has over 15 years experience in running large-scale computational projects and one of the most established open source text mining infrastructures (GATE and GATECloud); ii) Prof. Dino Pedreschi of UNIPI. The rest of the management structure is detailed in Section 3.2 and is designed to support the efficient bottom-up management of the various parts in the infrastructure integrating activity.
Specific work carried out within the task:
- Establish a high-profile Scientific Advisory Board
- Manage the funds and transfer them to the Partners
- Dealing with all contractual matters
- Preparing and leading of project meetings and reviews (propose agendas, prepare minutes) - Provide efficient financial management and timely payment procedures
Task leader: CNR. Participants: ALL PARTNERS.

T1.2 Consortium and Project Management.
This task will ensure that all partners share the collective mission of the consortium and that all partners are integrated in the general decision and scientific processes. It will coordinate the overall project with respect to networking, scientific and dissemination issues to ensure highest quality of service and completing deliverables according to the time frame and foreseen budget. It will prepare all the meetings of the project's management bodies and control the progress of all project work bodies, ensuring that they keep to schedule. For the external reporting and communication, the task will gather and compile all reports as requested by the Commission and all other external bodies as required.
Specific meetings: the General Partner Meetings include all partners and controls the work carried out at strategic and content level. There will be six monthly consortium meetings, involving all beneficiaries and directly managed by the Coordinator. The advisory board and other external consultants as seem relevant will be invited periodically, on as needed basis. Aligned to the general partner meetings or via teleconferences, there will be also regular Project Management Board meetings and on a more frequent basis meetings of the Steering Board. Finally, joint working groups and workshops will bring experts together from the interested communities to discuss specific issues of common interest and to synchronize activities budgets (see WP3 for details).
Task leader: CNR. Participants: ALL PARTNERS.

Participation per Partner

Partner number and short name	WP1 effort
1 - CNR	19.00
2 - USFD	5.00
3 - UNIPI	1.00
4 - Fraunhofer	1.00
5 - UT	1.00
6 - IMT LUCCA	1.00
7 - LUH	1.00
8 - KCL	1.00
9 - SNS	1.00
10 - AALTO UNIVERSITY	1.00
11 - ETH Zurich	6.00
Total	38.00

List of deliverables

Deliverable Number [14]	Deliverable Title	Lead beneficiary	Type [15]	Dissemination level [16]	Due Date (in months) [17]
D1.1	All Relevant Boards established	1 - CNR	Report	Public	3

Description of deliverables

Definition of the composition and the installation of the boards for the management of the project. The periodical management and financial report as required by the European Commission, namely report 1, report 2 and the final management and financial plan.

D1.1 : All Relevant Boards established [3]

Definition of the composition and the installation of the boards for the management of the project.

Schedule of relevant Milestones

Milestone number [18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification

Work package number [9]	WP2	Lead beneficiary [10]	12 - TU Delft
Work package title	NA1_Legal and Ethical Framework		
Start month	1	End month	48

Objectives

SoBigData has a strong ethical motivation. It aims at building an ecosystem where protection of personal information and the respect for fundamental human rights can coexist with a safe use of the same information for scientific purposes of broad and central societal interest. Extreme care is taken in all parts of the project and among all consortium partners not only to guarantee compliance with relevant laws, regulations, and codes of conduct, but also to prevent unethical outcomes of the envisaged cutting edge research and applications, through ongoing analysis of the moral problems that pose constant challenges for law, regulation and governance of information technology. As such, SoBigData pursues the views that the EU is developing on Responsible Research and Innovation, and it operationalizes the values that drive the ongoing reform of the EU Data Protection legislation. It does so in two ways. Firstly, by establishing the framework that strengthens research concerning the protection of personal data as a fundamental right, while at the same time boosting the free flow of personal data as a common good. Secondly, by applying the moral and legal results and findings to its own operations and activities. Consequently, SoBigData will establish a dedicated ethical and legal board with experts in Ethics of IT and IT Law, who will:

• Define and implement the legal and ethical framework of the SoBigData research infrastructure, in accordance with the European and national legislations (including data protection and intellectual property rights) as they develop
• Monitor the compliance of experiments and research protocols with the framework,
• Act as a continuous consultant for the JRAs on the development of big data analytics and social mining tools with Value-Sensitive Design and privacy-by-design methodologies, and
• Investigate, design and promote novel architectures, protocols and procedures for the safe and fair use of big data for research purposes, in order to boost excellence and international competitiveness of Europe's big data research.

Description of work and role of partners

WP2 - NA1_Legal and Ethical Framework [Months: 1-48]
TU Delft, CNR, USFD, UNIPI, LUH, ETH Zurich
T2.1 Ethics Board
This task will establish and manage an Ethics Board with experts in IT law and IT ethics and external specialists in data protection and research ethics. The board guarantees the good governance of data and the research integrity and academic ethics, compliance with ethical and legal framework (T2.2) and consultation with external experts when and where required. The board will be chaired by J. van den Hoven, professor of ethics at Delft University of Technology, while N. Forgò, professor for legal informatics and IT law at LUH, will act as Vice Chair and also as the legal counsel to the board.
Task leader: TUDelft. Participants: LUH

T2.2 Ethical and legal framework
This task will define and implement the legal and ethical framework of the SoBigData research infrastructure, in accordance with the European and national legislations (including data protection and privacy). A key legal and ethical issue will be the question under which provisions and codes of conduct collected data sets can be (re)used for secondary research purposes. Another important issue will be the question how to manage the collection of new social media and human data sets and their use for the Research Infrastructure. Collecting and using social data for scientific analyses is generally a secondary purpose to the original purpose for which the data was provided, collected and known to be collected e.g., publishing a post on a social media platform. The processing must be allowed under the Data Protection Directive either by a valid informed consent or a legal exception, such as "processing for scientific purposes", e.g. § 40 German Federal Data Protection Act (BDSG), and stipulations of the Art. 29 Group on the processing of statistical data for scientific purposes, as long as the data subject's interests, in particular its right to privacy, does not outweigh the interests of the data controller.
Also the reform of the Data Protection Directive, the proposed Data Protection Regulation and its impact on the processing of personal data for scientific purposes will be considered. The legal and ethical framework will be revised after year 2 of the project, in order to cover necessary modifications in the development of the project or changed legislation.

The task also defines the context of moral values of trust, security, fairness, equality and inclusion, transparency and accountability, diversity and democracy, and ethical behaviour that inspire the SoBigData project, also in reference to relevant European and international legislation. Suitable frameworks to evaluate the activities within SoBigData, such as the framework of Contextual Integrity, will be incorporated.
The results of this task will be used to develop the legal and ethical framework for the use of the platform concerning access to, use and offer of social data sets and social mining services and tools.
Task leader: TUDelft. Participants: LUH, CNR, UNIPI

T2.3 Privacy-by-design and Value-Sensitive Design in social mining
Whilst the goals in themselves may be well-intentioned and morally justified, any technology that collects and analyzes information about humans implies risks for their privacy. Without privacy- and security-by-design technology, ethical processes and procedures, collecting data from devices, services, and data related to users, and analyzing their behaviour, attitudes, and social interactions could potentially lead to the harming, disadvantaging, discrimination or disrespect of individuals. In the wrong hands or under regimes of differential access, it could give rise to unfair advantages and could lead to social sorting, where some citizens and consumers are not offered certain options and services, of which they may not even be aware that they existed. This task will study the possible legal and ethical issues, such as privacy and discrimination, identifying effective legal and technical mechanisms to detect and measure privacy violations, discrimination and exclusion risks related to each analytical activity in the SoBigData infrastructure. This task will support in a consultancy role the JRAs, especially WP9, for value-sensitive design (VSD) and privacy-by-design analytics to ensure that legal and ethical aspects are actually and demonstrably respected by the technical procedures used in collection and use of data. Possibly, new policies and protocols to establish future standards and adequate individual protection within big data analytics shall be elaborated. The concepts of ‘anonymity’ and ‘personal data’ will be carefully considered. The data protection rules are not applicable when personal data are anonymized, but clearly in big data analytics it will not suffice to delete direct identifiers (such as name, address etc.) only. Using different data sets and social media information for Big Data analyses are known to be able to help re-identify individual users. The risk of re-identification must be considered when researchers get access to data sets, especially for the assessment whether the data subject’s interests can outweigh the interests of the researchers. SoBigData will not track the behaviour or actions of identifiable individuals. Its aim is to better understand the macroscopic and statistical interdependencies within the highly complex techno-social-economic-environmental system on which we all depend. So the individual as such is not an object of study. Quite to the contrary, SoBigData constitutes an environment for the development of the next generation of advanced privacy-by-design technology and promotes the development of privacy-respecting data mining technologies that give users control over their data. It will strongly engage in preventing and counteracting the misuse of data and the Internet.
SoBigData will therefore only analyze data of individuals who have consented to sharing their information after having been fully informed of the context in which they do so. Or data will be collected that are, as much as this is technically possible, for all practical purposes very hard to link to individual persons (de-identified or anonymized according to standard industry best-practices) and, therefore, do not deal with personal data. Researchers will be required to alert public authorities to the ethical and practical implications of the ICT research outcomes, as and when particular issues become apparent within the research process. The Board (T2.1) engages in constant monitoring and support of the ethical and legal analysis.
Task leader: TUDelft. Participants: LUH, CNR, UNIPI

T2.4 Intellectual property and data ownership framework
The proposed research infrastructure includes already existing data sets as well as new data sets, that will be collected for this project. Furthermore it is planned that researchers can upload their data sets to the platform in the future which can contain Social Media Content (like tweets, blogs, etc). Collecting, using and sharing of such social media content raises questions about informed consent (T2.1-T2.3), but also about intellectual property and data ownership. Within this task the legal requirements concerning Intellectual property law will be analyzed, starting with the question whether social media content can achieve Intellectual Property protection (“high level creativity”) and how such possible protected content can be shared in a (virtual) research environment for scientific purposes. But also the ownership of new data and personal information generated through the social media analyses shall be analyzed and a specific focus will be attributed to sui generis protection of databases. Standard accounts of intellectual property that accommodate the idea that public data can be appropriated by enriching, enhancing them, or somehow adding value to them by means of investment of owned resources, energy or assets (e.g. algorithms, software, models), could act contrary to some of the basic ideas of data protection and privacy principles.
Task leader: LUH. Participants: TUDelft, CNR, UNIPI

Participation per Partner

Partner number and short name	WP2 effort
1 - CNR	4.00
2 - USFD	3.00
3 - UNIPI	2.00
7 - LUH	3.00
11 - ETH Zurich	2.00
12 - TU Delft	10.00
Total	24.00

List of deliverables

Deliverable Number [14]	Deliverable Title	Lead beneficiary	Type [15]	Dissemination level [16]	Due Date (in months) [17]
D2.1	Ethics Board Establishment	12 - TU Delft	Report	Public	6
D2.2	Legal and Ethical Framework for SoBigData 1	7 - LUH	Report	Public	12
D2.3	Legal and Ethical Framework for SoBigData 2	7 - LUH	Report	Public	24
D2.4	Legal and Ethical Framework for SoBigData 3	7 - LUH	Report	Public	36
D2.5	Value-Sensitive Design & Privacy-by-Design technologies for big data analytics 1	12 - TU Delft	Report	Public	12
D2.6	Value-Sensitive Design & Privacy-by-Design technologies for big data analytics 2	12 - TU Delft	Report	Public	24
D2.7	IP principles and business models 1	12 - TU Delft	Report	Public	12
D2.8	IP principles and business models 2	12 - TU Delft	Report	Public	24
D2.9	IP principles and business models 3	12 - TU Delft	Report	Public	36

Description of deliverables

Reports describing the composition of the ethics board and its plan of activities, governance framework, by laws, responsibilities of members and principles, schedule of meetings, reporting and event calendar.

D2.1 : Ethics Board Establishment [6]

Report describing the composition of the ethics board and its plan of activities, governance framework, by laws, responsibilities of members and principles, schedule of meetings, reporting and event calendar.

D2.2 : Legal and Ethical Framework for SoBigData 1 [12]

Report describing the evolving versions of the legal and ethical framework of the SoBigData ecosystem, together with recommendations and provisions for research procedures.

D2.3 : Legal and Ethical Framework for SoBigData 2 [24]

Report describing the evolving versions of the legal and ethical framework of the SoBigData ecosystem, together with recommendations and provisions for research procedures.

D2.4 : Legal and Ethical Framework for SoBigData 3 [36]

Report describing the evolving versions of the legal and ethical framework of the SoBigData ecosystem, together with recommendations and provisions for research procedures.

D2.5 : Value-Sensitive Design & Privacy-by-Design technologies for big data analytics 1 [12]

Report describing the evolving versions of the VSD and PbD methods, recommendations and provisions to be adopted in the SoBigData ecosystem, which will drive the integration and enhancement of social mining services and tools of JRA2 into the research infrastructure.

D2.6 : Value-Sensitive Design & Privacy-by-Design technologies for big data analytics 2 [24]

Report describing the evolving versions of the VSD and PbD methods, recommendations and provisions to be adopted in the SoBigData ecosystem, which will drive the integration and enhancement of social mining services and tools of JRA2 into the research infrastructure.

D2.7 : IP principles and business models 1 [12]

Report describing responsible IP principles and practices in SoBigData and listing of best practices that have evolved in and outside SoBigData so as to accommodate both legitimate IP claims in Data and privacy related rights.

D2.8 : IP principles and business models 2 [24]

Report describing responsible IP principles and practices in SoBigData and listing of best practices that have evolved in and outside SoBigData so as to accommodate both legitimate IP claims in Data and privacy related rights.

D2.9 : IP principles and business models 3 [36]

Report describing responsible IP principles and practices in SoBigData and listing of best practices that have evolved in and outside SoBigData so as to accommodate both legitimate IP claims in Data and privacy related rights.

Schedule of relevant Milestones

Milestone number [18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS2	SoBigData eInfrastructure and training launch	11 - ETH Zurich	12	- Repository of training materials established - Innovation and first services Integration - Ethical issues addressed - 2nd SoBigData Workshop - e-Infrastructure Deployment - Resource catalogue VRE - Completion of first series of educational initiatives: Datathon 1, Training HS 1 - e-Infrastructure Production Release

Work package number [9]	WP3	Lead beneficiary [10]	2 - USFD
Work package title	NA2_Dissemination, Impact, and Sustainability		
Start month	1	End month	48

Objectives

- Develop and manage the communication channels of the project
- Take concrete steps towards establishing the SoBigData starting community
- Disseminate project results to relevant scientific communities
- Put in place a sustainability plan
- Engage a wide range of stakeholders in order to guarantee the long-term sustainability of the project results
- Coordinate with WP5 on innovation activities, in order to maximize project impact

Description of work and role of partners

WP3 - NA2_Dissemination, Impact, and Sustainability [Months: 1-48]
USFD, CNR, UNIPI, Fraunhofer, UT, IMT LUCCA, LUH, KCL, SNS, AALTO UNIVERSITY, ETH Zurich, TU Delft
The work package will develop the dissemination and impact strategy for the entire project, by building on and integrating the dissemination strategies of all involved infrastructures. It will identify stakeholders and address these through a range of targeted communication means and channels. The aim is to take concrete steps towards establishing the SoBigData starting community, also working closely on that with WP4 and WP5. The second focus of our dissemination work will be to coordinate with WP5 on innovation activities, in order to increase the participation of industry partners and to develop a long-term sustainability strategy for the project. WP3 will be lead by USFD, who have extensive expertise from previous EC projects.

T3.1 Stakeholder Analysis and Dissemination Plan
This task first of all develops a comprehensive stakeholder analysis and plan when and how their needs are addressed. It will produce a dissemination plan that lists types and numbers of activities and how they can best target the wide range of stakeholders that the project entails.
Task leader: CNR. Participants: ALL PARTNERS

T3.2 Development of Dissemination, Communication and Impact channels
The project will have a range of dissemination channels from newsletters, to a website to a strong social media presence, including media and specialized journal. We will furthermore run a wiki- type environment for internal communication among the project partners. A project portal and management platform will be chosen to assist project planning and monitoring, and to allow discussion and exchange of draft documents between partners. This will be supplemented by a document repository containing final versions of deliverables, reports, presentations, publicity material etc. The project web site will also advertise the calls for transnational access projects, datathons, developer competitions, and all other project events (see WP4 for details). All the relevant infrastructures for communication and dissemination will be set up by M3 of the project.
Scientific dissemination and outreach to relevant research communities will be through high-profile journal and conference publication channels (e.g. ICWSM, SIGIR, WWW, WSDM, KDD, ECML PKDD, SDM, ICDM, ACL, COLING, EMNLP, all IEEE and ACM journal venues). This task will coordinate these actions and provide a regular overview to the SoBigData community on scientific publication channels and opportunities. This is currently not easily possible as the work on social data is spread out among multiple disciplines and organisations. We will also ensure compliance with the Horizon 2020 open access publication policy.
Task leader: CNR. Participants: ALL PARTNERS

T3.3 Sustainability Planning
The task will deliver a sustainability plan and strategy for the future development of the project results. The plan will detail how the pan-European infrastructure and services will be sustained after the end of the project, how the collaboration is enhanced and deepened and how project results are crucially embedded in all partner institutions. The task will monitor also that all development of services is done on an open source basis and published on relevant platforms such as GitHub and SourceForge – wherever possible.
Task leader: CNR. Participants: ALL PARTNERS

T3.4 Impact Coordination
In order to promote the rapid realisation of economic, policy and social benefits from the project, we will work closely to also attract policy makers and economic stakeholders, building on pre- existing contacts and past collaborations. This task will coordinate all related activities in WP4 and WP5 and produce an annual report on our impact progress, as well as make recommendations towards next priorities and regions that should be addressed. The report will be a priority item on the general partner meetings and will feed into targeting the innovation activities in WP5. The task will maintain a database of impact partners starting with the large number of existing ones from the project partners (e.g. Press Association, the British Library, BBC, National Statistics Bureau, e.g. ISTAT, national and regional governments such as Estonia and Tuscany, data protection commissions at European and National level, major telecom and telematics providers such as Orange Telecom, Telecom Italia, WIND, Octotelematics, the German Federal Association for Information Technology, Telecommunications and New Media) and ensure that this community is effectively addressed. Towards the end of the project, we will hold a larger event that will bring together academic and non-academic communities to concentrate impact.
Task leader: USFD. Participants: ALL PARTNERS

T3.5 Community Building and Outreach
The aim of this task is to establish the common vision and methodology at the basis of the project activities. In particular, it will establish which is the nature of data science as a new discipline by enlightening which are the specific facets that make it at the same time interdisciplinary and unique when considering methods, theories and applications. The identification of competence centers, besides the ones involved directly in the project, will be actively pursued through specific search and exploiting the relationships of all the partners. Engagement will be pursued through inviting prospect members of the community to specific workshops. Particular attention will be devoted to the existing EU research infrastructures and E-infrastructures and to the related strategic bodies, including relevant ESFRI projects (CLARIN, ESS, SHARE, CESSDA, DARIAH), OPENAIRE, EUDAT and RDA (Research Data Alliance). We plan five national workshops, one for each nation involved in the trans-national activities. They will be devoted to putting together academia, industry and the public sector with the goal of exhibiting the key project results and collecting the requirements and feedback of these communities.
We will also organise a mid-term and final conference to attract researchers to the project, engage them and get feedback on our results. The preparation group will be assembled in such as way that it represents all project partners and stakeholders.
Task leader: USFD. Participants: ALL PARTNERS

Participation per Partner

Partner number and short name	WP3 effort
1 - CNR	6.00
2 - USFD	6.00
3 - UNIPI	2.00
4 - Fraunhofer	2.00
5 - UT	2.00
6 - IMT LUCCA	2.00
7 - LUH	2.00
8 - KCL	4.00
9 - SNS	2.00
10 - AALTO UNIVERSITY	2.00
11 - ETH Zurich	2.00
12 - TU Delft	3.00
Total	35.00

List of deliverables

Deliverable Number [14]	Deliverable Title	Lead beneficiary	Type [15]	Dissemination level [16]	Due Date (in months) [17]
D3.1	Project web site, project presentation, and social media presence	1 - CNR	Demonstrator	Public	3
D3.2	Fact sheets aimed at different stakeholders	1 - CNR	Report	Public	3
D3.3	Initial dissemination and impact plan	2 - USFD	Report	Public	3
D3.4	Periodic dissemination and impact report and plan for following year 1	1 - CNR	Report	Public	18
D3.5	Periodic dissemination and impact report and plan for following year 2	1 - CNR	Report	Public	36
D3.6	Final dissemination and community building events	10 - AALTO UNIVERSITY	Report	Public	48
D3.7	Sustainability Plan	1 - CNR	Report	Public	47

Description of deliverables

Periodic dissemination and impact reports.

D3.1 : Project web site, project presentation, and social media presence [3]

This deliverable will contain the project presentation, the project website, and the project's social media presence. These comprise the main dissemination channels for promoting and communicating the project and relevant activities and achievements. The project web site, presentation, and social media presence will be updated continuously throughout the project.

D3.2 : Fact sheets aimed at different stakeholders [3]

These fact sheets will be published on the project web-site for download by users and others, and they may also serve as hand-outs at events such as workshops and conferences.

D3.3 : Initial dissemination and impact plan [3]

The plan concerns the first year of activity and matches the project dissemination goals with the available opportunities. It also outlines the activity for the second year according to the information available at the time of delivery. Finally it includes the initial publication plan.

D3.4 : Periodic dissemination and impact report and plan for following year 1 [18]

The annually updated report includes the description of dissemination and impact activities undertaken and scientific papers published, as well as the planning for the following year.

D3.5 : Periodic dissemination and impact report and plan for following year 2 [36]

The annually updated report includes the description of dissemination and impact activities undertaken and scientific papers published, as well as the planning for the following year.

D3.6 : Final dissemination and community building events [48]

The report includes the description of dissemination activities and of the events organized during months 37-42 and assesses all the dissemination activity carried out during the project, providing overall statistics and comparisons. It also includes the complete list of publications.

D3.7 : Sustainability Plan [47]

This report will provide a detailed sustainability plan for the new pan-European SoBigData infrastructure.

Schedule of relevant Milestones

Milestone number [18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS1	Inception and User Needs Gathering	8 - KCL	6	- Community Creation Started - 1st SoBigData Workshop - Advisory Panel installation
MS2	SoBigData eInfrastructure and training launch	11 - ETH Zurich	12	- Repository of training materials established - Innovation and first services Integration - Ethical issues addressed - 2nd SoBigData Workshop - e-Infrastructure Deployment - Resource catalogue VRE - Completion of first series of educational initiatives: Datathon 1, Training HS 1 - e-Infrastructure Production Release
MS3	Analytics service integration and data ecosystem creation	1 - CNR	24	- Services and Infrastructure Integration - Big Data Ecosystem and Analytics - 1st SoBigData Conference - Completion of second series of educational initiatives: Summer School 1, Training HS 2 & Datathons 2 - e-Infrastructure Enhanced with new services - e-Infrastructure Production Release

Work package number [9]	WP4	Lead beneficiary [10]	8 - KCL
Work package title	NA3_Training		
Start month	1	End month	48

Objectives

- To establish a joint training and education resource repository on big social data in the European research area
- To promote the education of the next generation of data science researchers, with emphasis on social data analytics
- To connect with research communities not yet linked to social big data resources
- To engage researchers in new research methodologies for social big data
- To support the SoBigData research infrastructure usage through user training

Description of work and role of partners

WP4 - NA3_Training [Months: 1-48]
KCL, CNR, USFD, UNIPI, Fraunhofer, UT, LUH, AALTO UNIVERSITY, ETH Zurich, TU Delft
The aim is to establish a unique, joint training and education resource centre on social big data in the European research area. We will engage all relevant stakeholder groups. We explore and develop conventional and unconventional training experience for master and PhD students and post- doctoral trainees. These experiences include project-oriented summer schools and datathons for matching research (and industrial) needs and people skills. Training courses oriented to create new start-up big data initiatives will be defined and offered, where we will cooperate with relevant local partners in order to reach out beyond the current boundaries of the project. Moreover, we develop a campaign for attracting young bright minds into data science (with special attention to gender issues), through a series of training events for pupils in schools illustrating the creativity and the opportunities of the data scientist job, and thus preparing the next generation of social data researchers. A Training Programme will be set up and updated by the work package leader in collaboration with the involved partners in order to harmonize the planning and execution of the SoBigData educational initiatives. The Training programme is yearly submitted for approval to the General Partner Meeting.

T4.1 Summer schools
We plan to organize at least 3 summer schools that are aimed at preparing the best PhDs in science and engineering to work as data scientists. In more detail, USFD and UNIPI will run a summer school on text mining of big social data; ETHZ and CNR - on data science in techno-socio- economic systems; and FRH on visual analytics. These summer schools will aim at introducing participants to techniques and methodologies for analyzing big data, in order to provide them with a solid background in the computational and mathematical theories behind algorithmic tools for empowering their future research. Participants will be selected on a competitive basis and are encouraged to complete their own mini projects during the summer school, in hands-on sessions making use of data and services provided by the SoBigData RI. Each partner will bring in their own methodological expertise as well as training infrastructure and invaluable experience from running relevant training events (e.g. DATASIM, MODAP, IDEAS, the Estonian CS summer school, the GATE annual training weeks).
Task leader: USFD. Participants: ETHZ, CNR, UNIPI, FRH

T4.2 Training Modules for stakeholders
In this task, we will deliver a range of open source training modules, that arise from training events. These will be integrated into the learning and teaching environments of the project partners and other interested parties. These modules will, firstly, integrate existing training and education activities (e.g. the UNIPI Big Data Master), and promote and sustain them by making them available as open educational resources. Secondly, and more importantly, T4.2 will develop new targeted modules for specific stakeholders that are currently not well-served by teaching material on big social data, such as social scientists and humanities researchers. The modules will be published in open educational standards for easy integration with relevant e-learning environments. A special subset of modules will include best practices and guidelines for data scientists to work with big social data resources provided by the RI, and to exploit the SoBigData RI analysis tools. In more detail, the providers of the integrated national infrastructures will be responsible for providing training materials in the respective thematic areas (see WP6 and WP7 for details). KCL will be responsible for tutorials aimed at the social sciences and humanities users of the SoBigData RI, while all other partners will contribute expertise in applied social data mining projects, to produce the other materials.
Task leader: KCL. Participants: ALL PARTNERS

T4.3 Series of Datathons

We will organize a series of events whose aim is to bring together young and bright minds (approx. 30 each) and provide theoretical skills and practical advice in using their technical skills and creativity to collect, analyze and visualize Big Data in order to solve important problems. These events will integrate an interactive classroom style with Datathons (= data+hackathons) and related community events, similar to events previously organized by ETHZ, for example hackdays 2013 to build a "Mobile Sensor and Gaming Platform" (see https://github.com/mosgap) or KCL's collaboration with Young Rewired State (http://big-social-data.net/2014/07/14/our-first-successful- hackathon/). Datathons will be designed to bring together computer programmers, storytellers, graphics designers and statisticians, who will form teams with the young participants in order to create novel intelligent services on top of our infrastructure's big social data analytics methods. Datathons will build the basis for tools and frameworks to gather data: (1) Data that is available online will be collected. (2) We will persuade organizations (both non-profit and for profit) to open up data they have generated or collected, and give them in return, access to data scientists and our infrastructure to extract insights from the data for the benefit of the organization and its business processes. Beyond the collection of data the students will receive training and practice in extracting useful information from these unstructured datasets and how to communicate this information effectively through visualization or other methods. All datathons will be characterized by a renowned advisory board, chaired by the organizing partner and including representatives of partners and international experts.
Task leader: ETHZ. Participants: KCL, CNR

T4.4 Addressing gender and diversity issues in data science through training
ICT, and Data Science in particular, currently fails to adequately address staff equality and diversity issues. For instance, women are still woefully underrepresented, but also certain migrant groups, etc. To tackle this issue, we will provide a series of training events tailored specifically for high- school students (including some aimed specifically at girls) and promoting the role of data scientists. The aim is to leverage existing networks (e.g. the women network at the European Association on Data Mining and Machine Learning), in order to raise awareness of the opportunities of the data scientist job and its real nature. We will highlight that it is not a purely technological ("nerdy") job, but one that requires bridging society and technology, requiring inter- disciplinary competences and revolving around human factors and ethical values. In particular, UNIPI and KCL will run workshops on Open Social Data and Gender. KCL will collaborate with Young Rewired State on hackathons, and UNIPI and CNR will run training events for school children and create promotional videos. All partners will carry out presentations at schools.
Task leader: UNIPI. Participants: ALL PARTNERS

Participation per Partner

Partner number and short name	WP4 effort
1 - CNR	6.00
2 - USFD	4.00
3 - UNIPI	4.00
4 - Fraunhofer	4.00
5 - UT	4.00
7 - LUH	4.00
8 - KCL	10.00
10 - AALTO UNIVERSITY	2.00
11 - ETH Zurich	6.00
12 - TU Delft	4.00
Total	48.00

List of deliverables

Deliverable Number [14]	Deliverable Title	Lead beneficiary	Type [15]	Dissemination level [16]	Due Date (in months) [17]
D4.1	Periodic Training Planning Reports 1	2 - USFD	Report	Public	18
D4.2	Periodic Training Planning Reports 2	2 - USFD	Report	Public	36
D4.3	Periodic Training Planning Reports 3	2 - USFD	Report	Public	48
D4.4	Data Scientist Training Materials 1	8 - KCL	Report	Public	20
D4.5	Data Scientist Training Materials 2	8 - KCL	Report	Public	34
D4.6	Data Scientist Training Materials 3	8 - KCL	Report	Public	46
D4.7	Training Programme 1	8 - KCL	Report	Public	6
D4.8	Training Programme 2	8 - KCL	Report	Public	24
D4.9	Training Programme 3	8 - KCL	Report	Public	36

Description of deliverables

Periodic training planning reports.

D4.1 : Periodic Training Planning Reports 1 [18]

This deliverable reports on the training activities carried out in this workpackage, at key milestones. This will include reports on summer schools, datathons, and training events for schools and underrepresented communities in data science. Attendance details and other statistics will be collected and reported.

D4.2 : Periodic Training Planning Reports 2 [36]

This deliverable reports on the training activities carried out in this workpackage, at key milestones. This will include reports on summer schools, datathons, and training events for schools and underrepresented communities in data science. Attendance details and other statistics will be collected and reported.

D4.3 : Periodic Training Planning Reports 3 [48]

This deliverable reports on the training activities carried out in this workpackage, at key milestones. This will include reports on summer schools, datathons, and training events for schools and underrepresented communities in data science. Attendance details and other statistics will be collected and reported.

D4.4 : Data Scientist Training Materials 1 [20]

This deliverable periodically reports and publishes the online training materials, as well as those created for the face to face training events. We will also include access statistics for the online training materials.

D4.5 : Data Scientist Training Materials 2 [34]

This deliverable periodically reports and publishes the online training materials, as well as those created for the face to face training events. We will also include access statistics for the online training materials.

D4.6 : Data Scientist Training Materials 3 [46]

This deliverable periodically reports and publishes the online training materials, as well as those created for the face to face training events. We will also include access statistics for the online training materials.

D4.7 : Training Programme 1 [6]

This deliverable contains the plan of the educational activities for every year.

D4.8 : Training Programme 2 [24]

This deliverable contains the plan of the educational activities for every year.

D4.9 : Training Programme 3 [36]

This deliverable contains the plan of the educational activities for every year.

Schedule of relevant Milestones

Milestone number [18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS1	Inception and User Needs Gathering	8 - KCL	6	- Community Creation Started - 1st SoBigData Workshop - Advisory Panel installation
MS2	SoBigData eInfrastructure and training launch	11 - ETH Zurich	12	- Repository of training materials established - Innovation and first services Integration - Ethical issues addressed - 2nd SoBigData Workshop - e-Infrastructure Deployment - Resource catalogue VRE - Completion of first series of educational initiatives: Datathon 1, Training HS 1 - e-Infrastructure Production Release
MS3	Analytics service integration and data ecosystem creation	1 - CNR	24	- Services and Infrastructure Integration - Big Data Ecosystem and Analytics - 1st SoBigData Conference - Completion of second series of educational initiatives: Summer School 1, Training HS 2 & Datathons 2 - e-Infrastructure Enhanced with new services - e-Infrastructure Production Release
MS4	Stepping up TA, VA, and networking activities, and community engagement	2 - USFD	36	- Analytics Services and Infrastructure Integration updated - 3rd SoBigData Workshop - Completion of third series of educational initiatives: Summer

Schedule of relevant Milestones

Milestone number [18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
				School 2, Training HS 3 & Datathons 3 - Use-cases VREs
MS6	Starting SoBigData Community Established	4 - Fraunhofer	48	- 4th SoBigData Workshop - Final SoBigData Conferences - Completion of last series of educational initiatives: Summer School 3, Datathons 4 - Use-cases VREs - Final evaluation and project end

Work package number [9]	WP5	Lead beneficiary [10]	11 - ETH Zurich
Work package title	NA4_Accelerating Innovation		
Start month	1	End month	48

Objectives

• To widen the project impact beyond the scientific communities, by undertaking a wide range of innovation activities aimed at industrial and other stakeholders (e.g. government bodies, non-profit organisations, funders, policy makers)
• To undertake social innovation activities through building partnerships with industry and other key stakeholders through knowledge transfer and consulting services; training data scientists in entrepreneurial skills; contributions to policy making; and involvement of relevant industrial and standardisation bodies as SoBigData stakeholders
• To develop an innovation accelerator, which helps measure the impact of the SoBigData research infrastructure on societal, economic, policy, and community levels

Description of work and role of partners

WP5 - NA4_Accelerating Innovation [Months: 1-48]
ETH Zurich, CNR, USFD, Fraunhofer, UT, KCL, SNS, AALTO UNIVERSITY
With market size of big data products and services estimated at $32.4 billion by 2017 by International Data Corporation (goo.gl/8f3ZJS) and hundreds of big data startups emerging, the SoBigData infrastructure will have the unique opportunity to improve the innovation capacity of companies with its offering of large social data and big data analytics services.
Companies have been directly involved and pushed some of innovations in big social data (analytics). Examples include larger companies such as Google's BigTable work or SMEs such as NEO Technology's development of the Neo4j graph database. The WP will investigate how to bring these efforts into our integration activities.

T5.1 Partnerships with industry and contributions to policy making.
T5.1 will focus on improving the innovation capacity and strengthening growth of European big data companies, through joint activities with SoBigData. For instance, USFD and their GATE text mining infrastructure have a successful track record of attracting code developers and contributors from SMEs and larger companies, who then also build and sell innovative products on top of the open-source infrastructure. We will take steps towards rolling out such research-industry partnerships at project-wide level. Consortium members will also contribute actively to policy making and pursue other socially important impacts (e.g. helping government organisations and NGOs to benefit from big social data analytics, in order to obtain knowledge from citizen forums). In particular, research outputs on social mining for participatory e-government (Task 9.8) will be used to formulate policy recommendations in regards to e-government service development.
Task leader: USFD. Participants: FRH, CNR, UNIPI, UT

T5.2 Knowledge Transfer and Consulting Services.
The project partners will set up knowledge transfer and consultancy services, aimed at SMEs and also bigger organisations. Often companies need to tailor generic big data analytics algorithms to the specifics of their vertical or product, so the project will be ideally placed to offer such customisation and knowledge transfer services, on top of the numerous open-source algorithms and free services provided by the SoBigData infrastructure. T5.2. will facilitate this process by also developing template non-disclosure agreements and consulting services contracts (in collaboration with WP2), in order to lower the barrier for SMEs and non-profits to benefit fully from the new European infrastructure.
Task leader: USFD. Participants: IGD, CNR

T5.3 Entrepreneurial skills for big data entrepreneurs.
It can be expected that in big data analytics many more start-ups will be launched and supported in innovative, viable and sustainable businesses. The goal of this task is to leverage well-structured and known entrepreneurial training courses and materials, currently offered by some of the consortium members (see e.g. PhD Plus at UNIPI which got a EU Leonardo Transfer of Innovation project 2012-14) in order to spur and/or boost the entrepreneurial spirit of MSc, as well as PhD students and post-doctoral researchers, by providing them with a series of seminars/lectures and coaching/mentoring periods for learning and practicing how to launch Big Data start-ups over challenging and sustainable businesses. Moreover, this task will investigate through consultations with all stakeholders how the results of the project can be leveraged by 3rd party startups.
Task leader: UNIPI. Participants: --

T5.4. Innovation Accelerator
The Innovation Accelerator is a web platform, already prototyped by ETHZ, which will be adapted in SoBigData to monitor, analyze, and enhance the social processes around data science and innovation. Central to that is Living Science, which is an interface to analyze and visualize social relations and communities in the network of (scientific) publications. These tools will be adapted and customised in the project, to help data scientists, funding bodies, and other decision makers to track the progress of the multiple disciplines that interface into data science; to measure funding impact and methods used; and to detect emerging technologies related to mining big social data.
Task leader: ETHZ. Participants: --

Participation per Partner

Partner number and short name	WP5 effort
1 - CNR	4.00
2 - USFD	6.00
4 - Fraunhofer	2.00
5 - UT	4.00
8 - KCL	2.00
9 - SNS	2.00
10 - AALTO UNIVERSITY	2.00
11 - ETH Zurich	10.00
Total	32.00

List of deliverables

Deliverable Number [14]	Deliverable Title	Lead beneficiary	Type [15]	Dissemination level [16]	Due Date (in months) [17]
D5.1	Preliminary Innovation and Action Plan	2 - USFD	Report	Public	9
D5.2	Final Innovation Report and Action Plan	2 - USFD	Report	Public	46
D5.3	Entrepreneurial skills materials and activities undertaken	3 - UNIPI	Report	Public	47
D5.4	Innovation Accelerator 1	11 - ETH Zurich	Demonstrator	Public	24
D5.5	Innovation Accelerator 2	11 - ETH Zurich	Demonstrator	Public	42

Description of deliverables

Innovation and Action Plan reports.

D5.1 : Preliminary Innovation and Action Plan [9]

This deliverable draws the plan of work for T5.1 and T5.2, including a concrete action plan for the innovation activities aimed at widening the impact of the SoBigData integrated research infrastructure within a 5 years horizon, i.e. including the second part of the project and a further period beyond the funded one.

D5.2 : Final Innovation Report and Action Plan [46]

The report updates the Innovation and Action Plan, keeping into account the progress of the SoBigData infrastructure and any change intervened in the general conditions during the second part of the project. It describes actions required in a 5-10 years horizon.

D5.3 : Entrepreneurial skills materials and activities undertaken [47]

It comprises the materials (slides, videos, etc.) and a report on activities undertaken as part of T5.3.

D5.4 : Innovation Accelerator 1 [24]

This software prototype deliverable consists of the customisation of the Innovation Accelerator web platform, including Living Science, as well as delivering impact and innovation metrics for publications and collaborations arising from SoBigData and other big social data analytics projects and key players.

D5.5 : Innovation Accelerator 2 [42]

This software prototype deliverable consists of the customisation of the Innovation Accelerator web platform, including Living Science, as well as delivering impact and innovation metrics for publications and collaborations arising from SoBigData and other big social data analytics projects and key players.

Schedule of relevant Milestones

Milestone number [18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS2	SoBigData eInfrastructure and training launch	11 - ETH Zurich	12	- Repository of training materials established - Innovation and first services Integration - Ethical issues addressed - 2nd SoBigData Workshop - e-Infrastructure Deployment - Resource catalogue VRE - Completion of first series of educational initiatives: Datathon 1, Training HS 1 - e-Infrastructure Production Release
MS4	Stepping up TA, VA, and networking activities, and community engagement	2 - USFD	36	- Analytics Services and Infrastructure Integration updated - 3rd SoBigData Workshop - Completion of third series of educational initiatives: Summer School 2, Training HS 3 & Datathons 3 - Use-cases VREs

Work package number [9]	WP6	Lead beneficiary [10]	2 - USFD
Work package title	TA1_Transnational Access		
Start month	13	End month	48

Objectives

Provision of on-site access is granted to six thematic clusters of facilities provided by seven national infrastructures, offering world-leading research expertise from multiple disciplines, as well as big data computing platforms, big social data resources, and cutting-edge computational methods.

Description of the infrastructure

Name of the thematic clusters:
1. Text and Social Media Mining (TSMM)
2. Social Network Analysis (SNA)
3. Human Mobility Analytics (HMA)
4. Web Analytics (WA)
5. Visual Analytics (VA)
6. Social Data (SD)

In the following table we summarize the national infrastructures with the thematic cluster covered:

Name of infrastructure Location Web site Annual operating cost Thematic clusters
SoBigData.it CNR, Pisa, IT
University of Pisa, IT www.sobigdata.it 100,000 € TSMM, SNA, HMA, WA

GATE USFD, Sheffield UK gate.ac.uk 280,000 € TSMM

IVAS Fraunhofer IGD,
Darmstadt, DE https://www.igd.fraunhofer.de
/en/Institut/Abteilungen/IVA 50,000 € VA

Alexandria LUH, Hannover, DE www.L3S.de 50,000 € WA

Aalto Helsinki, Finland SNA

E-GovData Tartu, Estonia www.cs.ut.ee/ SD

Living Archive Zurich, Switzerland SD

The national infrastructure expertise in the respective thematic cluster is described next, as well as the services offered during transnational access:

• SoBigData.it. Co-founded in 2013 by two institutes of the CNR (the Italian National Research Council) and the University of Pisa, the European laboratory on Big Data Analytics and Social Mining (www.SoBigData.it) is aimed at pursuing interdisciplinary research initiatives connected to the impetus that “big data” and the ICT’s are having on science, and the socio-economic sciences. One of the goals is the creation of a knowledge infrastructure for the acquisition and analysis of big data from different fields. CNR participates in the project through researchers from 5 groups: Knowledge Discovery and Data Mining Lab (KDD – ISTI), ‘Networked Multimedia Information System’ Lab. (NeMIS - ISTI), High Performance Computing Lab (HPC– ISTI), and the Web Applications for the Future Internet (WAFI-IIT) and Ubiquitous Internet groups (UI-IIT) - of the Institute for Informatics and Telematics (IIT).
o Thematic Cluster SNA - The KDD Lab. and UI Lab. at CNR are leading groups in the area of social network analysis: they have experience, develop and maintain methods and algorithms covering all the major research topics in this field, with particular attention to Structural Properties of Social Networks, Community Discovery, Link Analysis, Evolutionary Analysis and Multidimensional Network modeling. Based on existing collaboration agreements at national level, CNR will also aggregate expertise provided by the groups at IMT, Scuola Normale Superiore (SNS), and UNIPI, thus enhancing the TA opportunities to a broader repertoire of network data and analytics.
Services offered: The facility offered includes hundreds of state-of-the-art methods for extracting knowledge from networked data with particular emphasis on Social Network platforms. Among other, examples include: community

and dense sub-graph discovery algorithms (e.g., DEMON, TILES); algorithms to detect structural properties of Social Networks at the level of individual users (ego networks), and across users; methods to estimate tie strength information from interaction variables; tools and methods to crawl large scale SN platforms, such as Facebook and Twitter; approaches to analyze and forecast the behavior of dynamic interaction networks as well as find geographically coherent events in activity networks; multidimensional network modeling and analysis (e.g., UBIK, a multidimensional node ranking method); link analysis techniques and models aimed at analyzing epidemics and information diffusion processes; methodologies to analyse the impact of trust between users in SN on information dissemination and management; algorithms to analyse the relationships between human social interactions in cyber environments (i.e., SN platforms) and in the offline, physical world (i.e., human social networks not mediated by SN tools).

o Thematic Cluster HMA - The KDD Lab. is a leading research hub on mobility data mining, becoming a reference not only for the international research community but also for leading industrial and public operators, such as telecom providers (Orange, Wind, Telecom Italia) and mobility agencies of regional and municipal administrations.
Services offered: The facility includes state-of-the-art methods for mobility analytics such as: pre-processing and cleaning tools, integration of diverse mobility observations, advanced mobility data mining algorithms like Trajectory Clustering, Trajectory Patterns, Mobility Semantic Enrichment, Combination of network science methods and mobility. All these resources are components of the analytical platform MAtlas (www.m-atlas.eu), providing methods suitable for analysing vehicular GPS data and mobile phone data, as well as expertise on the definition and exploration of analytical processes based on mobility data. Many unique datasets of movement data at regional and national level will also be made available under various policies, as defined in WP8.

o Thematic Cluster WA - HPC Lab. conducts research in the area of Web Usage Mining by designing new models and tools for understanding user behaviour (e.g, queries, visited pages) and analyzing implicit user feedback (e.g., clicks) with the goal of improving the users' Web experience. This research mainly focuses on Web Search Engines, their ranking functions, efficiency vs. effectiveness trade-off and query log analysis.
Services offered: QuickRank is an open-source framework implementing state-of-the-art Learning-to-Rank (LtR) algorithms. LtR is gaining increasing attention also due to their use in Web Search Engines (WSE). QuickRank implements state-of-the-art approaches based on gradient boosted regression trees. The algorithms implemented by QuickRank can be successfully exploited in other scenarios, such as in learning ranking functions for recommendation tasks, or similarity functions for clustering tasks.

o Thematic Cluster TSMM - SoBigData.it has two groups with expertise in Text and Social Media Mining: HPC Lab and WAFI-IIT. The HPC Lab conducts research in various areas of High Performance Computing, including Cloud Computing, Cloud Federations, Large-scale Data Analysis, Information Retrieval, and Distributed Search Engines. The group, composed of around 20 people, is actively involved in several important European projects including MIDAS, InGeoClous, eCloud, Contrail. WAFI-IIT has followed the evolution of the World Wide Web from the success of markup languages, to the introduction of a semantic layer and the explosion of the social web content. WAFI-IIT studies the evolution of the Web in all its forms (data, communities, technologies, guidelines, recommendations, designs, software, tools). Currently, WAFI-IIT performs research and development activities on Social Media Analysis, spanning from Social Network Analysis to Social Sensing and Social Media Intelligence, Information Visualization on the Web, Mashup Applications, Open and Big Data, Web Mining, Semantic Web. It is also part of the Italian W3C office, involved in the organization of events and courses on Linked Open Data.
Services offered: Dexter (http://www.dxtr.it/) is an open-source framework implementing popular entity linking algorithms. The entity linking task aims at identifying all the small text fragments in a document referring to an entity contained in a given knowledge base, e.g., Wikipedia. Dexter already implements several state-of-the-art approaches for entity mention detection, disambiguation, and ranking. It allows to evaluate fairly the performance of existing methods in each sub-task of the entity linking process, and it allows to develop new entity linking algorithms by plugging in new custom components. Another platform available is the SoS-Social Sensing platform exploits the opportunistic crowdsensing, involving users in the sensing process in a minimal way, for social media emergency management purposes, in order to obtain a very fast, but still reliable, detection of emergency dimensions and estimate their consequences. The key value of the platform is that data mining and natural language processing techniques, burst detection algorithm and predictive analytics approaches are used together to create knowledge for social media data. The platform exploits several types of algorithms: from burst detection approaches for event detection, to natural language processing techniques and predictive analytics approaches for the damage assessment in the analysis phase. Currently the SOS platform offers dataset containing over 5 millions of tweets in English, Spanish and Italian and more than 40 distinct features useful for the study of earthquakes and their consequences.

• GATE - Thematic cluster TSMM - The Natural Language Processing (NLP) group at USFD is one of the largest and most successful research groups in text mining and language processing in the EU. USFD develops and maintains

the world-leading open-source GATE text mining and natural language processing infrastructure (http://gate.ac.uk) and its vibrant user community. GATE has thousands of users from hundreds of organisations, with the community growing annually through Sheffield-based summer schools and online training materials. GATE is currently the only NLP infrastructure which comes with an extensive set of algorithms for mining social media, using Linked Data for NLP, semantic search, and crowd-sourcing. In the past 5 years alone, GATE has underpinned over 15 European projects, including seven on analysing social media. We currently host over 5 self-funded visiting researchers in 2014 alone.
Services offered: the infrastructure offers over 150 models and algorithms for mining social media, Information Extraction (IE), ontology-based semantic annotation, machine learning for IE, and NLP evaluation tools. GATE accelerates NLP and text mining research through extensive toolset for modelling and analysis of social media content; gold-standard creation via expert- or crowd-sourcing; performance evaluation measurement; visualisation and editing of linguistic structures and results; a finite state transduction language for rapid prototyping and efficient NLP experimentation; pluggable machine learning algorithms. GATE is supplied with the TwitIE Information Extraction system, which has been adapted and evaluated extensively on social media content. During the SoBigData project, the GATE infrastructure will be extended and made inter-operable with other text and social media mining services and platforms from UNIPI, CNR, and others (see T9.3). It also offers state-of-the-art HPC and Hadoop computer clusters for running text and social media mining experiments, over USFD's 12TB of compressed Twitter data, going back to May 2009 and growing by 50 million tweets per day. There are also millions of web documents in the four main European languages, as well as human-annotated corpora for method evaluation.

• IVAS - Thematic cluster VA - The Competence Center for Information Visualization and Visual Analytics (IVA) at Fraunhofer IGD is a world-leading research centre for the interactive visualization of big data. They offer access to visual analytics technologies, solutions for semantics visualization and real-time solutions for the simulation, visualization and rendering of data.
Services offered: Their IVAS server-based infrastructure provides Visual Analytics-related data, algorithmic and visualization services. IVAS combines data mining and information visualization in an interdisciplinary way. IVAS will enable SoBigData researchers to visualize and interact with massive amounts of multi-dimensional big data. The core of IVAS's semantics visualization is the graphical presentation of information, to enhance decision making through visual processing of complex interdependencies. With IVAS, scientists can express their data analysis workflows explicitly while incorporating data from diverse inputs, such as web services, relational and non-relational databases, graph databases, etc. Finally the output of any node in the workflows can be visualized using the in-build visualization library. Visualizations can be created from scratch or adapted using the in-build editor. IVAS has been developed and used within three different FP7 projects to facilitate data sharing, providing analysts with data mining capabilities on social networks, and to make text mining results accessible to the general public.

• Alexandria - Thematic cluster WA - L3S provides access to innovative and cutting-edge datasets, methods and technologies for Web Science. The Alexandria infrastructure underpins novel research in Web Search (especially Entity-centric search), Web Information Management (including semantic technologies) and the Web of People (including Personalization and Social Web).
Services offered: The infrastructure offers services for data management (HBase, Hive), for search and retrieval (ElasticSearch), for the processing of very large versioned document collections (Hedera), large-scale machine learning (Mahout), and scripting support for the easy development of data-centric applications. With the integration of the Analytical Crawler (see WP8), the Web data set will regularly be extended with high quality Web content. During the project life time the Alexandria infrastructure will be extended with functionalities/scripts/libraries commonly requested by the data-centric research community (see WP9). In addition experimental tools and methods for the temporal analytics not covered by SoBigData, but currently under development within the ERC Grant Alexandria (ERC 339233), will be made available.

• AALTO - Thematic cluster SNA - it contributes to the project with participation from two research groups, the Sociophysics Laboratory from the department of Biomedical Engineering and Computational Science (BECS) and the Data Mining group from the department of Information and Computer Science (ICS). The research conducted in the Sociophysics Laboratory in BECS focuses on: (i) living and other complex systems, their measurement, analysis, modelling, understanding and control, (ii) detection of communities and social dynamics, with a focus on the dynamics of scientific interactions and human behaviour in social and information systems. The Data Mining group in ICS focuses on developing novel methods to extract knowledge from data, designing algorithms to summarize large volumes of data efficiently and effectively, and exploring new ways of using the extracted information.
Services offered: AALTO will offer a wide collection of state-of-the-art computational methods to analyze and extract knowledge from large social network datasets. Methods include finding dense sub-graphs in social networks and interaction networks, analyzing graphs with attribute information on vertices and edges, mining evolution of networks, mining city data to discover events and neighbourhood characteristics, analysis of movement traces to discover

backbones of traffic networks, understanding opinions and network dynamics, and more. Additionally, the unit has strong focus on analysis of bibliographic and citation data. The methods developed in AALTO combine experimental and computational approaches, and they are aimed to tackle major problems in well-being, medical diagnostics, energy, the society and the environment.

• E-Gov. data - Thematic cluster SD - Estonia is an early adopter of e-government services, thanks to an aggressive e-government strategy dating from the 1990s. Among others, Estonia is a vigorous adopter of e-prescription, e-identity, e-signature, e-voting, e-taxation, and a range of other e-government and e-health services for citizens and businesses. This makes Estonia a perfect laboratory to study the impact of e-government service deployment, adoption and usage on a range of societal issues. In this context, UT brings in a curated and inter-linked dataset of Estonian e-government and e-health service descriptions, detailed statistics of usage of said services, and data related to societal and economic development in Estonia over the past decade. The dataset is complemented with relevant automated analysis methods. This unique research infrastructure will allow researchers to study the impact of e-government service usage in Estonia on broader societal issues, as well as the potential impact of wider e-government service adoption in other countries.
Services offered: The facility offers access to Estonian e-government data over a period of 10 years, with an emphasis on e-government and e-health service usage. The data includes: detailed descriptions and usage statistics of Estonian e-government and e-health services; Estonian secondary education performance; Estonian hospital discharge summaries; and other e-government data. These datasets are accompanied by a toolkit that allows users to inter-link Web service descriptions and their usage information, and to analyze collections of Web services in terms of information diffusion and information redundancy. The toolkit has been used to detect information redundancy in the Estonian X-Road e-Government service collection [Küngas 2010], as well as to study information diffusion in this and similar e-service collections [Mokarizadeh 2012]. The infrastructure is currently used by researchers and PhD students at KTH Sweden and the University of Tartu.

• Living Archive for Open Data - Thematic cluster SD - The Department of Humanities, Social, and Political Sciences is situated at the leading university in continental Europe (ETHZ). They bring world-leading expertise in Sociology, in particular of Modeling and Simulation. As part of the FuturICT project, they worked on the intersection of natural, social and engineering sciences to address the challenges of the complex and strongly interdependent global techno-socio-economic-environmental systems and the big social data they generate. Access will be offered to the LivingArchive indexes of over 100.000 Open datasets, spanning across a large spectrum of domains. Living Archive and the other tools of the ETHZ Innovation Accelerator (see WP5) already implement many aspects of the objective to integrate large social data sets, in order to overcome fragmentation and provide interoperability.
Services offered: the facility provides a search engine for Open Data which is designed as a community-run catalogue of open datasets available on the web. Users can collect and bookmark links to datasets for themselves and others to use, or search for data that others have collected. Depending on the type of data (and its conditions of use), Living Archive may also be able to store a copy of the data or host it in a database, and provide some basic visualisation tools. Currently, Living Archive indexes over 100,000 datasets, spanning across a wide spectrum of scientific and social domains.

Description of work and role of partners

WP6 - TA1_Transnational Access [Months: 13-48]
USFD, CNR, Fraunhofer, UT, LUH, AALTO UNIVERSITY, ETH Zurich
Description of work
This Work Package (coordinated by USFD) will implement the activities related to Transnational Access (TA) offered by the infrastructure. It will specify the modality of access, including the offered support strategy, the target users and all the activities aimed to implement TA. TA will support primarily, but not exclusively younger scholars (e.g. PhD candidates and early career researchers) with limited resources and access to research infrastructures. Researchers from smaller institutions, newly associated states, or working independently with very limited infrastructure will be considered, regardless of career stages. Gender equality will also be an important factor.

Modality of access under this project:
TA management: TA management will be led by USFD and involve all WP6 partners, which jointly will constitute the TA Management Board. Activities will involve organising the TA (call publications, adverting, etc.) and the financial management of the resources. The TA Management Board will be responsible for the call publication and advertising and for setting up the international User Selection Panel. USFD will assign a day-to-day contact for the visiting researchers and the whole consortium to advise on cost eligibility, financial issues and official processes for fund transfers. USFD will also administer the TA travel re-reimbursement budget.

Guidance and assistance to researchers will be granted, especially for special kinds of activities concerning particular data and/or algorithms with restricted access (i.e., not sharable within virtual access due to stringent access policies) or involving challenging tasks which require expert's support.
It is planned that users or teams stay for a period (between a week and 2 months) at one of the installations, in order to interact with the local experts, discussing research questions they are interested in, and using the datasets and algorithms needed to perform their analysis. The visit agenda will include workshops with local experts, receiving their guidance, working on the data and discussing the results with them. The activity will consist of guided use of the infrastructure, autonomous use, and discussion in seminars and workshops. Users will benefit from the collaboration with highly experienced, world-leading research teams. A dedicated mentor will assist the visitors and interface with the experts.
At the end of the visit, the users will be required to provide a final report on the performed activity, which will also detail the data sets he/she will have worked with and associated challenges. These reports will satisfy a twofold purpose: (i) they will be collected and integrated in the annual dissemination activity reports; (ii) they will be used as a direct feedback on the installation activity to the consortium. Moreover, sharing these research experiences will be a way to foster multi-disciplinarity, as further means to community building.

Unit of access 1 week includes:
• Preparatory work by the user with the assistance of the host: documentation on the research problem, related works and tools to be used; special aspects and anticipated issues. Estimate: 2 person-days
• Remote training (whenever possible) on the tools to be used. Estimate: 3 person-days
• One week visit, with assistance of a dedicated tutor, including kick-off workshop, technical support, final wrap-up workshop. Estimate: 5 person-days (tutor) + 2 person-days (others)
• Remote post-visit assistance and guidance. Estimate: 3 person-days.
Total resource committed, per 5-days' visit: 15 person-days.
In some cases the activity may require longer visits of up to two months, depending on the research challenge faced. It is estimated that this happens in 25% of the cases.

Support offered under this project:
WP6 will offer support for two kinds of projects: theme-driven exploratory projects and blue-sky, open calls.
Exploratory projects are research experiments bound to multi-disciplinary themes specified in the calls, such as: social mining for measuring well-being, social mining for participatory e-government, social mining for city science, social mining for measuring and mitigation conflicts, social mining for understanding success in areas like sport, sciences, etc. The exploratory projects are expected to target data scientists from multiple disciplines (e.g., computer science, mathematics, physics, sociology, economy), and address the definition of needed infrastructure resources (datasets and/or methodologies and algorithms) and knowledge offered by specific hosting nodes. The schedule of the experiments, the expected results and research impacts will be defined on the basis of local availability of the experts at the installation site.
Open call, blue-sky projects will be targeted at researchers wishing to explore the infrastructure for their own research topics, without topic restriction in the call. Applicants will be asked to submit to the User Selection Panel experiment proposals exploiting datasets and/or algorithms hosted by different nodes of the infrastructure. The proposals can be related to both the analysis of datasets available at a specific installation and\or the evaluation of algorithms and analytical procedures with the support of local teams of experts.

Outreach to new users:
The project will advertise the availability of visiting researcher opportunities through its dissemination channels (web site, newsletter, etc.), as well as through its training and dissemination channels.
It is expected that there will be a very large demand for visits as a consequence of the much wider scope of the infrastructure that can be accessed. This will multiply the demand for TA, therefore periodic calls will need to be published and advertised, covering both exploratory and open calls.

Review procedure under this project:
The project's User Selection Panel (USP) will be appointed at the start of the project by the TA Management Panel. It will comprise leading international researchers from all key scientific disciplines brought together in SoBigData. The USP will hold virtual periodic meetings for ranking and selection of eligible proposals for access. Proposals will use a simplified form illustrating the activity proposed, the rationale of the visit, the objectives of the research, and the infrastructural facilities that they plan to make use of (data, computer facilities, and algorithms). The Panel will employ the support of independent experts asked to review proposals anonymously according to fair evaluation criteria published in the call for participation. It is anticipated that part of the available support may be reserved to help researchers

from under-represented categories (e.g. starting career, newly associated or candidate states, female data scientists or researchers with limited resources).
Task leader: USFD. Participants: ALL PARTNERS

Participation per Partner

Partner number and short name	WP6 effort
1 - CNR	18.00
2 - USFD	12.00
4 - Fraunhofer	5.00
5 - UT	4.00
7 - LUH	5.00
10 - AALTO UNIVERSITY	4.00
11 - ETH Zurich	1.00
Total	49.00

List of deliverables

Deliverable Number [14]	Deliverable Title	Lead beneficiary	Type [15]	Dissemination level [16]	Due Date (in months) [17]
D6.1	Periodic reports on TA activities 1	2 - USFD	Report	Public	18
D6.2	Periodic reports on TA activities 2	2 - USFD	Report	Public	36
D6.3	Periodic reports on TA activities 3	2 - USFD	Report	Public	48

Description of deliverables

Periodic reports on Transnational access activities.

D6.1 : Periodic reports on TA activities 1 [18]

It will report on calls initiated in the reporting period, user selection process and outcomes, TA visit reports, and self-evaluation. The M18 and M36 periodic reports will also contain plans for TA calls for the subsequent period.

D6.2 : Periodic reports on TA activities 2 [36]

It will report on calls initiated in the reporting period, user selection process and outcomes, TA visit reports, and self-evaluation. The M18 and M36 periodic reports will also contain plans for TA calls for the subsequent period.

D6.3 : Periodic reports on TA activities 3 [48]

It will report on calls initiated in the reporting period, user selection process and outcomes, TA visit reports, and self-evaluation. The M18 and M36 periodic reports will also contain plans for TA calls for the subsequent period.

Schedule of relevant Milestones

Milestone number [18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS3	Analytics service integration and data ecosystem creation	1 - CNR	24	- Services and Infrastructure Integration - Big Data Ecosystem and Analytics - 1st SoBigData Conference - Completion of second series of educational initiatives: Summer School 1, Training HS 2 & Datathons 2 - e-Infrastructure Enhanced with new services - e-Infrastructure Production Release
MS4	Stepping up TA, VA, and networking activities, and community engagement	2 - USFD	36	- Analytics Services and Infrastructure Integration updated - 3rd SoBigData Workshop - Completion of third series of educational initiatives: Summer School 2, Training HS 3 & Datathons 3 - Use-cases VREs
MS6	Starting SoBigData Community Established	4 - Fraunhofer	48	- 4th SoBigData Workshop - Final SoBigData Conferences - Completion of last series of educational initiatives: Summer School 3, Datathons 4 - Use-cases VREs - Final evaluation and project end

Work package number [9]	WP7	Lead beneficiary [10]	1 - CNR
Work package title	VA1_Virtual Access		
Start month	13	End month	48

Objectives

Provision of on-line access is granted to thematic clusters through platforms provided by seven national infrastructures, offering cutting-edge services for big data and social mining research.

Description of the infrastructure
Following the classification described in WP6, in the following table the installations available for the virtual access are summarized:

Name of infrastructure Name of virtual installation Location Web site Annual operating cost Thematic clusters
SoBigData.it · TagMe system,
· Elianto platform,
· Twitter streaming Monitor framework
· SNA Toolkit,
· MAtlas system CNR, Pisa, Italy
University of Pisa, Italy www.sobigdata.it 100,000 € TSMM, SNA, HMA
GATE GATECloud USFD, Sheffield UK cloud.gate.ac.uk 96,000 € TSMM

IVAS IVAS Platform Fraunhofer IGD
Darmstadt, DE https://www.igd.fraunhofer.de/
en/Institut/Abteilungen/IVA 50,000 € VA
Alexandria AlexandriaCloud LUH, Hannover, DE www.L3S.de 50,000 € WA

The detailed description of each installation and serviced offered is reported below:

• SoBigData.it provides several installations in different Thematic Clusters:

o TagMe system - Thematic Cluster TSMM - (http://acube.di.unipi.it/tagme) is a powerful text annotation tool, which identifies meaningful sequences of words in short texts, both in English and in Italian, and links them to pertinent Wikipedia pages in a fast and effective way. TagMe results are one of the best topic-annotators to date [Cornolti 2013] due to its high efficiency and effectiveness in annotating also very short texts, such as tweets and news. Since 2010 TagMe served more than 170 Million queries from research groups and companies worldwide. In addition to high usage statistics and numerous high profile publications, TagMe received two Google Research Awards: in 2010 for TagMe engineering and testing, and in 2013 for social-network analysis and query annotation using TagMe. In June 2014, TagMe became the basis of the SMAPH system, which was judged the best query-annotation tool in the ERD Challenge at ACM SIGIR 2014, beating around 20 other academic and industrial systems from around the world.
Services offered: It is an open-source software (released under Apache License 2.0) and is freely accessible over the Web via a RESTful API since June 2010 (http://tagme.di.unipi.it). The API allows users to annotate text in English or Italian and, in addition, to execute some useful basic functionalities for text mining: such as, Wikipedia graph exploration, computing the relatedness between two Wikipedia Pages, and returning the spots which point to a Wikipedia page and contain a specified term. These functionalities and TagMe itself are the building blocks upon which more sophisticated tools can be designed for solving more sophisticate text analysis and information retrieval applications. As part of the integration work in WP9 (T9.3), GATECloud and TagMe will be made tightly interoperable, in order to make it easier to carry out experiments by combining the different services and, thus, ultimately servicing more requests (up from the current tens of millions to a billion per annum). Quantitative usage metrics will be provided in D7.1.

o Elianto platform - Thematic Cluster TSMM - (http:// elianto.isti.cnr.it) is a platform hosting an open-source, user friendly and re-active Web interface to support the crowdsourced creation of gold standard datasets for entity linking and salient entities recognition.
Services offered: It supports human labelling of semi-structured documents through a guided two-step process. Collections of unstructured or structured documents can be uploaded on the platform and the guided annotation task easily monitored.

o Twitter Streaming Monitor framework - Thematic Cluster TSMM - is tool for gathering data from Twitter in a focused way. Using mechanisms offered by Social Media for collecting data not only useful information are retrieved. To reduce the consequent noise in the dataset the tool offers the possibility to specify gathering criteria and collect only data relevant for the experiment purposes.
Services offered: it helps data scientist to build datasets for their experiments, in real time and in a focused way. The gathering criteria can be expressed by the users in three ways: as one or more keywords, stating a user, or drawing a geographical area to monitor. In this way only tweets matching the gathering criteria are used for building the dataset. Moreover the tool offers different facilities to manage different gathering processes at the same time.

o SNA toolkit - Thematic Cluster SNA - will integrate the variety of network datasets and analytical methods offered by the partners, with particular attention to scalable methods for Community Discovery, Link Analysis, Evolutionary Analysis and Multidimensional Network modeling.
Services offered: The methods in the SNA toolkit are currently used by many students and scientists through the various interfaces provided by the individual partners. For instance, the DEMON software developed by CNR has become a reference in the area of community discovery in complex networks and it is widely used. However, the methods are not currently offered within an integrated virtual service, which is precisely the objective of this WP.

o MAtlas system - Thematic Cluster HMA - (http://m-atlas.eu/) is a widely used mobility data mining system, i.e., a querying and mining system centered onto the concept of trajectory. Besides the mechanisms for storing and querying trajectory data, M-Atlas has mechanisms for mining trajectory patterns and models that, in turn, can be stored and queried. The basic design choice is compositionality, i.e., querying and mining of trajectory data, patterns and models may be freely combined, in order to provide the expressive power needed to master the complexity of the mobility knowledge discovery process. Another design choice of the system is that all entities are represented in the object-relational data model, which is more suitable to tackle the structural complexity of spatio-temporal data w.r.t. simple tabular data.
Services offered: it provides state-of-the-art methods for mobility analytics such as: pre-processing and cleaning tools, integration of diverse mobility observations, advanced mobility data mining algorithms like Trajectory Clustering, Trajectory Patterns, Mobility Semantic Enrichment, Combination of network science methods and mobility. For datasets without sharing restriction as defined in WP8 (Big Data Ecosystem), the partner offers remote access to both methods and dataset. The access is provided by means of a remote instance of M-Atlas. The methods implemented by the framework can be configured locally on the client and then executed remotely on a central server. The VA guarantees the access to complex analytical tasks like, Origin-Destination Matrix extraction, Mobility Pattern visual exploration, Borders of Human Mobility, Urban Mobility Atlas.

● GATE - GATECloud - Thematic Cluster TSMM - (http://cloud.gate.ac.uk) is a unique, cloud-based infrastructure for large-scale, data-intensive Natural Language Processing (NLP) and text mining research. Important infrastructural issues are dealt with by the platform, completely transparently for the researcher: NLP algorithm distribution, load balancing, efficient data upload and storage, deployment on the virtual machines, security and fault tolerance. Another unique feature of this specialized NLP Platform-as-a-Service is its support for researchers who want to develop and run their own text mining/NLP pipelines on big data.
Services offered: it offers a growing number of NLP and text mining services for multiple european languages. Currently only around 30% of the algorithms and tools from the GATE infrastructure have been made available through GATECloud, and the number will continue to grow during the project lifetime. GATECloud offers a web interface for data upload, social media data collection, programming-less execution of the GATECloud text analytics services, and results download.

• IVAS - IVAS Platform - Thematic Cluster VA - is a server-based platform to provide Visual Analytics-related data, algorithmic and visualization services over an REST API to authorized clients. With IVAS, scientists can define workflows and include them in custom applications. They can work with data from diverse inputs like web services, relational and non-relational databases, graph databases, etc. [Andrienko 2011] IVAS can handle tabular (relational) data, document collections (non-relational), and graph data.
Services offered: the infrastructure has been developed to facilitate data sharing, providing analysts with data mining capabilities on social networks and to make text mining results accessible for the general public. In order to interact with the IVAS infrastructure, a client library for the Javascript programming language is provided by the server. Other programming languages can rely directly on the REST interface. Visualizations for use in browsers are provided by the server as well, however the user can choose to build their own as well.

• Alexandria - AlexandriaCloud - Thematic Cluster WA - enables virtual access to the Alexandria infrastructure. The Alexandria infrastructure is a Hadoop cluster provides all tools that are necessary for indexing, accessing, and analyzing

the German Web corpus. The Alexandria infrastructure is under development since March 2014 will be available at the latest end of 2014. As part of the ERC Grant Alexandria (ERC 339233) additional methods for the temporal analytics will regularly be added.[Papadakis 2012].
Services offered: it is currently under preparation and available at the end of 2014. Alexandria will provide the state of the art of Web analytic services on the German Web corpus, originally crawled by the Internet Archive (approx. 80TB). Due to copyright and licensing restrictions virtual access can only be given to extracted meta-information from the collection. By using Apache Pig in conjunction with Mahout and Giraph on this information, we will enable efficient analysis across time. Other services that will be made available are domain-based search via the Wayback machine, full-text search on extracted information, or the creation of topical sub-collections.

Description of work and role of partners

WP7 - VA1_Virtual Access [Months: 13-48]
CNR, USFD, UNIPI, Fraunhofer, LUH
Description of work
Modality of access under this project:
This WP will operate a Virtual Access e-Infrastructure (built in WP10) enabling sharing and reuse of existing and newly created social data mining resources. The e-Infrastructure is built as a "system of systems", i.e. it is equipped with a number of middleware services (called "enabling services": registry, subscription and notification of resource-related events, workflow management) for federating the resources offered by the infrastructures described above. The e-Infrastructure will provide access to: i) the catalogue of SoBigData resources; ii) the set of SoBigData Virtual Research Environments available and continuously updated; iii) the set of social networking facilities promoting collaboration with other users, e.g. enabling them to post or be notified about the availability of new resources, about the willingness to perform a given investigation, or to ask for expert supports and comments on a given idea or issue.
The resources catalogue provides users with a comprehensive view of the resources available within the e-Infrastructure. Each resource is equipped with a rich set of metadata capturing resource capabilities (e.g. methods), distinguishing features (e.g. temporal coverage), and usage policies (e.g. open access) according to data management plain in WP8 and methods' specification WP9. The catalogue enables two forms of discovery: a query based discovery and a subscription/ notification discovery. Moreover, it is oriented to serve both human users and non-human users, thus it is offered through a web based protocol.
Virtual Research Environments (VREs) are web-based working environments equipped with a number of applications enabling scientists to have access to the set of data, services, and algorithms needed to perform their investigation in a collaborative way. To this end, VREs embed the social networking facilities mentioned above, are created by aggregating resources federated via the underlying e-Infrastructure, and are available to the users.
The operation of the VA e-Infrastructure regards the efforts of deployment, maintenance, administration, and monitoring of a running production system including (1) the enabling services overlay, (2) the web portal and (3) the services providing social data storage and management functionality available at the different infrastructure sites. Such efforts are continuous along the life of the project and will be done in synergy with WP10 which is responsible of the planning and releasing of the e-infrastructure and VREs.

Support offered under this project:
The e-Infrastructure portal is equipped with online documentation and wizards guiding a user to help him/her exploit the offered facilities. Moreover, an help desk will be set up where users can post their requests for support that will be managed by the Virtual Access team. In addition to that, and thanks to the social networking facilities, users can post requests for support through the portal. The metadata offered via the resources catalogue contain both links to any documentation available and to contact persons per resource.

Outreach to new users:
This WP will liaise with dissemination and outreach WPs to attract new potential users for local infrastructure sites. To this aim, it will have a dedicated area in the project web site aiming at offering an up to date description of the e-Infrastructure capacities (e.g. number and typologies of resource federated) and its current exploitation (e.g. number of Virtual Research Environments created, number of registered users, outstanding results, latest posts). Moreover, scientists can register their resources and make them available to other scientists to discover, access, and discuss. Notification mechanisms can proactively inform interested (subscribed users) of the registration (hence availability) of new resources in the e-infrastructure.

In addition to that, the facility enabling user communities to ask for the creation of their own VRE is a very distinguishing feature that will be promoted to attract new clients and widening the exploitation of the community resources.

Review procedure under this project:
The VA e-Infrastructure collects a continuous and very detailed record of resources usage indicators, e.g. number of accesses, numbers of "likes". These indicators are immediately available to Virtual Access team through a web based console integrated in the web portal. With a bimonthly-based periodicity, aggregated data on resources usages will be posted on the web-based portal thus to make it possible for the scientists to assess them in a very open and collaborative way (social networking facilities will be used). In addition to this sort of open and community-based self-assessment done on aggregated data, the WP will set up and maintain a dedicated wiki page for European Commission to have access to the detailed indicators (D7.1).
Task leader: CNR. Participants: UNIPI, USFD, LUH, FRH.

Participation per Partner

Partner number and short name	WP7 effort
1 - CNR	9.00
2 - USFD	7.00
3 - UNIPI	4.00
4 - Fraunhofer	3.00
7 - LUH	5.00
Total	28.00

List of deliverables

Deliverable Number [14]	Deliverable Title	Lead beneficiary	Type [15]	Dissemination level [16]	Due Date (in months) [17]
D7.1	VA e-Infrastructure Service Provision and Operation Report 1	1 - CNR	Report	Public	18
D7.2	VA e-Infrastructure Service Provision and Operation Report 2	1 - CNR	Report	Public	36
D7.3	VA e-Infrastructure Service Provision and Operation Report 3	1 - CNR	Report	Public	48

Description of deliverables

Virtual Access e-Infrastructure operation reports.

D7.1 : VA e-Infrastructure Service Provision and Operation Report 1 [18]

Report on the e-Infrastructure operation activity since its last deployment, including a detailed set of usage indicators (e.g. number of resources, accesses to resources, usage of resources from scientists of different infrastructures, etc.). Indicators will be generated automatically by the VA e-infrastructure and their current status can be consulted at any moment from the VA e-infrastructure portal. This deliverable will include the assessment report provided by Project Advisory Board.

D7.2 : VA e-Infrastructure Service Provision and Operation Report 2 [36]

Report on the e-Infrastructure operation activity since its last deployment, including a detailed set of usage indicators (e.g. number of resources, accesses to resources, usage of resources from scientists of different infrastructures, etc.).

Indicators will be generated automatically by the VA e-infrastructure and their current status can be consulted at any moment from the VA e-infrastructure portal. This deliverable will include the assessment report provided by Project Advisory Board.

D7.3 : VA e-Infrastructure Service Provision and Operation Report 3 [48]

Report on the e-Infrastructure operation activity since its last deployment, including a detailed set of usage indicators (e.g. number of resources, accesses to resources, usage of resources from scientists of different infrastructures, etc.). Indicators will be generated automatically by the VA e-infrastructure and their current status can be consulted at any moment from the VA e-infrastructure portal. This deliverable will include the assessment report provided by Project Advisory Board.

Schedule of relevant Milestones

Milestone number [18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS2	SoBigData eInfrastructure and training launch	11 - ETH Zurich	12	- Repository of training materials established - Innovation and first services Integration - Ethical issues addressed - 2nd SoBigData Workshop - e-Infrastructure Deployment - Resource catalogue VRE - Completion of first series of educational initiatives: Datathon 1, Training HS 1 - e-Infrastructure Production Release
MS3	Analytics service integration and data ecosystem creation	1 - CNR	24	- Services and Infrastructure Integration - Big Data Ecosystem and Analytics - 1st SoBigData Conference - Completion of second series of educational initiatives: Summer School 1, Training HS 2 & Datathons 2 - e-Infrastructure Enhanced with new services - e-Infrastructure Production Release
MS4	Stepping up TA, VA, and networking activities, and community engagement	2 - USFD	36	- Analytics Services and Infrastructure Integration updated - 3rd SoBigData Workshop - Completion of third series of educational initiatives: Summer School 2, Training HS 3 & Datathons 3 - Use-cases VREs

Schedule of relevant Milestones

Milestone number [18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS5	SoBigData infrastructure - final deployment	1 - CNR	42	- e-Infrastructure Production Release
MS6	Starting SoBigData Community Established	4 - Fraunhofer	48	- 4th SoBigData Workshop - Final SoBigData Conferences - Completion of last series of educational initiatives: Summer School 3, Datathons 4 - Use-cases VREs - Final evaluation and project end

Work package number [9]	WP8	Lead beneficiary [10]	7 - LUH
Work package title	JRA1_Big Data Ecosystem		
Start month	1	End month	39

Objectives

Research on social mining relies on massive data sets of digital traces of human activities. Many big data sets are already available at the proposer's labs (see Section 2.2.2 for a detailed table), including call graphs from mobile phone call data, networks crawled from many online social networks, including Facebook and Flickr, transaction micro-data from diverse retailers, query logs both from search engines and e-commerce, society-wide mobile phone call data records, GPS tracks from personal navigation devices, survey data about customer satisfaction or market research, billions of tweets, and data from location-aware social networks. The partners will make such data available for collaborative research, by adopting various strategies, ranging from sharing the open data sets with the scientific community at large, to sharing the data with disclosure restriction within the consortium, also on a bilateral basis, or allowing data access within secure environments at each local installation.
In addition, the consortium will develop methods and ad-hoc campaigns for logging and collecting the digital footprints of multiple social dimensions generated by humans, when interacting in the physical environment by exploiting pervasive sensors. These will take the form of a wide variety of smart devices spread in the environment such as RFID readers, sensors and actuators, and sensor-rich smart phones as well as crowd-sensing campaigns leveraging participation of users as data producers. The access under VA and TA offered in SoBigData will concern both existing and newly collected datasets. The access trough VA will be granted for all those datasets whose policies allow open diffusion; conversely, for all the data set whose access is restricted due to licensing restrictions, access will be provided only through TA. Moreover, for some datasets, in order to avoid Term of Usage (ToS) infringements, access will be offered only through Virtual Data Collections (see T8.1) both in VA and TA. All the results of this WP will be served as input to T10.1 and integrated into the SoBigData RI.

Description of work and role of partners

WP8 - JRA1_Big Data Ecosystem [Months: 1-39]
LUH, CNR, USFD, UNIPI, Fraunhofer, UT, IMT LUCCA, KCL, SNS, AALTO UNIVERSITY, ETH Zurich
T8.1 Data Management and Integration Plan
The purpose of the data management plan is to define policies that will guide the partners in the collection, description, preservation and sharing of their data sets for VA and TA.
The description of the data set includes a unique reference, their nature, scale and available metadata (such as related scientific publications or similar resources). Moreover, the preservation procedures will describe how the partners store the data, which technology is used, and how long the data will be available.
This task will also connect to other relevant infrastructures and other social data providers as presented in the table in Section 1.3.3. Relevant data from their archives will be imported into SoBigData, after analysing the relevant APIs and developing mappings into the SoBigData RI. Special attention will be paid to a general approach that describes the interfaces between the distributed collections and the registry and the connectors between them.
Finally, a sharing restriction will be specified for each dataset among the following categories:

1. Publicly Available Data. Several datasets are made available by private and public entities (e.g. GeoLife by Microsoft, demographic data from public administration) and will be included in the list of resources of the project. The information about these datasets is not always complete or well described. For this reason a platform aimed at easily finding, annotating and discussing them within the research community is needed and will be part of the project as described in T8.4.

2. Restricted data. Within the consortium, we will make available proprietary datasets (i.e. GPS, GSM and Retail Market ones). Due to the restriction imposed by data owners, such collections will be made available prevalently on-site, through Transational Access. In such scenarios, the consortium will take care of simplifying the procedures needed to grant data access to the researchers who want to pursue experiments on them.

3. Virtual Collections from Social Media. Online human activities represent the majority of social data: Online Social Networks, Search Engines, as well as online Libraries are semantically rich data collections for data scientists. However, collecting and consolidating this kind of data is not a straight-forward task due to the various restriction, often imposed by their right holders and aimed at securing the valuable information their services supply. In order to comply with

increasingly stringent ToS (which frequently impose online analysis, denying the storage of raw data), within the consortium is planned to make available an ensemble of Virtual Data Collections (VDC) representing already analyzed data sets extracted by the major online services. A VDC is composed by the set of unique identifiers needed to query a specific online service, in order to build a specific, and already analyzed, data set. VDCs enable researchers to make analysis, which will be comparable with previous work without the need to store an entire data set, and thus without violating the ToS of the studied services.

The WP will also monitor that all the partners comply with the sharing restrictions declared for their data and will take care of the costs associated for their long-term preservation. The available and newly gathered data sets will be registered in the e-infrastructure following the specification defined in T10.2.
Task leader: CNR. Participants: ALL PARTNERS

T8.2: Participatory & Opportunistic Crowd Sensing
In the paradigm of crowd-sensing, the crowd becomes a distributed network of sensors that allows us to understand real life events at a quasi-real-time rate. In participatory sensing, individuals are actively involved in contributing sensor data (e.g. taking a picture, reporting a road closure). Opportunistic sensing is where the sensing is more autonomous and user involvement is minimal (e.g. continuous location sampling). Activities in this task will be focused on implementing the paradigm of crowd-sensing in these two distinct types of sensing. Regarding the opportunistic version, the infrastructure will offer methodologies for sensing the crowd during ongoing events (e.g. natural disasters). In this way it will be possible to retrieve data useful for the characterization of the event and the assessment of the consequences. For the participatory sensing the infrastructure will organize and support initiative aimed at stimulate user engagement for the collection of social data. In this way it will be possible to collect high quality, semantic enriched, data directly provided by aware human sensors.
Task leader: CNR. Participants: USFD

T8.3: Analytical Crawling
This task will develop and provision an adaptable integrated Web crawler, which is easy to use and supports the special requirements of data scientists. The flexible selection of crawling strategies will allow the different kinds of crawls (e.g. shallow observation crawls; focused; topical; or event crawls). Based on a semantic crawl specification, content analytics is used for guiding the crawler and to decide about relevant content. For achieving high coverage and quality, the crawler collects Web, Social Web and related Linked Data content in an integrated way. To achieve scalability, the analytical crawling framework will build upon Apache Nutch, a highly extensible and scalable open source Web crawler based on Hadoop. The analytical crawler will support intelligent crawl specifications, application aware link extraction for blogs and social media sites, fast semantic and Social Web analysis, entity and social graph generation from extracted contents, and provenance of information.
Task leader: LUH. Participants: CNR

T8.4 Integrating Open Data through the Living Archive platform.
Living Archive is a search engine for Open Data and has a community-run catalogue of interesting open datasets available on the web. It already implements many aspects of the objective to integrate large social data sets to overcome fragmentation and to provide interoperability. The aim of this task is to further extend the functionalities, with the purpose of including annotation tools, lenses to highlight specific social aspects, user-defined filters, data citation systems, and new dataset alert mechanisms. The Living Archive will support: a virtual centralized view of data distributed across remote repositories, integration of machine learning and text analysis methods (see WP9) to deal with incomplete or uncertain information in social datasets, and integration with complex network analysis tools from WP9 to understand the links and inter-dependencies within social data.
The Living Archive datasets catalogue will be integrated within the e-Infrastructure resource catalogue described in D10.6, following the guidelines described in D10.1.
Task leader: ETHZ. Participants: CNR.

Participation per Partner

Partner number and short name	WP8 effort
1 - CNR	18.00
2 - USFD	8.00

Partner number and short name	WP8 effort
3 - UNIPI	2.00
4 - Fraunhofer	7.00
5 - UT	4.00
6 - IMT LUCCA	4.00
7 - LUH	9.00
8 - KCL	4.00
9 - SNS	4.00
10 - AALTO UNIVERSITY	7.00
11 - ETH Zurich	8.00
Total	75.00

List of deliverables

Deliverable Number [14]	Deliverable Title	Lead beneficiary	Type [15]	Dissemination level [16]	Due Date (in months) [17]
D8.1	Data Management report	1 - CNR	Websites, patents filling, etc.	Public	3
D8.2	Crowd sensing platform	1 - CNR	Other	Public	12
D8.3	Analytical Crawling Platform	7 - LUH	Other	Public	24
D8.4	Living Archive Platform	11 - ETH Zurich	Other	Public	12

Description of deliverables

Data Management and Crowd sensing reports.

D8.1 : Data Management report [3]

The purpose of this deliverable is to provide an ongoing and up to date wiki containing the description of the available datasets in the consortium. The description includes statistics, metadata, sharing policies and archiving technologies as well as the preservation provisions and lifespan.

D8.2 : Crowd sensing platform [12]

The methods and technologies to collect data from the participatory and\or opportunistic crowd sensing.

D8.3 : Analytical Crawling Platform [24]

The final analytical crawling platform will be delivered and described with all its functionalities.

D8.4 : Living Archive Platform [12]

The extended Living archive platform will be delivered and described with all its functionalities.

Schedule of relevant Milestones

Milestone number [18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS3	Analytics service integration and data ecosystem creation	1 - CNR	24	- Services and Infrastructure Integration - Big Data Ecosystem and Analytics - 1st SoBigData Conference - Completion of second series of educational initiatives: Summer School 1, Training HS 2 & Datathons 2 - e-Infrastructure Enhanced with new services - e-Infrastructure Production Release

Work package number [9]	WP9	Lead beneficiary [10]	3 - UNIPI
Work package title	JRA2_Integrating Big Data Analytics Methods and Techniques		
Start month	1	End month	39

Objectives

This work package will integrate and advance the methods for mining social big data, within the SoBigData research e-Infrastructure. The expected outcomes are methods that are simultaneously principled (carry strong theoretical guarantees) and practically applicable, providing a broad statistical foundation for Big Data Analytics. The results of WP9 will endow and progressively enrich the RI, and thus the offerings of WP6 TA and WP7 VA, with interoperable scalable methods and services for analysing large-scale social data, as needed to produce comprehensive analyses of complex socio-economic phenomena.

The specific objectives are:
• To enhance the repertoire of efficient multi-dimensional analytical methods and algorithms for social mining.
• To integrate, systematize and make accessible analytical processes to understand social phenomena (e.g., personal behavioural profiles, the collective behavioural patterns, global patterns of diffusion and social influence, patterns of how sentiments and opinions vary in our societies.)

Description of work and role of partners

WP9 - JRA2_Integrating Big Data Analytics Methods and Techniques [Months: 1-39]
UNIPI, CNR, USFD, Fraunhofer, UT, IMT LUCCA, LUH, KCL, SNS, AALTO UNIVERSITY, ETH Zurich
T9.1. Social Network Analysis
Big social data often include networks with rich semantics, along several social dimensions. We will develop a framework of multi-dimensional/multi-layer network analytics, which integrates and enhances the plurality of methods of the national infrastructures (see WP6 for details), according to the following intertwined lines:

• Community discovery. Along this line, a variety of community discovery, link prediction and ranking techniques will be offered, which deal with multiple kinds of networks. Methods for discovering the correlation between personal user (ego) networks and community structures at global network level will also be integrated. This will allow to study how communities are formed from personal networks. Moreover, methods for exploiting the structure of communities and ego networks will be integrated, in order to learn and classify user behavioural patterns, such as the usage of online services. Particular focus will be on predicting changes in behavioural patterns by actors in a social network, through exploiting latent information within the actor's communities structure.

• Evolutionary pattern discovery. The infrastructure will be enhanced with algorithms to detect dynamic communities in social networks, based on consolidated frameworks (e.g. evolutionary and consensus clustering). The challenges to be addressed are in identifying meaningful snapshots, the development of suitable benchmarks, the identification of relevant time scales, and tracking of clusters over time, the interplay between information diffusion and community structure.

• Multi-dimensional/multi-layer network analysis. Extend and enrich analytics to deal with and exploit the semantic information attached to nodes and edges, which gives rise to multiple dimensions or layers of social networks. Along this line, a variety of methods and models for community and pattern discovery, ranking, search, analysis of social influence and diffusion, network dynamics will be enhanced towards multi-dimensionality. For the analysis of large socio-economic networks, in particular, we will integrate methods for Granger causality networks, where the presence and directionality of the link gives information on causality relations.

• Similarity-based networks. Analytic tools will be provided for the specific case when the presence of a link quantifies the degree of similarity between the entities represented by the connected nodes. A typical and important example is given when nodes represent time series of the activity of individual agents and the similarity is, for example, the correlation coefficient. We will also investigate the multi-layer and dynamic extension of these networks.
Task leader: AALTO. Participants: CNR, UT, SNS, IMT, UNIPI.

T9.2. Social Pattern Discovery

T9.2 will enhance the SoBigData e-Infrastructure with new methods and services for discovery of emergent patterns and rules, describing socio-economic behaviour of interesting sub-populations. In particular, methods for multi-dimensional evolutionary pattern discovery capable of connecting behaviours across different social dimensions: mobility, lifestyle, shopping behaviour, desire, opinions. The main challenges come from the integrated analysis of structured and unstructured data within the SoBigData infrastructure (e.g., social networks, mobility, text) and its use for scalable local pattern discovery. Visual analytics tools, such as FRH's One-Click-Mining for interactive pattern discovery, will be the basis for intuitive, visually supported co-active learning and mining process utilizing user feedback. The social pattern discovery methods will be used to classify and predict user behaviour, e.g., users' propensity for a given online service, based on correlating behavioural patterns with patterns extracted from user profiles and their social ties. To this end, in collaboration with T9.1, we shall integrate and extend methods for the discovery of patterns in large (social) networks (for example network motifs), a problem that critically depends on the null model for the network. Efficient methods to construct null models for large networks are very challenging. We will integrate computationally efficient methods to build null models based on the Maximum Entropy Principle, and then use these null models to mine interesting patterns in large socio-economic networks.
Task leader: UNIPI. Participants: CNR, SNS, FRH.

T9.3 Text mining and semantic annotation of big social data.
This task will enhance the SoBigData e-infrastructure and the underlying GATE TSMM infrastructure, by integrating a wide range of methods for clustering, mining and semantic analysis of unstructured social content, especially of short length such as Facebook posts, tweets, blogs, and even search-engine queries. In more detail, UT will enhance and integrate methods for feature-maximizing clustering of large-scale datasets, which are applicable to entity disambiguation in social networks and heterogeneous text clustering tasks. Another contribution will be the integration of UNIPI's widely used TagMe topic annotator, which will be also be delivered as a service for virtual access in WP7. We will also integrate Dexter (http://www.dxtr.it), the open source framework for experimenting with text annotation and entity disambiguation components. UNIPI will also integrate algorithms for space-time analysis of topics and actors in social media. Deep learning techniques for dealing with noisy social data will also be investigated. T9.3 will apply these integrated text and social media mining methods for entity-centric contextualization of social media content, Web archives and other textual content. A specific dimension of content contextualization is dealing with the strongly multilingual nature of social data. This will be addressed through enhancing GATE's entity recognition, linking, and semantic enrichment algorithms towards new European languages. Addressing multilingual content analytics is essential, since it brings better understanding of multicultural events and topics, offers complementary information and indicates contradictory views and statements supporting pluralism within the European society.
Task leader: USFD. Participants: KCL, LUH, CNR, UNIPI, UT.

T9.4 Mobility data analytics.
T9.4 will advance and integrate in SoBigData a rich repertoire of methods and services for the analysis of human mobility, based on the world-leading expertise of CNR, UNIPI, FRH and AALTO in mobility data mining. Methods will range from advanced mobility data mining algorithms (such as trajectory clustering, trajectory patterns, location prediction, semantic annotation, mobility network analysis) to complex mobility analytical processes, such as the estimation of origin-destination matrices, the visual exploration of mobility patterns through the Urban Mobility Atlas, the inference of the geographical borders of human mobility, the correlation between mobility patterns and social events or trends, the inference of statistically significant demographic information such as population of city users and flows from mobility data, the identification of unusual presence and mobility patterns.
Task leader: CNR. Participants: UNIPI, FRH, AALTO.

T9.5 Visual analytics.
T9.5 will advance and integrate in SoBigData a suite of scalable visual analytics methods for interactive visually-driven context-aware analysis of mobility and social network data, as well as visual exploration and interpretation of the results from the text analytics and social pattern discovery methods and services. FRH IGD will provide its visual analytics server (IVAS) as an integration platform for visual analytics tasks. The platform handles algorithmic tasks which need to be controlled directly by the user. IVAS provides support for data access, data transformations and mining as well as visualizations. FRH IAIS will develop and integrate methods to support understanding of the temporal aspect of the data and link visual analytics to social network analysis. Firstly, we will adapt methods for continuous GPS tracking to the specifics of longitudinal records of social network activities (e.g. twitter messages, shared photos). Secondly, FRH will develop and integrate novel methods for context-aware analyses of spatio-temporal data, which combine data records with additional knowledge about user location (e.g. extracted by text analytics methods from T9.3) and time-specific data, e.g. weather conditions, public events. In other words, these methods will enable joint, interactive analysis of mobility, social network, and social media data. CNR will advance and integrate the visual tool for mobility

analytics of MAtlas, in particular the Urban Mobility Atlas visual analytics system based on large GPS datasets, pursuing interoperability with the FRH IVAS platform.
Task leader: FRH. Participants: CNR.

T9.6 Social behaviour and human dynamics.
T9.6 will enhance the SoBigData infrastructure with methods and services for measuring and understanding social behaviour and human dynamics, as follows.

• Individual profiling: Mining the multi-faceted personal data of individual citizens enables the discovery of personal routines, diaries or profiles. While profiling is typically used for targetted marketing and recommendation, its generalization for social mining, within an ethical framework, will foster understanding of social behaviour, in that it represents a trade-off between the oversimplification of global measurements and the excessive diversity of individual behaviours. The methods will build on the social media mining methods from T9.3.

• Learning and monitoring the evolution of collective sentiments/opinions: New techniques of "quantification via classification" for detecting collective sentiment from surveys and social media data will be integrated. Quantification is the task of estimating correctly the distribution of the data items across the classes.

• Predictive analytics for social measurements: statistical learning and nowcasting models for the estimation of a wide range of phenomena from social media data (e.g., well-being or employment indicators, political sentiment and election outcomes, assessing the damage caused by a natural disaster).

• Models for multi-dimensional diffusion processes: Existing diffusion models will be generalized to multi-dimensional social networks, in order to study (i) patterns of information diffusion in evolving social networks, (ii) impact of multi-dimensionality on diffusion speed, virality and leadership. We will also provide tools for planning optimised information diffusion in human networks and for understanding the correlation between information diffusion and trust in social networks.

• Correlation between social structures in the cyber and physical world. We will advance and integrate methods to measure and understand the correlation between social structures that human maintain online (e.g., using Online Social Networks) and offline (e.g., through physical encounters). This will provide tools to better understand the specific role of cyber environments in social dynamics, and how to exploit them to facilitate social interactions.
Task leader: CNR. Participants: UNIPI, ETHZ, FRH, UT, AALTO.

T9.7 Engineering Massive Scale Social Mining.
We will address the major computational challenges of big data analytics and social mining, such as the high volume and dimensionality of data, the dynamics and velocity of information generation and the requirements of real-time reactions to changes in the environment. Given the variety of data sources used and the diversity of analysis conducted, the infrastructures will integrate different state-of-the-art solutions for high performance data mining and big data analytics services on dedicated architectures, including: large Hadoop platforms exploiting MapReduce, pig, Mahout, Spark; scalable storage services exploiting noSQL databases based on HBase, MongoDB, tokio cabinet, etc; specialized big-graph scalable algorithms using giraph, scala, hama; Storm-based streaming and concept-drifting analytics services to deal with extremely fast and large streams of data, e.g., Twitter Firehose. T9.7, in close collaboration with T2.4, will also pursue the objective that any analytical method developed and integrated in the infrastructure is designed and engineered according to the privacy-by-design principle, i.e., in such a way that the privacy requirements are embedded natively into the method. This activity builds on the vast experience in privacy preserving data mining of UNIPI and CNR and in value-sensitive-design of TUDelft.
Task leader: CNR. Participants: UNIPI, TUDelft, AALTO, LUH, UT.

Participation per Partner

Partner number and short name	WP9 effort
1 - CNR	22.00
2 - USFD	13.00
3 - UNIPI	10.00
4 - Fraunhofer	16.00

Partner number and short name	WP9 effort
5 - UT	8.00
6 - IMT LUCCA	5.00
7 - LUH	6.00
8 - KCL	11.00
9 - SNS	5.00
10 - AALTO UNIVERSITY	12.00
11 - ETH Zurich	11.00
Total	119.00

List of deliverables

Deliverable Number [14]	Deliverable Title	Lead beneficiary	Type [15]	Dissemination level [16]	Due Date (in months) [17]
D9.1	Social mining method and service integration 1	10 - AALTO UNIVERSITY	Other	Public	12
D9.2	Social mining method and service integration 2	1 - CNR	Other	Public	24
D9.3	Social mining method and service integration 3	1 - CNR	Other	Public	36

Description of deliverables

Reports on Social mining methods and service integration.

D9.1 : Social mining method and service integration 1 [12]
Design, integration specifications, and early prototypes.

D9.2 : Social mining method and service integration 2 [24]
Consolidated design and realization of first batch of integrated methods and services.

D9.3 : Social mining method and service integration 3 [36]
Consolidated design and integration of enhanced methods and services.

Schedule of relevant Milestones

Milestone number [18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS3	Analytics service integration and data ecosystem creation	1 - CNR	24	- Services and Infrastructure Integration - Big Data Ecosystem and Analytics - 1st SoBigData Conference - Completion of second series of educational initiatives: Summer

Schedule of relevant Milestones

Milestone number [18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
				School 1, Training HS 2 & Datathons 2 - e-Infrastructure Enhanced with new services - e-Infrastructure Production Release

Work package number [9]	WP10	Lead beneficiary [10]	1 - CNR
Work package title	JRA3_SoBigData e-Infrastructure		
Start month	1	End month	39

Objectives

The goal of this work package is to support the creation and operation of the SoBigData e-infrastructure that ingrates new and existing national infrastructures and services, and quantitatively and qualitatively enhances them. The activities comprise: (i) studies and definition of best practices/policies for the harmonization of federated resources available at the local infrastructure sites; (ii) adaptation of the existing resources to the identified best practices; and (iii) realization of VREs supporting scientists in benefitting from the integration of the federated resources and infrastructures.

Description of work and role of partners

WP10 - JRA3_SoBigData e-Infrastructure [Months: 1-39]
CNR, USFD, UNIPI, Fraunhofer, UT, IMT LUCCA, LUH, KCL, SNS, AALTO UNIVERSITY, ETH Zurich
This WP will create the SoBigData RI that will make the data and methods of WP8 and WP9 available to users through VA access in WP7. The first Virtual Research Environment (VRE) that will be developed in the e-infrastructure is the resource catalogue, implementing a smart search tool for the different kinds of resources available. New VREs will be created using the e-infrastructure workflow specification language, which describes data processing research activities and cross-relating resources available at different sites. The users will be able to define new VREs and share them with the community, thus enriching the system.

T10.1 e-infrastructure interoperability
Definition of the e-infrastructure resource model: types of resources (e.g. datasets, services, algorithms) and metadata properties of the resources (e.g. classifications, access parameters, hardware parameters, IPR issues, input/output data types). Specification of data models and formats for data exchange (e.g. data models as service parameters). Specification of common service APIs to enable service pipelining, subscription and notification, access to service functionality independently from implementation. Establishment of resource participation policies by resource typology, e.g. service SLA, QoS, quality and size of data.
These activities are continuous over the project lifetime and are performed in strict cooperation with WP8 and WP9, starting from their current status in existing national infrastructures (see WP7) and different research areas in order to align the community towards common and agreed-upon best practices.
Task leader: CNR. Participants: ALL PARTNERS.

T10.2 Integration to the e-infrastructure.
In this task, following the directives of T10.1, all WP10 partners will adapt the resources (both existing and newly enhanced in this project) available at their local infrastructures (i.e. services, algorithms, datasets), in order to permit their registration and re-use in the SoBigData e-infrastructure. In other words, this means that all datasets and tools described in WP8 and WP9 must be transformed into e-infrastructure resources and must be compliant with its specifications. Periodically a plan describing the set of actions needed to release the subsequent e-infrastructure version is reported in D10.2. Thanks to the e-infrastructure, every time a new resource is added, it will be accessible and may be used also in previously defined VREs, if it fits the requirements. An example of VRE is the “trajectory prediction process” describing the methodology for predicting the future position of users, given their current movement. This VRE is defined by the workflow specification language using a generic “prediction algorithm” operating on a specific data format. This means that the method will be discovered and instantiated dynamically by asking the scientist using the VRE to select it from the resources available in the e-infrastructure. In other words the e-infrastructure will provide a list of prediction algorithms suitable for the VRE, which may contain new services, which have been added to the RI after the VRE itself was defined.
A periodic task of this WP will be the description of the functionalities of the latest e-Infrastructure release which will be described in D10.4.
Task leader: CNR. Participants: ALL PARTNERS.

T10.3 e-infrastructure workflow specification language
Definition and realization of a workflow specification language able to express data processing workflows by taking full advantage of the local properties and features of SoBigData resources as defined in T10.1. The language will be inspired by state-of-the-art workflow languages (e.g. BPEL), but be strongly characterized by the resource model defined in

T10.1, through embedding in its logic the nature of social data analysis context. Its objective is to enable the definition of data-driven workflows in terms of combination of SoBigData resources, e.g. pipelining of services, execution of algorithms, and selection of data for services and/or algorithms. Therefore the workflow specification language will be used:
• to glue the different approaches to modelling together into a coherent framework;
• to establish the open platform principle, where scientists, developers as well as many other kinds of users can integrate and compose deep models, big-data-driven analytics to create new VREs;
• to design a rich repertoire of VREs to be shared between the different kinds of users supporting both (i) the expert data scientist or developer that can unleash the whole power of the framework and (ii) the human scientist or the policy maker that wants to perform complex simulation and what-if analyses with simple and reliable means.
The language will also be equipped with optimization tools, capable of validating the specified execution and proposing a more efficient surrogate based on metadata properties of datasets, services, and algorithms. The activities of this task will serve as feedback to the harmonization work of T10.1.
Task leader: CNR. Participants: --

T10.4 e-infrastructure VREs.
As mentioned in WP7, VREs are the means to deliver scientists the possibility to register/discover resources and configure/execute complex data processing workflows, via intuitive web user interfaces. Besides, they offer social networking tools allowing discussions about the overall research activities, e.g. registration of a resource, workflow executions, enquiries. This task will initially support the realization of the “resource catalogue” VRE that offers facilitated access to the e-infrastructure registry. It exposes social-oriented user interfaces enabling scientists to (i) (de-)register, discover and be notified of resources and (ii) opening discussion threads about resources or research activities. In the course of the project, other experiment-oriented VREs will be realized and deployed in cooperation with all the partners as real-case scenarios of cross-service integration. An example of a potential VRE is a text mining “benchmarking” suite. The VRE provides scientists with: (i) resource discovery functionalities for selecting a set of algorithms for text mining and a test dataset, (ii) select performance indicators, (iii) run the algorithms and visualize the results, and (iv) post the results and open a discussion thread via social networking tools.
Task leader: CNR. Participants: ALL PARTNERS.

Participation per Partner

Partner number and short name	WP10 effort
1 - CNR	16.00
2 - USFD	8.00
3 - UNIPI	8.00
4 - Fraunhofer	8.00
5 - UT	4.00
6 - IMT LUCCA	4.00
7 - LUH	5.00
8 - KCL	4.00
9 - SNS	6.00
10 - AALTO UNIVERSITY	7.00
11 - ETH Zurich	7.00
Total	77.00

List of deliverables

Deliverable Number [14]	Deliverable Title	Lead beneficiary	Type [15]	Dissemination level [16]	Due Date (in months) [17]
D10.1	Best practices and guidelines towards interoperability	1 - CNR	Websites, patents filling, etc.	Public	6
D10.2	SoBigData e-Infrastructure release plan 1	1 - CNR	Report	Public	6
D10.3	SoBigData e-Infrastructure release plan 2	1 - CNR	Report	Public	20
D10.4	SoBigData e-Infrastructure release plan 3	1 - CNR	Report	Public	38
D10.5	SoBigData e-Infrastructure software release 1	1 - CNR	Demonstrator	Public	11
D10.6	SoBigData e-Infrastructure software release 2	1 - CNR	Demonstrator	Public	23
D10.7	SoBigData e-Infrastructure software release 3	1 - CNR	Demonstrator	Public	39
D10.8	Resource adaptation to register to the e-infrastructure 1	1 - CNR	Other	Public	18
D10.9	Resource adaptation to register to the e-infrastructure 2	1 - CNR	Other	Public	36
D10.10	Resource adaptation to register to the e-infrastructure 3	1 - CNR	Other	Public	39
D10.11	Data processing workflow specification language	1 - CNR	Websites, patents filling, etc.	Public	12
D10.12	VRE specification and software 1	1 - CNR	Other	Public	12
D10.13	VRE specification and software 2	1 - CNR	Other	Public	36
D10.14	VRE specification and software 3	1 - CNR	Other	Public	39

Description of deliverables

Repots on SoBigData e-Infrastructure release plan.

D10.1 : Best practices and guidelines towards interoperability [6]

The deliverable has an “on-going” nature and organizes and presents the results of T10.1. It will serve as a deliverable and as release-driven guidelines for all SoBigData e-infrastructure stakeholders.

D10.2 : SoBigData e-Infrastructure release plan 1 [6]

Describes the set of actions to deliver subsequent releases of the e-Infrastructure as due by the milestones MS2, MS3 and MS5. Plans will be drafted in synergy with all WP8 and WP9 partners, and will drive the incremental integration of existing services in the e-infrastructure, resulting in the software releases indicated in D10.3. Initial plans will include deployment of enabling services and the portal of the e-infrastructure.

D10.3 : SoBigData e-Infrastructure release plan 2 [20]

Describes the set of actions to deliver subsequent releases of the e-Infrastructure as due by the milestones MS2, MS3 and MS5. Plans will be drafted in synergy with all WP8 and WP9 partners, and will drive the incremental integration of existing services in the e-infrastructure, resulting in the software releases indicated in D10.3. Initial plans will include deployment of enabling services and the portal of the e-infrastructure.

D10.4 : SoBigData e-Infrastructure release plan 3 [38]

Describes the set of actions to deliver subsequent releases of the e-Infrastructure as due by the milestones MS2, MS3 and MS5. Plans will be drafted in synergy with all WP8 and WP9 partners, and will drive the incremental integration of existing services in the e-infrastructure, resulting in the software releases indicated in D10.3. Initial plans will include deployment of enabling services and the portal of the e-infrastructure.

D10.5 : SoBigData e-Infrastructure software release 1 [11]

It describes and delivers the software release planned in D10.2 and to be deployed as due by MS2, 3 and 5.

D10.6 : SoBigData e-Infrastructure software release 2 [23]

It describes and delivers the software release planned in D10.2 and to be deployed as due by MS2, 3 and 5.

D10.7 : SoBigData e-Infrastructure software release 3 [39]

It describes and delivers the software release planned in D10.2 and to be deployed as due by MS2, 3 and 5.

D10.8 : Resource adaptation to register to the e-infrastructure 1 [18]

It delivers and describes the software release of all the adaptations of existing and newly developed resources. It describes also all the activities carried out by the partners in order to register those resources to the e-infrastructure.

D10.9 : Resource adaptation to register to the e-infrastructure 2 [36]

It delivers and describes the software release of all the adaptations of existing and newly developed resources. It describes also all the activities carried out by the partners in order to register those resources to the e-infrastructure.

D10.10 : Resource adaptation to register to the e-infrastructure 3 [39]

It delivers and describes the software release of all the adaptations of existing and newly developed resources. It describes also all the activities carried out by the partners in order to register those resources to the e-infrastructure.

D10.11 : Data processing workflow specification language [12]

The deliverable defines the workflow specification language, together with use-cases, and documentation for users.

D10.12 : VRE specification and software 1 [12]

It describes and delivers the VREs’ software, to be deployed as due by MS2, MS4 and MS6. In particular the first will concern the resource catalogue.

D10.13 : VRE specification and software 2 [36]

It describes and delivers the VREs’ software, to be deployed as due by MS2, MS4 and MS6. In particular the first will concern the resource catalogue.

D10.14 : VRE specification and software 3 [39]

It describes and delivers the VREs’ software, to be deployed as due by MS2, MS4 and MS6. In particular the first will concern the resource catalogue.

Schedule of relevant Milestones

Milestone number [18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS2	SoBigData eInfrastructure and training launch	11 - ETH Zurich	12	- Repository of training materials established - Innovation and first services Integration - Ethical issues addressed - 2nd SoBigData Workshop - e-Infrastructure Deployment - Resource catalogue VRE - Completion of first series of educational initiatives: Datathon 1, Training HS 1 - e-Infrastructure Production Release
MS3	Analytics service integration and data ecosystem creation	1 - CNR	24	- Services and Infrastructure Integration - Big Data Ecosystem and Analytics - 1st SoBigData Conference - Completion of second series of educational initiatives: Summer School 1, Training HS 2 & Datathons 2 - e-Infrastructure Enhanced with new services - e-Infrastructure Production Release
MS4	Stepping up TA, VA, and networking activities, and community engagement	2 - USFD	36	- Analytics Services and Infrastructure Integration updated - 3rd SoBigData Workshop - Completion of third series of educational initiatives: Summer School 2, Training HS 3 & Datathons 3 - Use-cases VREs
MS6	Starting SoBigData Community Established	4 - Fraunhofer	48	- 4th SoBigData Workshop - Final SoBigData Conferences - Completion of last series of educational initiatives: Summer School 3, Datathons 4 - Use-cases VREs - Final evaluation and project end

Work package number [9]	WP11	Lead beneficiary [10]	4 - Fraunhofer
Work package title	NA5_Evaluation		
Start month	13	End month	45

Objectives

• To provide an evaluation framework for algorithms related to Social Big Data Analytics
• To develop test data sets, tasks and metrics and publish these to the wider academic and non-academic communities
• To engage a wide range of stakeholders in the evaluation campaigns and provide a sustainable future for them
• To create cross-disciplinary social mining exploratories, demonstrating the benefits of SoBigData's cross-disciplinary approach to analysing social phenomena (e.g. well-being, sustainable cities).

Description of work and role of partners

WP11 - NA5_Evaluation [Months: 13-45]
Fraunhofer, CNR, USFD, UNIPI, UT, IMT LUCCA, LUH, KCL, SNS, AALTO UNIVERSITY, ETH Zurich, TU Delft
Evaluation and subsequent demonstration is key to the success of any large-scale effort to mine knowledge and data, while evaluation frameworks are essential components of community infrastructures for data and knowledge mining. Evaluation campaigns, such as TREC, MUC, ACE, and TAC, have guaranteed over the years that the information retrieval and text analytics communities were driven forward. Other recent successful examples include INEX, which evaluates semi-structured retrieval methods; MIREX for music information retrieval, which has similar issues in terms of the legal rights to test collections; and CLEF for multilingual evaluation. Some current tasks relate to social media analytics (e.g. TREC microblog retrieval, RepLab on online reputation management) and WP11 will be complementary and coordinate with them.
All these evaluation frameworks have led not just to the development of better algorithms and strong metrics but have also managed to build a strong community around them.
To our knowledge, there is currently no comparable joint effort around analysing big social data. WP11 will create a secure and independently organised benchmarking and evaluation framework, with regular tasks and workshops to bring together the different research communities and enable the comparison of new algorithms. Only like this can we address the specific challenges of analysing big social data, such as the complex relationship of structured and unstructured information, new language combinations, complex proprietary data sets and licensing terms, evaluations on large scale datasets, creation of relevant gold standard datasets, which are currently largely missing or are of insufficient size. A key novelty will be also the cross-disciplinary nature of the target participant community.
To complement the evaluation activities and engage the wider community, a number of cross-disciplinary social mining demonstrators will be developed. They will showcase how the SoBigData research infrastructure can be harnessed to address key cross-disciplinary research challenges. This work package will serve primarily the main stakeholders of this project, the research and industrial community of social analytics, but will also produce a reference point for stakeholders from other research communities, who would like to get involved (e.g. economics, digital humanities).
WP11 will be lead by FRH, building on their expertise in evaluation.

T11.1 Establish connections with existing relevant evaluation initiatives.
Wherever possible the project will connect to existing evaluation initiatives such as CLEF and organise joint evaluation campaigns. This task will determine how existing initiatives can be linked to our evaluation efforts in T11.2. Where appropriate, we will propose our evaluation campaigns as specialised labs for established initiatives like CLEF, organise joint workshops, etc. Finally, we will deliver a mapping and gap analysis that will determine where our own efforts will be most effective.
Task leader: KCL. Participants: ALL PARTNERS.

T11.2. Evaluation framework for social data mining.
This task will establish a dedicated framework for mining big social data. It will develop a series of test collections and accompanying task definitions, needed for comparison of algorithms and approaches. A particular challenge will be to respect the terms of service social networks, which will generally mean that participants will be responsible for downloading their own versions of the social media content. T11.2 will develop however tools that merge automatically

the gold standard annotations with the participant downloaded datasets, to allow participants to create the complete training and evaluation corpora. Over the course of the project, we plan to run at least 2 meetings to bring together researchers, algorithms and data sets and to develop reliable evaluation metrics. The calls for these meetings as well as the competitions and tasks they involve will be distributed widely and well in advance.
Task leader: FRH. Participants: ALL PARTNERS.

T11.3 Sustainability of the framework and engagement of commercial providers.
After the end of the project, the evaluation framework will hopefully be one of its lasting impacts on the community, as CLEF has been over the years for multilingual information access researchers. T11.3 will continue the work on the evaluation framework by engaging a range of stakeholders, going beyond the research community from T11.2, and towards the wider stakeholder community and funders. We will leverage the innovation activities and stakeholder connections from WP5, to pursue sponsorship deals with relevant industry partners, funders and standardization bodies, to ultimately make this effort sustainable.
Task leader: ETHZ. Participants: ALL PARTNERS

T11.4 Cross-disciplinary social mining exploratories
This task will create vertical, thematic environments, called exploratories, on top of the SoBigData RI, for performing cross-disciplinary social mining research. Example exploratories to be pursued in synergy with WP8 and WP9 are: Measuring Well-being; Participatory E-Government; Sustainable City and City Science; Measuring and Mitigating Social Conflict; Analysing Best Practices In Sport. Besides the exploratories proposed by the consortium partners, we shall solicit expressions of interests from the scientific community towards new exploratories. Exploratories will serve two purposes: on one hand, their definition and usage will be an important aspect of internal and external evaluation/ feedback of the effectiveness and usability of the research infrastructure; on the other hand, exploratories will provide the scientific context and research theme of the exploratory calls for projects of the Transnational Access activities in WP6.
Task leader: CNR. Participants: ALL PARTNERS

Participation per Partner

Partner number and short name	WP11 effort
1 - CNR	16.00
2 - USFD	9.00
3 - UNIPI	2.00
4 - Fraunhofer	6.00
5 - UT	2.00
6 - IMT LUCCA	2.00
7 - LUH	4.00
8 - KCL	7.00
9 - SNS	2.00
10 - AALTO UNIVERSITY	4.00
11 - ETH Zurich	6.00
12 - TU Delft	1.00
Total	61.00

List of deliverables

Deliverable Number [14]	Deliverable Title	Lead beneficiary	Type [15]	Dissemination level [16]	Due Date (in months) [17]
D11.1	Evaluation Framework Design and Positioning	8 - KCL	Report	Public	24
D11.2	Evaluation Framework Toolkit and Datasets 1	2 - USFD	Other	Public	30
D11.3	Evaluation Framework Toolkit and Datasets 2	2 - USFD	Other	Public	45
D11.4	SoBigData Evaluation Framework: Engagement and Sustainability Report	11 - ETH Zurich	Report	Public	45

Description of deliverables

Reports on Evaluation Framework

D11.1 : Evaluation Framework Design and Positioning [24]

This deliverable will report on the results of T11.1 and position the SoBigData evaluation framework in the context of other relevant established initiatives in the fields of text mining, natural language processing, social network analysis, human mobility analytics, web mining, and visual analytics. D11.1 will also contain the design specification of the SoBigData evaluation framework, based on the results from the first 6 months of work in T11.2. KCL will lead this deliverable, with contributions provided by all partners from T11.1& 2.

D11.2 : Evaluation Framework Toolkit and Datasets 1 [30]

D11.2 will contain the SoBigData evaluation data collection toolkit, which will enable the campaign participants to create automatically the evaluation datasets, as described in T11.2. In addition, D11.2 will comprise the materials and datasets created for the SoBigData evaluation campaigns, carried out as part of T11.2, and will report on the definition of the exploratories of T11.4. All five thematic areas covered by the SoBigData project will have their corresponding datasets: text and social media mining (USFD, UNIPI), social network analysis (CNR, AALTO), human mobility analytics (CNR), web analytics (LUH), visual analytics (FRH). FRH will have overall responsibility for coordinating the deliverable production.

D11.3 : Evaluation Framework Toolkit and Datasets 2 [45]

D11.2 will contain the SoBigData evaluation data collection toolkit, which will enable the campaign participants to create automatically the evaluation datasets, as described in T11.2. In addition, D11.2 will comprise the materials and datasets created for the SoBigData evaluation campaigns, carried out as part of T11.2, and will report on the definition of the exploratories of T11.4. All five thematic areas covered by the SoBigData project will have their corresponding datasets: text and social media mining (USFD, UNIPI), social network analysis (CNR, AALTO), human mobility analytics (CNR), web analytics (LUH), visual analytics (FRH). FRH will have overall responsibility for coordinating the deliverable production.

D11.4 : SoBigData Evaluation Framework: Engagement and Sustainability Report [45]

This deliverable will report on the activities of T11.3, and more specifically on stakeholder engagement, commercial sponsorship, and sustainability activities. ETHZ as T11.3 leader will be responsible for coordinating input and deliverable completion.

Schedule of relevant Milestones

Milestone number [18]	Milestone title	Lead beneficiary	Due Date (in months)	Means of verification
MS4	Stepping up TA, VA, and networking activities, and community engagement	2 - USFD	36	- Analytics Services and Infrastructure Integration updated - 3rd SoBigData Workshop - Completion of third series of educational initiatives: Summer School 2, Training HS 3 & Datathons 3 - Use-cases VREs
MS6	Starting SoBigData Community Established	4 - Fraunhofer	48	- 4th SoBigData Workshop - Final SoBigData Conferences - Completion of last series of educational initiatives: Summer School 3, Datathons 4 - Use-cases VREs - Final evaluation and project end

1.3.4. WT4 List of milestones

Milestone number [18]	Milestone title	WP number [9]	Lead beneficiary	Due Date (in months) [17]	Means of verification
MS1	Inception and User Needs Gathering	WP3, WP4	8 - KCL	6	- Community Creation Started - 1st SoBigData Workshop - Advisory Panel installation
MS2	SoBigData eInfrastructure and training launch	WP10, WP2, WP3, WP4, WP5, WP7	11 - ETH Zurich	12	- Repository of training materials established - Innovation and first services Integration - Ethical issues addressed - 2nd SoBigData Workshop - e-Infrastructure Deployment - Resource catalogue VRE - Completion of first series of educational initiatives: Datathon 1, Training HS 1 - e-Infrastructure Production Release
MS3	Analytics service integration and data ecosystem creation	WP10, WP3, WP4, WP6, WP7, WP8, WP9	1 - CNR	24	- Services and Infrastructure Integration - Big Data Ecosystem and Analytics - 1st SoBigData Conference - Completion of second series of educational initiatives: Summer School 1, Training HS 2 & Datathons 2 - e-Infrastructure Enhanced with new services - e-Infrastructure Production Release
MS4	Stepping up TA, VA, and networking activities, and community engagement	WP10, WP11, WP4, WP5, WP6, WP7	2 - USFD	36	- Analytics Services and Infrastructure Integration updated - 3rd SoBigData Workshop - Completion of third series of educational initiatives: Summer School 2, Training HS 3 & Datathons 3 - Use-cases VREs
MS5	SoBigData infrastructure - final deployment	WP7	1 - CNR	42	- e-Infrastructure Production Release
MS6	Starting SoBigData Community Established	WP10, WP11, WP4, WP6, WP7	4 - Fraunhofer	48	- 4th SoBigData Workshop - Final SoBigData Conferences - Completion of last series of educational initiatives: Summer School 3, Datathons 4 - Use-cases VREs - Final evaluation and project end

1.3.5. WT5 Critical Implementation risks and mitigation actions

Risk number	Description of risk	WP Number	Proposed risk-mitigation measures
R1	Staffing Levels and Recruitment Problems. Low risk	WP1, WP10, WP11, WP2, WP3, WP4, WP5, WP6, WP7, WP8, WP9	In most areas of vital importance to the project partners intend to use existing staff. Also, SoBigData has more than one partner with expertise in the key areas, which can be seen as risk balancing to avoid delays.
R2	Key personnel leaves while the project is running. Low risk	WP1, WP10, WP11, WP2, WP3, WP4, WP5, WP6, WP7, WP8, WP9	Given the interest of the key researchers in the topics addressed in this project, this risk can be considered very low. If it occurs, we will shift responsibilities to senior members and make a concerted effort to instruct newly recruited researchers on the status of their WPs, assisted by the leaving researcher.
R3	Organizational financial problems. Low risk	WP1, WP10, WP11, WP2, WP3, WP4, WP5, WP6, WP7, WP8, WP9	The Project Coordinator will monitor audit certificates and financial problems and will report to the management boards and the EC project officer.
R4	Key staff are ill during critical project time. Low risk	WP1, WP10, WP11, WP2, WP3, WP4, WP5, WP6, WP7, WP8, WP9	Partners have more than one person involved in critical elements of SoBigData. Important expertise is available from more than one partner.
R5	Partner leaves the project. Low risk	WP1, WP10, WP11, WP2, WP3, WP4, WP5, WP6, WP7, WP8, WP9	The consortium is highly motivated and committed to the project. In the unlikely event that a partner leaves, the remaining partners have the skills and capacity to compensate and re-allocate the work load.
R6	Low participation in the open workshop, datathons and summer schools. Low risk	WP3, WP4	Excellent and highly reputed experts within and out of the consortium will be invited and when possible the events will be co-located with major conferences or other events. We will also build on the strong user communities

Risk number	Description of risk	WP Number	Proposed risk-mitigation measures
			already existing around the national infrastructures being integrated in SoBigData.
R7	Infrastructure integration time is underestimated. Medium risk	WP10, WP11, WP8, WP9	Project milestones will help monitor and detect problems early, and take corrective action. The SoBigData RI development and service integration from national infrastructures will be done in parallel, but the latter can be re-scoped and re-timed to mitigate against RI delays. The partners involved have proven development experience and timely delivery from previous eInfrastructure projects. All key partners were chosen to have track record with EU projects.
R8	Software service deployment or integration fail or attain limited functionality. Low risk	WP10, WP11, WP8, WP9	The partners are all leaders in the research fields and national infrastructures covered by the project and have proven existing infrastructures, on which service development will be based. In addition, consortium members with complementary technologies have been allocated to these workpackages, thus ensuring technological robustness.
R9	Users fail to take up and/or understand the SoBigData RI. Medium risk	WP1, WP10, WP11, WP2, WP3, WP4, WP5, WP6, WP7, WP8, WP9	This risk will be mitigated in two ways. WPs 8-11 will provide, in addition to software, also code examples, documentation, and intuitive user and programmatic interfaces. WPs 3 - 5 will engage actively with the user community, run training and innovation events, and feed back the technical WPs with user feedback. To complement this, the TA and VA WPs will allow users to engage directly with the SoBigData RI

Risk number	Description of risk	WP Number	Proposed risk-mitigation measures
			developers and realise knowledge and expertise transfer.
R10	Unfortunate choice of standards. Low risk	WP10, WP8, WP9	The SoBigData RI will use the Software-as-a-Service approach, to mitigates against this risk, as new data input/output standards can easily be added. CNR, KCL, and USFD are already involved in relevant standardisation bodies, European initiatives, and infrastructure projects and will alert the consortium.
R11	Evaluation is hindered by lack of data. Low risk	WP10, WP11, WP8, WP9	The creation of gold-standards for evaluation is an essential part of SoBigData, as part of WP11. The partners already have sufficient data to trial the scalability of the big data analytics services.
R12	The big data analytics services cannot scale up. Medium risk	WP10, WP9	The national infrastructures underpinning SoBigData have already established support for scalability of the analytics methods and services. Also, through the analytical crawling work in WP8, tools will be delivered to filter out irrelevant content, this enabling efficient analysis while maintaining high recall.
R13	Risk related to Big Data Analytics platform design and implementation. Low risk	WP10	All partners involved in WP10 have a proven track record in software design and implementation, as well as experience and knowledge gathered from previous eInfrastructure projects. CNR, as the WP leader and main partner responsible for the new RI, is involved through their dedicated infrastructures research group. There is complementarity in expertise, so resource re-allocation would be

Risk number	Description of risk	WP Number	Proposed risk-mitigation measures
			possible, should it become necessary.
R14	Reliability of data provision and integration. Low risk	WP10, WP8	High quality data has been identified from the outset and substantial project resources have been allocated to creating the data ecosystem in WP8 and to implementing interoperability. The commitment of the infrastructure partners to interoperable infrastructures and standards are a mitigating factor.
R15	Privacy and copyright issues. Medium risk	WP2	Gathering, sharing, and analysing social data raises such issues, which will be dealt with extensively in the project, through a dedicated work package and involving partners with proven track record in this.
R16	Problems arising from the multi-disciplinary nature of SoBigData. Low risk	WP1, WP10, WP11, WP2, WP3, WP4, WP5, WP6, WP7, WP8, WP9	Substantial management, technical, and research effort will be dedicated to ensuring smooth cooperation between researchers from the five research fields brought together in SoBigData. All partners have expressed their commitment and enthusiasm for joint research. Moreover, we have a number of senior researchers in crucial positions, who have relevant expertise as a liaison between the different disciplines.
R17	Sustainability after the life of the project. Medium risk	WP1, WP10, WP11, WP2, WP3, WP4, WP5, WP6, WP7, WP8, WP9	Ample project resources are allocated to community building and sustainability, as well as for innovation activities. Several partners hold positions in key national and international initiatives, which will help promote the value of the infrastructure.

1.3.6. WT6 Summary of project effort in person-months

	WP1	WP2	WP3	WP4	WP5	WP6	WP7	WP8	WP9	WP10	WP11	Total Person/Months per Participant
1 - CNR	19	4	6	6	4	18	9	18	22	16	16	138
2 - USFD	5	3	6	4	6	12	7	8	13	8	9	81
3 - UNIPI	1	2	2	4	0	0	4	2	10	8	2	35
4 - Fraunhofer	1	0	2	4	2	5	3	7	16	8	6	54
5 - UT	1	0	2	4	4	4	0	4	8	4	2	33
6 - IMT LUCCA	1	0	2	0	0	0	0	4	5	4	2	18
7 - LUH	1	3	2	4	0	5	5	9	6	5	4	44
8 - KCL	1	0	4	10	2	0	0	4	11	4	7	43
9 - SNS	1	0	2	0	2	0	0	4	5	6	2	22
10 - AALTO UNIVERSITY	1	0	2	2	2	4	0	7	12	7	4	41
11 - ETH Zurich	6	2	2	6	10	1	0	8	11	7	6	59
12 - TU Delft	0	10	3	4	0	0	0	0	0	0	1	18
Total Person/Months	38	24	35	48	32	49	28	75	119	77	61	586

1.3.7. WT7 Tentative schedule of project reviews

Review number [19]	Tentative timing	Planned venue of review	Comments, if any
RV1	21	tbd	Mid-Term Review

1.3.8. WT8 Summary of transnational / virtual access provision per installation

Access provider short name	Short name of infrastructure	Installation		Installation country code[21]	Type of access[22]	Unit of access	Unit cost (€)	Min. quantity of access to be provided	Access costs[23]		Estimated number of users	Estimated number of projects
		number[20]	Short name						On the basis of UC	As actual costs		
1 - CNR	SoBigData.it	1	SoBigData.it-TA	IT	TA-ac	week		40.0		120000	24	24
1 - CNR	SoBigData.it	2	SoBigData.it-VA	IT	VA					80000		
2 - USFD	GATE Cloud	1	GATE Cloud TA	UK	TA-ac	week		33.0		100000	20	20
2 - USFD	GATE Cloud	2	GATE Cloud VA	UK	VA					60000		
3 - UNIPI	SoBigData.it	3	TagMe	IT	VA					32500		
4 - Fraunhofer	IVAS	2	IVAS	DE	VA					29500		
4 - Fraunhofer	IVAS	1	Fraunhofer IGD	DE	TA-ac	week		16.0		49875	10	10
5 - UT	eGovData	1	eGovData	EE	TA-ac	week		6.0		20000	4	4
7 - LUH	ALEXANDRIA	2	ALEXANDRIA	DE	VA					29750		
7 - LUH	ALEXANDRIA	1	L3S	DE	TA-ac	week		10.0		29750	6	6
10 - AALTO UNIVERSITY	AALTO	1	AALTO	FI	TA-ac	week		13.0		40000	8	8
11 - ETH Zurich	Innovation Accellerator	1	Innovation Accellerator	CH	TA-ac	week		4.0		12500	2	2

1.4. Ethics Requirements

Ethics Issue Category	Ethics Requirement Description
HUMANS	- Details on the procedures and criteria that will be used to identify/ recruit research participants must be provided
HUMANS	- Detailed information must be provided on the informed consent procedures that will be implemented.
HUMANS	- The applicant has to further clarify how passive actors are involved for the purpose of data collection, and what informed consent procedures will be implemented for that case. The applicants must provide a policy for dealing with re-identifiable data. Moreover, the applicants must provide a policy for dealing with findings which are unrelated to the proposed work.
PROTECTION OF PERSONAL DATA	- Copies of ethical approvals for the collection of personal data by the competent University Data Protection Officer / National Data Protection authority must be submitted
PROTECTION OF PERSONAL DATA	- Justification must be given in case of collection and/or processing of personal sensitive data
PROTECTION OF PERSONAL DATA	- Detailed information must be provided on the procedures that will be implemented for data collection, storage, protection, retention and destruction and confirmation that they comply with national and EU legislation
PROTECTION OF PERSONAL DATA	- Detailed information must be provided on the informed consent procedures that will be implemented
PROTECTION OF PERSONAL DATA	- The applicant must explicitly confirm that the existing data are publicly available
PROTECTION OF PERSONAL DATA	- In case of data not publicly available, relevant authorisations must be provided
PROTECTION OF PERSONAL DATA	- In the course of the project, it is possible that the consortium encounters new ethics issues. Any such new issues and how they have been addressed are to be documented to contribute to the understanding of ethics in big data research. The applicants must provide confirmation whether the proposed work requires local/ national ethics clearance in those countries where the research will take place.
NON-EU COUNTRIES	- The applicant must confirm that the ethical standards and guidelines of Horizon2020 will be rigorously applied, regardless of the country in which the research is carried out
NON-EU COUNTRIES	- The applicant must provide details on the material which will be imported to/exported from EU and provide the adequate authorisations

1. Project number

The project number has been assigned by the Commission as the unique identifier for your project. It cannot be changed. The project number **should appear on each page of the grant agreement preparation documents (part A and part B)** to prevent errors during its handling.

2. Project acronym

Use the project acronym as given in the submitted proposal. It can generally not be changed. The same acronym **should appear on each page of the grant agreement preparation documents (part A and part B)** to prevent errors during its handling.

3. Project title

Use the title (preferably no longer than 200 characters) as indicated in the submitted proposal. Minor corrections are possible if agreed during the preparation of the grant agreement.

4. Starting date

Unless a specific (fixed) starting date is duly justified and agreed upon during the preparation of the Grant Agreement, the project will start on the first day of the month following the entry into force of the Grant Agreement (NB : entry into force = signature by the Commission). Please note that if a fixed starting date is used, you will be required to provide a written justification.

5. Duration

Insert the duration of the project in full months.

6. Call (part) identifier

The Call (part) identifier is the reference number given in the call or part of the call you were addressing, as indicated in the publication of the call in the Official Journal of the European Union. You have to use the identifier given by the Commission in the letter inviting to prepare the grant agreement.

7. Abstract

8. Project Entry Month

The month at which the participant joined the consortium, month 1 marking the start date of the project, and all other start dates being relative to this start date.

9. Work Package number

Work package number: WP1, WP2, WP3, ..., WPn

10. Lead beneficiary

This must be one of the beneficiaries in the grant (not a third party) - Number of the beneficiary leading the work in this work package

11. Person-months per work package

The total number of person-months allocated to each work package.

12. Start month

Relative start date for the work in the specific work packages, month 1 marking the start date of the project, and all other start dates being relative to this start date.

13. End month

Relative end date, month 1 marking the start date of the project, and all end dates being relative to this start date.

14. Deliverable number

Deliverable numbers: D1 - Dn

15. Type

Please indicate the type of the deliverable using one of the following codes:

R	Document, report
DEM	Demonstrator, pilot, prototype
DEC	Websites, patent fillings, videos, etc.
OTHER	

16. Dissemination level

Please indicate the dissemination level using one of the following codes:

PU	Public

CO	Confidential, only for members of the consortium (including the Commission Services)
EU-RES	Classified Information: RESTREINT UE (Commission Decision 2005/444/EC)
EU-CON	Classified Information: CONFIDENTIEL UE (Commission Decision 2005/444/EC)
EU-SEC	Classified Information: SECRET UE (Commission Decision 2005/444/EC)

17. Delivery date for Deliverable

Month in which the deliverables will be available, month 1 marking the start date of the project, and all delivery dates being relative to this start date.

18. Milestone number

Milestone number:MS1, MS2, ..., MSn

19. Review number

Review number: RV1, RV2, ..., RVn

20. Installation Number

Number progressively the installations of a same infrastructure. An installation is a part of an infrastructure that could be used independently from the rest.

21. Installation country

Code of the country where the installation is located or IO if the access provider (the beneficiary or linked third party) is an international organization, an ERIC or a similar legal entity.

22. Type of access

VA	if virtual access,
TA-uc	if trans-national access with access costs declared on the basis of unit cost,
TA-ac	if trans-national access with access costs declared as actual costs, and
TA-cb	if trans-national access with access costs declared as a combination of actual costs and costs on the basis of unit cost.

23. Access costs

Cost of the access provided under the project. For virtual access fill only the second column. For trans-national access fill one of the two columns or both according to the way access costs are declared. Trans-national access costs on the basis of unit cost will result from the unit cost by the quantity of access to be provided.

History of Changes

Date	Editor	Description
09/04/2015	CNR	Updated column "*Min quantity of access to be provided"* into Table 3.2c.
09/04/2015	CNR	Inserted as footnote the explanation about the calculation of the visit cost.
10/04/2015	CNR	Updated tables in "*Appendix B. Access cost calculation sheets*" regarding the calculation of Transnational Access: - modified the Table in Section 1.1.1 for CNR: corrected the description of the costs. - corrected the Table in Section 1.1.6 for AALTO: increased 2 P/Ms for part B as required by WP6 and not appropriately reported consequently (it was 2 instead of 4) decreased the costs in part A. In both cases no total access cost has changed.
10/04/2015	UNIPI	Updated tables in "*Appendix B. Access cost calculation sheets*" regarding the calculation of Virtual Access: - modified the Table in Section 1.2.3 for UNIPI: corrected the description of the costs. No total access cost has changed.
10/04/2015	CNR	Updated tables in "*Appendix B. Access cost calculation sheets*" regarding the calculation of Virtual Access: - modified the Table in Section 1.2.1 for CNR: corrected the description of the costs. No total access cost has changed.
13/04/2015	CNR	Updated "*Section 3.2 Management Structure and Procedures"* including the Project Administrative Manager.
13/04/2015	CNR	Updated figures in "*Section 3.4 Resources to be Committed"*: - corrected the amount of "*TA – Direct eligible cost for providing access*" for AALTO University. - corrected the amount of total PM for TA.
14/04/2015	CNR	Added "*Section 5.1.6 Ethics requirements Summary*" in "*Section 5. Ethics and security*".
23/04/2015	CNR	Corrected the deadline for deliverables D10.10, D10.14, D11.3, D11.4 before month 48, for work packages 10 and 11 respectively. The scheduled deadlines for the deliverables exceed the deadline of the work packages.
23/04/2015	CNR	Corrected the type from DEM to DEC of deliverable "*D10.1 Best practices and guidelines towards interoperability*".
23/04/2015	CNR	Corrected the type from DEM to OTHER of deliverable "*D9.1 Social mining method and service integration 1*".
04/05/2015	CNR	Restored column "*Min quantity of access to be provided"* into Table 3.2c to the proposal ones.
04/05/2015	CNR	Updated and Corrected Section 5.1 "Ethics".
11/05/2015	CNR	Updated description of Alessandro Chessa for IMT Key Staff.
20/05/2015	CNR	Corrected the amount of "Other goods and services" for KCL in Table 3.2b as originally requested by the partner.

20/05/2015	CNR	Updated table in "*Appendix B. Access cost calculation sheets*" for AALTO.
20/05/2015	CNR	Updated Table 3.2b by adding the amount for "TA other direct costs" as reported in part A of the calculation sheets for CNR, USFD, FHR, UT and AALTO.
20/05/2015	CNR	Corrected *table of budget allocation* in Section 3.4 "*Resources to be Committed*".

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

Table of Contents

1. EXCELLENCE 5
1.1 OBJECTIVES 5
1.2 RELATION TO THE WORK PROGRAMME 10
1.3 CONCEPT AND APPROACH 12
1.3.1 Concept 12
1.3.2 Pan-European Infrastructure Creation 14
1.3.3 National/International Research and Innovation Activities 14
1.3.4 Overall Approach and Methodology 17
1.3.5 Gender Diversity 18
1.4 AMBITION 18
1.4.1 Social Sensing: State-of-art and Research Roadmap 19
1.4.2 Social Mining: State-of-art and Research Roadmap 21
1.4.3 Progress Beyond the State-of-art 23

2. IMPACT 25
2.1 EXPECTED IMPACT 25
2.1.1 Impact as Set Out in the Work Programme 25
2.1.2 Improving Innovation Capacity 27
2.1.3 Social and Other Impacts 29
2.2 MEASURES TO MAXIMISE IMPACT 30
2.2.1 Dissemination and Innovation Activities and Impact Planning 30
2.2.2 Research Data Management 31
2.2.3 Open Source Software 33
2.2.4 Knowledge Management and Protection 34
2.2.5 Communication Activities 35

3. IMPLEMENTATION 36
3.1 WORK PLAN — WORK PACKAGES, DELIVERABLES AND MILESTONES 36
3.2 MANAGEMENT STRUCTURE AND PROCEDURES 41
3.2.1 Management Tasks and Procedures 44
3.2.2 Decision Making Processes 46
3.2.3 List of Milestones 46
3.2.4 Critical Risks for Implementation 47
3.3 CONSORTIUM AS A WHOLE 48
3.3.1 Swiss Participation 51
3.4 RESOURCES TO BE COMMITTED 52

4. MEMBERS OF THE CONSORTIUM 58
4.1. PARTICIPANTS 58
4.1.1 Consiglio Nazionale delle Ricerche – Italy (CNR) 58
4.1.2 The University of Sheffield – UK (USFD) 63
4.1.3 Università di Pisa – Italy (UNIPI) 66
4.1.4 Fraunhofer IAIS and IGD – Germany (FRH) 69
4.1.5 Tartu Ulikool – Estonian (UT) 72
4.1.6 Scuola IMT (Istituzioni, Mercati, Tecnologie) Alti Studi di Lucca – Italy (IMT) 74
4.1.7 Gottfried Wilhelm Leibniz Universitaet Hannover – Germany (LUH) 76
4.1.8 King's College London – UK (KCL) 79
4.1.9 Scuola Normale Superiore di Pisa – Italy (SNS) 81
4.1.10 Aalto University – Finland (AALTO) 84
4.1.11 ETH Zurich – Switzerland (ETHZ) 87
4.1.12 Technische Universiteit Delft – The Netherlands (TUDelft) 90
4.2. THIRD PARTIES INVOLVED IN THE PROJECT (INCLUDING USE OF THIRD PARTY RESOURCES) 91

5. ETHICS AND SECURITY 92
5.1 ETHICS 92
5.1.1 Proactive approach 92
5.1.2 Data protection 93
5.1.3 Human participants 94
5.1.4 Integrating the gender dimension in SoBigData 94
5.1.5 Ethical Review 94
5.1.6 Ethics requirements summary 95
5.2 SECURITY 97
BIBLIOGRAPHIC REFERENCES 98

1. Excellence

Mathematics and ICT - Starting Communities

☑ Distributed, multidisciplinary European infrastructure on Big Data and social data mining.

1.1 Objectives

MOTIVATION AND OVERALL GOAL

One of the most pressing and fascinating challenges is understanding the complexity of the globally interconnected society we inhabit today. Ours is also a time of opportunity to observe how our society works: the **big data** originating from the digital breadcrumbs of human activities, sensed as a by-product of the ICT systems that we use every day, promise to let us scrutinize the ground truth of individual and collective behaviour at an unprecedented detail and at a global scale. Multiple dimensions of our social life have "big data" proxies: our desires, opinions and sentiments leave their traces in our web pages, blogs, tweets, social media, query logs; our social links leave their traces in the network of our phone or email contacts, in the friendship links of our social network; our shopping patterns and lifestyles leave their traces in the records of our purchases; our movements leave their traces in the records of our mobile phone calls and GPS tracks. This increasing wealth of data is a chance to disentangle social complexity and face the challenges of our world, provided we can rely on **social data mining**, i.e., adequate means for accessing big data and extracting useful knowledge from them. Examples range from monitoring well-being and resilience of society based on behavioural patterns extracted from big data, to discovering surprising correlations between the socio-economic status and big data measurements, to *nowcasting* happiness, trust, the patterns of human mobility, and so on.

There is an urgent need to start harnessing these opportunities for scientific advancement and for the social good, compared to the currently prevalent exploitation of big data for commercial purposes (e.g. user profiling and behavioural advertising) or, worse, social control and surveillance. The main obstacle to this accomplishment, besides the scarcity of data scientists, is the lack of a large-scale open ecosystem where big data and social mining research can be carried out.

At an EU level, a critical mass of research centres exists, which developed in the last 20 years: (i) multiple social sensing methods to procure, manage and curate big data sources, with reference to a variety of legal, ethical, commercial and technical constraints; (ii) multiple social mining methods to make sense of these data and extract relevant social knowledge from them; (iii) multiple visualization and story-telling tools for sustaining the discovery and the interpretation of social knowledge and turning it into actionable policies. The complete analytical processes, from raw big data to actionable knowledge, have been created in diverse domains, ranging from mobility patterns for smart cities, monitoring well-being and social progress, social contagion and diffusion, etc.

The problem is that all this expertise and tools are fragmented, due to the difficulties in sharing experimental data and methods. This in turn, has posed severe limitations on the aggregated impact, which at present is confined to a multitude of isolated experiments. In practice, the big data research revolution is in its infancy, and there is still a big gap between the opportunities and the status quo. There are many regulatory, business and technological barriers to harnessing the true power of big data for social mining, so that scientists, and in the future any other interested party, can pursue the knowledge opportunities. Big data reside in diverse repositories, often inaccessible, due to proprietary and legal constraints. We still lack the ability to capture social diversity and complexity with patterns, models and indicators that narrate realistic stories. Breakthrough advances in social mining will only be possible by empowering interdisciplinary teams of scientists – computer,

complexity and social scientists, in particular – with the ability of performing large-scale social mining experiments on big heterogeneous data.

This is why we propose to establish **SoBigData**, the **Social Mining & Big Data Ecosystem:** a research infrastructure (RI) for ethic-sensitive scientific discoveries and advanced applications of social data mining to the various dimensions of social life, as recorded by "big data". The research community will use the **SoBigData** RI facilities as a "secure digital wind-tunnel" for large-scale social data analysis and simulation experiments. **SoBigData** will serve the wide cross-disciplinary community of data scientists, i.e., researchers studying all aspects of societal complexity from a data- and model-driven perspective, including data and text miners, visual analytics researchers, socio-economic scientists, network scientists, political scientists, humanities researchers, and more. The RI will provide the common ground for numerous H2020 projects aimed at tackling societal challenges like demographic change and well-being; secure, clean, efficient and integrated transport; Europe in the changing world; digital human. The RI is based on three pillars:

1. An ever-growing, distributed data ecosystem for procurement, access and curation of big social data, to underpin social data mining research within an ethic-sensitive context. This will be based on innovative strategies for acquiring social big data for research purposes, using both opportunistic means offered by social sensing technologies and participatory means based on user involvement as prosumers of social data and knowledge.

2. An ever-growing, distributed platform of interoperable, social data mining methods and associated skills: tools, methodologies and services for mining, analysing, and visualising complex and massive datasets, harnessing the techno-legal barriers to the ethically safe deployment of big data for social mining.

3. A starting community of scientific, industrial, and other stakeholders (e.g. policy makers), supported by transnational and virtual access activities, and brought together by extensive networking and innovation actions (e.g. workshops, summer schools, datathons, training resources in social data mining, knowledge transfer, industrial partnerships). In particular, the training events and resources are aimed at creating a new generation of multi-disciplinary social data scientists, with focus on early career researchers and promoting gender balance.

To pursue this overarching goal, **SoBigData** adopts an approach, that is sustainable and effective in the short, medium, and long term. In the short term, we aim at:

- Integrating key national infrastructures and centres of excellence at European level in big data analytics and social mining to create a networked, virtual ecosystem, the **SoBigData** RI. **SoBigData** will leverage these rich scientific assets (big data, analytical tools and services, and skills), to enable cutting-edge, multi-disciplinary social mining experiments;

- Granting access (both virtual and trans-national on-site) to the **SoBigData** RI to multi-disciplinary scientists, innovators, public bodies, citizen organizations, SMEs, as well as data science students at any level of education.

In the medium to long term, RI sustainability and growth will be pursued through opening the RI for contributions from members of the **SoBigData** stakeholder community (both as new datasets and methods), adopting an open source and open science approach, and ultimately to establish an advanced community in social data mining.

As part of the proposed integrating activity, **SoBigData** will enable trans-national access of researchers to data centres, tools and guidance, and the creation of new web-based, virtual access

services, based on common interfaces to data repositories, reference data sets, and innovative big data analytics methods.

Consequently, **SoBigData** will open up new research avenues in multiple research fields, including mathematics, ICT, and human, social and economic sciences, by enabling easy comparison, re-use and integration of state-of-the-art research data, methods, and services, into new research. It will contribute to the creation of a new community of researchers – data scientists – ready to exploit the opportunities of big data and to incorporate it in data-driven science and innovation. In addition, as an open research infrastructure, **SoBigData** will promote repeatable and open science.

To achieve its high level goal, the project will build on expertise and recent advances in eInfrastructures, relevant European data infrastructures, and seven established national infrastructures organized in six thematic clusters that arise from different research fields of social mining and big data analytics (*text and social media mining – TSMM*; *social network analysis – SNA*; *human mobility analytics – HMA*; *web analytics – WA*; *visual analytics – VA*; and *social data – SD*). The following infrastructures will be integrated and enhanced:

- **SoBigData.it**, the knowledge infrastructure of the European Laboratory on Big Data Analytics and Social Mining, funded by two institutes of CNR and the University of Pisa, offering data, services and expertise on TSMM, SNA, HMA, WA
- **GATE Cloud**, the cloud-based infrastructure for large-scale, data intensive natural language processing and text mining at the University of Sheffield, offering data, services and expertise on TSMM
- **IVAS**, the server-based infrastructure for Information Visualization and Visual Analytics at the Fraunhofer competence centre, offering data, services and expertise on visual analytics
- **Alexandria**, the infrastructure for Web Science at Leibniz Universitaet Hannover, offering data, services and expertise on WA
- **AALTO**, the Sociophysics and Data Mining laboratories at Aalto University in Helsinki, offering data, services and competences on SNA
- **E-Gov.data**, the centre at University of Tartu curating the Estonian e-government and e-health data, offering access and services to these unique dataset over 10 years (SD)
- **Living Archive for Open Data**, a search engine for Open Data offered by the Department of Sociology at ETH Zurich (SD).

SoBigData has a strong ethical motivation. We aim at building an ecosystem where protection of personal information and the respect for fundamental human rights can coexist with a safe use of the same information for scientific purposes of broad and central societal interest. We will establish an infrastructure that strengthens research concerning the protection of personal data as a fundamental right, while at the same time boosting the free flow of personal data as a common good. Extreme care is taken in all parts of the project and among all consortium partners not only to guarantee compliance with relevant laws, regulations, and codes of conduct, but also to prevent unethical outcomes of the envisaged cutting edge research and applications by ongoing analysis of the moral problems that pose constant challenges for law, regulation and governance of information technology. As such, **SoBigData** pursues the views that the EU is developing on Responsible Research and Innovation, and it operationalizes the values that drive the ongoing reform of the EU Data Protection legislation. **SoBigData** will establish a dedicated ethical and legal board, which will define and implement the legal and ethical framework of **SoBigData**.

Objective	Brief Description
Overcome tool and data fragmentation	**SoBigData** will promote repeatable and open science in the newly emerging, inter-disciplinary field of large-scale social data mining. The project will integrate key large social data sets with cutting-edge analytics for social mining. We will provide a visual analytics layer over the extracted knowledge, in order to help data scientists with result interpretation. Ultimately, **SoBigData** will deliver a virtual big social data analytics platform, which underpins the trans-national and virtual access activities and enables novel trans-disciplinary research. Open science and repeatability will be promoted through shared data sets, open source algorithms, an evaluation framework for social data analytics, and an ethical and legal framework.
Unified data formats and service interfaces	The project will create a unique, new European e-infrastructure, which will ultimately comprise a virtual research environment for data scientists, large datasets in unified format, and hundreds of distributed, interoperable analytics and visualisation services, accessible through common interfaces.
Development of innovative tools and services	The new **SoBigData** integrated infrastructure will stimulate the creation of new data sets and innovative services, based on multi-disciplinary approaches and methodologies. In particular, the project will develop new multi-dimensional analytical methods and algorithms for social data mining, to process and interpret big data from diverse sources. They will be built on top of the state-of-the-art privacy-aware social sensing, data mining, text analytics, pattern discovery and network analysis methods, offered by the underlying national infrastructures.
Multidisciplinary Community building	The creation of a multidisciplinary community is a substantial component of the **SoBigData** research infrastructure. The project will act as a catalyst and a focus point for a community of practice fostering the collection, analysis, and visualisation of big social data. SoBigData will involve all major actors from the five thematic clusters, already present in the consortium, as well as many other leading European and international research and industrial organisations, which will be brought together through the project's outreach, innovation, and community building activities. Networking activities such as training, dissemination, and innovation actions, will foster infrastructure usage, according to a well-planned growth strategy. The **SoBigData** community will bring together for the first time researchers from multiple disciplines studying all aspects of societal complexity from a data- and model-driven perspective. This will include, inter alia, data and text miners, social network analysts, visual analytics researchers, socio-economic scientists, political scientists, humanities scientists, legal and ethics researchers, and more.
Attract critical mass of users	The consortium already has a critical mass of data and manages a number of established national infrastructures and services, which will form the basis for the proposed integrated **SoBigData** infrastructure. Thus, from its very inception, **SoBigData** will be a usable and valuable resource for the research community and industry alike. Since this is a starting community, during the project lifetime, there will also be substantial further enhancements, enlarging the scope and content of the new **SoBigData** infrastructure, and hence its usefulness for the data science research community. The project will also train and support data scientists through transnational and virtual access to key high quality social data sets, coupled with the necessary hardware and software infrastructure for social data analytics. **SoBigData** will resolve legal, ethical, methodological and infrastructural issues arising from working with social data, in order to help data scientists to focus on research itself.
Creating a new generation of data scientists	The project will foster the development of a new generation of data scientists, skilled to work in multi-disciplinary settings and capable of drawing on different computer science fields, when designing new methods and experiments. Moreover, these

	researchers will be trained to exploit the advantages of research infrastructures for collaborative and complex, data-intensive research. Young researchers are addressed by all project training and dissemination actions, for example, through a ring-fenced budget in the Transnational Access activities.
Innovation for improving SME competitiveness and addressing societal challenges	**SoBigData** will undertake a wide range of innovation activities aimed at industrial and other stakeholders (e.g. government bodies, non-profit organisations, funders, policy makers). In particular, we will build partnerships with industry and other key stakeholders through knowledge transfer and consulting services; training data scientists in entrepreneurial skills; contributions to policy making; and involvement of relevant industrial and standardisation bodies as **SoBigData** stakeholders. The project will also build several demonstrators of how the new **SoBigData** RI can be harnessed in addressing major societal challenges (e.g. measuring well-being, smart cities).

Key to project success is stakeholder engagement, where **SoBigData** will target:

1. **Big data analysts and social informatics researchers**, who want to enhance their algorithms to deal with social data, gain multi-disciplinary research skills, harmonise existing data and analytics infrastructures, and engage other research communities in the development of these key enabling technologies for the future digital economy and society;
2. **Economists, social science and humanities researchers, journalists, policy and law makers**, who have to analyse the avalanche of (big) social data, in order to gain insight and actionable knowledge.
3. **Researchers in related communities**, who would like to use the algorithms, the analytical competences and data infrastructure; and
4. **The public as a whole**, who would like to understand their role in the production, consumption and value-creating of social data.

SUCCESS INDICATORS

The achievement of the above objectives will be measured against the following indicators.

Objectives	Action	Success Indicator
Overcome fragmentation and provide interoperability	Integrate the national infrastructures, with their data and services	• Number of datasets integrated via common data formats and APIs • Number of services integrated across the national infrastructures • Service and data interoperability across national infrastructures • Number of new integrated services provided
	Adoption of relevant data format and interoperability standards	• Number of datasets made accessible in standard formats (e.g. Linked Data) • Number of methods and services made interoperable through common interfaces and standards
	Build a common Virtual Research Environment (VRE)	• Number of datasets and services available in the **SoBigData** RI • Number of interactive visual analytics tools provided
Innovation	Stakeholder involvement	• Number of SMEs, bigger companies, and other institutional stakeholders engaged through innovation activities

Critical mass	Attract users	• Number of users registered • Data usage statistics • Service usage statistics • **SoBigData** RI usage statistics • RI promotion campaign statistics
Community building	Training	• Number and types of users trained
	Dissemination	• Extent of audience reached
	Transnational Access	• Number of users requiring access • Number of users given access • Number of experiments carried out
	Virtual Access	• EC assessment of VA success, via VA e-Infrastructure operation reports
Creating a new generation of data scientists	Training and access to the Research Infrastructure	• Number of young researchers trained • Number of young researchers requiring Trans-national Access • Number of registered young researchers to on-line Infrastructure services • Gender balance statistics • Country breakdown statistics • Datathon and bootcamp statistics

1.2 Relation to the Work Programme

This project addresses the objectives and expected impacts of topic INFRAIA-1-2014/2015: Integrating and opening existing national and regional research infrastructures of European interest, from call H2020-INFRAIA-2014/2015. **SoBigData**'s high-level goal is to deliver a distributed, pan-European, multi-disciplinary infrastructure for big social data analytics, coupled with the creation of a strong European research community of big data scientists. By bringing together and opening data and research methods from different research fields, the next generation trans-disciplinary data science will be enabled. **SoBigData** will build on already established research infrastructures and initiatives (e.g. Research Data Alliance, EUDAT, CLARIN, CESSDA, DASISH), where we plan to collaborate regarding re-use of data and other resources. To this end, partners from relevant ESFRI and other e-infrastructure projects have been integrated into the project (iCORDI, RDA Europe, CLARIN, DARIAH, OpenAIREplus, D4Science).

The main result of **SoBigData** will be a Starting Community in Mathematics and ICT on "Distributed, multidisciplinary European infrastructure on Big Data and social data mining". Mining big data, and social data in particular, is a very hot research area, with many active research and industry players. However, due to the starting community nature of this project and the associated funding limits, it has not been possible to involve all of these as partners. Instead, the community is seeded with a core set of internationally leading research labs and their established research infrastructures, and the rest will be drawn in as active contributors to and users of the SoBigData RI, through the transnational and virtual access activities, the networking and training events, innovation actions, and other collaborative activities. The following challenges, as defined on p. 19 of the Work Programme, will be addressed:

- ***Integrate key national research infrastructures of European interest:*** seven established research infrastructures from six scientific areas covering social mining and big data analytics (*text and social media mining – TSMM*; *social network analysis – SNA*; *human mobility analytics – HMA*; *web analytics – WA*; *visual analytics – VA*; and *social data – SD*) will be integrated, in order to enable novel multi-disciplinary approaches to social data mining.

- ***Provide access to large social data repositories***: **SoBigData** will provide access to a diverse set of big social data, including mobile and sensor data, social networks, social media data (including a Twitter gardenhose archive dating back to 2009, Facebook, Skype, and Foursquare content); open social data and relevant Linked Open Data resources; and other social data. In addition to integrating datasets from consortium partners, **SoBigData** will also integrate relevant social datasets from related e-infrastructure projects.

- ***Integrate social data mining methods and services***: **SoBigData** will integrate and make interoperable methods and services from several complementary scientific disciplines and national infrastructures, in order to enable new, multi-disciplinary methods for analyzing big social data. Novel research will be stimulated further through the creation of a common evaluation framework for social data analytics methods.

- ***Harness (super-)computing facilities for large-scale social data analytics***: The consortium partners providing trans-national and virtual access for running large-scale experiments all have substantial HPC facilities and Hadoop clusters. Some of the national infrastructures (e.g. GATECloud) also support on-demand scale-out to public clouds.

The joint research activities in the project are in line with the target aim of "*performing complex processes to extract social knowledge*". The expected impact on the way data scientists will analyze and visualize big social data will translate itself into research on new cross-cutting topics. That is currently not feasible, due to the high degree of research field fragmentation, geographic dispersal of the various social data sets, and lack of training in big data analytics. Within big social data analytics, it is the current lack of cross-disciplinary knowledge and ease of integrating methods from complementary research fields that inhibits research activity. By unifying data and methods within a common pan-European infrastructure, **SoBigData** will substantially shift the research boundaries, and provide much needed training for the next generation, pan-European data scientists.

The joint research activities are complemented by equally extensive networking activities, that will bring together scientists from diverse research communities (data mining, text mining, human mobility analytics, visual analytics, complex systems), industries (both SMEs and enterprises) and other stakeholders (e.g. government organisations, policy makers, funders). Since social data analytics has very wide applications and policy relevance, **SoBigData** will aim to create a vibrant beneficiary community, including inter alia intelligent transport management, healthcare, digital journalism, intelligent knowledge management, political science, eGovernment, customer-relation management. Innovation activities (including partnerships with industry and knowledge transfer) are a key impact and sustainability driver. Clustering and coordination actions with relevant European and national research infrastructure and e-Infrastructure projects are also envisaged.

Last, but not least, in order to enable a truly pan-European big social data research community, substantial resources are allocated to trans-national and virtual access. Over the course of the project, a number of calls for novel research on emerging social data analysis challenges will be publicised. These will provide opportunities to individual researchers from across Europe to work on unique, large social data sets, using a variety of social data analytics services from across the **SoBigData** infrastructure.

1.3 Concept and Approach

1.3.1 Concept

Figure 1 depicts the conceptual architecture of the **SoBigData** research infrastructure (RI). It comprises several kinds of resources, namely social datasets (e.g. foursquare check-ins, tweets), social data processing algorithms (e.g. mobility analytics, text analytics), and social data management services (e.g. data anonymization, algorithm parallelisation, large data storage).

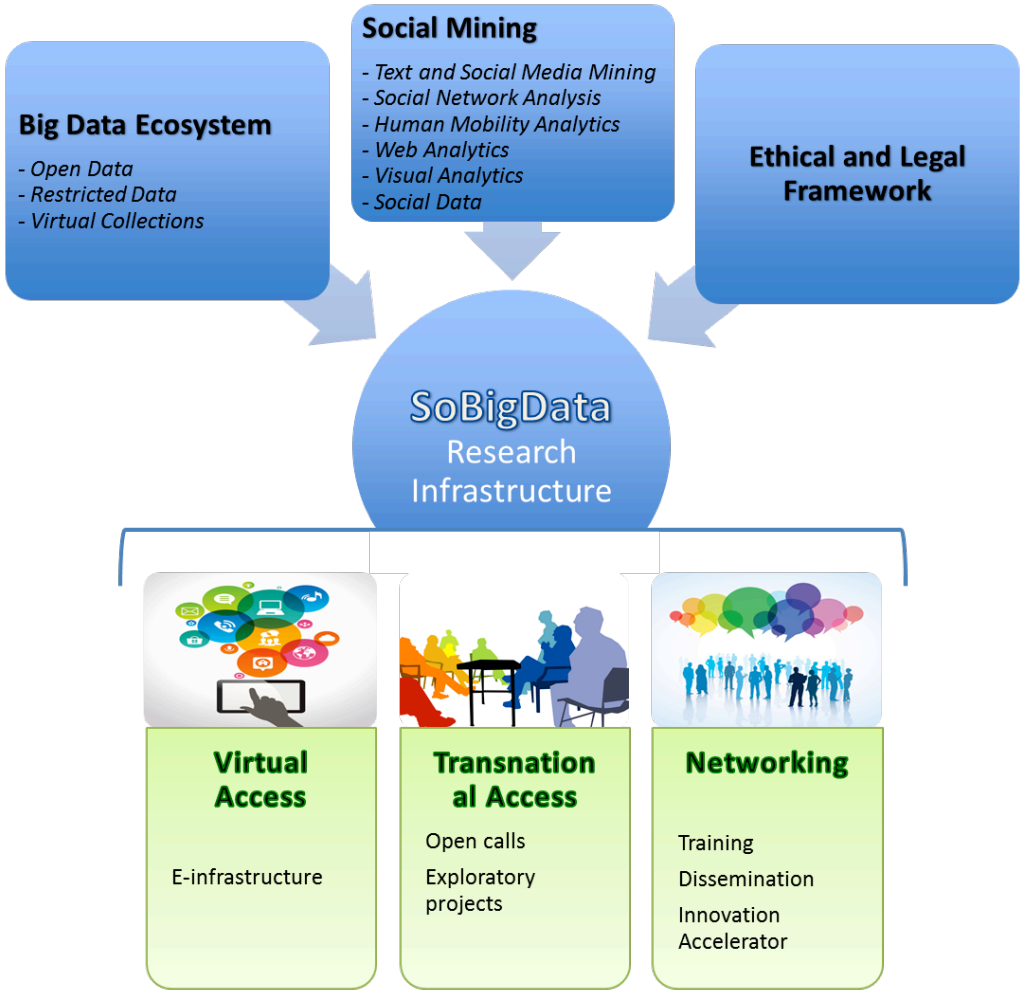


Figure 1. Conceptual architecture of the SoBigData research infrastructure.

The RI will take care of the legal, ethical, methodological, and infrastructural issues arising from working with social data, in order to enable data scientists to focus on research itself. It will provide access to the following key types of social data:

- Mobile and sensor data
- Social networks data
- Social media data, including Twitter, Facebook, and FourSquare content, organised into topic- and problem-specific social media virtual collections
- Mobility data, e.g. London Transport Oyster Card records and vehicular GPS trajectories
- Open social data and relevant Linked Open Data resources
- Other social data (such as one of the largest databases of Pinterest records)

The **SoBigData** RI will be built as a "system of systems", federating the resources offered by the underlying national infrastructures. The integrated methods from the national infrastructures will be organized into six thematic clusters, corresponding to key research fields and scientific expertise present in the consortium. The RI will also promote the use of open data and open source algorithms, in order to facilitate data sharing, open science, and repeatability.

The new RI will also include a new, extensive ethical and legal framework, which will be defined in accordance with European and national legislations (including data protection and intellectual property rights). It will ensure the development and application of policies for the collection and analysis of large-scale datasets in an ethical manner by-design.

As shown in Figure 1, the **SoBigData** RI underpins three different types of activities:

- **Virtual Access:** The Virtual Access (VA) is realized thanks the e-Infrastructure which will offer a web front-end comprising of **a catalogue of SoBigData resources** (both data and services) and **a set of SoBigData Virtual Research Environments (VREs)** dynamically created to address the requirements of specific research experiments. The web front-end will also provide scientists with a set of **social networking facilities** promoting researcher collaboration with other users in an innovative and effective way, e.g. enabling them to post or be notified of news about the availability of a new resource, to post news about planned experiments or ask for expert support and comments on a given idea or issue. Virtual Research Environments (VREs) are web-based working environments equipped with a number of applications, enabling scientists to have access to the set of data, services, and algorithms needed to perform their investigation in a collaborative way. To this end, VREs embed the social networking facilities mentioned above; are created by aggregating resources federated via the underlying e-Infrastructure, and are available to registered users. An essential part will be the visual analytics tools, which will help the social data scientists to aggregate and interpret the automatically extracted knowledge.
- **Transnational Access:** on-site access services granted by seven national infrastructures, offering world-leading research expertise from multiple disciplines, as well as big data computing platforms, big social data resources, and cutting-edge computational methods. The access will be granted through project calls in one of the following forms: (i) *Exploratory projects* are research experiments bound to multi-disciplinary themes specified in the calls and are expected to target data scientists from multiple disciplines (e.g., computer science, mathematics, physics, sociology, economy), and address the definition of needed infrastructure resources (datasets and/or methodologies and algorithms) and knowledge offered by specific hosting nodes; (ii) *Open call, blue-sky projects* will be targeted at researchers wishing to explore the infrastructure for their own research topics.
- **Networking:** the set of community building activities, which will enable **SoBigData** RI users and developers to co-operate on different analytical approaches, share methods and data resources, and achieve a truly multi-dimensional, distributed social mining infrastructure in the European Research Area, bridging across the diverse areas of expertise of the participants. Particular focus will be placed upon: establishing the innovation agenda, promoting standardization and open-access policies, disseminating good practices, evaluating impact, taking care of sustainability of the research infrastructure services, and promoting a wide educational network for a new generation of data scientists.

At the start of the project, the **SoBigData** RI will be bootstrapped from the large number of social datasets and analytical methods currently available from the underlying national infrastructures and it will be able to offer, since the very beginning, an extensive range of transnational and virtual accesses. As the joint research activities the partners deliver new datasets, tools, and analytical methods, these will be integrated continuously into the **SoBigData** RI (as transnational and/or virtual resources) and will thus become immediately available for experimentation.

Last but not least, the **SoBigData** RI will be aligned to and made interoperable with other relevant, complementary e-infrastructures and initiatives, as detailed in Section 1.3.3 below.

The breakdown of these activities into workpackages is detailed in Section 3.1, whereas the inter-dependencies and interactions between them is shown in Figure 2 in the same section.

1.3.2 Pan-European Infrastructure Creation

While researchers at big research labs and corporations (e.g. Google, Yahoo, Microsoft, IBM) have access to both Web-scale social data and vast computing infrastructures, scientists from smaller research groups and universities in newly associated member states are faced with major technological challenges when carrying out cutting edge big social data analysis experiments.

To redress this imbalance, the project will create the **SoBigData** RI that will make the datasets and methods of the national research infrastructures, available to users anywhere, through virtual access. The first service that will be developed in the **SoBigData** RI is the *resource catalogue*, implementing a smart search tool for all the different kinds of resources available. There will also be a workflow specification language for combining data analytics services and cross-relating distributed resources. Users will be also be able to enrich the system with new datasets and services.

In order to ensure interoperability, we will define the e-infrastructure resource model: types of resources (e.g. datasets, services, algorithms) and their metadata properties (e.g. classifications, access parameters, hardware parameters, IPR issues, input/output data types). We will also specify data models and formats for data exchange (e.g. data models for service parameters). Where other infrastructures (e.g. CLARIN) have already specified relevant data models and/or W3C/ISO/other standards exist, these will be reused, and, if necessary, adapted or extended.

Tools and algorithms from the national infrastructures will be made available online following the software-as-a-service paradigm. Then service inter-operability will be ensured through the definition of common service APIs. These will enable service pipelining, subscription and notification, and access to service functionality independent from its implementation.

These integration activities will be continuous over the project lifetime and will be performed in strict cooperation with the JRA WPs. The starting point will be the existing national infrastructures (see WP 7) and the different thematic areas identified there, in order to align the different research communities involved in **SoBigData**, towards common and agreed-upon best practices.

1.3.3 National/International Research and Innovation Activities

SoBigData plans to re-use results and knowledge from a number of already completed projects, and to share data, algorithms, knowledge, and dissemination activities with concurrently running projects. Where applicable, **SoBigData** consortium members involved are indicated.

Project Name and URL	**Project Overview**
EUDAT www.eudat.eu	EUDAT is building a pan-European collaborative data infrastructure, aimed at storage, sharing, and sustainability of research data. **SoBigData** will seek cooperation with EUDAT and sign a memorandum of understanding. **SoBigData** will also benefit directly from the knowledge and services offered by EUDAT with respect to management of the research data generated by this project.
Research Data Alliance rd-alliance.org/	(CNR) RDA is a worldwide initiative for research data sharing, currently funded by the European Commission, USA, and Australia. RDA is developing policy, standards, and best practice guidelines for research data sharing. CNR, as one of the RDA Europe partners, will build on their guidelines when creating the **SoBigData** research

	infrastructure. We will establish a memorandum of understanding with RDA and contribute through involvement in working and interest groups.
CLARIN www.clarin.eu	(USFD, UT, KCL): **SoBigData** fits into the vision of the pan-European initiative CLARIN, which is to enable lower thresholds for accessing multicultural and multilingual content. **SoBigData** will show how such shared datasets would be combined with **SoBigData**'s content analytics services to extract, aggregate and visualize results, in various domains and use cases.
DASISH dasish.eu	(KCL, UT): DASISH has brought together all relevant players in the SSH ESFRI domain to develop joint services and approaches. He has resulted in a range of enhanced community activities and joint research work relating to the SSH data domain and also solutions relating to legal and ethical aspects of SSH data. While DASISH will come to an end at the end of 2014, there is a range of of follow-on projects like PARTHENOS (h-Cluster) that **SoBigData** will collaborate with.
DARIAH www.dariah.eu	(KCL, CNR) DARIAH is an ESFRI infrastructure to support transnational research in all arts and humanities disciplines and related social sciences (such as political communication or media research). We will collaborate with DARIAH on social history and media data sets, once it has become an ERIC is September 2014. Tobias Blanke (KCL) is one of the designated first-time directors.
CODATA www.codata.org	The Committee on Data for Science and Technology is working towards better practices for scientific data management and use. It promotes open data and repeatable science through policies and frameworks for data management, also working closely with the Research Data Alliance. As with RDA, **SoBigData** will seek to adopt best practice policies and frameworks arising from CODATA and, in particular, from the task group on information commons for science.
Data without Boundaries (DwB) www.dwbproject.org	DwB is building a data infrastructure for secure virtual access to official statistical microdata (e.g. survey and census results). **SoBigData** intends to collaborate actively with DwB. Ultimately, the **SoBigData** RI will aim to combine such kind of data, to complement the social media, social network, mobility, and web data that is the core focus here.
EHRI www.ehri-project.eu	(KCL) The European Holocaust Research Infrastructure is an Integrating Activity to bring together Holocaust material across Europe. It has had a key impact on constituting transnational contemporary history research in Europe. We will collaborate with EHRI in discovering patterns of racist language and radicalisation in contemporary social media data sets.
PERICLES http://pericles-project.eu/	(KCL) PERICLES is addressing the challenge of ensuring that digital information (including research data) remains accessible in an environment that is undergoing continual change. **SoBigData** will seek cooperation with PERICLES and sign a memorandum of understanding. **SoBigData** will also benefit directly from the knowledge and tools developed by PERICLES as regards digital preservation of the research data generated by this project.
MLI mli-project.eu	(USFD) MLI is a support action defining the operational specifications for a European MultiLingual data & services Infrastructure, focused specifically on unstructured content. The design of the **SoBigData** RI will take MLI's specifications into account, but will ultimately address the bigger challenge of analysing all kinds of big social data, including also mobility data, structured data, and social networks, as well as the inter-connections between them.
ARIADNE	(CNR) ARIADNE is a digital infrastructure for archaeological research and is

www.ariadne-infrastructure.eu/	interesting to us as there is a growing interest in digital archaeological approaches to social media mining
CESSDA www.cessda.org/	CESSDA is an ESFRI infrastructure to integrate social science data archives across Europe. In the second phase of the project we will expand our efforts to mine social media data sets to other related social data sets and especially work with CESSDA towards a comprehensive preservation strategy of our derived data sets.
AnnoMarket annomarket.com	(USFD) A recently launched cloud-based text annotation infrastructure and marketplace. It enables running large-scale content analytics services on-demand. The AnnoMarket cloud-based platform-as-a-service will underpin the large-scale services for mining unstructured social media content, and also serve as an additional exploitation vehicle, for SMEs who wish to use these services in a pay-as-you-go, on-demand manner.
META-NET www.meta-net.eu	(USFD): META-NET aims at an open infrastructure of resources, especially sharing of large repositories of processed and annotated language data. META-NET and its related projects, will serve as a source of multilingual resources and tools for **SoBigData**, as well as a dissemination network.
GATE Cloud bit.ly/gate-cloud-jisc-project	(USFD): A UK-funded project, which delivered an infrastructure for deploying text analytics applications on the cloud. The resulting GATECloud infrastructure will now be the leading text and social media mining infrastructure for VA in **SoBigData**.
E-CLOUD pro.europeana.eu/web/europeana-cloud	(CNR, KCL) Europeana Cloud is a Best Practice Network, submitted under Objective 2.1.a and coordinated by the Europeana Foundation, designed to establish a cloud-based system for Europeana and its aggregators. Europeana Cloud will provide new content, new metadata, a new linked storage system, new tools and services for researchers and a new platform – Europeana Research.
DATASIM www.uhasselt.be/datasim	(CNR, FRH) Data Science in the Era of Electric Vehicles. It aims at providing an entirely new and highly detailed spatial-temporal micro-simulation methodology for human mobility, grounded on massive amounts of Big data of various types and from various sources, e.g. GPS, mobile phones and social networking sites, with the goal to forecast the nation-wide consequences of a massive switch to electric vehicles, given the intertwined nature of mobility and power distribution networks.
LIFT	(CNR, FRH) Using Local Inference in Massively Distributed Systems is an FP7-ICT FET Open project on design and analysis of privacy-preserving frameworks in distributed environments where the data are generated in streaming.
MODAP www.modap.org	(CNR) Mobility, Data Mining, and Privacy is an European Coordination Action, which will help with dissemination of research results on Mobility data mining and privacy.
EINS www.internet-science.eu/	(CNR, ETHZ, UDelft) The goal of this Network of Excellence in Internet Science is integrating current research initiatives aimed at achieving a deeper multidisciplinary understanding of the development of the Internet as a societal and technological artefact. In particular, one of the focuses is the analysis of the structures formed by humans in virtual environments such as Online Social Networks, and how they correlate with structures found in "offline" social networking environments.
ENFORCE enforce.di.unipi.it	(CNR) Computer science and legal method for enforcing the personal rights of non-discrimination and privacy in ICT systems (ID Progetto: RBFR081L58) is an Italian project founded from MIUR FIRB "Futuro in Ricerca" program. Contribution: Study and design of techno-legal solutions for the discovery and the prevention of discrimination and for the preservation of individual privacy.

CAPER www.fp7-caper.eu	(CNR, FRH) CAPER will support collaborative multilingual analysis of audiovisual content and biometrics information, based on Visual Analytics and Data Mining technologies. The integration of database technologies, application workflow and semantic modelling of processes and legal limitations will permit participating partners share knowledge. The CAPER platform will be built in close collaboration with the users in order to fulfil their current and forthcoming needs. The project is clearly focused on the fusion and real validation of existing state of the art to solve current application bottlenecks.
SCAMPI www.ict-scampi.eu	(ETHZ, CNR, AALTO) The SCAMPI project envisioned that the Future Internet will be characterised by an increasing diffusion of devices with heterogeneous capabilities and resources. Analysis of users' behavior in Online Social Network was carried out to understand patterns of cooperation and sharing, and ways to incorporate information about social relationships in the design of Future Internet mobile networks.
PETRA www.petraproject.eu	(CNR) Personal Transport Advisor: an integrated platform of mobility patterns for Smart Cities to enable adaptive transportation systems.
MIDAS www.midas-project.eu	(CNR) The subject of MIDAS (Model and Inference Driven, Automated testing of Services architectures) is an integrated framework for SOA testing automation that will be available as a Platform as a Service (PaaS) on a cloud infrastructure.
TAGME acube.di.unipi.it/tagme	(UNIPI) The TAGME project developed a "topic annotator" that is able to identify meaningful sequences of words in a short text and link them to a pertinent Wikipedia pages. Google Research Awards 2010 and 2013, Italian MIUR-FIRB project , Italian MIUR-PRIN project
SOCIALNETS www.social-nets.eu	(CNR) The SOCIALNETS project explored how social networks can be exploited for the delivery and acquisition of content, including issues of security and trust. This covers both online social networks and opportunistic (mobile peer to peer) wireless networks
NOMAD nomad-project.eu	(Fraunhofer) Nomad will provide decision-makers with fully automated solutions for content search, acquisition, categorisation and visualisation that work in a collaborative form in the policy-making arena. Multi-faceted aggregates, including multi-dimensional geo-demographic features will be extracted and presented intuitively using Visual Analytics techniques.
EU Community project.eucommunity.eu	(FRH) EU Community provides decision makers with better policy options by combining social media interactions, qualified contributors, document curation, visual analysis plus online and offline trust-building tools. The results will be open source platforms, and the data itself will be open to re-use by others.

1.3.4 Overall Approach and Methodology

The high-level goal of **SoBigData** is to deliver an innovative research infrastructure for analysing large volumes of social data, by bringing together state-of-the-art methods and services from complementary research fields. Coupled with that will be the creation of a vibrant community of multi-disciplinary data scientists, working on social data analytics.

In order to realise this ambitious goal, **SoBigData** will adopt an agile methodology, which closely couples researcher requirements, iterative technology development and integration, evaluation, and deployment. The work programme is structured around 3 iterative cycles, reflected in our

milestones and deliverables (including RI creation and deployment, data and service integration, and innovative tool/service development).

The consortium's strong track record in e-infrastructures, gathering and processing big social data, multilingual text analysis, data mining, human mobility analytics, social network analysis, web mining, web science (including Linked Open Data), and information visualization, provides a solid starting point. The community engagement workshops, training events, and datathons will ensure that the requirements of researchers from all these disciplines are factored in continuously, as well as gathering their feedback once the **SoBigData** RI becomes operational. A substantial part of the work programme is dedicated to networking activities and virtual and trans-national access, in order to create a vibrant user community around the new pan-European infrastructure, as well as attract new contributors of datasets and methods, which they integrate in the open **SoBigData** RI.

In order to promote business impact and adoption by SMEs and enterprises, there will be many innovation activities, aimed at these stakeholders. Many of the tools and services to be integrated in the **SoBigData** RI are open source, using commercially friendly licenses. Similarly, the newly developed or enhanced services created during the project, will be released with similar open licenses. Far reaching impact and sustainability beyond this starting community funding are a key priority. In this regard, the project will build on successful experience with creating and maintaining the national infrastructures and their large, active user communities over the past 15 years.

1.3.5 Gender Diversity

In addition to building a new research infrastructure and an associated starting community in big social data analytics, **SoBigData** aims to address gender diversity.

Firstly, two of the three members of the senior management team of the **SoBigData** consortium are female: the coordinator Fosca Giannotti (CNR) and one of her deputies - Kalina Bontcheva (USFD). Together, they will ensure that gender diversity and equality issues are dealt with fairly, both in the consortium and in the **SoBigData** starting community as a whole.

Secondly, **SoBigData** and data science in general, have a unique opportunity to reconcile technological research with social research (society and ICT), and therefore this opens new opportunities and motivations for the involvement of women, who tend naturally to gravitate towards human factors and social research. We have dedicated training and communication activities, aimed specifically at young girls and underrepresented, early career researchers. In particular, the project will communicate the unique opportunities and trans-disciplinary nature of data science, demonstrating that it is not the preserve of socially awkward geeks, as stereotyped in books and TV shows (e.g. the Big Bang theory).

1.4 Ambition

The integration at European scale of a research infrastructure dedicated to big data analytics and social mining aims at providing an open, yet regulated, ecosystem where social data mining can be pursued by a plurality of actors and where the state of art of big data analytics and social mining can be steadily advanced, up to providing a middle-layer technology for making the knowledge and semantics hidden in the big data available for addressing the big questions about social complexity [Giannotti 2012, Pentland 2012, Lazer 2009][1].

We briefly discuss the level of ambition of **SoBigData** to advance research and technology in social sensing and social mining. In particular, we concentrate on the roadmap of research needed to pursue this objective, beyond the limitations of the state-of-art in big data analytics.

[1] A reference list is provided in Appendix C of INFRAIA-1-2014-2015-SoBigData_Part_B_sections_4-5.pdf

[2] See Article 29 Working Party (2013). Opinion 03/2013 on purpose limitation. Available at http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

1.4.1 Social Sensing: State-of-art and Research Roadmap

We use the term *social sensing* to denote the novel ICT techniques for collecting the digital footprints generated by humans when interacting with the surrounding techno-social ecosystem and making such big data available for mining/analysis at properly high abstraction levels. The complexity of social sensing arises both from the variety of forms of data, and from the variety of acquisition devices and methods available. Data formats range from text, graph, logs, streams, structured, and semi-structured data. Acquisition sources range from social networks, web logs, web content, links, transactions, search engines, social media etc. to the wide variety of smart devices spread in the environment (such as RFID readers, sensors and actuators, sensor-rich smart phones, or wireless communication technologies such as WiFi, Bluetooth and cellular networks).

In all cases, the low-level representation of human activities in sensed data makes it difficult to observe human behaviour. The major demand for social sensing is to complement the raw data with additional information conveying concepts of human behaviour and social semantics, either by automated annotation and enrichment methods that extract social semantics along several dimensions from otherwise raw data, or by advanced participatory systems, that explicitly involve people into the picture, by means of crowd computing, serious games, participatory sensing. Such a big picture may be mapped onto the four following research challenges.

- **Semantic Enrichment of Digital Breadcrumbs.** Digital breadcrumbs are structured data that record basic events underlying complex ICT applications or sensor networks, and contain historical and collective information of basic human actions. They are stored in centralized repositories or in smart personal devices, or they may fast pass trough a data-streaming window. To make such breadcrumbs usable for analysis, it is needed to inject back the contextual knowledge of human actions that got lost during sensing. So, discretizing, normalizing, aligning, sampling are preliminary steps towards making such data closer to the phenomena to be observed. More properly, those are steps of a process aimed at adding context information to raw data by augmenting, tagging, and/or classifying data in relationship to dictionaries and/or other reference sources. The literature is rich in proposals for semantic enrichment in specific application domains. For instance, in case of mobility data, a basic enrichment is the extraction of *trips* obtained by segmenting movements into intervals of time when the object is considered as stationary (stop) and moving (move). Another dimension is enriching trips with concepts related to the *activities* performed, representing the purpose of the movement. This implies the recognition of how stop/move patterns correlate to specific activities (home, work, shopping, leisure, bring/get, etc.) [Parent 2013]. In case of query logs, the basic enrichment is the extraction of *query sessions*, i.e., specific sets/sequences of queries submitted by a user while interacting with a search engine to perform tasks [Lucchese 2011]. In case of tweets, enriching user profiles with concepts related to topics and entities implies to exploit prior social knowledge about bags of entities and categories, such as those provided by Wikipedia and DBpedia. At the edge of research, this mapping allows to crosslink the enhanced tweets to other sources such as news, by mapping both tweets and news in the same entity-based coordinate system. The mixing of languages – traditional idioms and tech idioms such as emoticons – is another big challenge. The overall repertoire of semantic enrichment methods needs to adapt with respect to the continuous progress of sensing devices, to improve its capability to map human behaviour along multiple dimensions, and to become increasingly intelligent with use, by incorporating machine learning techniques.

- **Human-level understanding of text at Web scale**. We need systems that progressively acquire the ability to interpret textual communications and to extract structured information (semantic objects) from various online sources, including text and other unstructured data, which is exchanged by people through the web, in pages, blogs, news, tweets, etc. This project will

advance the state of the art in human-level understanding of documents, which is instrumental for analyzing social data. The new methods, to be developed in **SoBigData**, will combine information present in individual documents with relations among documents and people using those documents. The work is based on combined natural language processing (NLP) and statistical techniques, which are fundamental to deal with open-domain acquisition from disparate sources, for collecting and contrasting evidence at various level of linguistic knowledge: terminological and syntactical. Discriminative analysis is needed to recognize emerging syntactic patterns. Though many approaches to computational content analysis of big social data components meet the challenge of scalability and relative domain independence, there are still issues, notably an often large number of incoherent and uninformative items extracted, which are related to the specific linguistic and structural composition of big social data. Furthermore, while some techniques work in on smaller datasets, they do not scale well. A good example is the new area of identifying contradicting opinions, which is an effective way for focusing on the interesting parts of big social data. For example, we identify that suddenly the opinions of people on some topic, such as nuclear power, have changed from positive to negative, or that a new popular topic, such as fiscal reforms, is causing both positive and negative sentiments. These results could be the basis for the further analysis of the data, leading to a better understanding of the causes, as well as the future social trends.

- **Analytical crawling**. Analytical crawling improves current technology by adopting a goal-driven approach capable of combining semantic enrichment and machine reading techniques to the aim of gathering semantic-rich data linked across diverse web resources. Crawling techniques so far have been designed to support search tasks and research has mainly focused on improving the indexing structure, both in efficiency and efficacy. The new challenge is to push part of the intelligence of social sensing as far as possible into the gathering process and make it capable of constructing, maintaining and using multi-dimensional network of semantic objects. A simple example is the study of relationships of important persons from around the world: this might be the network constructed by crawling their co-occurrence on the same article (news, blogs etc.) bound by positive or negative terms, phrases, or opinions. Another example is the study of relationships among international institutions and organizations: again this might be the network constructed by crawling their co-occurrence on the same article, bound by collaboration ties. The crawling process will construct new nodes or new edges annotated with relevant concepts by analysing on the fly the web content, extracting the concepts by matching with the current network, with external sources and with the analytical goals. The crawling process will continue by following pertinent URLs, selected on the basis of the semantic network constructed incrementally; such network, at the end of the process, is the final output and input for the subsequent social mining analyses.

- **Participatory sensing.** Another challenge is to directly involve people as active/passive actors in data collection, by developing tools to elicit facts, opinions and judgments from crowds. Participatory sensing aims at building, on demand, a sensor network infrastructure (as for example the mobile phone network, or wearable electronic device network or a social network) based on people collaboration towards a common sensing task (*crowdsensing*). There might be different levels of user involvement in participatory social sensing, ranging from participation in *ad-hoc survey campaigns*, up to active engagement in *human computation* and *crowdsourcing* [Buckingham-Shum 2012; Bontcheva 2014] Survey campaigns, the basic instrument of data collection for data analysts, statisticians, social scientists etc., has been completely renovated by the proliferation of ICT technology: from web, to smart phones to sensor networks. On the other hand, the social web has prompted the emergence of the new paradigm of human computation, where the interaction among users is harnessed to help in the cooperative solution of tasks. Crowdsourcing is a facilitator of human computation: it addresses the distributed assignment of work to a vast community of executors in a structured platform [Howe 2006; Bontcheva 2014].

At the edge of research we have today the idea of analytical crowdsourcing systems, capable of engaging vast populations of participants in performing social responses, such as liking, ranking, and tagging, organized into social networks and collaborating communities, within incentives models that foster the emergence of collaborative behaviour [Helbing 2009]. This enhanced view of participatory sensing is based on the idea that aware participation in collective sensing campaigns will enable the creation of social data of higher quality and realism, capable of portraying concurrently different dimensions of social life. Finally, the ability of combining data sensed both by opportunistic and participatory methods is expected to open new avenues for learning systems, capable of recognising and mapping human behaviour at global and local scale.

1.4.2 Social Mining: State-of-art and Research Roadmap

Social Mining aims to provide the analytical methods and associated algorithms needed to understand human behaviour by means of automated discovery of the underlying patterns, rules and profiles from the massive datasets of human activity records produced by social sensing. Although data mining and statistical learning from traditional databases are relatively mature technologies [Kumar 2006, Hastie 2009], the emergence of big data of human activities, their networked format, their heterogeneity and semantic richness, their magnitude and their dynamical nature pose new exciting scientific challenges. Human Data Mining, Reality Mining: these are the new keywords witnessing the flourishing of novel data mining, statistical learning and network science methods centred on the digital footprints of human activities. The following fields are the ground where the different disciplines are converging within the proposed **SoBigData** RI:

- **Social network analysis** refers to the study of interpersonal relationships, with the purpose of understanding the structure and the dynamics of the fabric of human society. It is currently a very hot and attractive research area with increasing presence in the major journals and conferences of the aforementioned disciplines. Roughly, we may distinguish two major lines: the statistical laws regulating statics and dynamics of complex networks [Watts 1998, Barabasi 1999, Caldarelli 2007, Newman 2010, Easley 2010], and the methods aimed at discovering patterns, evolutionary rules, community structure and the dynamics in large social network: i) *community discovery* [Fortunato 2010, Coscia 2011, Tumminello 2010, Tumminello 2011, Gurtner 2014]; ii) *information propagation* [Kleinberg 2000, Kempe 2003, Pastor 2001, Keeling 2005]; iii) *link prediction* [Liben-Nowell 2003, Kashima 2009, Leskovec 2010]; iv) *temporal evolution* [Leskovec 2005, Holme 2011, Mucha 2010, Berlingerio 2010, Bringmann 2010]; v) *multilevel networks* [Jianxi 2012, Berlingerio 2012, Tang 2009]. [Ghani 2013] presents a visual analytics techniques designed for and together with social scientists. They describe the adaption process from standard techniques known from the visualization field to a scientist's tool to address specific question in social networks.

- **Social media mining** methods address heterogeneous data, especially text, sensed from different on line sources (tweets, mails, blogs, web pages, etc.) to the purpose of extracting the hidden semantics from them. A key topic is *opinion/sentiment mining*, [Pang 2008] which has come to play a key role in customer relationship management, consumer attitude detection, brand/product monitoring, and market research in general [Esuli 2010]. Interest in these applications has spawned a new generation of companies and products devoted to online reputation management, market perception, and online content monitoring in general. Historically, one of the most important incarnations of opinion mining has been *sentiment classification*, the task of classifying a given piece of natural language text (be it a short remark, or a blog post, or a full-blown product review) not according to the topic it is about (as in standard text classification) but according to the opinions expressed in it. In [Bosch 2013] others harness visual analytics is described: a monitoring system for the effective retrieval of

interesting microblog messages, where a microblog classifier is tailored to the specific interests of a user. The system also meets a fundamental requirement for "useful" analysis: shielding the user from the complexity of the underlying model.

- **Understanding human mobility** emerged as a vital field, leveraging the spatio-temporal dimensions of big data (such as mobile phone call records and GPS tracks) to the purpose of mining mobility patterns, daily activity patterns, geographic patterns. Network scientists concentrated on the models of human movements, showing the emergence of power laws both in the probability of a banknote or a person to traverse a certain distance, and in the distribution of people's radius of gyration, the characteristic distance travelled by a person [Brockmann 2006, Barabasi 2008, Song 2010]. Models of pedestrian behavior have also been developed and validated on empirical data, capable to predict crowd disasters [Moussad 2011]. Computer scientists put forward *mobility data mining*, aimed at discovering patterns of interesting mobility behaviors of groups of moving objects [Giannotti 2008]. Examples of this process are the mining of frequent spatio-temporal patterns, trajectory clustering, and prediction of future movements [Trasarti 2011, Giannotti 2007, Giannotti 2011, Monreale 2009]. A recent trend in research focuses on extracting mobility profiles of individuals from their traces, trying to address one of the fundamental and traditional question in the social science: how human allocate time to different activities as part of the spatial temporal socio-economic system, and how groups or individuals interact with different places, at different time of the day in the city [Trasarti 2001, Jiang 2012, Ferrari 2012, Zimmermann 2002].

- **Multi-dimensional social mining.** At the edge of research there is the *convergence of data mining and social network analysis* towards the definition of novel mining models and patterns that combine the different dimensions of human activities: social ties, mobility, opinions and preferences, socio-economic behaviour etc. What is the interplay between the social and mobile network? How can we automatically identify strong and weak ties in virtual human relationships by analysing the human interaction patterns? How can we use the social and mobile networks to answer question about the socio-economic structure of a city? We expect to observe an explosion of social mining methods along this path, as richer and more comprehensive data will be harvested by social sensing. An example of early work in this direction is [Wang 2011]. Indeed, social mining is aimed at extracting multi-dimensional patterns and models from multi-dimensional data. To this aim, social mining needs novel concepts and techniques of *multi-dimensional pattern discovery* and of *multi-dimensional social network analysis*. Examples include: (i) multi-dimensional, evolutionary community detection and link prediction driven by similar socio-economic behaviour (homophily), (ii) definition of multi-dimensional measures of network centrality, relevance and trust; (iii) correlations analysis between diverse social dimensions, e.g., proximity metrics in the evolving network of physical contacts derived from user mobility traces and proximity metrics defined with respect to the social space of people profiles; (iv) discovery of correlation patterns connecting behaviours across different social dimensions; (v) multi-dimensional evolutionary pattern mining from social and mobility networks, from streaming data with uncertainty; (vi) discovery of evolutionary patterns for trust and risk measurement. Based on the novel social mining techniques, it will become possible to address in a systematic way many challenging questions about multi-dimensional social phenomena, ranging from the individual sphere of personal behavioral profiles, to the collective sphere of group or subpopulation behavior, to the global phenomena related to diffusion and spreading in the social network, to the analysis of how sentiments and opinions vary in our societies.

- **Computational aspects of social mining**. With the size and complexity of big data, also *new computational challenges* emerge. Researchers have worked with graphs for many years, but the graphs now being plotted from social networks or sensor networks are of an unprecedented

scale, they are produced continuously and massively distributed. Dealing with graphs of that size and scope, and applying modern analytic tools to them, calls for better models, scalable algorithms and other innovations. The mainstream paradigm for analysis today is: gather data first, and analyse them later. This paradigm falls short in view of the challenges posed by social sensing and mining, which requires for real-time reactions to changes in the techno-social ecosystem. This is a perspective that calls for fundamental advances in many fields of data mining and machine learning, such as: concept drift and management of model evolution/obsolescence, learning from continuous data streams, real-time and interactive visual analytics and learning (putting users within the loop), massively distributed data mining, high-performance data mining and big graph analytics. For example, explorative visual analysis of such large datasets calls for database tools able to provide results in real time, for an interactive experience [Fekete 2012]. [Keim 2010] provide a comprehensive view on the challenges related to the visual analysis of large datasets.

1.4.3 Progress Beyond the State-of-art

The creation of the **SoBigData** research infrastructure is the ground breaking attempt to construct a novel ecosystem, where social mining and big data research can be carried out in a multi-disciplinary fashion, in order to pursue the advances towards the above research roadmaps. If successful, **SoBigData** will provide a unique environment for scientists to perform big data and social mining research, previously impossible, due to many technological, legal, commercial and organizational barriers. To reach this ambitious objective, a combination of *networking*, *trans-national and virtual access*, and *joint research* activities needs to be put in place.

- The **networking activities** will foster the creation of the **SoBigData** user community, which will bridge across the diverse areas of expertise of the participants from the different research fields, and achieve truly multi-disciplinary collaborations in the European Research Area. Particular focus will be placed upon: training a new generation of data scientists, establishing the innovation agenda, promoting standardization and open-access policies, disseminating good practices, evaluating impact, taking care of sustainability of the research infrastructure, and promoting the growth of the **SoBigData** starting community.

- The **trans-national and virtual access activities** will leverage the complementarities and multi-disciplinary perspectives into social mining, to establish a unified entry point into the **SoBigData** RI for scientists and their research questions, enabling such users to fully exploit the distributed ecosystem and conduct high quality research. Scientists will have access to, and will be guided in the use of, an innovative RI for working on specific research questions involving social data mining. Particular focus will be placed upon establishing multi-disciplinary research opportunities by means of thematic calls (dubbed *exploratory projects*), endowed with specific installations of the **SoBigData** ecosystem, such as: social mining for measuring well-being, social mining for participatory e-government, social mining for city science, social mining for measuring and mitigating conflicts, and social mining for understanding success in sport or science. Due to licensing restrictions on certain kinds of social data (e.g. tweet redistribution), transnational access activities are particularly well-suited, as they enable scientists from smaller research labs and newly associated states to carry out cutting edge research through visiting leading European host institutions. These will be complemented by a wide range of virtual access services to the national infrastructures, which are being integrated in the **SoBigData** virtual research infrastructure.

- The proposed **joint research activities** will address the advancement of social sensing and mining methods and tools along the research roadmap discussed above, in order to progressively expand the capabilities and scope of social mining and big data analytics methods offered by the **SoBigData** RI. The long-term vision of **SoBigData** goes beyond a platform for sensing and

mining social big data, towards a *techno-social ecosystem* offering a goal-oriented knowledge discovery framework, combining data, methods and expertise. It will be able to configure itself towards the aim of answering questions about the status of the global society. Given an analytical request, such as for instance the measurement of multi-dimensional indices of societal well-being, the **SoBigData** infrastructure will activate a process composed of a variety of interconnected tasks exploiting the full repertoire of available methods for social sensing and mining within the ecosystem provided by the networked infrastructures.

An overview of actions and effects beyond the state-of-art is reported in the table below.

Activity Type	**Action**	**Effect**
Networking Activities: Establishing a the **SoBigData** starting community and driving the enhancement of services through user involvement	• Liaison with related national and international initiatives • Direct communication and events with user communities of the RIs • Online information, user alerts and interaction tools in the virtual research environment, monitoring mechanisms for services and usage	• Community building and communication • Promote wide take up and research use of the **SoBigData** RI
	• Identify current and emerging stakeholder needs and requirements to design and enhance the **SoBigData** RI and services • Ensure that user needs are met and effectively served by the RIs	• Understand and serve real user needs • Foster the **SoBigData** starting user community • Provide valuable big data research services, promoting involvement • Reinforce sense of community and collaborative spirit
	• Raise awareness of the **SoBigData** RI and its benefits • Define good practice and guidelines • Provide training in big social data analytics and how to benefit from the open **SoBigData** RI and its services (e.g. publications, training schemes)	• Promoting the **SoBigData** RI and the best practices and methodology for analysing large social datasets • Foster a new generation of multi-disciplinary data scientists
	• Partnerships with industry and other non-academic stakeholders • Knowledge transfer and consulting services; • Training data scientists in entrepreneurial skills; • Contributions to policy making; • Involvement in relevant industrial and standardisation bodies	• Promote and accelerate innovation in social data analytics • Support startup and spin-off creation
Transnational and virtual access: new opportunities for the next generation of data scientists	• Transnational access to large social data collections and leading national infrastructures from complementary research fields	• Enable young researchers and users to co-work with experts in critical phases of their research • Allow effective use of the research infrastructure • Support innovative multi-disciplinary research

	• Virtual research environment, tailored to the requirements of social data scientists • Online access to virtual collections of social data and over 100 analytics services	• Enable researchers to avail effectively of accumulated data and re-use them • Allow effective use of the research infrastructure
Joint Research Activities: Integrated services for big data collection, analytics, and result evaluation	• Providing integrated service interfaces for social data access and collection	• Unified access to and collection of social data • Improved dataset sharing and reuse • Promotion of repeatable research results through shared datasets
	• Creation of the open **SoBigData** research infrastructure • Integration of leading national infrastructures from complementary research fields	• Creation of a novel European RI • Integrated access to over 100 different social data analytics services, spanning multiple research fields
	• Development and integration of new analytics services, in five thematic clusters (see WP9)	• Bigger set of **SoBigData** services available, to promote multidisciplinary research and take up of the RI
	• Creation of a new evaluation framework	• Promoting repeatability of social data analytics experiments • Community fostering and engagement through open evaluation initiatives

2. Impact

2.1 Expected Impact

2.1.1 Impact as Set Out in the Work Programme

The high level goal of **SoBigData** is to create a novel, multi-disciplinary, pan-European research infrastructure, enabling data scientists to experiment with large-scale social data analytics. In order to ensure that the new RI will last well beyond the project bounds, it will be complemented with the creation of a starting user community, and a wide range of innovation and training activities.

This infrastructure will not be created from scratch, but will instead, be an integrating activity which will bring together a number of already established national infrastructures, as detailed in WP6 and WP7. The new **SoBigData** RI will also build upon the state of the art in e-infrastructures for data and software-as-a-service, and work in cooperation with relevant e-infrastructures, as detailed in Section 1.3.3 above. As shown in Section 1.3.3, the relevant key EU projects and initiatives are directly represented within **SoBigData** by one or more partners. Moreover, the project community building activities include establishing coordination with the other relevant existing or forthcoming national, European or international initiative.

In this way, **SoBigData** will aim to realise the target impact sought by Commitment n. 4 of the Innovation Union flagship initiative, namely "*Opening of Member State operated research infrastructures to the full European user community*".

As a result, for the first time in big social data mining research, researchers, industrial users, and other stakeholders will be offered the opportunity to access a rich social data ecosystem, in a unified manner, and to conduct large-scale, multi-disciplinary experiments on extracting social knowledge. They will draw on cutting edge open-source methods offered by leading European research centres and their infrastructures. Thanks to the open nature of the **SoBigData** RI, it will be possible to have it developed and expanded further, in a sustainable fashion, by users and developers in the **SoBigData** community, formed during the project's lifetime. Since the creation of such a vibrant, sustainable community takes time, the **SoBigData** project is structured so that the new RI is developed and released early in the project, allowing for sufficient time over its four year duration for user engagement, training, collaboration, innovation, and other community building activities.

Being unfeasible to have a very large number of stakeholders as partners in a starting community project with a limited budget, the project has chosen the approach of associating them, with management mechanisms to coordinate their involvement. Formal association at institutional level will require the signature of a Memorandum of Understanding, but individual users will simply be able to join the **SoBigData** community and contribute without further formalities. The benefits from institutional level association will be in demonstrating commitment; better visibility through the **SoBigData** web portal; and also being better represented at user consultations and project events. Such associated institutions will also be assisted with the integration of their own datasets and social data mining services within the **SoBigData** RI. This openness to new contributors will thus enable the new RI to grow organically, alongside its starting user community.

Since the field of big data research is very dynamic and hot, one of the aims of **SoBigData** is to act as a catalyst for innovation and collaboration, by bringing together scientists across different disciplines and giving them a common experimental platform, thus contributing to the development and harmonisation of the European Research Area.

In particular, **SoBigData** will contribute towards realising the impacts listed in the work programme, many of which also align with the five ERA key priorities:

1. **Better access to the best national research infrastructures**, integrated under the umbrella of the common **SoBigData** RI. In this way, currently disjoint hubs of knowledge will be made accessible to researchers irrespective of their nationality, country of work, and facilities at their organisation. This will benefit in particular researchers from smaller research institutions, newly associated states, and early career researchers, all of whom are currently facing difficulties in competing against researchers working in big European centres and industrial research labs. The obstacles to be addressed by **SoBigData** are both in terms of access to big social data and also scalable methods and services for their analysis.

2. **Training a new generation of mobile researchers**: The transnational and virtual access activities in **SoBigData** will offer new research and pan-European collaboration opportunities, to help develop a new generation of data scientists. Through exposing them to mobility opportunities and transnational collaborations, coupled with making available a rich set of training materials and events, the project will ultimately aim to improve the employment opportunities of European data scientists.

3. **More effective national research systems**: By integrating complementary datasets and big social mining algorithms from five different research fields, **SoBigData** will contribute towards improving the efficiency of the underlying national research infrastructures by eliminating the need for each of them to implement such a wide range of software tools and services. There will also be efficiency benefits realised through the integration and harmonisation of the data and service interfaces across these national infrastructures. The

joint research activities will avoid unnecessary duplication of effort and promote common, inter-disciplinary research and development methodologies.

4. **Optimal transnational cooperation**, realised through streamlined and coordinated collaboration, not only between project partners, but also including the new pan-European starting community of big social data scientists. Thanks to the transnational access, joint research, and training activities, truly pan-European cooperation will arise, focused on tackling the hot research challenges in big social data mining, detailed in Section 1.4. **SoBigData** will also stimulate multi-disciplinary research and cooperation, including sharing of good practices in research infrastructure deployment, data standards and management of massive data, provision of high-quality services, and novel experiments.

5. **Accelerating innovation through partnerships with industry**: Many of the underlying national infrastructures have a strong track record in promoting innovation and knowledge transfer through open science and partnerships with industry. Successful previous and ongoing collaborations include, inter alia, the BBC, the Press Association, Financial Times, Thompson Reuters, Elsevier, the British Library, the UK National Archives, Nokia, Vodafone, Microsoft, Google, Yahoo Research, and Skype. The new RI will facilitate this process further, as detailed in Section 2.1.2 below.

6. **Effective data, method, and knowledge sharing**: From its inception, the **SoBigData** RI will focus on answering current and emerging user needs, around effective sharing of datasets, analysis methods, and best practices in social sensing and social mining research. Relevant standards, initiatives, and e-Infrastructures will play further enabling role.

With respect to **scientific impact**, **SoBigData** will enable the big social data research communities to be at the forefront of data-intensive research, working on research questions that could not be addressed before; drawing on volumes and types of social data they had hitherto been unable to analyse; and allowing for collaborations between previously fragmented national infrastructures and initiatives across Europe and beyond. In particular, a new generation of multi-disciplinary data scientists will be trained to exploit in the best way the new research infrastructure, with its wealth of datasets, social data analytics services, and visual research environments.

Expected impacts in terms of research productivity and innovation comprise, for example, easy access to distributed virtual collections, drawing on datasets currently scattered across different countries and institutions, and promoting open, repeatable science through shared data and open methods. This will be fostered further by the implementation of a much needed benchmarking and evaluation framework, that will bring together researchers and help them work on challenging shared tasks.

Going beyond data scientists, **SoBigData** will likely impact on other research fields, including but not limited to web science, digital humanities, social informatics, and economics. They will benefit from the substantial quantity of data and analytical services integrated within **SoBigData**.

2.1.2 Improving Innovation Capacity

The project will also focus on improving the innovation capacity and strengthening growth of European big data companies. For instance, USFD and their GATE text mining infrastructure have a successful track record of attracting code developers and contributors from SMEs and larger companies, who then also build and sell innovative products on top of the open-source infrastructure. **SoBigData** will roll out such research-industry partnerships at project-wide level.

Secondly, by open-sourcing most of **SoBigData**'s results, we will support SMEs and other companies in adopting and benefiting from the new RI. The RI will help both with their new

research and development activities, and also, provide building blocks in innovative products and services. More specifically, SMEs and bigger companies will be attracted during the project, capitalising specifically on the strong commercial partnerships already in place through project partners and users of their open-source tools and national infrastructures. This currently includes over 50 SMEs and entrepreneurs worldwide, as well as the bigger organisations listed above.

The project partners will set up knowledge transfer and consultancy services, aimed at SMEs and bigger organisations. Often companies need to tailor generic big data analytics algorithms to the specifics of their vertical or product, so the project will be ideally placed to offer such customisation and knowledge transfer services, on top of the numerous open-source algorithms and free services provided by the **SoBigData** infrastructure.

In more detail, there is currently a large, untapped potential for developing new affordable, big social data analytics products and services, since many companies in diverse areas (e.g. business intelligence, market research, campaign and brand reputation management, customer relationship management, enterprise search and knowledge management) are analysing and comparing big social data, often in a labour intensive and expensive manner. This manual approach is quickly becoming infeasible, due to the continued growth of data volumes.

The range of vertical domains that can be targeted through innovation activities in **SoBigData** is very wide. In the rest of this section we discuss some of these, but this list is by no means exhaustive and is likely to grow during the project's lifetime.

Political Analysis, PR, campaign, campaign and communication professionals: A big social data analytics platform will offer companies the chance to offer cheaper services and expand their offerings to a greater variety of customers, through being able to analyse larger data volumes. Additionally, the automated analysis can highlight to customers when trends are changing so that they can commission work at the right time. Communications consultancies could benefit similarly with respect to lower costs, faster services, and more in-depth analysis of big social data. The project is also relevant to political consultants, especially those working on pan-European level.

Customer Engagement and Market Research: Social media streams are an extremely valuable resource for companies, who stand to gain a direct insight about what customers think of their products. Information is currently gathered from focus groups or questionnaires or manually by employees whose job is to read Twitter and user forums. This suffers from high cost, poor coverage and bias. Therefore we will promote the automated **SoBigData** services to companies interested in using intelligent technologies, as well as companies offering bespoke products in this market. We will seek to attract further significant industrial interest and generate exploitation opportunities, firstly by presenting project results to our existing industrial contacts and secondly, by targeting new ones through webcasts, YouTube presentations, and participation in relevant industry-oriented forums (e.g. Intelligent Content, Text Analytics, KDNuggets).

Financial and Business Intelligence: Another target are companies building products and services for financial professional communities, investors, financial authorities, and analysts. The project partners will promote **SoBigData** results to the business intelligence community, through existing contacts (e.g. XBRL Europe, LIST spa, MPS bank, CreditReform, Unicredit), as well as through relevant business conferences, and online communities (e.g. Business Intelligence Network, LinkedIn BI group).

eGovernment, Digital Libraries and Archives: USFD's and KCL's existing text mining tools are already used in electronic archives and digital libraries to enable semantic search, e.g. for person names and locations. For instance, recently the UK National Archives awarded USFD a contract to develop a bespoke 'intelligent discovery tool' to improve search over archived UK Government

websites (around 7TB of data). In that context, the **SoBigData** RI will be a very cost-effective solution, which would benefit many similar institutions. USFD and LUH have also participated in the ARCOMEM project, which involved working with a number of digital libraries and archives. Further impact in this area will be achieved through KCL's digital humanities partners, and the users of the DARIAH and CLARIN e-Infrastructures, with which **SoBigData** will cooperate.

Official statistics: national statistical institutes, such as ISTAT, and EUROSTAT, are beginning to explore novel ways to produce high-quality statistical information using big social data sources. This trend is motivated by achieving faster and more accurate measurements of socio-economic indicators, such as employment, poverty, inequality and the various dimensions of well-being, in order to achieve a better monitoring of our societies at smaller geographical scale. CNR and UNIPI have initiated pilot projects on this line with ISTAT. **SoBigData** will represent an ideal environment where this high-impact line of research and development can take place, especially for the ability to access multiple big data sources and to integrate multi-disciplinary competences.

Smart cities and smart communities: big data and social mining at urban scale, when merged with traditional infrastructures, allow for new understanding of urban problems; for effective and feasible ways to coordinate urban technologies; for new models and methods for using urban data across spatial and temporal scales; for developing new forms of urban governance and organisation; for discovering critical problems relating to cities, transport, and energy; and identifying risk, uncertainty, and hazards in the smart city. CNR, UNIPI and FRH are frontrunners in the exploitation of big data analytics for smart cities, in collaboration with industry and public administration. The proposed **SoBigData** infrastructure has the potential to become the place for sophisticated data-driven modelling and simulation of urban innovation projects, offering novel opportunities to local and regional governments, policy makers, and citizen organizations.

2.1.3 Social and Other Impacts

The project will develop an **innovation accelerator** (Task 5.4)**,** which will measure the impact of the **SoBigData** research infrastructure at societal, policy, and community levels. It will monitor, analyze, and enhance the social processes around data science and innovation. These tools will help data scientists, founders, and other decision makers to track the progress of the multiple disciplines that interface into data science; to measure funding impact and methods used; and to track the impact of emerging topics and technologies related to mining big social data.

SoBigData is also well positioned to make contributions to **policy making**. Consortium members (e.g. FRH, UT, CNR) will contribute actively to policy making and pursue other socially important impacts (e.g. helping government organisations and NGOs to benefit from big social data analytics, in order to obtain knowledge from citizen forums). In particular, research outputs on social mining for participatory e-government (Task 9.8) will be used to formulate policy recommendations in regards to e-government service development.

Smart disaster management: FRH is the technical partner of the Federal Office of Civil Protection and Disaster Assistance (BBK), working on early detection of relevant critical events in near real-time. **SoBigData** methods are relevant here, including distributed event recognition based on various data sources such as Twitter, infusion of the human-as-a-sensor principle, analysis of social groups and impacts during nation-wide disasters.

The project will also aim to impact relevant **industrial and standardisation bodies and international initiatives**, including inter alia:

- the European Research Data Alliance (RDA)
- the Ontology-Lexica Community Group of the World Wide Web Consortium (W3C)

- the Federal Association for Information Technology, Telecommunications and New Media, Bitkom e.V.
- the Smart Data Innovation Lab (SDIL) - a German industry network on big data intelligence
- the Networked European Software and Services Initiative (NESSI)

2.2 Measures to Maximise Impact

2.2.1 Dissemination and Innovation Activities and Impact Planning

For maximum impact, **SoBigData** will adopt a multi-channel, multi-target dissemination approach. Due to **SoBigData** being a research infrastructure, the main dissemination targets will be social data scientists, which currently span multiple research communities (data mining and pattern discovery, text and social media mining, mobility analytics, social network analysis, visual analytics, etc.). We will also complement these with innovation actions and events aimed at companies in our target verticals (especially SMEs), government organisations, and policy makers. Since the project will produce a new research infrastructure, as well as new open-source tools and services, training and cross-disciplinary community building is also a target.

There will be a full programme of scientific papers and presentations at technical and scientific conferences. These will be aimed at sharing the results of **SoBigData** with the scientific community, to encourage the take-up of the new research infrastructure and enable the growth of the **SoBigData** starting community with new research labs and scientists. Example conferences to be targeted are:

- **Language Processing and Text Mining**: (E)ACL, COLING, CONLL, EMNLP, CIKM, SIGIR, ECIR
- **Online Social Networks**: ACM COSN, WWW, SocialCom, ACM SIGCOMM, IEEE INFOCOM, ACM CHI
- **Social Media and Social Informatics**: ICWSM, WSDM, CIKM, SIGIR, WebScience, Social Computing, ECIR, Social Informatics, Digital Humanities
- **Data Mining and Pattern Discovery**: CIKM, IEEE-ICDM, ACM-SIGKDD, PKDD, SDM
- **Information Visualisation**: CHI, IUI, Inf. Visualisation, TVCG, CG&A, EuroVis, VAST, IEEE InfoVis

Complementing this, there will be a programme of papers and articles in the information technology and general business literature, as well as presentations at IT and business seminars and conferences. We will also target key stakeholder groups for the project's innovation activities, split by target verticals, as discussed in Section 2.1.2. Again, this will be directed towards the European community and associated countries. Targeted events will include:

- **Big Data**: Analytics and Big Data Congress, Big Data TechCon, Big Data Innovation Summit, Big Data Analytics
- **General IT, language technologies**: Online Information, SemTech, Text Analytics Summit, European Semantic Technology Conference, CeBIT
- **Law & technology**: CPDP - Computers, Privacy and Data Protection Conference
- **Digital media**: Media Pro, Henry Stewart, Online Information, FT Digital Media, DigitalMedia Europe

The project will also set up and maintain a **public project web site**. It will function both as a project dissemination tool and as a web-front end of the **SoBigData** research infrastructure. The website will also provide training materials, code examples, API documentation, and online

tutorials explaining how the research infrastructure can be used to run new experiments; be integrated in applications; or be used as component services by SMEs and other companies.

SoBigData will also harness **social media as a dissemination channel** and will actively promote the new research infrastructure and its services to the large big data research community. We will engage with relevant groups on LinkedIn, Facebook, and blogs in the fields of data mining and pattern discovery, social network analysis, text and social media analysis, semantic web, business intelligence, new media. The project will also setup an active Twitter dissemination account.

Within the limitations imposed by appropriate protection of intellectual property, **all scientific results of this project will be made available to the research community**. In addition, a number of big social data analytics tools and resources from **SoBigData** will be made available as **open-source** to facilitate take-up.

Another vital part of the dissemination and innovation activities will be the **support of a community of early adopters** of the research infrastructure, through the web site, online user collaboration tools, code examples, training events, datathons, etc.

Training materials resulting from the project will be gathered and used by the academic partners as part of their course materials at postgraduate level, as well as, used by all partners to provide courses to companies and other non-scientific stakeholders. In addition, the networking, transnational, virtual, and joint research activities in the project will involve PhD students and early career researchers, thus aiming to enlarge the social data analytics research community.

SoBigData will also produce **results relevant to other on-going projects**, with which **it will cooperate**. In addition, project partners have already established links to complementary projects in all key RTD areas, as detailed in Section 1.3.3.

The innovation activities (see WP5 for details) will widen the project impact beyond the target scientific communities, to include **outreach towards businesses in diverse sectors (especially SMEs),** as well as **other stakeholders** (e.g. government bodies, funders, policy makers). For instance, we will provide consultancy to business and contractual research and development, train data scientists in entrepreneurial skills, contribute to policy making, and participate in relevant industrial and standardisation bodies. To complement these, **SoBigData** will develop an innovation accelerator, which will help measure impact of the research infrastructure at scientific, societal, economic, policy, and user community levels.

A range of other instruments will also be employed for dissemination including press releases and media engagement, further participation in appropriate networks of excellence, and activities aimed at both the academic and industrial sectors (e.g. delivery of lectures at summer schools, industry-oriented dedicated workshops, etc). For details see WP3, WP4, and WP5.

2.2.2 Research Data Management

Terabytes upon terabytes of big social data from thousands of datasets will be made available by the consortium partners, via transnational, and where possible, virtual access. Moreover, new big social datasets will be acquired during the project from a variety of sources, as detailed in WP8 and the table below. This reports the different categories of data already available in the consortium, and new datasets that we plan to collect and organize, as well as the partners who, at the moment, own these datasets.

Category	Description	Partners
Social Networks	Several kinds of Social Network data are available such as Google+ circles and friendships, collaborative writing interactions (20Lines), file sharing on FileList.org (80 GB); 68.7 million Pinterest users and 3.8 billion directed edges; social contacts in location-based social networks (BrightKite)	CNR, ETHZ, KCL, USFD, LUH
Social Media	Big data sets from social media are provided, including 12 TB compressed Twitter data, starting from May 2009; also country-specific tweet collections (e.g. 380k Italian users, ~300M of tweets; 250k UK users and followers) and event specific tweets (e.g. world cup, earthquakes, London Olympics); Foursquare checkins (over 220,000 locations and 22 million checkins); public Facebook pages, blogs (over 2TB of data), Flickr images (245GB of metadata + 400GB of images); Youtube data (146 GB of user profiles, network, published clips, etc.); social tagging datasets (Delicious and Flickr – 3GB of compressed metadata).	USFD, AALTO, CNR, UNIPI, ETHZ, KCL, LUH
Mobility data	In the mobility category, among the others, are provided data about: GPS tracks of 170,000 private cars in Italy over a month (approx. 8% circulating cars in Tuscany); 2 Billion of Oyster Card records; bike sharing in Barcelona; US flights; NY taxis; OpenStreet maps (400GB).	CNR, AALTO, ETHZ
Mobile Phone data	Several large mobile phone (GSM) datasets are available at national and regional scale, including Wind and Orange Call Detail Records (CDR) datasets (approx. 10 million users over several months); 759 million call records; 243 million text messages;	AALTO, CNR, ETHZ, KCL
E-government	Detailed usage records of Estonian E-government and E-health services (e-prescriptions, e-identity, e-signature, e-voting, e-taxation)	UT
Web crawls and logs	Crawl of world wide web services and national level domains (e.g. 80TB compressed crawl of .de; 25 TB uncompressed of web pages in 10 different languages; 100s of millions of news articles); search queries and associated clicks (e.g. ~50M queries with their associated clicks from several search engines)	LUH, ETHZ, USFD, UT, CNR
Retail Market Data	Detailed records of supermarket purchase transactions in Italy (2 million users, 150 stores in the country, 6 years coverage; around 2 billion data points).	CNR
Text Archives and annotated Corpora	Linguistically annotated corpora of chats, reviews, emails, blogs, comments, web pages, and newsgroups in English, Italian, and other languages (1.8 million words annotated; 1.5 trillion words un-annotated).	USFD, CNR, ETHZ, UNIPI
Rating Data and Trends	Trends about news, epidemics and events (i.e. Google Flu Trends (50 million of the most common search queries), We Feel Fine, Cyber Emotions)	ETHZ

While all these datasets will be available via Transnational Access, only a subset of them will be made available via Virtual Access, due to sharing restrictions and privacy issues. Moreover, in order to comply with the Terms of Service defined by data providers (i.e. Twitter, Facebook) some datasets will be made available as Virtual Collections: only unique identifiers and queries needed to reconstruct the dataset will be released.

During the project two types of activities will be pursued: harvesting relevant data and producing newly generated data. Concerning the harvesting step, the data collection approaches are described in WP8. We also plan work on data archiving and preservation, where the ground rules will be set in the data management plan (a deliverable in WP8). Both pre-existing and newly acquired datasets will be made available through web services part of the **SoBigData** virtual RI, where this does not infringe privacy, copyright, and terms and conditions of the data provider.

Concerning the newly produced data, the consortium will rely on the state-of-the-art standards for data generation, maintenance and exploitation. Thus, the data will be maintained in XML, but it will be also RDF-ized and mapped to an adequate ontology, following best practices in this area.

Concerning data access and storage, data will be stored according to the source requirements and if deletion is required (e.g., tweet deletion), we will remove the content from our data storage as soon as we receive the request. To prevent data leakage, data will be accessible for the consortium behind a firewall and if applicable virtual private networks will be created. For some special types of data (defined by the source as such) that must not be stored on unencrypted removable devices, all partners will be required to encrypt the hard drives storing any such data.

In addition, all data will be accessed, stored and used according to the EU information privacy and community recommendations, and in general, the **SoBigData** legal and ethical framework (WP2).

2.2.3 Open Source Software

The Open source software brought by the consortium partners via their national infrastructures is summarized below, and detailed in WP6 (TA) and WP7 (VA). The algorithms and services will be continuously enhanced, invented and integrated as part of WP9.

Category	Algorithms	Partners
Social Network Analysis	• Community Discovery • Evolutionary Network analysis • Information Diffusion • Multidimensional Networks • Network and Mobility • Link analysis • Fake account detection	CNR, AALTO, SNS, IMT, ETHZ, UT
Data Collection	• Social Media Crawling • Indexing and Search	CNR, AALTO, LUH, ETHZ
Text Mining	• Language identification • Part-of-speech tagging and shallow parsing • Named entity recognition • Information Extraction	USFD, CNR, LUH, KCL, UNIPI, ETZH

	• Entity Linking and Disambiguation • Sentiment analysis • Data enrichment and annotation • Text classification • Evolution • Text summarisation	
Mobility	• Mobility Profiling • Location Prediction • Mobility Borders • Presence analysis • Recommendation • Trajectory pattern mining	CNR, FRH
Privacy	• Anonymization • Differential Privacy • Risk analysis	CNR, KCL
Data Mining	• Classification • Quantification • Image Classification • Clustering • Learning to Rank	UNIPI, CNR, LUH, UT, KCL

2.2.4 Knowledge Management and Protection

The consortium partners will bring in their world-leading expertise and knowledge to make the **SoBigData** project a success. Knowledge created during the project will be distributed within the consortium - and where not subject to confidentiality, beyond the consortium - to enable a targeted and coordinated progress towards the project objectives. This also requires active knowledge exchange between the project partners and the **SoBigData** user community, which will be fostered through the community building, dissemination, training, and transnational access activities. In addition to creating an open, creative and flexible project environment, which is built on mutual trust and respect, knowledge exchange will be fostered by meetings, teleconferences, technology demonstrations, the publication of information on the project web site, and scientific papers.

As shown, the **SoBigData** consortium includes many partners who bring in open source software for social data analytics. The new methods and services to be developed as part of the joint research activities will also be made open source under suitable licenses. Since the project is seeking to maximise take up by SMEs and other stakeholders outside of academia, commercially friendly licenses will be considered or, where this is not suitable, a dual licensing option will be offered.

Open source alone is not enough - the infrastructure and new services will be supported, maintained and promoted, and the communities of developers nurtured and organised. Many **SoBigData** partners have a proven track record in this area, which will serve as a solid basis for starting and growing the **SoBigData** user community. To maximise take up of the **SoBigData** RI, it will be promoted firstly to the existing user communities of the underpinning national infrastructures. More importantly, the RI will also be provided as easy-to-use web services for virtual access by any European researcher, anywhere. Accompanying documentation, training materials, and example code will also be available, as well as training events and datathons. With respect to the datasets

shared through the **SoBigData** RI, we will aim to maximise their take up through adoption of relevant Open Data standards and promotion and publication in other relevant data repositories.

In accordance with EC policies, scientific publications will be made publically available through open access. Most scientific partners already have in place policies and infrastructure for self-archiving and gold open access, depending on the publication venue. As an alternative, open access papers can be deposited via the OpenAIRE portal (www.OpenAIRE.eu).

More formal issues of Management of Knowledge will be addressed in the Consortium Agreement, which will be signed before the start of the project.

2.2.5 Communication Activities

Based on the identification of key stakeholders and the associated key messages for each target groups, **SoBigData** will develop a set of tools and techniques that will be used to carry out diverse and novel activities and create compelling events and workshops that live-up to the wide-ranging challenges arising in big social data analytics. The material's content and design will be adapted to the needs, expertise and focus of the different target groups and dissemination channels as described above. Our dissemination and communication tools will include a balanced mix of traditional (workshops, summer schools, conferences and scientific publications) and innovative methods (datathons, online videos, social media, specialist web front end) to promote take-up and reach widest possible impact. Diverse media engagement will be an essential component, including press releases, social media (i.e. Twitter, Facebook, LinkedIn, project news and blog), online videos (e.g. Youtube), and other promotional materials, such as a project logo, presentation and factsheet/brochure.

A **SoBigData** Dissemination and communication kit will be created and updated at key milestones:

- **Project logo**: needed to establish the visual identity of the project, to appear on all project presentations, online videos, web pages, training materials, etc.
- **Project Presentation**: A public, high-level presentation of the project's aims and objectives; research infrastructure scope, design, and positioning; transnational and virtual access offered; networking activities and how to get involved in the **SoBigData** starting community.
- **Presentation template**: A presentation template, to include the project logo and visual identity, as well as funding acknowledgements. Will be used by all partners when creating dissemination and training materials arising from the project.
- **Project poster**: with an appealing graphical design and presenting the key aim, objectives, and results of the project. The poster will also promote the project website and be presented at workshops and other dissemination events.
- **Project brochure/factsheet**: Similar content to the project presentation, but in an appealing foldable brochure design.
- **Project website**: This will be the web face of the project, where stakeholders will find information about the project, publications, public deliverables, training materials, open and past calls for transnational access, networking events, and ways to join the **SoBigData** community. The web site will also provide an attractive entry point to the **SoBigData** virtual research infrastructure, as described in WP7 and WP10. The web site will be designed following standard Web technology, including RSS feeds, social sharing and bookmarking, news and blog, tagging. There will also be a private area, which is only available to the consortium partners, the EC, and project reviewers. This will act as a document sharing and

collaboration platform, including all non-public deliverables, administrative and financial information, minutes from project meetings, and a project management wiki.

- **Press releases and media coverage**: A press release will be made by the coordinator at the beginning of the project, which will then be adapted, translated, and released by each of the partners. Subsequent, periodic press releases will be made at key milestones and/or events, e.g. on launching the **SoBigData** research infrastructure; on signature of important collaboration agreements with ESFRI infrastructures or major international initiatives; on achieving critical mass of the **SoBigData** community. Media coverage will be tracked and reported on the project website, social media channels, and dissemination reports.
- **Scientific papers**: See section 2.2.4 for details on open access.
- **Social media presence**: Within a month of project start, we will setup Twitter and Facebook accounts. These communication means will be promoted through the more traditional channels (factsheet, web site, presentations, posters) and all partners will ensure that active the project has active social media engagement. Conversely, social media will be used to promote actively the other communication tools, e.g. web site, publications, deliverables, software releases, press releases, calls for transnational access projects, blog posts, and other news.

Other publicly facing communication activities, such as summer schools, datathons, and training events, are described in more detail in Section 2.2.1, as well as WP3 Dissemination and Community Building, WP4 Training, and WP5 Innovation. The mechanisms for publicising the calls for transnational access and stakeholder-oriented promotion of the free virtual access offered by the **SoBigData** infrastructure were described in WP6 and WP7 respectively.

With respect to project-internal communication, please see Section 3.2 on project management.

The success of the communication activities will be monitored closely and reported in the periodic activity deliverables arising from WP3, WP4, and WP5. In more detail, download statistics for the various materials from the dissemination and communication kit will be reported, as will be statistics for web site access, presentations given, peer-reviewed papers published, media coverage of press releases made, and social media activity reach (e.g. number of Twitter followers, re-tweets of key messages, number of Facebook group members, number of posts and discussions). These will complement the user statistics for the **SoBigData** RI, to be compiled and reported as part of WP7 on virtual access.

3. Implementation

3.1 Work Plan — Work Packages, Deliverables and Milestones

The work plan groups activities according to the three axes: community building (including innovation and networking activities), building the social mining research infrastructure and making it accessible to its users (provided by virtual and transnational access). In particular, as shown in Figure 2, WP2, WP3, WP4, WP5, WP11 are aimed at constructing the community composed by excellence centres, other academic and industrial users, and trainee data scientists. WP8, WP9 and WP10 concur to set up the core of the project, which will deliver the actual research infrastructure by integrating and enhancing the underlying national infrastructures. The third group of work packages (WP6 and WP7) aims at transferring the results, i.e. to have an impact on society by providing virtual and on site access for the RI stakeholders. Figure 2 evokes how the three components continuously feedback in a self-reinforcing loop, which progressively yields higher impact.

A short description of the scope of each WP follows:

- Community building:

- **WP2 - NA1_Legal and Ethical Aspects:** the main task of this work package is to define and implement the *legal and ethical framework* of the **SoBigData** research infrastructure, in accordance with the European and national legislations (including data protection and intellectual property rights), and to ensure the development and application of policies for the collection and analysis of large-scale datasets in an ethical manner by-design.
- **WP3 - NA2_Dissemination, Impact, and Sustainability:** this work package will develop the dissemination and impact strategy for the entire project, by using and integrating the dissemination strategies of all involved infrastructures. It will identify stakeholders and address these through a range of targeted communication means and channels.
- **WP4 - NA3_Training:** the aim of this work package is to establish a joint training and education domain on social big data in the European research area. It explores and develops conventional and unconventional training experiences for MSc and PhD students and early career post-doctoral researchers. Moreover, it proposes campaigns aimed at high schools for attracting bright young minds into data science, with special emphasis on gender issues.
- **WP5 - NA4_Accelerating Innovation**: The aim of this work package is to widen impact beyond scientists, through innovation activities aimed at industry and other stakeholders (e.g. government bodies, non-profit organisations, funders, policy makers), including building partnerships with and transferring knowledge to stakeholders. It will also develop an innovation accelerator, which helps to monitor and measure the impact of **SoBigData** at societal, economic, policy, and community levels.
- **WP11 - NA5_Evaluation** develops a new benchmarking and evaluation framework for big data analytics and social mining methods, for use by the main stakeholders of this project, the research and industrial community of social analytics. It will also produce a reference point for stakeholders from other research communities, who would like to get involved.

- Building the social mining research infrastructure:

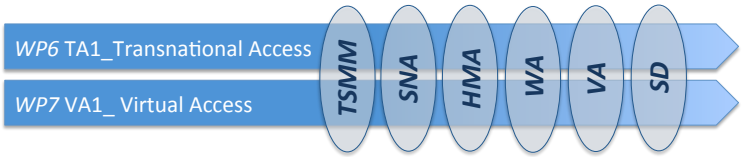
- **WP8 - JRA1_Big Data Ecosystem:** research on social data relies on massive data sets of digital traces of human activities. This work will make the wealth of social data sets of the consortium available for collaborative research adopting various strategies, ranging from sharing the open data sets with the scientific community, to sharing the data with disclosure restriction among the infrastructure users, or allowing data access within secure environments at each local installation. In addition, the consortium will develop methods and ad-hoc campaigns for crowd sensing digital footprints.
- **WP9 - JRA2_Integrating Big Data Analytics Methods and Techniques** will integrate within the **SoBigData** RI and enhance the analytical methods for mining social big data. The expected outcomes are methods that are simultaneously principled and practically applicable, providing a broad multi-disciplinary foundation for big data analytics.
- **WP10 - JRA3_SoBigData e-Infrastructure**: the goal of this work package is to design and create the **SoBigData** RI, that will integrate existing datasets and services from the national infrastructures, as well as the new datasets and methods from WP8 and WP9 respectively.

- Virtual and on site (transnational) access

- **WP6 - TA1_Transnational Access**: provide on-site access to the infrastructures and the accompanying world-leading research expertise. Calls for on-site access will be launched for two types of projects: *Exploratory projects* are research experiments bound to multi-disciplinary themes specified in the calls; *Open call, blue-sky projects* are targeted at researchers wishing to explore the infrastructure for their own research topics.
- **WP7 - VA1_Virtual Access**: This WP will offer virtual access to the **SoBigData** RI from WP10, incl. sharing and reuse of existing and newly created social data mining resources.

Transactional and Virtual access is organized into six thematic clusters of common competences and services of the seven national infrastructures, to avoid duplication and stimulate integration:

1. *Text and Social Media Mining* (TSMM)
2. *Social Network Analysis* (SNA)
3. *Human Mobility Analytics* (HMA)
4. *Web Analytics* (WA)
5. *Visual Analytics* (VA)
6. *Social Data* (SD)



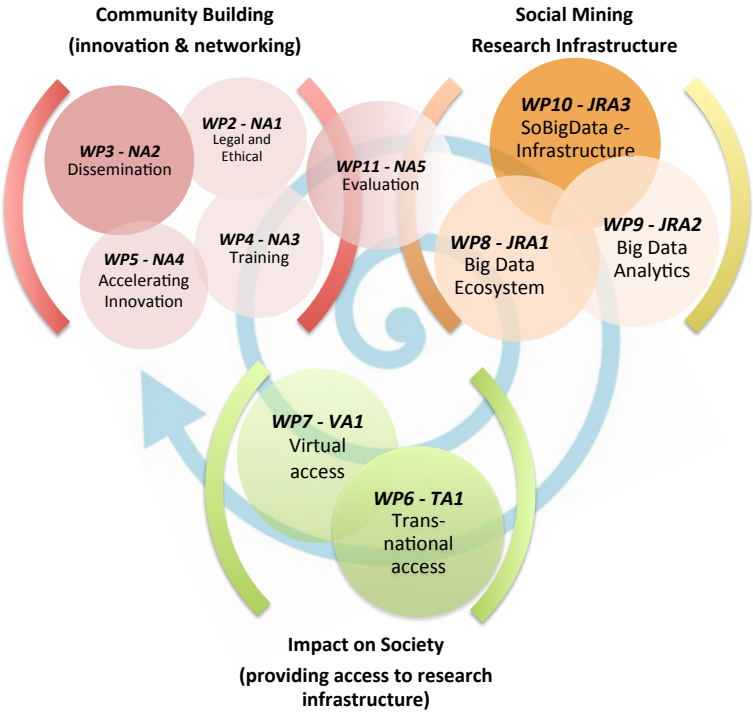


*Figure 2. **SoBigData** components and its self-reinforcing spiral.*

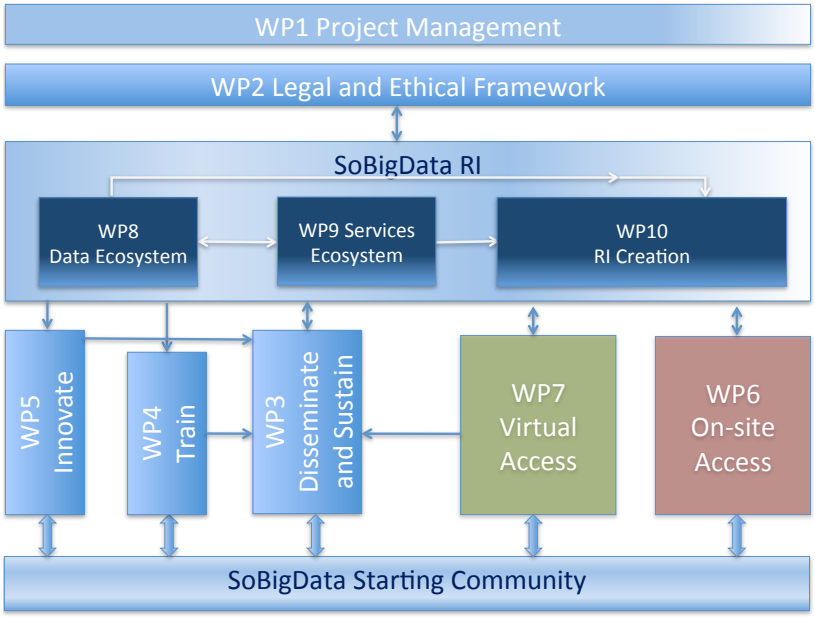


*Figure 3. PERT diagram showing the interdependencies of **SoBigData** WPs.*

Figure 3 shows **SoBigData**'s PERT diagram, displaying a functional view of the interdependencies among the work packages.

Figure 4 describes the timeline of the planned **SoBigData** events, while Figure 5 outlines the structure of WPs into Tasks together with the Gantt chart of temporal distribution of WPs and Tasks over the project span of 48 months (16 quarters).

	1	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48
Project Management																	
periodical report							■						■				■
Meetings																	
kick off	■																
project meeting			■		■		■		■		■		■		■		■
Workshops & Conferences																	
workshops			■		■						■				■		
final conference																	■
confererence							■										
Schools																	
summer schools						■				■				■			
training high-schools					■				■				■				
Special Training Events																	
datathons				■				■				■				■	

*Figure 4. Timetable of **SoBigData** planned events.*

SoBigData work plan			Gantt Chart (activity periods of tasks across all WPs)															
			Year 1				Year 2				Year 3				Year 4			
ID	Title	Lead	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
WP1	*Management*	CNR																
T1.1	Project administration and financial management																	
T1.2	Consortium and Project Management																	
WP2	*NA1 Legal and Ethical Aspects*	TUDelft																
T2.1	Ethics Board																	
T2.2	Ethical and legal framework																	
T2.3	Privacy-by-design and Value-Sensitive Design in social mining																	
T2.4	Intellectual property and data ownership framework																	
WP3	*NA2 Dissemination, Impact, Community Building, and Sustainability Activities*	USFD																
T3.1	Stakeholder Analysis and Dissemination Plan																	
T3.2	Development of Dissemination, Communication and Impact channels																	
T3.3	Sustainability Planning																	
T3.4	Impact Coordination																	
T3.5	Community Building and Outreach																	
WP4	*NA3 Training and Best Practice Guidelines*	KCL																
T4.1	Summer schools																	
T4.2	Training Modules for stakeholders																	
T4.3	Series of Datathons																	
T4.4	Addressing gender and diversity issues in data science through training																	
WP5	*NA4 Accelerating Innovation*	ETHZ																
T5.1	Partnerships with industry and contributions to policy making																	
T5.2	Knowledge Transfer and Consulting Services																	
T5.3	Entrepreneurial skills for big data entrepreneurs																	
T5.4	Innovation Accelerator																	
WP6	*TA1 Transnational Access*	USFD																
WP7	*VA1 Virtual Access*	CNR																
WP8	*JRA1 Big Data Ecosystem*	LUH																
T8.1	Data Management Plan																	
T8.2	Participatory & Opportunistic Crowd Sensing																	
T8.3	Analytical Crawling																	
T8.4	Integrating Open Data through the Living Archive platform																	
WP9	*JRA2 Big Data Analytics Methods and Techniques*	UNIPI																
T9.1	Social Network Analysis																	
T9.2	Social Pattern Discovery																	
T9.3	Text mining and semantic annotation of big social data																	
T9.4	Mobility data analytics																	
T9.5	Visual analytics																	
T9.6	Social behaviour and human dynamics																	
T9.7	Engineering Massive Scale Social Mining																	
WP10	*JRA3 e-Infrastructure for Social Analytics*	CNR																
T10.1	e-infrastructure interoperability																	
T10.2	Integration to the e-infrastructure																	
T10.3	e-infrastructure workflow specification language																	
T10.4	e-infrastructure VREs																	
WP11	*NA5 Evaluation*	FRH																
T11.1	Establish connections with existing relevant evaluation initiatives																	
T11.2	Evaluation framework for social data mining																	
T11.3	Sustainability of the framework and engagement of commercial providers																	
T11.4	Cross-Disciplinary Social mining Demonstrators																	

Figure 5. Gantt chart of temporal distribution of WPs and Tasks.

3.2 Management Structure and Procedures

On the basis of the experience gained by the **SoBigData** partners in previous multi-disciplinary and multi-national projects, the management structure of the project is based on the roles/mechanisms explained below. All project management activities will be supported by e-facilities to minimise travel and to speed up the decision making process.

In particular, the overall objective of **SoBigData** is to strengthen and consolidate a network of leading research centres and infrastructures through Europe and to enhance the integration of scholars, data providers, and users within the **SoBigData** stakeholder community.

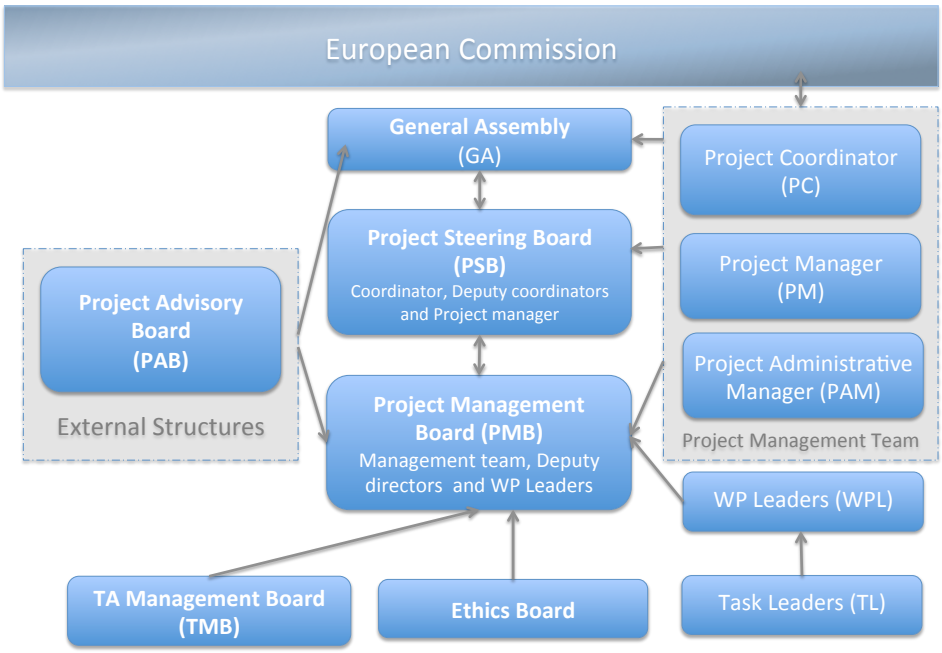


*Figure 6. **SoBigData** Management Structure.*

The **SoBigData** management structure, shown in Figure 6, is designed for accurate and timely implementation of the activities that will be undertaken in the project. The management structure secures the coordination and integration of these activities towards the **SoBigData** objectives of the project and will thus help to reach the desired synergistic effects across partners and infrastructures.

In order to achieve our objectives, the work has been divided into 11 work packages. Each WP is led by investigator/researchers with an established track record and solid experience in running projects of this scale. It is a deliberate design choice to have a mix of senior researchers and younger ones in the increasing stage of their careers, in order to blend experience and enthusiasm.

Three key management roles are defined for the project:

1. **The Project Coordinator (PC),** *Fosca Giannotti*, director of research at CNR, will be in charge of the overall scientific coordination and project implementation, as well as administration. The coordinator is the contact person for all interaction and communication with the European Commission and all legal aspects, monitoring the implementation and progress as well as the achievements of the project and reporting to the Commission. The PC acts as chair of the Project Management Board, the Project Steering Board and the Project Management Team. Two Deputy coordinators will assist the PC: *Kalina Bontcheva* (USFD) and *Dino Pedreschi* (UNIPI).

2. **The Project Manager (PM)** will report to the Project Coordinator and will be responsible for administrative management tasks (organisation of meetings and events, coordinating the preparation of reports, monitoring and maintaining the content on the website, facilitating communication within the consortium, and preparing publicity materials). The PM will be responsible for resolving financial and other administrative issues and monitoring the completion of technical and business objectives. The PM will also participate in facilitating dissemination and training activities in collaboration with the responsible WP leaders. The PM will be *Valerio Grossi* (UNIPI). The PM will be supported by the Project Administrative Manager (PAM), *Chiara Falchi* (currently at CNR) for monitoring and resolving financial and other administrative issues, taking care of the reporting on Sygma portal supervising the activity of the overall consortium and interacting with the central administration of CNR.

3. **The Work Package Leaders (WPL)** are key investigators who will manage and monitor the scientific and technical objectives, deliverables and milestones of their respective work package and report to the Project Management Board. WP Leaders are responsible for maintaining an archive of documents, including draft documents, deliverables and meeting minutes, which are stored in the private area of the project website. The leaders of the work packages are:

    - WP1: Management. *Fosca Giannotti*, CNR
    - WP2: Ethics. *Jeroen Van Den Hoven*, TUDelft
    - WP3: Dissemination. *Hamish Cunningham*, USFD
    - WP4: Training. *Tobias Blanke*, KCL
    - WP5: Innovation. *Dirk Helbing*, ETHZ
    - WP6: Transactional Access. *Kalina Bontcheva*, USFD
    - WP7: Virtual Access. *Roberto Trasarti*, CNR
    - WP8: Big Data EcoSystem. *Wolfgang Nejdl*, LUH
    - WP9: Big Data Analytics methods and techniques. *Dino Pedreschi*, UNIPI
    - WP10: SoBigData e-Infrastructure. *Paolo Manghi*, CNR
    - WP11: Evaluation, *Natalia Andrienko*, FRH

4. **The Task Leaders (TL)** are investigators who will manage and monitor the scientific and technical objectives, effort expenditure, deliverables and milestones of their respective tasks.

Management bodies

The main management bodies responsible for the strategic direction, management and administration of the project are defined as follows:

- The **Project Management Board (PMB)** is composed of the Project Coordinator, the Project Manager and the Work Package Leaders. Meetings will be chaired by the Project Coordinator. This committee is the primary executive decision making body. Each area of work will discuss the specifics of their technical input, progress and effort expenditure within their respective work packages and the PMB will balance these inputs, along practical lines, within the overall technical direction agreed in the original Description of Work. Ethics and Gender issues will also be monitored by the PMB. There will be meetings of the PMB at least every six months and more frequently, if required. These will be virtual or face-to-face, as appropriate.

- The **Project Advisory Board (PAB)** will comprise recognised experts of international standing from a range of academic and non-academic organisations. Members will be senior researchers, business executives, educators and decision makers from both technical and non-technical backgrounds. The PAB will review "*VA Service Provision and Operation"* reports and produce the assessment report, helping to identify problems and possible remedies. The PBA will be

invited to attend some selected meetings to offer the project the benefit of their domain and technical expertise and help position the technical advances of the project in a wider scientific and commercial context. The PAB will meet at least three times during the course of the project. The Project Coordinator will chair PAB's meetings. Members will be required to sign appropriate confidentiality agreements covering their involvement.

- The **Project Management Team (PMT)** will provide a management structure that monitors the progress and quality of the project and resources. The PMT will consist of the Project Coordinator, the Project Manager and the Project Administrative Manager, supported by the CNR European Office (LEAR). The Project Coordinator may co-opt into the PMT other representatives as necessary to assist her in the day-to-day running of the project. This Team is the central locus for the ultimate success of the project, so particular attention will be paid to monitoring the Gantt chart of the Description of Work for time planning, project reporting and financial control to ensure the project is delivered on time and to budget.

- The **Project Steering Board (PSB)** is composed of the Project Coordinator (chair), the Project Manager and the two deputy project coordinators. Its primary focus is on operationalizing the scientific vision of the project issues. The PSB is a smaller body than the PMB and is more flexible to meet whenever either the Project Coordinator or a member of the PMB considers this necessary to ensure focused and flexible action. In the first phase of the project the PSB will meet twice per month. When the project is well on track, this will be reduced to four to six times per year. The meetings will be a mix of teleconferences and face-to-face meetings as appropriate. The PSB will regularly discuss scientific progress, dissemination, intellectual property rights, exploitation, legal, ethical and regulatory issues and measures to be tabled to consortium bodies.

- The **General Assembly (GA)**: All the partners are member of the GA (one person (= vote) per partner). The Coordinator acts as chair. The GA decides on: the yearly technical and financial plan, the plan for use and dissemination, the strategy and procedure for publications, the Training Programme. All beneficiaries will be invited to participate in the meeting and to ask questions, give advice and propose actions. The General Assembly also play a key role in assessing and deciding whether milestones have been achieved.

- The **TA Management Board (TMB)**: led by *Kalina Bontcheva* (USFD) with the involvement of all WP6 partners, the TMB will be responsible for the call publication, advertising, management and for setting up the international User Selection Panel (USP).

- The **Ethics Board**: led by *Jeroen Van Den Hoven*, professor of ethics at TUDelft, the Ethics Board will guarantee the good governance of data and the research integrity and academic ethics, compliance with ethical and legal framework and consultation with external experts when and where required. *Nikolaus Forgò*, professor for legal informatics and IT law at LUH, will act as Vice Chair and also as the legal counsel to the board.

Note that the responsibilities of the PMT differ from those of the PMB; the PMB is responsible for managing the successful execution of the project in terms of the research and other activities described in the Description of Work, and will have members from all work packages. The PMT, on the other hand, addresses administrative, legal and financial matters, and is based entirely at the lead institution (CNR). The PMT will be supported by the CNR EU Research Funding Office, which will be responsible for legal issues (e.g. Grant Agreement negotiations, Consortium Agreement drafting) and financial issues (e.g. receipt, transfer, and accounting of payments, collection of audit certificates). Each beneficiary will be responsible for their local financial and administrative obligations and will report to the Project Coordinator. Secretarial support will be provided by existing departmental staff as required.

3.2.1 Management Tasks and Procedures

Project coordination. The Project Coordinator will define and implement the management functions of **SoBigData**, in line with the Consortium Agreement, drawn up and signed by all partners prior to the project start. The Project Coordinator will define a Management Charter that defines the practical steps for implementing the Consortium Agreement, in case of events such as changes in the PSB membership, loss of a partner, and dealing with under-performing partners. The Project Coordinator will be the main interface between EC officials and the project. The Project Coordinator will monitor dissemination and exploitation activities in collaboration with the responsible WP leaders, and provide guidance and direction as appropriate.

Project Management Board meetings. Scheduled meetings of the PMB will be announced at least four weeks in advance unless called under an ad-hoc request (see below). The agenda will be prepared by the Project Manager in consultation with the other members of the PMB. Agendas will be sent out to all participants via email at least two weeks in advance (except for PMB ad-hoc meetings). Minutes will be circulated within a week from the meeting taking place and partners will have up to fifteen days to comment on them. The PMB will monitor progress and take the necessary corrective measures, in case there are deviations from the agreed effort or work plan.

Ad-hoc meetings. If required, any member of the PMB or PSB has the right to call for an ad-hoc meeting by notifying the other members by email or phone. These meetings will be organised as soon as practicable. Virtual meetings will be organised within one week of the request.

Project Management Team tasks. The PMT will make operational decisions necessary for the smooth running of the project. All major decisions need approval by the PMB. PMT tasks include:

- Day-to-day management of the project, ensuring smooth communication.
- Collecting from all partners the cost and other statements for submission to the EC.
- Preparing all progress and financial reports and documents required by the Commission
- Ensuring prompt delivery of all data identified as deliverables in the Description of Work or requested by the EC for reviews and audits, including the results of the financial audits prepared by independent auditors.
- Allocation of budgets to the partners in accordance with the EC contract.
- Taking measures in the framework of controls/audit procedures.
- Reviewing and proposing budget reallocations to the partners (to be approved by the PMB).
- Making proposals to the partners for the review and/or amendment of the terms of the EC contract and the consortium agreement (to be approved by the PMB).

Website and document management. The public website will be used for dissemination of knowledge generated by the project as well as relevant links to other sites and publications. Content added to the site will be monitored by the Project Manager.

In addition to the public website, there will be a private wiki, protected by access management controls, that will be used to hold internal documents, draft versions of reports and other deliverables, and for internal project discussions. Access to certain areas of the private site will be granted to EU representatives, to enable them to download deliverables ready for review.

Reporting. At the requested time points during the project, reports will be submitted to the EC. All annual reporting will be undertaken by the Project Coordinator. WP leaders will be responsible for collecting and aggregating information relating to their work package and submitting this to the

Project Manager. Members of the PMT will assemble the report, which will be approved by the Project Coordinator prior to submission to the EC. The private area of the project website will act as a repository for reporting information and for the exchange of reporting information.

Quality assurance. The project will implement review procedures for internal and public deliverables, to ensure a consistent level of quality and scientific excellence and compliance with EC policies, such as open access submissions for project papers. In particular:

- Internal reports must be approved by the WP leader prior to upload to the approved area of the private web site.
- Deliverables must be approved by the WP leader concerned before being forwarded to the PMB for final approval. Drafts of deliverables will be required to be submitted prior to the final deadline to allow sufficient time for internal review.
- Key selected deliverables will be forwarded to the PAB members for review. The feedback collected will be used to guide the strategic direction of the project. The Operational Plan prepared by the Project Coordinator at the start of the project will define all the relevant reporting and operational procedures.

Intellectual property management. The Project Coordinator, and representatives from the beneficiaries will confer on a regular basis with the aim of carrying out technology audits of the projects results as they occur during the lifetime of the project, with a view to determining what advances are capable of commercialisation or other exploitation and will agree arrangements (within the terms agreed in the Consortium Agreement) for any protection of arising information and knowledge. Recommendations will be made for approval by the PMB. An IPR management plan will be prepared, to define specific procedures to be followed within the project.

Risk management. The consortium has designed the work plan in a way that prevents the most common risks right from the start. This is achieved by proactive risk management:

- modular work packages,
- clear responsibilities,
- minimization of critical paths; if one work package is delayed, other packages can still continue to achieve their results,
- project board members have proven skills and track record in international projects and research organization,
- no risk of bankruptcy of co-ordinator, since it is a publicly funded research organization,
- one person devoted full-time to financial and other project administration.

There is already a preliminary contingency plan defined for the major risks (Table 3.2b). This will be updated in each management reporting deliverable.

Risk analysis and management will be performed by the PSB and PMB on a regular basis and taken into account in key decision making. **SoBigData** is a highly innovative project, so changes to user requirements, business needs, time schedules and costs may occur. The PSB and PMB will make an evaluation of risks at each milestone, in order to determine whether the level of risk is acceptable and appropriate for the project. Appropriate contingency planning will be carried out accordingly.

Consortium agreement. The project organisation will be governed by a Consortium Agreement that details the financial aspects for implementation, sub-contracting particulars, Intellectual

Property Rights and Access Rights of each beneficiary. The Project Coordinator is responsible for preparing, managing and updating the agreement.

Addition of beneficiaries during the lifetime of the project. The consortium will be open to new Partners if one of the founding Partners should withdraw from the Consortium or new tasks should be identified that cannot be fulfilled by any of the existing Partners. In this case, the rules for inclusion of new Partners will be followed and all existing legal documents will be adjusted according to the requirements of the Grant Agreement. In such cases the process will be managed by the EU Research Funding Office of the Coordinator.

3.2.2 Decision Making Processes

Voting and decision making. Each beneficiary has one vote in the consortium body of which it is a member. Nevertheless unanimous consensus is obviously the first option to be explored at all times in decision making. In cases where no consensus can be reached the decision will be taken by majority vote. In case of equal votes, the chairperson's vote is decisive.

- When a decision has an impact on several WPs the decision is taken at the level of the PMB.
- In case of disagreement by one of the partners affected by the decision, the final decision will be taken at PMB level.

Dispute resolution. If any conflict, whether technical, managerial or financial, should occur during the course of the project, it will be resolved by the PMB either at its next meeting or by other means of communication, in accordance with the procedure detailed in the Consortium Agreement and endorsed by the PMB. These dispositions will be in line with the current recommendations of the European Commission policy and the model Grant document.

Communication. Internal communication will be open and transparent to ensure that partners are kept fully informed of progress, results, developments and decisions. The private area of the project website and wiki will function as a day-to-day interface, as well as a pathway to distribute and exchange information. Internal project communication will be facilitated by mailing lists. Periodic newsletters documenting project achievements will be prepared and circulated by the PMT.

To facilitate project internal communication, an initial workshop will be held near the project launch date, with the aim of promoting a creative collaborative environment for the project and establishing direct lines of communication between participants.

External communications will be facilitated through the public area of the project website, which will be used to host all public deliverables, publicity materials, links to publications, conference reports and information on the results achieved. The project will arrange a presence at major conferences and exhibitions as well as participating in liaison and clustering activities with other related projects.

3.2.3 List of Milestones

The six milestones of **SoBigData** denote the completion and integration of work, forming a coherent set of results needed to start a next major phase of the project. In the table below the results to be achieved by each milestone are detailed as a set of key activities, which are also reflected in the deliverables due by the respective dates. The successful completion and acceptance of these corresponding deliverables will be the means to verify that each milestone has been reached successfully.

Project progress will be checkpointed at months 6, 12, 24, 36, 42, and 48, plan maintenance and review carried out and any newly identified risks addressed.

In addition to this high-level structure, each work-package has defined deliverable dates at a higher level of granularity.

3.2.4 Critical Risks for Implementation

A comprehensive state-of-the-art Risk Assessment and Management Plan will be implemented within the first six months of the project and will address different kinds of risk (external, internal, strategic, operational, other). This work, which is part of the management activity, will:

- Identify risks of any nature that might occur during the project and assess their probability and potential impact on the project;
- Plan concrete actions to prevent the occurrence of the risk;
- If problems do occur, then swiftly implement the associated contingency measure to minimize the impact.

In more detail, since the **SoBigData** project mainly deals with software development, carried out by several partners, it is exposed to risks in the following key areas:

- **Technology, eInfrastructures, and Standards** – **SoBigData** partners are involved in the relevant standardisation bodies and are at the leading edge of technology in their areas. Many partners are also heavily involved in other eInfrastructure projects and have experience with building, deployment, and integration of infrastructures.
- **People** – **SoBigData** has world-leading partners in all areas of expertise needed in the project. Many of them also have an established history of previous collaborations.
- **Organisational** – the management structure has been designed to be appropriate for the size of the consortium and the type of project. The coordinator together with the EC will monitor how well the management structure functions and changes will be introduced swiftly.
- **Architecture and Tools** – **SoBigData** builds on state-of-the-art national infrastructures, standards and tools, developed in successful national and international initiatives. The partners who will be developing the novel **SoBigData** RI have a strong track record in building successful innovative solutions and in implementing big data analytics projects.
- **Requirements** – the requirements of the RI will be chosen carefully to correspond to real user needs, including not only researchers, but also SMEs, larger companies, and other stakeholders. There will be dedicated events during the project inception stage, which will create an initial set of requirements and usage scenarios. These will then evolve continuously, through networking, transnational and virtual access activities.
- **Effort and budget planning** – the project management team will monitor regularly the resources spent on each task and work-package. The plan for the next six months will also be reviewed and adjustments made if necessary.

3.3 Consortium as a Whole

The **SoBigData** consortium consists of 12 partners from 6 member countries of the European Union (Italy, UK, Germany, Estonia, Finland, The Netherlands) and Switzerland. The core consortium has been designed to encompass the leading European research groups that contributed to shape the big social data analytics arena in the past decade. The core partners represent a mix of research centers that focus on big data from multiple perspectives: e-infrastructures (CNR, KCL, USFD), data mining and knowledge discovery (CNR, UNIPI, AALTO, FRH, UT), text and social media mining (USFD, KCL, UNIPI, CNR, LUH), social network analysis (CNR, KCL, AALTO, UT, ETHZ, SNS, IMT), web intelligence (LUH, CNR), visual analytics (FRH), complex network science (AALTO, SNS, IMT, ETHZ), computational sociology and social simulation (ETHZ).

All partners have extensive experience in acquiring, managing and analyzing real-world big data from heterogeneous sources and have the capacity to transform their findings into tools and technologies usable by the scientific community at large. They will also transfer to the **SoBigData** stakeholder community their national infrastructures, expertise, and methodologies. Therefore, the consortium as a whole has the necessary capability to establish the **SoBigData** RI (including big data, tools and skills), coupled with community building and transnational access activities.

The consortium is also powered by leading researchers on IT ethics and IT law: J. van den Hoven, professor of ethics at TUDelft, and N. Forgò, professor for legal informatics and IT law at LUH, who will complement the scientific-technical partners to face the tremendous ethical and legal repercussions of big data analytics.

The partner CNR coordinates **SoBigData** on behalf of the Laboratory on Big Data Analytics and Social Mining, founded in Pisa in 2013 by CNR and UNIPI and recently extended with partners SNS and IMT, in order to team up a critical mass of multi-disciplinary scientists that pursued pioneering work on the various facets of big data and social mining. The goal of CNR and UNIPI within this initiative is to promote a European-wide research infrastructure on Big Data Analytics and Social Mining, capable of boosting research and innovation in the deployment of big data analytics, social mining and privacy/trust technologies to face global challenges. Fosca Giannotti (coordinator), Andrea Passarella and Raffaele Perego are the key scientists at CNR, while Dino Pedreschi and Paolo Ferragina are the key scientists at UNIPI. This initiative has been sponsored by the regional government of Regione Toscana that is also endorsing the **SoBigData** project (see enclosed letter in appendix A) with commitment of pursuing synergies with regional ESIF funds.

Experience in running research infrastructures: The InfraScience team at CNR has long standing involvement in Research Data Infrastructures and E-Infrastructures for several scientific communities (iMarine, D4science, DRIVER). They also have an active role in promoting the Research Data Alliance, which is defining the roadmap of research and technological innovation required by the emerging data intensive modern science. In addition, KCL is major hub for research infrastructures in the SSH domain and is the organiser of the humanities and social sciences track of the annual IEEE Big Data conference (http://cci.drexel.edu/bigdata/bigdata2014/). It has played a key role in setting up the ESFRI projects CLARIN and DARIAH, and now Tobias Blanke is one of the designated DARIAH directors. KCL also participated in 2 more FP7 projects (EHRI and CENDARI), which are integrating services and data in various humanities disciplines and has been representing DARIAH in the joint SSH cluster project DASISH. Its pioneering role in digital approaches to humanities and social sciences research is recognised worldwide.

Previous collaborations and endowment from FuturICT: Dirk Helbing (ETHZ), Fosca Giannotti (UNIPI), Dino Pedreschi (UNIPI) and Jeroen van den Hoven (TUDelft) collaborated within the FuturICT Flagship Pilot initiative, in particular working towards shaping the vision of: a) a digital nervous system enabling big data analytics for social good and for managing social complexity; b)

the ethical dimensions of data science and big data analytics; c) accelerating innovation. The following open access papers detail this vision, which inspires the approach adopted in **SoBigData**:

- F. Giannotti, D. Pedreschi, A. Pentland, P. Lukowicz, D. Kossmann, J. Crowley and D. Helbing. A planetary nervous system for social mining and collective awareness. *Eur. Phys. J. Special Topics* 214, 49-75 (2012). http://dx.doi.org/10.1140/epjst/e2012-01688-9

- J. van den Hoven, D. Helbing, D. Pedreschi, J. Domingo-Ferrer, F. Giannotti and M. Christen. FuturICT – The road towards ethical ICT. *Eur. Phys. J. Special Topics* 214, 153-181 (2012). http://dx.doi.org/10.1140/epjst/e2012-01691-2

Other scientific competences of the consortium crucial to the success of the initiative:

- Big data analytics in the social sciences, humanities, and economics (IMT, SNS, ETHZ, CNR, UNIPI, KCL), big data analytics in demographic change and well-being (ETHZ, UNIPI, CNR, UT), big data analytics for the smart city of the future (CNR, ETHZ, SNS), big data analytics for official statistics (UNIPI), big data analytics for social simulation and policy support (ETHZ, FRH, UT, CNR)

- High-performance and scalable big data analytics, big data algorithms and data structures for compression, indexing and search, stream computing, big data architectures (UNIPI, CNR, KCL, USFD);

Pioneering scientists: The key scientists in the consortium are pioneers in the cornerstones of what is today referred to as big social data analytics: H. Cunningham (USFD) is a pioneer in open source infrastructures for text analytics, D. Helbing (ETHZ) is a pioneer in social simulation and modelling, F. Giannotti (CNR), and D. Pedreschi (UNIPI) are pioneers of mobility data mining and privacy-preserving data mining, G. and N. Andrienko (FRH) are pioneers of visual analytics, S. Fortunato (AALTO), G. Caldarelli (IMT) and F. Lillo (SNS) are pioneers in network science., J. Van Den Hoven (TUDelft) is a pioneer of ethics of data science, W. Nejdl (LUH) is a pioneer of Web Science.

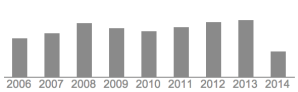
Industrial involvement capacity: the partners exhibit a long strand of strategic collaborations with industries on big data analytics. It is worthwhile to mention just few significant examples: i) the collaboration between UT and Microsoft/Skype aimed at analyzing the structural properties of the global Skype social graph (800+ million nodes, 6+ billion edges); ii) the strategic partnership of FRH with Vodafone, Orange, Nokia, Siemens and Bosch on service usage, behavioural patterns, and trajectory analysis; iii) the strategic partnership of CNR with various telecom and telematics providers such as Orange, Wind, Telecom Italia, KDDI Japan, Toyota MC, Octotelematics (the world's largest provider of vehicular GPS trajectory data); iv) the partnership of USFD with content providers that are end users of big data analytics tools: BBC, Press Association, Financial Times, Thompson Reuters, Elsevier, the British Library, the UK National Archives; v) the partnership of USFD with SME providers of big data analytics tools such as Ontotext, Daedalus, Text Mining Solutions, WebLyzard, Digital Pebble; vi) the partnership of UNIPI and CNR with research centres of Google, Yahoo!, Tiscali aimed at design of efficient search algs and data structures, and semantic search., vi) the partnership of CNR with IBM and ENEL for Big data analytics and smart city and smart grid. Least but not last, FRH is Member of the Federal Association for Information Technology, a German industry network on big data intelligence aggregating more than 20 German companies, CNR is member of EIT ICT Labs, (http://www.eitictlabs.eu/) one of the three Knowledge and Innovation Communities (KICs) selected by the European Institute of Innovation & Technology (EIT).

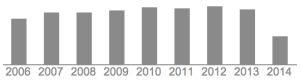
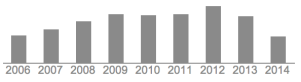
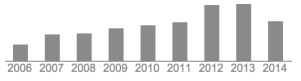
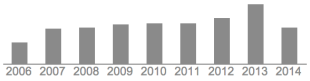
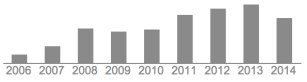
Public body involvement capacity: the partners exhibit a long strand of strategic collaborations with public bodies on big data analytics. Among the many local and national authorities and governments it is worthwhile to mention the collaboration of UT with the Estonian government that will provide to the project the records of all e-gov transactions in the country, the collaboration of CNR with the Italian Statistical Institute (ISTAT) which has activated a project on measuring well-being with Big Data, and the collaboration of UNIPI with the Italian Authority on Data Protection which is interested in privacy preserving analytics.

Training. The partners have already a track record in training big data scientists:

- UNIPI has given rise to the Master degree (M.Sc.) in Business Informatics with a track in Data Mining already in year 2002; starting 2014, this master degree has a Big Data Analytics track. A Master in Smart Cities has been activated in 2014, and a Master in Big Data Analytics and Social Mining will start in January 2015 as an initiative of the SoBigData.it Lab with CNR.
- USFD runs an annual summer school in text analytics, established in 2009 and attended by 40-50 participants each, including PhD students, researchers, and company participants.
- KCL has been teaching MA's in Digital Humanities, Digital Culture and Digital Asset and Media Management for many years and is now launching two new Masters in Big Data (MSc) and Big Social Data (MA), with distinct pathways for statisticians and computer scientists.
- UT offers courses on big data mining within its Master in Computer Science.
- AALTO offers courses on advanced data mining within its Master on Computer Science.
- ETHZ offers a Master on Advanced Studies in Applied Statistics, which includes a wide-ranging selection of statistical methods for data analysis and modelling.
- FRH offers technology-transfer short courses of Data Science on basic analytics, text analytics, visual analytics, big data architectures. FRH organized a number of industry workshops for the EU and German government.

The scientific excellence of the consortium is also reflected in the high citation counts and increasing paper citation trends of our key scientists:

Scientist	**Citations**	**h-index**	**i10-index**	**Trends**
FRH, Natalia Andrienko	4,916	35	97	
USFD, Hamish Cunningham	7,087	36	91	
KCL, Tobias Blanke	298	9	9	
ETHZ, Dirk Helbing	24,877	72	213	
CNR, Fosca Giannotti	3,385	28	69	

UNIPI, Dino Pedreschi	4,119	33	72	
LUH, Wolfgang Nejdl	13,171	56	240	
UT, Marlon Dumas	15,410	57	129	
AALTO, Aristides Gionis	7,558	36	71	
SNS, Fabrizio Lillo	3,543	31	54	
IMT, Guido Caldarelli	5,638	35	82	
TUDelft, Jeroen Van Den Hoven	638	15	25	

3.3.1 Swiss Participation

Due to recent political developments, Switzerland has lost its status as associated partner of Horizon 2020 and is considered as an entity from an industrialised third country. The **SoBigData** consortium rooted in established collaborations in the context of the FP7 Flagship Proposal FuturICT with prof. Dirk Helbing at ETHZ. ETHZ is in charge of the realization of: a series of "datathons" to be held in different European cities, one summer school and is responsible for the realization of the tasks "Innovation accelerator" and "Living Archive". Moreover, prof. Helbing brings in the liaisons of the large FuturICT network composed by more than 300 European institutions, fundamental for disseminating and attracting trainees. We decided to keep the organizational structure despite this change of status of Switzerland within Horizon 2020, as Swiss institutions are still eligible to participate as long as funding does not originate from the EU and their work is fundamental for project success.

The Swiss government has decided on March 7, 2014 to directly support the Swiss contribution of a successful Horizon 2020 consortium proposal. The legal basis of this decision is Article 2 of the "BundesbeschlusseszurFinanzierung der SchweizerBeteiligung an den Rahmenprogrammen der Europäischen Union in den BereichenForschung und Innovation 2014–2020". The relevant information can be found on the website of the Swiss State Secretariat for Education, Research and Innovation (http://www.sbfi.admin.ch/index.html?langen). We are therefore in a position to guarantee the financing of the activities that are covered by the Swiss participants of the consortium.

In this way, a substantial part of the budget of **SoBigData** (~20%) is covered from non-EU sources. This allowed us to expand the scope and activities of **SoBigData** without exceeding the budgetary limits of the call.

3.4 Resources to be Committed

The total budget for the project is 5,917,500 EUR with requested EC contribution of 5,000,000 EUR, as detailed in the following table of budget allocation per participants and cost type.

	Direct personnel costs	Other direct costs	TA - Direct Personal costs	TA - Direct elegible cost for providing access	VA - Direct Personal costs	VA - Direct elegible cost for providing access	Total Direct personnel costs	Total Other direct costs	Total + Overhead	EC Requested contribution
1 - CNR	664000	44000	93000	3000	54000	10000	811000	57000	1085000	1085000
2 - USFD	402900	258000	78000	2000	41000	7000	521900	267000	986125	986125
3 - UNIPI	208000	54000			21000	5000	229000	59000	360000	360000
4 - FRH	382000	48000	36400	3500	18600	5000	437000	56500	616875	616875
5 - UT	134000	32000	13000	3000			147000	35000	227500	227500
6 - IMT	108000	32000					108000	32000	175000	175000
7 - LUH	220000	58000	23800		23800		267600	58000	407000	407000
8 - KCL	255441	46559					255441	46559	377500	377500
9 - SNS	108000	32000					108000	32000	175000	175000
10 - AALTO	264000	44000	24144	7856			288144	51856	425000	425000
11 - ETH	576000	148000	10000				586000	148000	917500	0
12 - TUDelft	116000	16000					116000	16000	165000	165000
total	3440900	810000	278344	19356	158400	27000	3875085	858915	5917500	5000000

We divided the effort and budget between the different types of activities so that solid work can be done under each of the activity types and that synergies between them are fully exploited, as detailed in the following table of budget breakdown by activity type and cost type.

SoBigData overall budget by Activity Type	Personnel	PM	% Effort	Other Costs including Training Events	TA-Travel & Subsistence	Total costs	Total budget including overhead	% Budget
Managenent	€ 252010	38	6.5%			€ 252010	€ 315013	5.3%
Networking	€ 1334440	200	34.1%	€ 390700		€ 1725140	€ 2156425	36.4%
TA	€ 296350	49	8.4%	€ 32178	€ 240000	€ 568528	€ 710660	12.0%
VA	€ 179530	28	4.8%	€ 27000		€ 206530	€ 258163	4.4%
JRA	€ 1803992	271	46.2%	€ 177800		€ 1981792	€ 2477240	41.9%
Total	**€ 3866322**	**586**		**€ 627678**	**€ 240000**	**€ 4734000**	**€ 5917500**	

As can be seen, there is a good balance between networking activities (including community building) on one hand and JRA on the other. The budget and effort for the latter is somewhat higher, due to this being a starting community, which needs to invest effort in the creation and enrichment of the **SoBigData** research infrastructure.

In more detail, the effort for networking activities is broken down as follows:

- Legal and ethical framework - 24 PMs (4.1% of the total effort)
- Dissemination, impact and sustainability activities - 35 PMs (6% of the total effort)
- Training is 48 PMs (8.2% of the total effort), as one of the key community building activities
- Innovation activities is 32 PMs (5.5% of the total effort), which will ensure that significant impact is generated through outreach and collaborations with industry and other stakeholders
- Community-oriented evaluation campaigns - 61 PMs (10.4%), which also includes effort for collection, annotation, and distribution of gold standard datasets, the organisation of community-oriented campaigns, and the synthesis and publication of their results.

With respect to management, previous experience has allowed streamlining the efforts required for project management (6.5%), while at the same time ensuring that resources are adequate both in qualitative and quantitative terms.

Apart from being broken down by activity type, the direct costs in the budget can be broken down into personnel costs (58.7%), travel and event organization (12%), consumables (1%). The rest are costs for management, and trans-national and virtual access. Personnel costs are calculated on the basis of person-months allocated to the project by the partners and the corresponding average person-months rates of each partner. The travel costs cover all participants travel and daily

allowance that will be likely needed to assure adequate participation of partners in bilateral and consortium meetings and working sessions, as well as in key industrial and scientific conferences for dissemination purposes. In general, travel costs have been reviewed and minimized as far as practically possible. In addition, USFD has 200,000 euros in travel costs ring-fenced for reimbursing the travel and subsistence costs of the external researchers, who participate in trans-national access. Lastly, there is no budget requested for hardware, since all partners have already significant computational resources, including HPC infrastructure they can use.

The budget of the coordinator (CNR) is 18% of the project budget, but this includes not only project management (only 4% of the total budget), but also the leadership of two of the most critical work packages - virtual access (WP7) and the implementation of the **SoBigData** RI (WP10). CNR also consists of four separate groups, with complementary research expertise, all of which is critical for the success of the community building and joint research activities in the project. Similarly, USFD's budget is 16%, but when the ring-fenced travel and subsistence TA fund is subtracted (250,000 euros when overheads are included), the actual budget is only 12.6%. This is very reasonable since they are also responsible for two key work packages (transnational access and dissemination and sustainability), act as deputy scientific coordinators, and contribute significantly to training, innovation, and joint research activities.

There is also a substantial third-party (Swiss) contribution to the project. The budget of ETHZ (917,500 euros) will be covered by Swiss funding, which amounts to nearly 16% of the total project budget or nearly 20% of the requested EC contribution. For details see Section 3.3.1.

The overall budget has been carefully constructed so that it is both adequate for the work packages and robust for the full 4-year duration of the project. The budget will be reviewed annually, in the light of progress and as part of risk management, to ensure that all tasks are successfully completed and the coherence of **SoBigData** is maintained throughout.

Finally, we remark no funds have been allocated to equipment as all partners possess well-established high performance computing facilities.

Table 3.4b: 'Other direct cost' items (travel, equipment, other goods and services, access costs)

1 – CNR	Cost (€)	Justification
Travel	36000	Participation to planned project events, review meeting and dissemination at conferences
Equipment		
Other goods and services	8000	Consumables, Data, costs for dissemination on various media including open access, translations
Access cost (if applicable)	13000	Maintenance, utilities, consumables and communication for VA, TA other direct costs
Travel and subsistence for trans-national access (if applicable)		
Total	57000	

2 – USFD	Cost (€)	Justification
Travel	24000	Participation to planned project events, review meeting and dissemination at conferences
Equipment		
Other goods and services	34000	- Consumables, Data, costs for dissemination on various media including open access, translations - Funds to cover the planned initiatives: a. Summer School (1): 16000 € b. Training Event for High-Schools (1): 10000 €
Access cost (if applicable)	9000	Consumables (cloud computing fees) for VA, TA other direct costs
Travel and subsistence for trans-national access (if applicable)	200000	Estimated-Cost for supporting the selected researchers for on site access: 100 weeks
Total	267000	

3 – UNIPI	Cost (€)	Justification
Travel	24000	Participation to planned project events, review meeting and dissemination at conferences
Equipment		
Other goods and services	30000	- Consumables, Data, costs for dissemination on various media including open access, translations - Funds to cover the planned initiatives: a. Datathon (1): 16000 € b. Training Event for High-Schools (1): 10000 €
Access cost (if applicable)	5000	Maintenance, utilities, consumables and communication for VA
Travel and subsistence for trans-national access (if applicable)		
Total	59000	

4 – FRH	Cost (€)	Justification
Travel	24000	Participation to planned project events, review meeting and dissemination at conferences
Equipment		
Other goods and services	24000	Consumables, Data, costs for dissemination on various media including open access, translations - Funds to cover the planned initiatives: a. Workshop (1): 16000 €
Access cost (if applicable)	8500	IVAS computing costs for VA, TA other direct costs
Travel and subsistence for trans-national access (if applicable)		
Total	56500	

5 – UT	Cost (€)	Justification
Travel	12000	Participation to planned project events, review meeting and dissemination at conferences
Equipment		
Other goods and services	20000	- Consumables, Data, costs for dissemination on various media including open access, translations - Funds to cover the planned initiatives: a. Datathon (1): 16000 €
Access cost (if applicable)	3000	TA other direct costs
Travel and subsistence for trans-national access (if applicable)		
Total	35000	

6 – IMT	Cost (€)	Justification
Travel	12000	Participation to planned project events, review meeting and dissemination at conferences
Equipment		
Other goods and services	20000	- Consumables, Data, costs for dissemination on various media including open access, translations - Funds to cover the planned initiatives: a. Workshop (1): 16000 €
Access cost (if applicable)		
Travel and subsistence for trans-national access (if applicable)		
Total	32000	

7 – LUH	Cost (€)	Justification
Travel	24000	Participation to planned project events, review meeting and dissemination at conferences
Equipment		
Other goods and services	34000	- Consumables, Data, costs for dissemination on various media including open access, translations - Funds to cover the planned initiatives: a. Summer School (1): 16000 € b. Training Event for High-Schools (1): 10000 €
Access cost (if applicable)		
Travel and subsistence for trans-national access (if applicable)		
Total	58000	

8 – KCL	Cost (€)	Justification
Travel	24000	Participation to planned project events, review meeting and dissemination at conferences
Equipment		
Other goods and services	22559	- Consumables, Data, costs for dissemination on various media including open access, translations - Funds to cover the planned initiatives: a. Workshop (1): 16000 €
Access cost (if applicable)		
Travel and subsistence for trans-national access (if applicable)		
Total	46559	

9 – SNS	Cost (€)	Justification
Travel	12000	Participation to planned project events, review meeting and dissemination at conferences
Equipment		
Other goods and services	20000	- Consumables, Data, costs for dissemination on various media including open access, translations - Funds to cover the planned initiatives: a. Workshop (1): 16000 €
Access cost (if applicable)		
Travel and subsistence for trans-national access (if applicable)		
Total	32000	

10 – AALTO	Cost (€)	Justification
Travel	24000	Participation to planned project events, review meeting and dissemination at conferences
Equipment		
Other goods and services	20000	- Consumables, Data, costs for dissemination on various media including open access, translations - Funds to cover the planned initiatives: a. Workshop (1): 16000 €
Access cost (if applicable)	7856	TA other direct costs
Travel and subsistence for trans-national access (if applicable)		
Total	51856	

11 – ETHZ	Cost (€)	Justification
Travel	40000	Participation to planned project events, review meeting and dissemination at conferences
Equipment		
Other goods and services	68000	- Consumables, Data, costs for dissemination on various media including open access, translations - Funds to cover the planned initiatives: a. Datathon (2): 32000 € b. Summer School (1): 16000 € c. Workshop (1): 16000 €
Access cost (if applicable)		
Travel and subsistence for trans-national access (if applicable)	40000	Estimated-Cost for supporting the selected researchers for on site access: 20 weeks
Total	148000	

12 – TUDelft	Cost (€)	Justification
Travel	12000	Participation to planned project events, review meeting and dissemination at conferences
Equipment		
Other goods and services	4000	Consumables, Data, costs for dissemination on various media including open access, translations
Access cost (if applicable)		
Travel and subsistence for trans-national access (if applicable)		
Total	16000	

4. Members of the consortium

4.1. Participants

4.1.1 Consiglio Nazionale delle Ricerche – Italy (CNR)

Consiglio Nazionale delle Ricerche CNR	

ORGANISATION DESCRIPTION

Consiglio Nazionale delle Ricerche, www.cnr.it

CNR, the Italian National Research Council, is the main public Italian research institution, it employs more than 4000 researchers in 100 Institutes across Italy. CNR participates to this project with two institutes: the **Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo"** (ISTI) and the **Istituto di Informatica e Telematica** (IIT) both located in Pisa. Before presenting the two institutes, it is worth mentioning that ISTI and IIT in 2013 co-funded the European laboratory on Big Data Analytics and Social Mining (www.SoBigData.it), one of the pillars of this project, aimed at pursuing interdisciplinary research initiatives connected to the impulse that "big data" and ICT are having on (socio-economic) sciences. One of the goal is the creation of a knowledge infrastructure for the acquisition and analysis of big data from social media, online social networks and other sources aimed at the realization of large scale data analytics and simulation projects. CNR participates in the project with the researchers of 5 different groups presented below.

KDD - ISTI (http://www-kdd.isti.cnr.it): founded in 1996, is aimed at pursuing fundamental research, strategic applications and higher education in the area of Knowledge discovery and Data Mining and is a joint initiative with University of Pisa. It is today a leading research hub on mobility data mining and privacy preserving data mining, becoming a reference not only for the international research community but also for leading industrial and public operators, such as telecom providers (Orange, Wind, Telecom Italia, Toyota) and mobility agencies of regional and municipal administrations. KDD Lab. has also created novel methods for mining social networks, aimed at discovering patterns, evolutionary rules, community structure and predictive rules. KDD Lab. led and participated to a stream of FET-Open projects on big data and social mining (see project section).

Nemis- ISTI participates with its InfraScience Research Group (lead by Dr. Donatella Castelli), has a long track of EC funded projects aimed at designing and operating data infrastructures. Members of this team have scientifically coordinated the projects DILIGENT, D4Science, D4Science-II, iMarine that have played a central role in the definition and implementation of data infrastructures and Virtual Research Environments. Currently, they are also coordinating the technical development of OpenAIRE, the Open Access Infrastructure for Research in Europe. The team is also a partner in the Research Data Alliance Europe initiative and in ENVRI, the cluster of ESFRI Research Infrastructures on Environment.

HPC - ISTI (http://hpc.isti.cnr.it) carries out research on algorithms and systems addressing computational and data-intensive problems arising in scientific, business, social, and knowledge-based applications. Our advanced solutions exploit high-performance techniques and tools to deal with the exponential growth of users, data, knowledge and services that need to be efficiently stored, managed, analysed and retrieved. The main research topics are: Web Search, Data and Web Mining, Cloud Computing. The group is composed of about 20 people involved in several projects including the running EC projects Midas, Contrail, IngeoClouds, eCloud, and the national projects Secure!, RIS and MOTUS.

WAFI-IIT has followed the evolution of the World Wide Web from the success of markup languages, to the introduction of a semantic layer and the explosion of the social web content. WAFI-IIT studies the evolution of the Web in all its forms (data, communities, technologies, guidelines, recommendations,

designs, software, tools). Currently, WAFI-IIT performs research and development activities on Social Media Analysis, spanning from Social Network Analysis to Social Sensing and Social Media Intelligence, Information Visualization on the Web, Mashup Applications, Open and Big Data, Web Mining, Semantic Web. It is also part of the Italian W3C office, involved in the organization of events and courses on Linked Open Data.

UI-IIT research activities are focused on the Future Internet design and evaluation, mobile and online social networks being one of the key research topics the group is focusing on. UI-IIT research also focuses on analysis of users' behaviour in Online Social Networking environments, collection and analysis of large-scale datasets in these environments, and development of models of social relationships. Moreover, UI-IIT researchers feature a broad expertise in the design, modelling, performance evaluation through analysis, simulation and experiments of innovative Internet architectures and protocols in various types of networking environments. UI-IIT members are actively participating to relevant expert groups at the national and European levels, and have been involved in the EC FIRE consultations since the beginning.

Role in the Project

CNR coordinates the project and leads WP7 virtual access and WP 10 on the SoBigData RI. It will contribute to all three kind of activities with special focus on integrating the many methods for social sensing and social mining from the SoBigData.it national infrastructure; creating the new pan-European SoBigData RI; its enrichment with new research methods through JRAs; and on providing transactional and virtual access.

Infrastructure brought into the project

CNR in 2013 has promoted the European laboratory on Big Data Analytics and Social Mining SoBigData, dedicated to the creation of a knowledge infrastructure for the acquisition and analysis of big data from social media, online social networks and other sources aimed at the realization of large scale data analytics and simulation projects. The infrastructures of SoBigData.it are entirely brought into the project. They are articulated on various themes: network analytics, mobility analytics, social sensing, text mining, indexing and search. Moreover, they comprise a vast repertoire of big data resources, including mobile phone call records, social networking records, GPS tracks from navigation devices, supermarket transaction records, web search query logs, etc.

Relevant Projects (5 CNR-ISTI and 5 CNR-IIT)

DataSIM - FP7 FET- OPEN – Data Science in the Era of Electric Vehicles. http://www.uhasselt.be/datasim (2011-2014)
SEEK (2012-2014), FP7 PEOPLE Marie Curie - SEmantic Enrichment of in Knowledge discovery. Coordinated by CNR-KDD Lab. http://www.seek-project.eu/
OpenAIRE & OpenAIREPlus (2011 – 2015). -FP7 e-Infrastructure. Open Access Infrastructure for Research in Europe, www.openaire.eu
iMarine (2011-2014). -FP7 e-Infrastructure. Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources. www.i-marine.eu
Yahoo Faculty Research and Engagement Program (FREP), awarded to Raffaele Perego, 2014.(http://labs.yahoo.com/news/faculty-research-and-engagement-program-2014-recipients-selected). The project involves the design and the experimentation of models to predict the tasks accomplished by users through their smartphones
CAPER – (FP7, partner) **–** Collaborative information acquisition, processing, exploitation and reporting for the prevention of organised crime http://www.fp7-caper.eu/
OpeNER (FP7, partner) - provide a set of ready to use tools to perform some natural language processing tasks, free and easy to adapt. http://www.opener-project.eu/
SOCIALNETS, (FP7, partner), - measurement and analysis of large-scale data from Online Social Networks, in order to better characterize the social behavior of the users in such virtual networking environments. http://www.internet-science.eu

EINS Network of Excellence in Internet Science - (FP7/partner) - EINS is studying Virtual Communities, and specifically how social relationships between users of Online Social Networks are formed and evolve over time
SCAMPI - (FP7/partner) Facebook tools to monitor interactions between users, which have been used to infer the strength of social relationships in this Social Network. http://www.ict-scampi.eu/

Relevant Publications (5 CNR-ISTI and 5 CNR-IIT)

F Giannotti, M Nanni, D Pedreschi, F Pinelli, C Renso, S Rinzivillo, R Trasarti. Unveiling the complexity of human mobility by querying and mining massive trajectory data. The VLDB Journal 20(5), 695-719 (2011)

D. Wang, D. Pedreschi, C. Song, F. Giannotti, A.-L. Barabasi. Human mobility, social ties, and link prediction. Proc. ACM SIGKDD Int. Conf. on Knowledge discovery and data mining, p. 1100-1108. (2011)

M. Berlingerio, M. Coscia, F. Giannotti, A. Monreale, D. Pedreschi. Multidimensional networks: foundations of structural analysis. World Wide Web 16(5-6): 567-593 (2013)

Castelli D., Manghi P., Thanos C. A vision towards Scientific Communication Infrastructures. In: International Journal on Digital Libraries, vol. 13 (3-4) pp. 155 - 169. Springer, 2013.

Lucchese, C., Orlando, S., Perego, R., Silvestri, F., & Tolomei, G. (2013). Discovering tasks from search engine query logs. ACM Transactions on Information Systems (TOIS), 31(3), 14. (ACM Computing Reviews Notable Article 2013).

M. Avvenuti, S. Cresci, A. Marchetti, C. Meletti, M. Tesconi "EARS (Earthquake Alert and Report System): a real time decision support system for earthquake crisis management", 2014, 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining

M. Avvenuti, S. Cresci, M. N. La Polla, A. Marchetti, M. Tesconi "Earthquake emergency management by Social Sensing", 2014, 2014 IEEE International Workshop on Social and Community

V. Arnaboldi, M. Conti, M. La Gala, A. Passarella, F. Pezzoni, "Information Diffusion in OSNs: the Impact of Nodes' Sociality", 29th ACM Symposium On Applied Computing (ACM SAC 2014), Gyeongju, S. Korea, March 24 - 28, 2014

V. Arnaboldi, A. Guazzini, A. Passarella, "Egocentric Online Social Networks: Analysis of Key Features and Prediction of Tie Strength in Facebook", *Elsevier Computer Communications*, Volume 36, Issues 10–11, June 2013, Pages 1130–1144,

A. Passarella, R.I.M. Dunbar, M. Conti, F.Pezzoni, "Ego network models for Future Internet social networking environments", *Elsevier Computer Communications*, Vol. 35, Issue 18, December 2012

Overview of Key Staff

Fosca Giannotti (female) is Research Director at ISTI-CNR, Pisa. She serves as general chair (2012-2015) of the Steering Committee of the European Association for Machine Learning and Knowledge Discovery. She has spent more than three years visiting various Research Institutions in US and EU. She has been involved in many European research projects since 1990, and coordinated the GeoPKDD FP6 FET-Open project (www.geopkdd.eu) awarded in 2010 as one of the top results achieved within the FET-Open program and presented at the European parliament in Strasbourg. She is the author of more than 200 publications and served in the scientific committee of the main conferences in the area of Databases and Data Mining. She chaired ECML/PKDD 2004, the European Conf. on Machine Learning and Knowledge Discovery in Data Bases, and ICDM 2008, the IEEE Int. Conf. on Data Mining. Her current research interests include spatio-temporal data mining, privacy preserving data mining, social network analysis and data mining query languages.

Donatella Castelli (female) is Senior Researcher at CNR-ISTI and leader of the InfraScience research group. Her research interests include digital libraries and data infrastructure content modeling, interoperability and architectures. She is author of several research papers in such fields. Under her supervision, the

InfraScience team coordinated and participated in several EU and nationally funded projects on Digital Libraries and Research Data Infrastructures. In particular, she has been the technological coordinator of the DRIVER, DRIVER-II and OpenAIRE projects. In this role she has contributed to the development of the D-Net software toolkit. D. Castelli is currently technical manager of the OpenAIREPlus project and the scientific director of the iMarine project and of the D4Science infrastructure. She is also a member of the RDA Europe Expert Group that promotes research and cross-infrastructure coordination at global level and member of the COAR Strategic Committee.

Paolo Manghi (male), Researcher at CNR-ISTI since 2006, member of the InfraScience research group. His research focuses on data models for Digital Library Management Systems, data curation in digital libraries, and autonomic service-oriented data infrastructures, with special care to the realization and technological transfer of open source software. He is author of several research papers in such fields. Moreover, he has participated and coordinated the research and technical activities of several EC projects, such as DRIVER, DRIVER-II, OpenAIRE, HOPE, EFG, and several others whose focus was on the realization of aggregative data infrastructures. He is currently coordinating research and development activities supporting the production data infrastructures for the OpenAIREplus project and the EAGLE project. Finally, he is conceiver and designer of the D-NET Software Toolkit (http://www.d-net.research- infrastructures.eu), a framework for the construction of autonomic and sustainable aggregative data infrastructures.

Raffaele Perego (male) (http://hpc.isti.cnr.it/~raffaele/) is a senior researcher at ISTI, where he leads the High Performance Computing Lab (http://hpc.isti.cnr.it/) carrying out research on algorithms and information systems addressing computational-, and data-intensive problems arising in scientific, business, social, and knowledge-based applications scenario. His main research interests include: Web information retrieval, data mining, machine learning, parallel and distributed computing. Raffaele Perego co-authored more than 120 papers on these topics published in journals and in the proceedings of peer reviewed international conferences. He participated and coordinated activities in several EC projects. He is the recipient of a 2014 Yahoo! Faculty Research and Engagement Program (FREP) award.

Claudio Lucchese (male), PhD, (male) (http://hpc.isti.cnr.it/claudio) researcher at ISTI-CNR. His main research activities are in the areas of data mining techniques for information retrieval, large-scale data processing and cloud computing. He published more than 60 papers on these topics in peer reviewed international conferences and journals. He teaches courses on data mining and parallel computing at the Computer Science dept. of the University of Florence. He participated to and coordinated activities in the EC projects Sapir, S-Cube, Assets and InGeoClouds.

Roberto Trasarti, PhD (male) researcher at ISTI-CNR. He obtained the Ph.D. in Computer Science at the School for Graduate Studies "Galileo Galilei", (University of Pisa) in June 2010 presenting his work "Mastering the Spatio-Temporal Knowledge Discovery Process". He published in several top conferences on data mining and database technologies (KDD, ICDM, EDBT, etc.) as well as important journals in the fields (VLDB, SIGMOD records, etc.). Currently (from 2011) he has a position as researcher at CNR (ISTI - Italy) as member of Knowledge Discovery and Delivery Laboratory (KDD Lab). His interests regard data mining, mobility data analysis, artificial intelligence and automatic Reasoning, databases, and data mining query languages.

Maurizio Tesconi (male) is a researcher in Computer Science at WAFI-IIT. His research interests include social web mining and social media analysis, semantic Web, and natural language processing. He participated to the EU co-funded KYOTO and PANACEA projects. He is currently participating in the EU CAPER project, dealing with the creation of a common platform for the prevention of organized crime, and the OpeNER project, that has the goal of reusing and repurposing exiting language resources and data sets to provide a set of opinion mining and sentiment analysis tools. He is responsible of the SoS project,

an internal project aiming to exploit the user of social media platforms as sensors in the detection of occurring real events. Dr. Tesconi is also part of the permanent team of the European Laboratory on Big Data Analytics and Social Mining (www.sobigdata.eu), performing advanced research and analyses on the emerging challenges posed by big data and he is teaching Visual Analytics at the Master Programme in Business Informatics.

Andrea Passarella (male) is a Researcher at UI-IIT. Before joining UI-IIT he was a Research Associate at the Computer Laboratory of the University of Cambridge, UK. Currently, he is mainly interested in Online and Mobile Social Networks, opportunistic and delay-tolerant networks, content-centric mobile networks. His specific focus in the area of Social Networking is on the characterisation of social structures in large-scale OSNs, modelling and evaluation of information diffusion, and design of solutions for Mobile Social Networking. He published 100+ papers in international journals and conference proceedings, receiving the best paper award at IFIP Networking 2011 and IEEE WoWMoM 2013. He was a member of the EC FIRE expert group on Future Internet research and experimentation. He is Co-Chair of ACM CHANTS 2014, and was Program Co-Chair for IEEE WoWMoM 2011, TPC Vice-Chair for REALMAN 2005, MDC 2006 and REALMAN 2006, TPC co-chair of ACM MobiOpp 2007 and Workshop Co-Chair of IEEE AOC 2009. He was Workshops Co-Chair for IEEE PerCom 2010 and IEEE WoWMoM 2010. He is in the Editorial Board of the Elsevier Pervasive and Mobile Computing Journal, and of Inderscience Intl. Journal of Autonomous and Adaptive Communications Systems. He is the Vice-Chair of the IFIP WG 6.3"Performance of Communication Systems".

Andrea Marchetti (IIT) works at the Institute of Informatics and Telematics since 1999, leading the Web Applications for the Future Internet Group since 2009. He is responsible of the activity "Social and Semantic Web". Over the past years, he has published many articles in the field of semantic web, dealing with the management of lexical resources. His latest research interests focus on web of data, digital humanities and social media analysis. He participated to the EU co-funded KYOTO and PANACEA projects. He is currently participating in the EU CAPER project, dealing with the creation of a common platform for the prevention of organized crime, and the OpeNER project, that has the goal of reusing and repurposing exiting language resources and data sets to provide a set of opinion mining and sentiment analysis tools. He is also leading the ASIA project, having the objective to develop a platform for investigating on potential tax evaders by automatically collecting, analyzing and visualizing publicly available information from the Web. He is participating in the activities of the Multilingual Web initiative, led by W3C.

Mariantonietta N. La Polla (female) is a post-doc research fellow and is part of the WAFI-IIT since 2010. Her research interests include Network Analysis and his application to social media data, Social Media Intelligence, in particular targeted to support Law Enforcement Agencies (LEAs) activities, and Social Sensing. She is participating to the EU CAPER project. She is also part of the team of the SoS project that exploit the user of social media platforms as sensors in the detection of occurring real events. She is part of several PCs (ISDA2012/2013, COOL2014, CRIMENET2014) and serves as reviewer for Collective Behaviors and Networks (EPJ Data Science, Springer) and Knowledge Based System (Elsevier).

4.1.2 The University of Sheffield – UK *(USFD)*

THE UNIVERSITY OF SHEFFIELD USFD	 The University Of Sheffield.

ORGANISATION DESCRIPTION

The University of Sheffield www.sheffield.ac.uk

The Natural Language Processing (NLP) group at the University of Sheffield is one of the largest and most successful research groups in text mining and language processing in the EU. The group is based in the Department of Computer Science, which also includes world-class teams in the areas of speech, knowledge and information processing, biotechnology, and machine learning for medical informatics.

The NLP group has world-leading research record in the fields of NLP infrastructures (GATE), information extraction, standardisation, machine learning methods for NLP, dialogue systems, question answering, terminology extraction, NLP methods for Knowledge Management and the Semantic Web. We will build on the results of previous EC and national projects, including AnnoMarket (cloud-based NLP services marketplace), CLARIN (language resources infrastructure), GateCloud (UK JISC project), KHRESMOI (biomedical text mining), EnviLOD (UK JISC project on mining scientific collections with the British Library), NEON (mining FAO content). USFD also has extensive experience in innovation and knowledge transfer activities, in particular collaborations and consulting for companies (both SMEs and large corporates), government bodies, and other organisations.

Role in the Project

USFD will assist CNR with project management, as well as lead WP6 on trans-national access and WP 11 on building a new evaluation framework. They are also leading T3.4 on impact coordination; T5.1 on partnerships with industry; and T5.2. on knowledge transfer, which builds on their experience with generating significant impact with the GATE infrastructure. Also, they are responsible for T4.1 summer schools, based on their track record with running the annual GATE training school. In the JRAs, USFD leads T9.3 on enriching the SoBigData RI with new text and social media mining services.

Infrastructure brought into the project

USFD develops and maintains the world-leading open-source GATE text mining infrastructure (http://gate.ac.uk) and its vibrant user community (41,000 software downloads in the past 12 months alone and 265,000 – in the past 9 years). GATE has a repository of over 150 text mining and NLP models and algorithms, including Information Extraction (IE), biomedical text mining, ontology-based semantic annotation, machine learning for IE, and NLP evaluation tools, as well as many 3rd party provided text mining plugins. USFD also maintains two cloud-based deployments of GATE as text mining platforms-as-a-service: GATECloud (aimed at researchers wanting to run GATE-provided or their own text mining pipelines on big data) and AnnoMarket (aimed at companies and other users, who want to use pre-packaged, highly scalable GATE text mining web services).

Relevant Projects

AnnoMarket – http://www.annomarket.eu (FP7, coordinator) **–** Cloud-based text mining services marketplace

TrendMiner – http://www.trendminer-project.eu (FP7, partner) - Text mining and summarisation of large-scale social media streams

PHEME - http://www.pheme.eu (FP7, coordinator) – developing text mining methods for computing veracity of social media content
Arcomem - http://www.arcomem.eu (FP7, coordinator) – deployment of GATE text mining pipelines on Hadoop platforms
uComp – http://www.ucomp.eu/ (CHIST-ERA, coordinator) – crowdsourcing corpora

Relevant Publications

Tablan, V., Roberts, Cunningham, H. and Bontcheva, K. GATECloud.net: a Platform for Large-Scale, Open-Source Text Processing on the Cloud. Philosophical Transactions of the Royal Society A, 371(1983), 2013 doi:10.1098/rsta.2012.0071.

Tablan, Bontcheva, Roberts, Cunningham, Dimitrov. AnnoMarket: An Open Cloud Platform for NLP. Proceedings of ACL 2013.

K. Bontcheva, L. Derczynski, A. Funk, M.A. Greenwood, D. Maynard, N. Aswani. TwitIE: An Open-Source Information Extraction Pipeline for Microblog Text. Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2013).

L. Derczynski, D. Maynard, G. Rizzo, M. van Erp, G. Gorrell, R. Troncy, K. Bontcheva. Analysis of Named Entity Recognition and Linking for Tweets. Information Processing and Management (To appear).

Cunningham H, Tablan V, Roberts A, Bontcheva K (2013) Getting More Out of Biomedical Documents with GATE's Full Lifecycle Open Source Text Analytics. PLoSComputBiol 9(2): e1002854. doi:10.1371/journal.pcbi.1002854

Overview of Key Staff

Hamish Cunningham (male) is head of the 15 strong GATE team and leads the research on the GATE infrastructure and its components. Another strength is multilingual Information Extraction (IE), where he has developed novel annotation graph-based finite state transduction IE front-end. He has published some 140 peer-reviewed articles, cited more than 3000 times on Google Scholar. Since 1997, he has, singly or jointly, secured and directed 32 research grants worth over £9Million. He was a founding member of the Information Retrieval (IR) Facility along with many of the most influential figures in the IR community world-wide. Prof. Cunningham is the coordinator of the AnnoMarket STREP and the ARCOMEM Integrated Project.

Kalina Bontcheva (female) is a senior research scientist and the holder of a prestigious EPSRC career acceleration fellowship, working on text mining and summarisation of social media. Dr. Bontcheva received her PhD on the topic of adaptive hypertext generation from the University of Sheffield in 2001. She has been a leading developer of the GATE text analytics infrastructure since 1999. Her main interests are software infrastructures for NLP, information extraction, natural language generation, and text summarization. Kalina Bontcheva is currently coordinating the PHEME FP7 project on computing veracity of social media content, as well as leading the Sheffield teams in TrendMiner, DecarboNet, and uComp. Previously she coordinated the EC-funded TAO STREP project on transitioning applications to ontologies and contributed to the MUSING, SEKT, and MI-AKT projects. Dr Bontcheva is co-organiser of the bi-annual conference “Recent Advances in Natural Language Processing”, co-chair of the Information Extraction track of ACL’2010 and EMNLP’2010, a demo co-chair for ACL’2014, an area co-chair for UMAP’2014, and a PC co-chair for UMAP’2015. She has published extensively in high profile journals and conferences and delivered invited talks and tutorials.

Mark A. Greenwood (male) received his PhD from the University of Sheffield in 2006 for work on Open-Domain Question Answering from large text collections. Since 2005 he has been a research associate on a number of natural language processing projects including RESuLT, X-Media, LarKC, Khresmoi, and most recently the EU funded ForgetIT project. Since 2009 he has also been a member of the GATE team at Sheffield making contributions to the development of both the core framework and plugins. Mark has been involved in the planning and running of two international workshops on Information Retrieval for Question Answering (IR4QA) as well as reviewing for numerous conferences, workshops and journals and examining MSc and PhD thesis. He has also published a number of journal articles and conference and workshop papers relating to his own work in a number of areas of natural language processing.

4.1.3 Università di Pisa – Italy (UNIPI*)*

UNIVERSITÀ DI PISA UNIPI	UNIVERSITÀ DI PISA

ORGANISATION DESCRIPTION

Dipartimento di Informatica, Università di Pisa, Pisa, Italy. http://www.di.unipi.it/

The University of Pisa (www.unipi.it) is one of the earliest in Europe, active since the eleventh century and formally established in 1343. Today, it is a large, research-oriented university, whose faculty of 1,500 permanent scientists covers all disciplines, ranking in the top 100 world universities in physics, math and computer science. With its 50,000+ students in a city of 100,000, it offers students and scholars a unique learning and research experience. U. Pisa had a pioneering role in the international ICT scene since the 1950's, when one of the first computers was realized in Pisa on a suggestion by Enrico Fermi, in collaboration between the university, the National Research Council (CNR) and Olivetti. The CS department, Dipartimento di Informatica (www.di.unipi.it), has a faculty of 60 permanent scholars, with pioneers and influential scientists in most traditional and novel areas of the computer and information sciences. The CS department, and the university as a whole, are pursuing many interdisciplinary research initiatives connected to the impulse that "big data" and the ICT's are having on science, and the socio-economic sciences in particular. U. Pisa researchers are involved in many projects at global scale, and the department of computer science maintains strong liaisons with the international ICT research at the highest levels, including the research centers of the major ICT, Web and telecom corporations. U. Pisa has a long lasting partnership with CNR, the Italian National Research Council, especially between the CS department and the ICT institutes ISTI (Information Science and Technologies Institute) and IIT (Istituto di Informatica e Telematica) located in Pisa, with many common projects and laboratories. Recently the three institutions have given rise to the European Laboratory on Big Data Analytics and Social Mining www.sobigdata.eu in order to team up a critical mass of multi-disciplinary scientists that pursued pioneering work on the various facets of big data and social mining, as well as their big data and tools and systems and networks of international and interdisciplinary collaborators in academia, research and industry. The goal of this initiative is properly the one to create a European-wide research infrastructure on Big Data Analytics and Social Mining, capable of boosting research and innovation in the deployment of big data analytics, social mining and privacy/trust technologies to face global challenges. The founding labs within the CS dept are: **Knowledge Discovery and Data Mining Laboratory** – KDD LAB, DI-UNIPI (PI: Dino Pedreschi), **Advanced Algorithms and Applications Laboratory** – ACUBE LAB, DI-UNIPI (PI: Paolo Ferragina)

Role in the Project

UNIPI are leading the JRA WP9 on enriching the SoBigData RI with new methods and services for social data mining and social sensing. They are also leading T4.4 on addressing gender and diversity issues in data science through training and outreach to high school students (an area where they have carried out workshops and school training events before). They also lead T5.3 on entrepreneurial skills training for data scientists, building on their EU Leonardo Transfer of Innovation project. UNIPI lead the JRA T9.2 on enriching the RI with new methods for social pattern discovery, while also contributing methods to other tasks in WP9. They are also involved in the virtual access activities and contributing to WP11 on evaluation.

Relevant Projects

Google Research Award 2010 and Google Research Award 2013 (http://research.google.com/university/relations/research_awards.html).
"On-the-fly annotation of short texts (by Wikipedia pages), with applications", awarded to Prof. Paolo Ferragina. This project allowed the development of TagMe tool for annotating short texts with links to Wikipedia pages; TagMe is now the best worldwide annotator to date [Cornolti et al., WWW 2013].
(http://research.google.com/university/relations/research_awards.html).
"A novel graph for social-network analysis and search built by entity-annotators, and its applications", awarded to Prof. Paolo Ferragina. This project is undergoing and has already led to design SMAPH, a system for annotating search-engine queries and that recently won the Short Track in the ERD Challenge 2014 within the ACM SIGIR Conference 2014.

Yahoo Faculty Research and Engagement Program (FREP), awarded to Rossano Venturini, 2014.(http://labs.yahoo.com/news/faculty-research-and-engagement-program-2014-recipients-selected). The project involves the design and the experimentation of compressed data structures for indexing users' personal collections of documents which grow and change frequently in the time.
GEOPKDD (2004-2009) - FP6 FET-Open – Geographic Privacy-aware Knowledge Discovery and Delivery – Coordinated by KDD Lab (Fosca Giannotti). Selected to be demonstrated et European Parliament in Strasbourg in 2010. http://www.geopkdd.eu.
DataSIM (2011-2014) FP7 FET-Open – Data Science in the Era of Electric Vehicles. http://www.uhasselt.be/datasim (PI: Fosca Giannotti).
ICON (2012 – 2014) – FP7 FET-Open. Inductive Constraint Programming. http://www.icon-fet.eu/ (PI: Dino Pedreschi).

Relevant Publications

P. Ferragina, U. Scaiella. Fast and accurate annotation of short texts with Wikipedia pages. IEEE Software, 29(1): 70-75, 2012.

M. Cornolti, P. Ferragina, M. Ciaramita. A framework for benchmarking entity-annotation systems. Procs World Wide Web Conference (WWW), Rio de Janeiro (Brasil), 249-260, 2013.

G. Ottaviano, R. Venturini. Partitioned Elias-Fano Indexes. Procs ACM SIGIR, Gold Coast (Australia), 273-282, 2014. (Winner of the Best Paper Award)

D. Wang, D. Pedreschi, C. Song, F. Giannotti, A.-L. Barabasi. Human mobility, social ties, and link prediction. Proc. ACM SIGKDD Int. Conf. on Knowledge discovery and data mining, p. 1100-1108, 2011.

Giannotti, F., Pedreschi, D., Pentland, A., Lukowicz, P., Kossmann, D., Crowley, J., & Helbing, D. A planetary nervous system for social mining and collective awareness. The European Physical Journal Special Topics, 214(1), 49-75, 2012.

Overview of Key Staff

Dino Pedreschi (male) is Professor of Computer Science at the University of Pisa, and a pioneering scientist in mobility data mining, social network mining and privacy-preserving data mining. His research focus is on big data analytics and mining and their impact on society. He is a founder of the Business Informatics MSc program at Univ. Pisa, a course targeted at the education of interdisciplinary data scientists. Dino has been a visiting scientist at the Center for Complex Network Research of Northeastern University, Boston (2009-2010), and earlier at the University of Texas at Austin (1989-90), at CWI Amsterdam (1993) and at UCLA (1995). In 2009, Dino received a Google Research Award for his research on privacy-preserving data mining. D.P. served as vicerector at Univ. Pisa (2000-2002), in charge of implementing a comprehensive reform of the educational system at undergraduate and graduate level.

Dino is the author of many influential publications (more than 200 in total, with an h-index of 32).

Paolo Ferragina (male) is Professor of Computer Science and, currently, Vice-Rector on "Applied Research and Innovation" at the University of Pisa. He is also the President of the IT Center of the University of Pisa, which is indeed competence center about Cloud and HPC for Dell and Intel, Xeonphi Centre for Intel, and recently Transform Data Center immersion for Microsoft. Ferragina got his PhD (1996) in Computer Science from the University of Pisa, and his Post-doc from the Max-Planck Institut fur Informatik (Saarbrucken, 1997-98). His promotion to full professor was sponsored by Yahoo! Research. His research is devoted to the design and software implementation of algorithms and data structures for large text repositories. His research results originated two US Patents (one owned by Lucent and the other one by University of Pisa and Rutgers), four other US patents are pending (owned by Yahoo! and NYU), and he also got some international awards, three are mentioned above from Google and Yahoo. He has been keynote speaker of several international conferences, among the last ones: ESA/ALGO 2010, and in the Industral Track of ECIR 2012; he also was co-chair of worldwide conference in IR and algoithms: the latest ones being ACM WSDM 2013 and ALGO-ESA 2012. He co-authored more than 120 publications in international refereed conferences and journals on Theoretical Computer Science and Algorithmics. Gaining an h-index of 22 (from 2009) and 31 (overall). He is serving in the Editor Board of the Journal of Graph Algorithms and Applications (JGAA), and one of the Area Editors of the Encyclopedia of Algorithms (Springer, Editor Ming-Yang Kao, 2008) for the topics "Data compression, String Algorithms and Data Structures".

Franco Turini (male) was born in 1949 in Italy. He graduated in Computer Science (Laurea in Scienze dell' Informazione) summa cum laude in 1973, at the University of Pisa. He is currently a full professor in the Department of Computer Science of the University of Pisa. In 78/80 he has been a visiting scientist of the Carnegie-Mellon University (Pittsburgh) and of the IBM Research Center S. Jose, afterwards. In 92/93 he has been visiting professor at the University of Utah. His research interests include programming languages design, implementation, and formal semantics, knowledge representation and knowledge discovery. He is one of the principal investigators of the research lab KDDLAB (http://www-kdd.isti.cnr.it/) formed by researchers of the Computer Science Department of the University of Pisa and researchers of institute ISTI of the National Research Council in Pisa. Franco Turini is a member of both ACM and IEEE, and currently serves as chairman of the Department of Computer Science of the University of Pisa.

Valerio Grossi (male) holds a Ph.D. in Computer Science from the University of Pisa and he is currently a research fellow at the Department of Computer Science, University of Pisa. He has been a researcher at Department of Pure and Applied Mathematics, University of Padova. His research interests focus on the analysis of massive and complex data, including mining data streams, ontology-driven mining, business intelligence and knowledge discovery systems. He took part in several European research projects as a member of the UNIPI research group, among which figure BRITE (Business Register Interoperability throughout Europe), MUSING (MUlti- industry, Semantic-based next generation business INtelliGence) and ICON (Inductive Constraint Programming), where he focused on the development of brand new data mining and knowledge discovery applications and on the research for the development of new business intelligence approaches.

Rossano Venturini (male) is an assistant professor of Computer Science at University of Pisa. He received his PhD. from the University of Pisa in 2010 discussing his thesis titled "On Searching and Extracting Strings from Compressed Textual Data". His research interests are mainly focused on the design and analysis of algorithms and data structures with special attention to problems of indexing and searching large textual collections. He is coauthor of "System and Method for String Processing and Searching using a Compressed Permuterm Index", patent pending owned by Yahoo! Inc. He is author of more than 30 papers published on the major conferences and journals of theoretical computer science. He is the recipient of a 2014 Yahoo! Faculty Research and Engagement Program (FREP) award, and he is (co-)author of the paper "Partitioned Elias-Fano Indexes", winner of the best paper award at ACM SIGIR 2014.

4.1.4 Fraunhofer IAIS and IGD – Germany *(FRH)*

FRAUNHOFER – IAIS/IGD FRH	

ORGANISATION DESCRIPTION

FRAUNHOFER – IAIS/IGD http://www.fraunhofer.de/

The Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V. is with more than 20,000 employees and 2 Billion Euro annual research budget one of the leading organizations for applied research world-wide. Fraunhofer participates in the project with two institutes: the Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS), and The Fraunhofer Institute for Computer Graphics Research (IGD).

The Fraunhofer **IAIS** is the Fraunhofer Society's "data institute" and one of the most renowned research institutions in the data science area, with an extremely strong track record in data mining, machine learning, visual analytics, semantic technologies, information retrieval and software engineering. In addition to the strong position in research IAIS has an outstanding track record in exploitation and transfer of research results. In addition to application-oriented European research projects, this is demonstrated in particular through contractual research performed for customers in telecommunications (Vodafone, Nokia Siemens), financial services and insurance (Allianz, Commerzbank), publishers (Springer, Zeit), retail (Rewe, Douglas) and logistics (Deutsche Post). Through its business consultants networks and CRM database IAIS has strong ties to all large German companies and a significant share of Germany's SMEs and startups. The Knowledge Discovery Department (KD) at the Fraunhofer Institute IAIS in Sankt Augustin, Germany, is a research group (about 50 scientists) operating in the field of Machine Learning, Data Mining, and Visual Analytics

The Fraunhofer **IGD** is the world's leading institute for applied research in Visual Computing. Visual Computing is image- and model-based information technology and includes computer graphics, computer vision, as well as virtual and augmented reality. Fraunhofer IGD develops prototypes and complete solutions based on end-user-specific requirements. Their research and development projects are directly applicable to current problems in industry. Visual analytics and semantics visualization are the topics of the Competence Center for Information Visualization and Visual Analytics (IVA) at Fraunhofer IGD. The interdisciplinary team around Dr. Jörn Kohlhammer conducts research and develops solutions for the interactive visualization of big data, so called visual analytics technologies, solutions for semantics visualization and real-time solutions for the simulation, visualization and rendering of data.

Role in the Project

FRH are contributing their IVAS visual analytics infrastructure, to be integrated in the SoBigData RI and offer both transnational and virtual access. In addition, they are leading the JRA activities in T9.5 on enhancing the SoBigData RI with new visual analytics methods. They are also involved in a wide range of networking activities, namely summer schools, workshops, innovation and impact generating activities with industry users; impact through policy bodies.

Infrastructure brought into the project

Fraunhofer IGD will contribute their expertise in the integration of disparate tools into cohesive analytics pipelines. Their major infrastructural contribution (the visual analytics framework) was born out of a need to systematise the integration process. In addition, the researchers in the Competence Centre for IVA have experience in user-centred design, engineering and deployment of visual analytics solutions, which

they will bring into the design and evaluation work packages.

IAIS has a cluster of 35 high-performance servers connected to a living lab, which its uses to develop big and linked data prototype systems. Fraunhofer maintains a large scale Customer Relationship Management system with tens of thousands of business contacts in all major German and European companies and a significant part of German SMEs.

Relevant Projects (5 IAIS and 5 IGD)

MODAP FET-Open Coordination Action (Mobility, Data Mining and privacy, 2009-2012)

VisMaster FET-Open Coordination Action on Visual Analytics (2008-2010)

GeoPKDD FET-Open project (Geographic Privacy-aware Knowledge Discovery and Delivery, 2005 - 2009)

DataSim FET-Open project (Data Science for Simulating the Era of Electric Vehicles, 2011 - 2014)

FERARI (Flexible Event pRocessing for big dAta aRchItecures), http://www.ferari-project.eu. EU FP7-ICT, 2014 – 2017

VisMaster (FET Open CA) project (over 20 partners) for the establishment of a visual analytics research community in Europe.

VIS-SENSE (STREP) project, in which visual analytics solutions for the improvement of network security

CAPER (IP) project, in which a visual interactive application for the combination and analysis of open and closed source datasets was realized for law enforcement agencies.

NOMAD (STREP) project a visual analytics solution was created to assist politicians in the assessment of citizen sentiment in policy decision making.

Relevant Publications (5 IAIS and 5 IGD)

Visual Analytics of Movement, G. Andrienko, N. Andrienko, P. Bak, D. Keim, S. Wrobel. Springer-Verlag, 2013, ISBN 978-3-642-37582-8

A Visual Analytics Framework for Spatio-temporal Analysis and Modelling, N. Andrienko and G. Andrienko, Data Mining and Knowledge Discovery, 2013, v.27(1), pp.55-83

Identifying Place Histories from Activity Traces with an Eye to Parameter Impact, G. Andrienko, N. Andrienko, M. Mladenov, M. Mock, C. Poelitz, IEEE Transactions on Visualization and Computer Graphics, 2012, v.18(5), pp.675-688

Spatial Generalization and Aggregation of Massive Movement Data, N. Andrienko, G. Andrienko. IEEE Trans. on Visualization and Computer Graphics, 2011, v.17 (2), pp.205-219

Challenging Problems of Geospatial Visual Analytics, G. Andrienko, N. Andrienko, D. Keim, A. MacEachren, S. Wrobel, Journal of Visual Languages and Computing, 2011, v.22 (4), pp.251-256

Bernard, Wilhelm, Krüger, May, Schreck, Kohlhammer; *MotionExplorer: Exploratory Search in Human Motion Capture Data Based on Hierarchical Aggregation*; IEEE Transactions on Visualization and Computer Graphics, Vol.19, 12, pp. 2257-2266; 2013.

Burkhardt, Nazemi, Stab, Steiger, Kuijper, Kohlhammer; *Visual Statistics Cockpits for Information Gathering in the Policy-Making Process*; 9th International Symposium on Visual Computing (ISVC), Springer, Berlin, Heidelberg, New York, LNCS 8034, pp. 86-97; 2013.

Steiger, May, Davey, Kohlhammer; *Visual Analysis of Expert Systems for Smart Grid Monitoring*; International EuroVis Workshop on Visual Analytics. Goslar: Eurographics Association, pp. 43-47, 2013.

Stab, Burkhardt, Breyer, Nazemi; *Visualizing Search Results of Linked Open Data*; Semantic Models for Adaptive Interactive Systems, Springer, Berlin, Heidelberg, New York, pp. 133-149, 2013.

Bernard, Ruppert, Scherer, Kohlhammer, Schreck; *Content-Based Layouts for Exploratory Metadata Search in Scientific Research Data*; Proceedings: 12th ACM/IEEE-CS Joint Conference on Digital Libraries. New York: ACM, pp. 139-148; 2012.

Overview of Key Staff

Gennady Andrienko (male) and **Natalia Andrienko** (female) are lead scientists responsible for the visual analytics research at Fraunhofer IAIS and full professors (part time) at City University London, UK. They co-authored monographs "Exploratory Analysis of Spatial and Temporal Data" (Springer, 1996) and "Visual Analytics of Movement" (Springer, 2013), more than 60 peer-reviewed journal papers, 20 book chapters, and about 100 conference papers. Since 2007, Gennady Andrienko is chairing the Commission on GeoVisualization of the ICA - International Cartographic Association. He co-organized scientific events on visual analytics, geovisualization and visual data mining, and co-edited 11 special issues of major journals. Since 2012, Gennady Andrienko is associate editor of two journals, Information Visualization and IEEE Transactions on Visualization and Computer Graphics. Gennady and Natalia Andrienko received best paper awards at AGILE 2006 and IEEE VAST 2011 and 2012, honorable mention award at IEEE VAST 2010, VAST challenge award 2008, and best poster awards at AGILE 2007 and ACM GIS 2011 conferences.

Georg Fuchs (male) is a postdoctoral research scientist (both basic and applied research) and project manager at Fraunhofer IAIS co-affiliated with the machine learning group at Bonn University working in the field of visual analytics, with a strong emphasis on spatio-temporal data analysis. His research interests include information visualization in general and visualization of spatio-temporal data in particular, visual analytics methodologies, task-driven adaptation of visual representations and Smart Visual Interfaces, as well as computer graphics and rendering. He has co-authored 38+ peer-reviewed research papers and journal articles, and received a best short paper award at Smart Graphics 2008.

Hendrik Stange (male) is a research scientist at Fraunhofer IAIS and senior project manager of many research and industrial big data projects with international consortia. He is the project manager at Fraunhofer as well as the liaison officer to the Federal Office of Civil Protection and Disaster Assistance within the EU project INSIGHT. His current research interests include learning on massive streams of heterogeneous data, visual analytics and big data infrastructures. Hendrik Stange has been principal investigator in several recent industry funded projects on mobile analytics and social mining. He has been an invited speaker at several conferences on public security, critical infrastructures and big data infrastructures.

Dr.-Ing. Jörn Kohlhammer (male) studied Computer Science at the Ludwig-Maximilian-University of Munich, Germany, where he received his MS (Dipl.-Inform.). From 1999 he worked as a full-time researcher at the Fraunhofer Center for Research in Computer Graphics (CRCG) in Providence, RI, USA. He received his Doctor's degree from the Technische Universität Darmstadt (TU Darmstadt) for his research on *Knowledge Representation for Decision-Centered Visualization*. Since 2004 he has headed the Fraunhofer IGD Competence Center for IVA with more than 15 employees. He has access to the extensive experience of his institute and the Fraunhofer-Gesellschaft in efficiently and successfully managing European projects.

Dr.-Ing. Thorsten May (male) studied Mathematics at the TU Darmstadt, Germany, where he received his MS (Dipl.-Math.). He joined the Competence Center for IVA in 2001. His research interests include high-dimensional data visualization, interactive feature subspace and data subset exploration, and visual cluster analytics. He received his Doctor's degree from the TU Darmstadt in 2011 for his research on *Models and Methods for the Coupling of Automated and Visual-Interactive Methods for Data Analysis*.

4.1.5 Tartu Ulikool – Estonian *(UT)*

UNIVERSITY OF TARTU UT	

ORGANISATION DESCRIPTION

University of Tartu http://www.ut.ee/

University of Tartu (UT) is the largest Estonian University (17000 students and 1800 academic staff) and the highest-ranked University in the Baltic States according to the Times Higher Education survey. UT's Institute of Computer Science conducts research in the fields of data mining, software engineering, distributed systems, bioinformatics and computational neuroscience. The Institute is the main research partner of the Estonian Software Technology and Applications Competence Centre (STACC), an R&D centre that conducts industry-driven research in the fields of software engineering and data mining in collaboration with 9 industry partners. Among other topics, the centre conducts research in the field of social network analysis (with Microsoft/Skype) and with the Estonian e-health foundation.

Role in the Project

UT will contribute to the RI e-government data, as well as being involved in trans-national access to help European researchers to make best use of their expertise in working with this kind of data. They will also act as the SoBigData contact point for researchers from the Baltic states and Eastern Europe, who are some of the target beneficiaries of transnational and virtual access, due to the limited resources they tend to have at their own institutions. UT will also contribute methods to the SoBigData RI and be involved in training and community building events.

Infrastructure brought into the project

UT's Institute of Computer Science will provide an inter-linked dataset of Web service descriptions covering, among others, the entire Estonian e-government infrastructure (X-Road). The institute will also provide data on service usage of Estonian e-government and e-health services and data related to societal and economic development in Estonia during the past decade. This research infrastructure will allow researchers in the network to study the actual impact of e-government service usage in Estonia on broader societal issues, as well as the potential impact of wider e-government service adoption in other countries.

Relevant Projects

ACSI – http://www.acsi-project.eu (FP7, partner) **–** Use of data mining methods to analyse event logs extracted from Web service hubs.
LiquidPub - http://liquidpub.wordpress.com/ (FP7, partner) – Social network analysis methods to assess reputation and impact in networks of scientists
CLARIN (ESFRI, Partner) **-** Common LAnguage Resources and technology Infrastructure, the UT node provides text and speech resources (corpuses and analysis methods) for Estonian language.
Data-Driven Management of Software and Business Processes (Estonian Research Council Institutional Grant) – Data mining methods for analysis of business process execution logs and Web service APIs.

Relevant Publications

S. Mokarizadeh, P. Küngas, M. Matskin: A framework for evaluating semantic annotations of Web services: A network theory based approach for measuring annotation quality. Web Intelligence and Agent Systems 12(1):15-34, 2014.

P. Küngas, M. Dumas, S. Mokarizadeh, M. Matskin: Analyzing Web Services Networks: Theory and Practice. In Advanced Web Services, Springer, 2014.

V. Floreskul, K. Tretyakov, M. Dumas: Memory-Efficient Fast Shortest Path Estimation in Large Social Networks. Proceedings of ICWSM 2014.

P. Küngas, S. Karus, S. Vakulenko, M. Dumas, C. Parra, F. Casati: Reverse-engineering conference rankings: what does it take to make a reputable conference? Scientometrics 96(2):651-665, 2013.

S. Mokarizadeh, P. Küngas, M. Matskin, M. Crasso, M. Campo, A. Zunino: Information Diffusion in Web Services Networks. Proceedings of ICWS 2012.

Overview of Key Staff

Marlon Dumas (male) is Professor of Software Engineering at UT. He is one of the five most highly cited software engineering researchers in Europe according to Google Scholar (15000+ citations, H-index of 57). Prof. Dumas has been co-recipient of three best paper awards (ETAPS'2006, BPM'2010, BPM'2013), three best student paper awards with his PhD students (EUC'2005, EEE'2006, CEC'2009) and a "10-years most influential paper award" at MODELS'2011. He is co-inventor of four granted US patents, two granted EU patents, and five pending US patent applications. He is also co-author of a textbook that is used in over 70 universities worldwide (Fundamentals of Business Process Management, Springer, 2013). Throughout his career, Prof. Dumas has been principal investigator, work-package leader or equivalent in two FP7 projects (ACSI and LiquidPub), two Estonian national projects and six Australian Research Council projects, and recipient of a prestigious Queensland State Fellowship (2004-2007). Since 2009, he is research leader at the Estonian Software Technology and Applications Competence Centre (STACC), a six-years, 10-million euros R&D centre jointly funded by EU Regional Development Funds and 9 private companies.

Dr. Peep Küngas (male) is a senior research fellow at UT. He holds a PhD in Information and Intelligence Science from Norwegian University of Science and Technology (2006) and worked as a post-doctoral fellow at the Department of Microelectronics and Information Technology of the Royal Institute of Technology (KTH) from 2007 to 2009. Peep has extensive experience in EU projects, including RoboSWARMs (FP6 ICT), LiquidPub, ACSI and REMICS (FP7 ICT). Over the years, he has developed expert knowledge in the areas of Web service composition, analysis of Web service networks, semantic interoperability, linked data and e-government services. He has published about 70 scientific peer-reviewed and more than 20 popular-science papers. Peep is member of Estonian National IT Interoperability Working Group and actively contributes to the development of e-Government strategies and policies.

Jean-Charles Lamirel (male) is senior lecturer at UT's Institute of Computer Science. Prior to this appointment, he was a research member at INRIA Nancy (CORTEX and TALARIS projects) and Assistant Professor (Maître de Conference) at University of Strasbourg. His areas of research are text mining, social network analysis and neural networks. He is the creator of the concept of Data Analysis based on Multiple Viewpoints (MVDA), which has been fruitfully implemented in the MultiSOM and MultiGAS models. His other main topics of research relate to Big Data Mining via numerical approaches, visualization methods for Big Data Mining, and data mining methods for novelty detection in social networks. His tools are currently used in several European projects on Webometrics and Data Analysis, e.g. EISCTES project. His research work has led to the defense of 8 Ph.D. theses and ca. 150 publications in international conferences and journals.

4.1.6 Scuola IMT (Istituzioni, Mercati, Tecnologie) Alti Studi di Lucca – Italy *(IMT)*

IMT Alti Studi Lucca IMT	

ORGANISATION DESCRIPTION

IMT Institute for Advanced Studies Lucca, https://www.imtlucca.it

IMT Institute for Advanced Studies Lucca is an international Graduate School, Institute of Technology and Institute for Advanced Studies that acts as a research university with the aim of forming human capital - specifically with regard to its Ph.D programs - in disciplines characterized by their high potential for concrete application. IMT strives to reach the fusion of theoretical comprehension and practical relevance.

Role in the Project

IMT will integrate in the SoBigData RI methods for social network analysis (WP10), as well as participate in the JRA activities in T9.1 on this topic, building on their expertise and methods for analyzing complex networks and systems (see projects next). They will also be involved in the networking and community building activities in the project, mainly through training and impact generation.

Relevant Projects

Financial systems simulation and policy modelling (SIMPOL) - Oct 2013 – Oct 2016, EU 7th Framework Programme, To make fundamental advances in financial policy modelling by bridging network science, big data and ICT. http://www.simpolproject.eu
Foundational research on multilevel complex networks and systems (MULTIPLEX)
Foundational research on multilevel complex networks and systems (MULTIPLEX) EU 7th Framework Programme, To develop a framework for multilevel complex systems. http://www.multiplexproject.eu. Partner/Co-PI; Nov 2012 – Oct 2017
CRISIS LAB, National Research Council of Italy,To define new concepts, methods and metrics for the quantitative analysis of trade networks in a strongly interacting, globalized market.
Partner/Co-PI; Oct 2013 – Sep 2016
FOC, EU 6th Framework Programme, Application of financial networks tools for the analysis of precursors of financial crises. http://www.focproject.eu, Partner/Co-PI; Oct 2013 – Sep 2016

Relevant Publications

G Caldarelli, “Scale-Free Networks: Complex Webs in Nature and Technology”, Oxford University Press (2007)

S. Battiston, G. Caldarelli, Co-Pierre Georg, R. May, J. Stiglitz, “Complex Derivatives”, Nature Physics 9 123 (2013)

W. Quattrociocchi, G Caldarelli and A Scala, Opinion dynamics on interacting networks: media competition and social influence, Nature Scientific Reports 4, Article number: 4938, doi:10.1038/srep04938 (2014)

W. Quattrociocchi, G. Caldarelli and A Scala, Self-healing networks: redundancy and structure, Plos One 9, e87986 http://dx.doi.org/10.1371%2Fjournal.pone.0087986 (2014)

G. Caldarelli, A. Chessa, A. Gabrielli, F. Pammolli, M. Puliga, “Reconstructing a Credit Network” Nature Physics 9 125 (2013)

Overview of Key Staff

Guido CALDARELLI (male) will act as Team Leader for research, and will work on the theoretical part related to the network analysis. Italian, born in 1967, he is Full Professor in theoretical physics in the institute IMT Alti Studi Lucca. He is author of about 100 publications on international journals (Nature, Nature Physics, Physics World, Physical Review Letters and Physical Review E, EPL, EPJB) and conference proceedings. Coordinator of the European project COSIN (http://www.cosinproject.org) one of the first and seminal acitivities in the European area of complex networks. He is presently coordinator of FOC (http://www.focproject.net). From January 1st 2012 he will build an observatory of financial risk financed by Italian Government. He served in the European Society for Complexity and he is author of one book on scale-free networks (published by Oxford University Press 2007) and editor of two other books (World Scientific Press 2007 and Cambridge University Press 2010) on network theory. He is also author of a "Very Short Introduction to Complex Networks" Oxford University Press (2012).

Walter Quattrociocchi (male) is research associate at LIMS and Associate Professor at IMT Lucca, where he is the director of the Computational Social Science Laboratory http://csslab.imtlucca.it/. He obtained the Bachelor Degree in Computer Science at the University of Siena, the Master Degree in Computer Science (Summa cum Laude) at the University of Parma, and the Ph.D in Logic and Computer Science at the University of Siena with the thesis "Computational Aspects of Social Dynamics". His research interests include dynamic networks, graph algorithms, dynamic processes on complex networks (social and biological). From 2007 to 2010 he worked at the Institute of Cognitive Sciences and Technologies (ISTC) of the Italian National Research Council (CNR). In 2012 he joined Northeastern University as Postdoctoral Research Associate at the Mobs Lab of Alessandro Vespignani. He was the technology coordinator of different Italian national projects among which Cnipa Projects servicomuni.it and intema2000.it. He serves as teacher and/or collaborator in courses of Algorithms, Programming, Distributed Computing, Social Networks for graduate and undergraduate students. Recently his research has focused on the information (and misinformation) diffusion in online social network and their relation with the opinion evolution. He has more than 30 publications.

Alessandro CHESSA (male) graduated in Physics and got a PhD in theoretical Physics at the University of Cagliari (Italy). He has been research associates in the Physics Department of Boston University, and Scientific Consultant at the International Center for Theoretical Physics (ICTP, Trieste). He has been chief Business Development Operator for Abbeynet S.p.a., and acted as a private Internet Marketing consultant. He has been university professor in a course of "Data structure and algorithms". In 2008 found Linkalab a private Complex Systems Computational Laboratory. In the year 2012 he has been adjunct researcher in the Institute for Complex Systems (CNR), 'La Sapienza' Rome, doing research in the field of Complex Network Theory. At present he is Assistant Professor in Statistical Physics in IMT, Institute for Advanced Studies, Lucca (Italy).

4.1.7 Gottfried Wilhelm Leibniz Universitaet Hannover – Germany *(LUH)*

Leibniz Universität Hannover – L3S Research Center LUH	

ORGANISATION DESCRIPTION

Leibniz Universität Hannover - L3S Research Center, http://www.L3S.de/

Founded in 1831, Leibniz Universität Hannover (www.uni-hannover.de) is one of the nine leading Institutes of Technology in Germany (TU 9) with around 23.500 students. More than 2.600 academics and scientists work in nine faculties with around 160 departments and institutes.

The L3S Research Center is a joint research institute of several universities in Lower Saxony, Germany formally associated with Leibniz Universität Hannover (LUH). Work at L3S focuses on innovative and cutting-edge methods and technologies for Web Science, a key topic for the European Information Society, investigating the Future of Information and Communication in the Web Age. Projects investigate innovative scenarios and solutions in the Web Science areas of Web Search (especially Entity-centric search), Web Information Management (including semantic technologies), Middleware for Web Infrastructures (including Security), the Future Internet and the Web of People (including Personalization and Social Web solutions). The L3S team builds upon technologies and approaches from various areas such as information retrieval, machine learning, natural language processing, Semantic Web technologies, data management, and Distributed Information Systems. LUH investigates the topics Web Archive and Social Web analysis, diversity and temporal aspects in information, metadata management, event-centric, entity-centric and privacy preserving information access to information in the context of various European and national projects like the ERC Advanced Grant Alexandria, or the projects Cubrik, Qualimaster, ARCOMEM or LivingKnowledge.

The Institute for Legal Informatics (IRI), being part of LUH's School of Law and member of L3S, was established in 1983 and is the first Institute dedicated to scientific research on all issues of Information and Communication Technologies at a German University. With currently more than 50 people staff IRI is one of Europe's largest institutions in the field and is actively involved in about 10 European research projects with a focus on data protection, data security and intellectual property. Nikolaus Forgó has been leading IRI in cooperation with Prof. Dr. Metzger since 2008.

Role in the Project

LUH will lead the JRA in WP8 on building the big data ecosystem, in order to enrich the SoBigData RI with new data through analytical crawling methods. This builds on the Alexandria infrastructure, which will be integrated and offered both through transnational and virtual access. LUH will also be involved in WP2, building on their expertise in legal informatics. In particular, they will lead T2.4 Intellectual property and data ownership framework. In WP9 they will work on enhancing the SoBigData RI with text mining methods, whereas in WP11 they will contribute with datasets and protocols for evaluating analytical crawling and engaging with the community around this challenge. Also training, community building, and innovation activities in this area.

Infrastructure brought into the project

LUH contributes the Alexandria infrastructure to the project that consists of 10 Nodes and 80 cores with 10GB per core. The cluster has gross storage capacity of 720 TB (net 261 TB). It comprises a Hadoop cluster with tools for indexing, accessing, and analyzing large Web corpora. In particular, the German Web corpus covers all Web pages of the .de domain from 1996 – 2013 originally crawled by the Internet Archive (approx.. 80TB). The infrastructure also provided virtual access via the AlexandriaCloud.

Relevant Projects

Alexandria - http://www.alexandria-project.eu/ (ERC Advanced Grant, Coordinator**) -** The ALEXANDRIA project aims to develop models, tools and techniques necessary to explore and analyze Web archives in a meaningful way.

CubRIK - http://www.cubrikproject.eu/ (FP7, Partner) - The CUBRIK project aims to develop an open source multimedia search framework which helps search application designers, technology providers, and content owners to manage the complexity of constructing and evolving scalable, multimodal, real-time media search and retrieval applications.

QualiMaster - http://qualimaster.eu/ (FP7, Coordinator) - The QualiMaster projects aims at making high volume real-time data processing a highly opportunistic process that flexibly exploits data sources, reconfigurable hardware and families of approximate algorithms in a configurable, quality-driven and adaptive way.

ARCOMEM - http://www.arcomem.eu/ (FP7, Partner) – AROMEM aims at developing social and semantic aware technologies for preserving community memories.

Optimis – http://www.optimis-project.eu/ (FP7, Partner) **–** Optimized Infrastructure Services: The high-level objective of the OPTIMIS project is to enable an open and dependable Cloud Service Ecosystem that delivers IT services that are adaptable, reliable, auditable and sustainable (ecological and economical)

Relevant Publications

Diaz-Aviles, E.; Drumond, L.; Schmidt-Thieme, L.; Nejdl, W.: Real-time top-n recommendation in social streams. ACM RecSys, Dublin, Ireland, 2012

Minack, E.; Siberski, W.; Nejdl, W.: Incremental diversification for very large sets: a streaming-based approach. ACM SIGIR, Beijing, China, 2011

Papadakis, G.; Ioannou, E.; Palpanas, T.; Niederée, C.; Nejdl, W.: A Blocking Framework for Entity Resolution in Highly Heterogeneous Information Spaces. Transactions on Knowledge and Data Engineering, 2012

Siersdorfer, S.; Chelaru, S.; Nejdl, W.; Pedro, J.S.: How useful are your comments?: analyzing and predicting YouTube comments and comment ratings. WWW, Raleigh, USA, 2010

Forgó, N.; Djemame, K.; Barnitzke, B.; Corrales, M.; Kiran, M.; Jiang, M.; Armstrong, D.; Nwankwo; I.: Legal issues in clouds: towards a risk inventory. Philosophical Transactions of the Royal Society A, A 2013 371, 20120075

Overview of Key Staff

Wolfgang Nejdl (male) has been full professor of computer science at the University of Hannover since 1995. He heads the L3S Research Center, as well as the Distributed Systems Institute / Knowledge Based Systems, and does research in the areas of search and information retrieval, information systems, data mining and recommender systems, semantic web technologies and digital libraries. Recent projects in the L3S context include the EU projects ARCOMEM and ForgetIT, investigating Web and organization archive management and advanced search in such archives, the EU project CUBRIK, developing algorithms and

systems for human-enhanced time-aware multimedia search, and Qualimaster, focusing on real time data stream analytics. He has just started an ERC Advanced Grant, ALEXANDRIA, to work on foundations for temporal retrieval, exploration and analytics in Web archives.

Thomas Risse (male) is the deputy managing director of the L3S Research Center and research group leader at L3S in Hannover. He received a PhD in Computer Science from the Darmstadt University of Technology, Germany in 2006. Before he joined the L3S Research Center in 2007 he lead a research group about intelligent information environments at Fraunhofer IPSI, Darmstadt. He worked in several European and industrial projects. He was the technical director of the European funded project BRICKS (about decentralized digital library infrastructures) and ARCOMEM (about Social and Semantic based Web archiving). He was the project coordinator of the EU FP7 project Living Web Archive (LiWA). Thomas Risse's research interests are Semantic Evolution, Digital Libraries, Web Archiving, Data Management in Distributed Systems and Self-organizing Systems.

Nikolaus Forgó (male) is a full Professor for Legal Informatics and IT-Law at the University of Hanover and leads together with Prof. Dr. Metzger the Institute for Legal Informatics. He specialises in e-commerce, internet and data protection law, areas in which he has published extensively. Prof. Forgó has substantial teaching and consulting experience in all fields of IT-law, legal informatics, civil law, legal history and legal theory.

Dipl.-Jur. Wolfgang Rottwinkel (male) is a Lecturer on law for engineers and law-students at Leibniz University of Hanover (LUH). He has studied law in Hanover. He is specialized in copyright law, intellectual property law and data protection law. At the Institute for Legal Informatics (IRI) he is taking care of all financial and administrative issues.

4.1.8 King's College London – UK *(KCL)*

KING'S COLLEGE LONDON KCL	

ORGANISATION DESCRIPTION

King's College London, Social and Cultural Informatics Group, Department of Digital Humanities, http://www.kcl.ac.uk/artshums/depts/ddh/index.aspx

King's College London is a research university located in London and a constituent college of the University of London. KCL is represented in ESDAI by a collaboration between the King's Social and Cultural Informatics group within the Department of Digital Humanities and the King's Digital Society Lab. We are an international leader in the application of technology in research in the arts and humanities, and in the social sciences. Key research areas are the theory and practice of digital libraries, digital archives and digital asset management; knowledge organisation and digital information management, through all stages of the digital lifecycle; researcher practices in the digital domain, including citizen and community engagement; ICT methods for digital scholarship and research, and the formalisation of research methods through the use of ICT; and e-research infrastructures and environments.

The Department of Digital Humanities has generated over £17 million in research grants over the past 7 years, and is involved in more than 30 major research projects, with funding from the Arts and Humanities Research Council (AHRC), the European Commission, the Engineering and Physical Sciences Research Council (EPSRC) and organisations like the Leverhulme Trust. The Department has extensive experience in the field of research infrastructures. It has built up a world class reputation in the management and use of digital content, including standards and guides to best practice for data creation, use and preservation. The Department is playing a leading role in the FP7 Integrating Activities EHRI, DAHISH and CENDARI, and has become the major hub for research infrastructure projects in the arts and humanities in the UK.

Role in the Project

KCL will lead T4.2 on training materials for stakeholders, where a particular asset is their strong connection to the digital humanities and social informatics users of social data analytics methods and datasets. They will also lead T11.1 to build connections to existing evaluation initiatives from related fields. KCL will also be responsible for organizing one of the datathons and some of the gender and diversity training activities. With respect to JRAs, they will be involved in enhancing the SoBigData RI with text and social media mining methods.

Recent Relevant Projects

PERICLES - Promoting and Enhancing Reuse of Information throughout the Content Lifecycle taking account of Evolving Semantics, (FP7, Coordinator). Development of new digital perseveration and curation techniques for research data

Our data ourselves, (AHRC, Investigator). Big Data Analytics on mobile phone data connecting to youth developer communities

Data Service Infrastructure for the Social Sciences and Humanities, (FP7, Member). Infrastructure cluster to connect all SSH ESFRI projects

European Holocaust Research Infrastructure, (FP7, Joint Research Coordinator). Integration of archives on the Holocaust; development of VRE

gMan project (finished in 2011), (JISC, Investigator). General-purpose VRE for document-centric research.

Recent Relevant Publications

Blanke, Tobias; Bryant, Michael; Hedges, Mark / Back to our data—Experiments with NoSQL technologies in the Humanities. 2013 IEEE International Conference on Big Data. Vol. N/A N/A. ed. N/A : IEEE, 2013.

Blanke, T.; Hedges, M. / Scholarly primitives: Building institutional infrastructure for humanities e-Science, In: Future Generation Computer Systems, Vol. 29, No. 2, 2013.

Hedges, Mark; Blanke, Tobias / Digital Libraries for Experimental Data : Capturing Process through Sheer Curation. Research and Advanced Technology for Digital Libraries, TPDL 2013. Berlin, Springer Berlin Heidelberg, 2013.

Blanke, T.; Candela, L.; Hedges, M.; Priddy, M.; Simeoni, F. / Deploying general-purpose virtual research environments for humanities research. In: Philosophical Transactions of the Royal Society A, Vol. 368, No. 1925.

Anderson, Sheila; Blanke, Tobias / Taking the Long View: From e-Science Humanities to Humanities Digital Ecosystems. In: Historical Social Research, Vol. 37, No. 3, 2012.

Overview of Key Staff

Tobias Blanke (male) is a Senior Lecturer in the Department of Digital Humanities at King's College London. His academic background is in philosophy and computer science. Before joining King's, he has held various lead development and analyst positions in the digital industries, from small-scale media companies to an adventure as a data warehouse analyst at a multi-national investment bank. Tobias can look back to over 10 years of leadership in research and development projects. Tobias has been leading on several projects in the computational analytics, from network analysis, open linked data, scholarly primitives to document mining and information extraction for research. His particular expertise is text mining and information retrieval on heterogeneous textual collections. His text analysis work is extensively published and has won best paper awards and runners-up best papers in international conferences.

Tobias works on several international projects and committees. Most notably, he is one of the directors of the Digital Research Infrastructure for Arts and Humanities (DARIAH), a European ESFRI project to create an integrated research infrastructure for digital arts and humanities, a key initiative for the upcoming Horizon 2020 framework and shared between 15 European countries.

Giles Greenway (male) is a Researcher in the Department of Digital Humanities at King's College London. Giles has a physics background, with a Master's degree in Computational Physics from the University of Salford and a Ph.D. in Polymer Physics from the University of Reading. Giles has over 8 years of experience in industrial and academic research. His recent research interests include graph databases, machine learning, natural language processing and enabling the teaching and learning of programming, particularly through mobile app. Giles is currently leading the research and development work on KCL's 'Our Data Ourselves' project to mine social data from mobile phone environments. From September 2014, he will be collaborating with the UK's Open Data Institute on a project to develop open social data.

4.1.9 Scuola Normale Superiore di Pisa – Italy *(SNS)*

SCUOLA NORMALE SUPERIORE SNS	

ORGANISATION DESCRIPTION

Scuola Normale Superiore di Pisa (www.sns.it), Group of Quantitative Finance (http://mathfinance.sns.it)

Scuola Normale Superiore di Pisa (SNS) is a public higher learning institution in Italy. It was founded in 1810 by Napoleon decree as a branch of the École Normale Supérieure of Paris. It operates a highly selective student admission procedure. Students are provided of full tuition, college life and international curriculum**.** It has roughly 300 undergraduate students, 200 PhD students and 110 professors and researchers. It is organized in two areas, the Faculty of Arts and the Faculty of Sciences. The last one includes Mathematics, Physics, Biology, Chemistry, and Information Science**.** SNS hosts a PhD school in Mathematics for the Industry and for Finance.

The group of Quantitative Finance (http://mathfinance.sns.it) is composed by three faculty members, 5 postdocs and 7 PhD student. Its research activity is focused on the analysis and modeling of socio-economic systems with tools of mathematics, statistics, statistical physics, network theory, and machine learning. Moreover part of its research focuses on air traffic modeling, by investigating large and complete databases of the European traffic.

Role in the Project

SNS will contribute to the SoBigData RI methods and expertise in social network analysis and pattern discovery. Existing methods will be integrated early in the project and new ones will be added as part of the JRA activities in WP9. SNS will also contribute to the dissemination and innovation activities in these fields, as well as to the evaluation framework in WP11.

Relevant Projects

Empirically grounded agent baSed models for the future ATM scenario (ELSA) - http://complexworld.eu/wiki/ELSA, (SESAR, partner) Big Data analysis and agent based modelling of air traffic in Europe

ComplexWorld. http://complexworld.innaxis.org/index.html (SESAR, partner), Complex System Research in Air Traffic Management (Managed by F. Lillo for the University of Palermo)

Complexity Research Initiative for Systemic InstabilitieS (CRISIS) - http://www.crisis-economics.eu (FP7, partner) **–** Large scale agent based modelling and data analysis for systemic risk in financial and economic systems

New tools in the credit network modelling with agents' heterogeneity - http://ineteconomics.org/grants/new-tools-credit-network-modeling-heterogenous-agents (INET, partner) Agent based and network modeling, large dataset analysis of databases of credit between banks and firms

Relevant Publications

G. Gurtner, S. Vitali, M. Cipolla, F. Lillo, R.N. Mantegna, S. Micciché, S. Pozzi, Multi-scale analysis of the European airspace using network community detection, *PLoS ONE* **9**(5): e94414. doi:10.1371/journal.pone.0094414 (2014)

M. Tumminello, S. Micciché, F. Lillo, J. Piilo, and R.N. Mantegna, Statistically validated networks in bipartite complex systems, *PLoS ONE* **6**(3): e17994. doi:10.1371/journal.pone.0017994 (2011).

M. Zanin and F. Lillo, Air Transport Network: a short review, *European Journal of Physics - Special Topics* **215** 5-21 (2013).

M. Tumminello, F. Lillo, J. Piilo, and R.N. Mantegna. Identification of clusters of investors from their real trading activity in a financial market. *New Journal of Physics* **14** 013041 (2012).

M. Tumminello, F. Lillo, R. N. Mantegna, Correlation, hierarchies, and networks in financial markets. Journal of Economic Behavior and Organization 75, 40-58. (2010)

Overview of Key Staff

Fabrizio Lillo (male) is Master in Physics and PhD in Physics at the University of Palermo (Italy). He is currently Junior Research Professor of Mathematical Finance at the Scuola Normale di Pisa and Assistant Professor in Physics at the University of Palermo (on leave). He is also External Faculty at the Santa Fe Institute, after being member of the Resident Faculty for three years (2009-2012). Formerly he has been postdoc (1999-2001) and then researcher of the National Institute of the Physics of Matter (2001-2003). After that he has been postdoc (2003) and member of the External Faculty (2004-2009) of the Santa Fe Institute. He has been awarded the Young Scientist Award for Socio- and Econophysics of the German Physical Society in 2007. He is author of more than 70 referred scientific papers. The ISI papers have received more 1552 citations and his ISI h-index is 23. He is also member of the editorial board of the Journals of Statistical Mechanics, Fluctuation and Noise Letters, Complex Adaptive System Modeling., Market Microstructure and Liquidity Besides other projects, he is responsible of one of the units of the FP7 funded European project CRISIS (Complexity Research Initiative for Systemic InstabilitieS) and of an INET grant, both focused on financial systemic risk. His research is focused on the application of methods and tools of statistical physics to economic, financial, and biological complex systems. Recently he has been interested in the microstructure of financial markets and high frequency finance, in the quantification and modeling of dependencies among economic quantities by using techniques of network theory, and models of systemic risk.

Stefano Marmi (male) studied physics at the University of Bologna, where he obtained a Master degree in June 1986 and a Ph.D. in Mathematical Methods of Physics in 1990. Since 1/11/2003 I have been full professor of dynamical systems at the Scuola Normale Superiore, Pisa. Previously he was full professor at the University of Udine and assistant professor at the University of Florence. Presently he is the director of the Centro di Ricerca Matematica “Ennio De Giorgi”, the director of the Fibonacci Laboratory UMI 3483 of the CNRS, the coordinator of the PhD school in Mathematical Finance at the Scuola Normale Superiore. During 2006 he was a visiting professor at the Mathematics Department of Princeton University. He is the author of several books on dynamical systems and analytical mechanics published in Italian and English by Bollati Boringhieri, Springer-Verlag and Oxford University Press. He has written over 50 publications on peer-reviewed international journals on topics ranging from dynamical systems to number theory and statistical mechanics. He is a specialist of dynamical systems and his contributions have been mostly on the topic of the transition from quasi-periodic to stochastic behavior and the interaction between dynamics and number theory. In the last few years he has also been interested in applications of dynamics and time-series analysis in economics and finance. In 1999 he was awarded the ISAAC prize at the Fukuoka Institute of Technology in Japan and he has been a Séminaire Bourbaki speaker in Paris on the chaotic dynamics of the Solar System.

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

Giacomo Bormetti (male) is Assistant Professor in Mathematical Finance at the Scuola Normale Superiore in Pisa since May 2013. Previously, he was post-doctoral fellow at the Scuola, Faculty of Sciences, working under the scientific supervision of Stefano Marmi. In 2007 he earned a PhD in Theoretical and Mathematical Physics at the University of Pavia discussing a thesis about the applications of Statistical Mechanics to the dynamics of financial markets. Then, he moved to the Institute for Advanced Studies in Pavia and for three years since 2008 was fellow of the Italian National Institute of Nuclear Physics. His expertise covers a range of areas, including theoretical physics, mathematical modeling and numerical methods. Since his graduation in Physics he has also dealt with different topics ranging from probability and statistics to quantitative finance, financial engineering and econometrics. Over the years he has published several academic articles on peer reviewed journals, and he is referee of various international journals. After completing his PhD he has been involved in several projects with the industry delivering tailored solutions to both Italian and international financial companies.

4.1.10 Aalto University – Finland *(AALTO)*

AALTO UNIVERSITY UH	Aalto University

ORGANISATION DESCRIPTION

Aalto University (AALTO), www.aalto.fi

Aalto University is a newly created university from the merger of three Finnish universities: Helsinki School of Economics, University of Art and Design Helsinki, and Helsinki University of Technology – all leading and renowned institutions in their respective fields and in their own right. Aalto University will contribute with the expertise and international recognition of the Department of Biomedical Engineering and Computational Science (BECS) and the Department of Information and Computer Science (ICS).

BECS is an interdisciplinary research environment, and has hosted the Academy of Finland Center of Excellence of Computational Complex Systems Research (COSY), in the period 2000-2011. The research conducted at BECS focuses on living and other complex systems, their measurement, analysis, modelling, understanding and control. The scale of the systems of interest ranges from quantum dots to broad networks: nanosystems, molecular interactions, neurons, the brain and the heart, cognition, information technology systems and social and economic networks. BECS combines experimental and computational methods and develops algorithms and new technologies to tackle major problems in human well-being, medical diagnostics, energy, the society and the environment.

The department of ICS focuses on research on advanced computational methods for modelling, analyzing, and solving complex tasks in technology and science. The research aims at the development of fundamental computer science methods for the analysis of large and high-dimensional datasets, and for the modelling and design of complex software, networking and other computational systems. Currently, the department hosts three national Centres of Excellence (COIN, SyMMyS, and ReSoLVE). The department also contributes to the Helsinki Institute for Information Technology HIIT and the Helsinki node of the EIT ICT Labs. The department is a member of Informatics Europe, the association of European computer science departments. The department of ICS has long-standing tradition in the areas of machine learning and data mining, and it attracts a continuous stream of bright Ph.D students and postdoctoral researchers from Finland and abroad.

Role in the Project

AALTO are providing transnational access methods to a social network analysis infrastructure for analyzing and extracting knowledge from large social network datasets. AALTO are also leading WP3 on dissemination, impact, and community building. In particular, they lead T3.5 on community building. In the JRAs in WP9, AALTO will lead T9.1, which will develop and integrate methods for exploration of social ego-networks, information flow on the networks, opinion dynamics in social networks, analysis of dynamic interaction networks, team formation in social networks.

Relevant Projects

MULTIPLEX (FP7, from December the 1st 2013 till November the 30th 2017 http://www.multiplexproject.eu/) Multiplex is a project gathering together the prominent research institutions at the intersection within *Network Science*, Complex Systems, *Algorithms*, *Statistical Physics* and so forth. The main aim of this project is oriented to use the mathematical framework of Complex Networks and Algorithmics to establish a theoretical basis for the understanding, prediction and possibly control of the Complex Systems. This would be obtained by reconstructing their Dynamics from the huge

and heterogeneous datasets available at different levels. The methods used will be derived from the approach developed in Graph Theory, Algorithmics and Statistical Physics. Validated Models will also be proposed.

Relevant Publications

S. Fortunato, Community detection in graphs, Physics Reports 486, 75-174 (2010). Most cited article in Physics Reports (since 2009), http://www.journals.elsevier.com/physics-reports/most-cited-articles/

C. Castellano, S. Fortunato, V. Loreto, Statistical physics of social dynamics Reviews of Modern Physics 81, 591-646 (2009). New Hot Paper in Physics, Thomson Reuters (ISI Web of Science), November 2010. http://sciencewatch.com/dr/nhp/2010/10novnhp/10novnhpCast/

Gionis, H. Mannila, P. Tsaparas, Clustering aggregation, ACM Transactions on Knowledge Discovery from Data (TKDD) 1(1), 4, (2007)

G. De Francisci Morales, A. Gionis, M. Sozio: Social Content Matching in MapReduce. Proceedings of the VLDB Endowment (PVLDB) 4(7): 460-469 (2011)

J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A.-L. Barabási, Structure and tie strengths in mobile communication networks, Proceedings of the National Academy of Science of the USA 104 (1), 7332-7336 (2007).

Overview of Key Staff

Aristides Gionis (male) is an Associate Professor in Aalto University, Finland, and the director of the Algorithmic Data Analysis (ADA) programme in Helsinki Institute for Information technology (HIIT). Previously he has been a senior research scientist in Yahoo! Research. He received his Ph.D. from the Computer Science department of Stanford University in 2003. He is currently serving as an associate editor in the Transactions of Knowledge and Data Engineering (TKDE). He has served in the PC of numerous premium conferences, including being the PC co-chair for WSDM 2013 and ECML PKDD 2010. Professor Gionis focuses on developing data-analysis algorithms for problems motivated by large-scale and real-world applications, and he has extensive experience in academic and industrial research. His expertise spans a wide range of areas, such as clustering problems, social-network analysis, web mining and usage mining, sequence segmentation, similarity search, streaming computation, and privacy preservation. According to Google scholar he has collected 7500 citations and an h-index of 36.

Santo Fortunato (male) has joined BECS in October 2011. Previously, he has held positions at the University of Bielefeld (Germany), at Indiana University (USA), and at the Institute for Scientific Interchange in Turin (Italy), where in 2009 he became leader of the Sociophysics Laboratory. He is the recipient of the Young Scientist Award for Socio- and Econophysics 2011, a prize given by the German Physical Society. His interests span from the theory and modelling of networks, with a focus on the detection of communities, to social dynamics, with a focus on the dynamics of scientific interactions and human behaviour in social and information systems. Recent projects include theoretical developments on graph clustering, e.g. the issue of detectability and testing of algorithms, and the characterization and modelling of citation and collaboration networks of scientists. He has authored over 70 peer-reviewed papers, which have collected more than 9,000 citations (Google Scholar), with an H-index of 35. His survey *Community detection in graphs* (Physics Reports 486, 75-174, 2010), the first exhaustive review article in the field, is the most cited paper in network science of the last years. He has delivered over 30 invited talks at international conferences and workshops, including several keynote talks, and has been in the PC of nearly 30 international events.

Kimmo Kaski (male) is Supernumerary Fellow of Wolfson College, Oxford and Fellow of APS, IOP, Finnish Academy of Sciences and Letters, Finnish Academies of Technology, and Academia Europeae. He was the director of the Center of Excellence COSY (2000-2011). He was two five-years terms Academy Professor (1996-2006) by the Academy of Finland and has held professorial level positions at Oxford and Temple

University. His research interests are in computational science and engineering, including computational physics, computer science/information technology, and complexity in economic, biological, and social systems. He has produced over 300 papers, with about 9,000 citations (Google Scholar) and an H-index of 47.

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

4.1.11 ETH Zurich – Switzerland *(ETHZ)*

Swiss Federal Institute of Technology Zurich ETHZ	ETH Eidgenössische Technische Hochschule Zürich Swiss Federal Institute of Technology Zurich

ORGANISATION DESCRIPTION

ETH Zurich – Swiss Federal Institute of Technology https://www.ethz.ch/

ETH Zurich was founded in 1855 and is the leading university in continental Europe. It hosts more than 400 professors from all over the world. Its main areas of research and education are engineering, natural sciences, mathematics, architecture, system-oriented sciences, management and social sciences. With these fields of expertise and broad competences, ETH Zurich addresses global challenges, such as sustainable energy production or systemic risks. 1 Turing award was earned here, and 21 Nobel laureates studied, taught or conducted research at ETH Zurich. ETH Zurich had 550 FP7 projects and among these, there had been 80 ERC projects. ETH Zurich engages in fundamental and goal-oriented research of lasting value. ETH Zurich has recently established a strategic focus also in Integrated Risk Management, which will closely link to the financial and insurance cluster in the Greater Zurich area. ETH Zurich leads the international Climate-KIC, an initiative to drive innovation in environmental change adaptation and mitigation, and many other remarkable initiatives such as the Future Cities Programme in collaboration with the two main universities in Singapore. Finally, ETHZ will greatly benefit from its Data Centre and the CSCS Swiss National Supercomputing Center.

The **Chair of Sociology**, in particular of Modeling and Simulation (SOMS), held by Professor Dr. Dirk Helbing, in the Department of Humanities, Social and Political Sciences (GESS) at ETHZ, focuses on social systems from the point of view of the complexity science, especially issues of emergence, self-organization, and network-related dynamics. Research at SOMS is extremely interdisciplinary and includes social modeling, computational social science, and big data analytics in demographic change and well-being. Recent research directions are also conducted in the field of social sensing: passive and participatory acquisition and procurement of big data in digital systems. Further focus areas are human interaction and coordination in space and time, quantitative sociodynamics, behavior of crowds and masses, pedestrian dynamics and panic, multi-agent simulation, self-organization and collective intelligence in complex systems, artificial societies, game theory, behavioral and decision models, robust organizations and disaster response management.

Role in the Project

ETHZ will contribute to the SoBigData RI data and methods for searching and analyzing Open Data (WP6). Their role in networking and community building activities will be in leading T4.3 on datathons, as well as contributing to T4.1 on summer schools. ETHZ will lead also T5.4 on building an innovation accelerator, which will enable funding bodies, policy makers, industry, and other stakeholders to measure and track the project impact and discover new trends and the latest methods in mining big social datasets. ETHZ will also lead T11.3 on sustainability and community engagement into the SoBigData evaluation framework, building on their extensive international research connections and industrial ties. With respect to JRAs, ETHZ lead T8.4. on integrating open data in the SoBigData RI, as well as integrating methods and services for analyzing social behaviour in T9.6.

Infrastructure brought into the project

The **Innovation Accelerator** (http://www.inn.ac/) is a general platform designed to monitor the latest developments in science, detect emerging trends and innovations, find the best experts for projects, and support the distributed generation of new knowledge, thereby promoting innovation. To that aim, an ecosystem of interconnected, open tools has been developed. Among these tools, LivingArchive indexes over 100.000 datasets (most of the data is openly licenced), spanning across a large spectrum of scientific domains. VirtualJournal performs massive data mining on open access papers. The dataset consists of PLOS (Public Library of Journals) journals (with full-text and references), amounting up to 1000000+ open access publications.

Relevant Projects

QLectives (EU project) - Socially Intelligent Systems for Quality Collectives (EU Integrated Project, Partner) http://www.qlectives.eu/

FuturICT Pilot (EU project) - Participatory Computing for Our Complex World http://www.futurict.eu/

Coping with Crises in Complex Socio-Economics Systems (ETH Zurich and ETH Foundation) http://www.ccss.ethz.ch/

Visioneer (EU project) Envisioning a Socio-Economic Knowledge Collider: From Computational Social Science to Social Computing (EU Support Action, completed) http://www.visioneer.ethz.ch/

Momentum (ERC Advanced Investigator Grant) - Modeling the Emergence of Social Complexity and Order: Order: How Individual and Societal Complexity CoEvolve

Relevant Publications

D. Helbing and S. Balietti (2011) How to create an innovation accelerator. *EPJ Special Topics* 195, 101–136.

D. Helbing and S. Balietti (2011) From social data mining to forecasting socio-economic crises. *EPJ Special Topics* 195, 3–68.

R. Conte, N. Gilbert, G. Bonelli, and D. Helbing (2011) FutureICT and social sciences: Big data, big thinking. *Zeitschrift für Soziologie* 40(5), 412–413.

F. Giannotti, D. Pedreschi, A. Pentland, P. Lukowicz, D. Kossmann, J. Crowley and D. Helbing (2012) A planetary nervous system for social mining and collective awareness. *EPJ Special Topics* 214, 49–75.

D. Brockmann and D. Helbing (2013) The hidden geometry of complex, network-driven contagion phenomena. *Science* 342(6164), 1337–1342.

Overview of Key Staff

Prof. Dirk Helbing (male) has worked in natural, engineering and social science departments and is currently a Full Professor in Sociology, in particular of Modeling and Simulation. After receiving his PhD in Theoretical Physics, he became Managing Director of the Institute for Transport & Traffic at Dresden University of Technology. This multi-disciplinary background and reputation of his work (with publications in Nature, Science and PNAS), allows him to oversee and lead scientific endeavours of wide scope. Mr. Helbing was elected as member of the German Academy of Sciences 'Leopoldina', the European Academy of Sociology, and the Complexity Council of the World Economic Forum. He is cofounder and chair of the Physics of Socio-Economic Systems Division of the German Physical Society, founder of the Competence Centre 'Coping with Crises in Complex Socio-Economic Systems', and cofounder of ETH Zurich's Risk Centre. He has coordinated the FuturICT Initiative (http://www.futurict.eu), which focuses on the understanding of techno-socio-economic systems, using Smart Data. His work is documented by hundreds of scientific articles, keynote lectures and media reports worldwide. His research ranges from fundamental science to practical applications. He has advised the International Risk Governance Council, and edited a book on Managing Complexity. He scientifically guided the reorganisation of the Muslim

pilgrimage to Mecca in 2007, accompanying the construction of the New Jamarat Bridge, a 1 billion dollar investment. Mr. Helbing's team also developed a novel traffic assistance system to mitigate traffic congestion and patented a paradigm-changing self-control of urban traffic lights, which has received the Golden Idea Award of Idee Suisse. On January 10, 2014, he received a honorary PhD from the TU Delft jointly from the Faculty of Technology, Policy and Management and the Faculty of Civil Engineering and Geosciences.

Dr. Izabela Moise (female) is currently a postdoctoral researcher at the Chair of Sociology, in particular Modeling and Simulation at ETH Zurich. Izabela received her PhD degree in Computer Science from INRIA Rennes, France, the National French Institute for Research in Computer Science and Automatics. In the past, she has worked as a research engineer in the context of two european projects: EIT ICT labs, the "location privacy" action line and the Quaero project.
Izabela was responsible for developing GEPETO(GEoPrivacy Enhancing TOolkit), a platform that enables users to conduct a privacy analysis on large-scale geolocated datasets. The main challenge was to provide efficient solutions for algorithms that process Big Data (clustering and indexing tools). All algorithms are designed as MapReduce applications and developed based on Hadoop. During the PhD studies, Izabela has attained a deep knowledge in the domain of distributed systems and algorithms with a main focus on constructing scalable and reliable distributed systems.

Dr. Evangelos Pournaras (male) is a postdoctoral researcher in the Chair of Sociology, in particular Modeling and Simulation at ETH Zurich. Evangelos has earlier been a postdoctoral researcher at Delft University of Technology working on the NWO project "RobuSmart", focusing on managing robustness in Smart Grids using large-scale demand data. He continued and completed his PhD studies in 2013 at Delft University of Technology after moving from VU University Amsterdam. He holds a MSc with distinction in Internet Computing from University of Surrey, UK and a BSc on Technology Education and Digital Systems from University of Piraeus, Greece. He has also been a visiting researcher at Ecole Polytechnique Federale de Lausanne in Switzerland in 2013 and an intern at IBM T.J. Watson Research Center in the USA in 2010. He serves the program committees of several international conferences and has several publications in area of self-management of large-scale decentralized systems, such as Smart Grids. He is currently working on the Planetary Nervous System project that involves a large-scale distributed collection and analysis of data for self-regulating socio-technical systems.

4.1.12 Technische Universiteit Delft – The Netherlands *(TUDelft)*

Technische Universiteit Delft TUDelft	TU Delft Delft University of Technology

ORGANISATION DESCRIPTION

Technische Universiteit Delft, http://www.tudelft.nl/en/

TU Delft is one of three technical universities in the Netherlands, renowned for the high quality of its research and teaching. Recently, TU Delft has been ranked on the 42nd place in the global reputation ranking list of universities, the World Reputation Rankings of Times Higher Education magazine. With the new ranking, TU Delft is now the highest ranked Dutch university and the third highest European university of technology.

Within the **Faculty of Technology, Policy and Management** (TPM) there exists a co-operative relationship between philosophy, the social sciences and the exact sciences/technology, in order to be able to make a significant contribution to sustainable solutions for social problems in which technology plays an important role. Important research themes at TPM are 'responsible innovation' and 'value sensitive design' (which includes 'privacy by design'), and 'social system interaction'. The TPM Faculty actively participates in the 3TU.Centre for Ethics and Technology (3TU.Ethics), which is a collaboration of the philosophy sections of the 3 Dutch Universities of Technology. Their research programs on the philosophy of science, technology and engineering are recognized as internationally leading in these fields. 3TU.Ethics studies ethical issues in the development, use and regulation of technology – including ICT and energy.

TU Delft will provide the research context of a newly established Big Data Centre, the Delft Institute for Safety and Security and the context and resources of the new cybersecurity chair at TU Delft. Crucial is the working group on VSD that is host to a number of PhD students who all work on designing IT applications for Moral Values.

Role in the Project

TUDelft will lead WP2, addressing the critical challenge of establishing the legal and ethical framework for the SoBigData RI, in accordance to the EU policy for responsible research and innovation. They will also establish and chair the project's ethics board, which will ensure that all JRA activities are in line with ethical guidelines. They will also work on privacy-by-design and value-sensitive social data mining.

Relevant Projects

ETICA: Ethical Issues of Emerging ICT Applications (7th Framework Programme, ref 230318). Duration: April 2009 - May 2011. Jeroen van den Hoven was the principal investigator.
STARS: Sensor Technology Applied in Reconfigurable systems (www.starsproject.nl). Duration: July 2010 – January 2015. Jeroen van den Hoven is leader of the work package on ethics.
SURVEILLE: Surveillance: Ethical Issues, Legal Limitations, and Efficiency (7th Framework Programme, ref 284725). Duration: February 2012 – April 2015. Jeroen van den Hoven is advisor of the project.
FuturICT (7th Framework Programme, FET Flagship Pilot, ref 284709). Duration: May 2011 – April 2012. Jeroen van den Hoven has been member of the preparatory research group.
Expert Group on the State of Art in Europe on **Responsible Research and Innovation** (DG Research and Innovation). Final report published in 2013. Jeroen van den Hoven has been chair of the expert group.

Relevant Publications

van den Hoven, J. (2013). "Value Sensitive Design and Responsible Innovation". In: Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society. Ed. by M. H. Richard Owen John Bessant. Wiley Online Library, pp.75–83.
van den Hoven, J., D. Helbing, D. Pedreschi, J. Domingo-Ferrer, F. Gianotti, and M. Christen (2012). FuturICT—The road towards ethical ICT. The European Physical Journal - Special Topics 214(1), 153–181.
Manders, N. and J. van den Hoven (2009). "The need for a value-sensitive design of communication infrastructures". In: Evaluating new technologies: methodological problems for the ethical assessment of technology developments (International library of ethics, law and technology, 3). Ed. by P. Sollie and M. Düwell. New York/Berlin/Dordrecht: Springer, pp. 51–60.
van den Hoven, J. (2008). "Information technology, privacy, and the protection of personal data". In: Information technology and moral philosophy. Ed. by J. van den Hoven and J. Weckert. Cambridge, New York: Cambridge University Press, pp.301–321.
van den Hoven, J. and J. Weckert, eds. (2008). Information Technology and Moral Philosophy. Cambridge: Cambridge University Press.

Overview of Key Staff

Jeroen van den Hoven (male) is professor of moral philosophy at TU Delft and until quite recently scientific director of 3TU.Ethics. He is editor-in-chief of Ethics and Information Technology (Springer), and Founding Chair of the CEPE conference (Computer Ethics Philosophical Enquiry). He is member of the editorial board of Information, Computers and Society (Routledge), Journal of Information, Communication and Ethics in Society and consulting editor of Episteme. He is also member of the board of the International Society for Information Ethics (INSEIT). Van den Hoven has received several grants from the Netherlands Organization for Scientific Research (NWO) on ethics and IT, and related subjects. He is member of the IST Advisory Group to EU Commissioner Reding for ICT and New Media. He has been advisor to the Dutch Government in various roles, has worked together with IT industry, and has published numerous articles on ethics and ICT. He has previously contributed to EU-project ETICA (Ethical issues of emerging ICT applications).

4.2. Third parties involved in the project (including use of third party resources)

Neither third parties, nor subcontractors are involved in this project.

5. Ethics and security

5.1 Ethics

Due to the nature of the proposed research infrastructure, **SoBigData** needs access to digital records of personal activities that potentially contain sensitive information: user mobility data, online social network and other web and internet data, handling of personal data in web pages, transaction records, etc. In this context, **SoBigData** has a commitment to promoting and adopting legally and ethically grounded collection, management and analysis of data, as demonstrated by effort of the project in considering and investigating ethical and privacy issues in the definition of big data analytics and social mining technologies (see NA1 work package). This aspect is shown also by the adoption of privacy enhancing tools. Also gender issues are considered proactively into the project work plan.

5.1.1 Proactive approach

The **SoBigData** consortium does not consider ethics as simply an external constraint on its research, but proactively addresses ethical issues as part of the research activities. In order to address ethical issues in an adequate manner, a number of measures (see work package NA1) will be taken: An ethics board will be installed, an ethical and legal framework to be used within the project will be designed, and issues related to privacy and data ownership in the context of **SoBigData** will be pro-actively explored.

The ethics board serves as an internal review process and provides guidance to researchers as to how to address issues concerning research ethics and data protection. All experiments, including the analysis of existing datasets that include personal information, will be reviewed by the ethics board and it will be ensured that adequate safeguards are in place. Where applicable, all experiments will be furthermore submitted for review at the corresponding ethics board of the research institutions conducting the experiment.

The development of an ethical and legal framework as part of work package NA1, but also the application of the value-sensitive design methodology to the **SoBigData** objectives, add further pro-active elements to these safeguards. Issues that are known to be problematic with regard to data protection and the application of relevant legal provisions will be actively explored as part of the research objectives. These will in turn inform the research activities that are executed as part of other work packages (e.g. JRA2).

Besides, we consider Value Sensitive Design (VSD) as a normative base for **SoBigData** and for the education and research of data scientists, reflecting the idea that legal norms and moral values – such as privacy and justice – should be construed as non-functional requirements of any product of software engineering, architecture building, modelling, simulations and serious gaming and should be clearly and precisely represented in software and system development. VSD should allow for a systematic, sustained and seamless collaboration between all involved Partners on the articulation of moral principles of a just information society and their effective implementation. The formulation of a robust notion of VSD requires:

- A detailed account of how values and moral considerations can be given their due place in software engineering (methodology)
- A detailed account of how they can be used as requirements in the design of ICT tools, applications and services.

Instruments to increase societal ethical awareness and participation: **SoBigData** will develop and implement a research infrastructure that requires ethical awareness, which should be coupled to the ability of scientists to act responsibly and to consider possible consequences. We will therefore

adopt appropriate training for allowing individual researchers to reflect, compare and deliberate upon their values and upon their resources and competences to deal with moral issues.

5.1.2 Data protection

Applicable data protection regulations will be investigated and measures to enforce privacy will be implemented as an integral part of the systems that will be developed in the project. However, scientists within the consortium and outside it (within TA) will manage potentially re-identifiable data for the purpose of conducting the planned research. While doing so, all institutions and individual researchers in the **SoBigData** consortium shall comply with the applicable legislation on EU and national level, in particular Directive 95/46/EC (Data Protection Directive). Furthermore, relevant guidelines such as the "code of conduct applying to processing of personal data for statistical and scientific purposes" set forth by the Italian National Data Protection Commission will be considered.

Whenever possible, data subjects will be informed about the data processing and asked for their consent. In cases where existing datasets are analysed (e.g. for the crawling of web-data and the import of data sets), the guidelines set forth by the Article 29 Working Party on *purpose limitation* will be taken into account[2]. This means that data processing for the goal of statistical processing takes place separated from the original purposes for which this data was collected. The results of such processing are not linked back to the data subjects and all data is anonymized as early as possible.

Researchers will be trained in applying the necessary procedural safeguards and the project will take the necessary steps to ensure that applicable legislation concerning data protection is respected, and the steps taken during the project lifetime to ensure conformance to established guidelines will be documented. To this end, a specific *SoBigData Ethics Board* will be established (see NA1 work package).

It should be noted, besides, that a number of privacy issues that could arise when handling the above data, do not arise within **SoBigData**; specifically:

1. It is not planned to process, to combine or to re-link the data to be analyzed with personal data such as name, address, social security ID etc., and other information that can be used for directly identifying persons. If directly identifiable data is collected (e.g. as part of user provided datasets), such data will be removed or anonymized as early as possible. All data that is used for analysis can therefore at most be used to indirectly identify individuals.
2. The data sets and the potentially sensitive information contained in then will be handled within the data management provisions of the Big Data Ecosystem (see JRA1 work package), securing the different sharing/diffusion constraints of each data source. All data sets will be stored with appropriate security restrictions in place. This comprises access restrictions to IT systems and storage such as password protection, encryption, etc., complying with all national and EU legislation.
3. **SoBigData** investigates the data sets solely for the purpose of research. Publications need not and will not disclose specific privacy-relevant information.

[2] See Article 29 Working Party (2013). Opinion 03/2013 on purpose limitation. Available at http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

5.1.3 Human participants

Some of the planned activities include the direct involvement of human participants in the collection of data. This concerns, for example, the possibility to upload social network datasets (work package NA5) or participatory sensing methods (work package JRA1). To the extent that those activities constitute human subjects research, relevant principles, such as those set forth in the Declaration of Helsinki, will be applied. The experiments will furthermore be reviewed by the **SoBigData** ethics board and by review committees of the corresponding research institutions.

When describing issues relating to informed consent, it will be necessary to illustrate an appropriate level of ethical sensitivity. All participating users will have given signed consent, following the EU requirements using a standardized form, which will be printed in the appropriate language.

Prior to the start of the process, the partners will agree the method of approach to the individual, the handling of the completion of forms, the confidentiality of the process, the ability of the user to decline, and the handling, access, and storage of data will all follow a predetermined design at all sites. All users will also be informed of the process of withdrawal from the project. Each user will be given a user identity number and only this will appear on all other documentation during the project. This will assist in the anonymity of all users.

5.1.4 Integrating the gender dimension in SoBigData

Following the European policy of equal opportunities between women and men, the Commission has adopted a gender mainstreaming strategy by which each policy area, including that of research, must contribute to promoting gender equality. The consortium is highly concerned with this issue and all the partners are committed to promote women's participation in SoBigData.

We will specifically address gender aspects in recruitment. The **SoBigData** consortium is also engaged to take actions to increase the involvement of women scientists in order to reach a 40% minimum target. The main actions will be articulated around these ideas:

- Women participation in research must be encouraged both as scientists/technologists and within the evaluation, consultation, and implementation processes,
- There should be a gender balance between men and women in laboratories, in senior management, which reflects their roles in society.

We note that, specifically, **SoBigData** has the potential and the necessary instruments to promote excellence through mainstreaming gender equality since it has:

- a female global coordinator (Fosca Giannotti – CNR)
- a female deputy-coordinator (Kalina L. Bontcheva – USFD) who is also work package leader of TA1 – Transnational Access
- leading female researchers, such as Natalia Andrienko (FRH) and Donatella Castelli (CNR).

The European Machine Learning and Knowledge Discovery community, chaired by F. Giannotti, has already adopted a proactive approach in promoting gender equality, while taking into account the specific circumstances and interests of the different women scientists active in the community, and will serve as a channel for recruiting female scientists both in the consortium and as **SoBigData** users.

5.1.5 Ethical Review

SoBigData involves the groups at TUDelft and LUH as internal IT ethics and IT law experts for consultation and is open to ethical review by the Commission. The consortium shall ensure that the

research carried out under the project fully complies with the national legal and ethical requirements of the countries where the tasks raising potential ethical issues are to be carried out. To this purpose, the consortium shall submit an ethics self-assessment, where we shall explain in detail how to address the ethical issues, if any, raised by the research objectives, methodology and results.

5.1.6 Ethics requirements summary

SoBigData has a dedicated WPs NA1 (Legal and Ethical framework) that addresses the ethical requirements. In particular, **SoBigData** will establish a dedicated ethical and legal board with experts in Ethics of IT and IT Law, who will:

- Define and implement the legal and ethical framework of the **SoBigData** research infrastructure, in accordance with the European and national legislations (including data protection and intellectual property rights) as they develop
- Monitor the compliance of experiments and research protocols with the framework,
- Act as a continuous consultant for the JRAs on the development of big data analytics and social mining tools with Value-Sensitive Design and privacy-by-design methodologies, and
- Investigate, design and promote novel architectures, protocols and procedures for the safe and fair use of big data for research purposes, in order to boost excellence and international competitiveness of Europe's big data research.

The following table summarize how the Ethics requirements will be tackled.

Category HUMANS	
1. *Details on the procedures and criteria that will be used to identify/recruit research participants must be provided*	See section 5.1.3 (Human participants) The **SoBigData** Ethics Board will review the research procedures to monitor their compliance with Declaration of Helsinki.
2. *Detailed information must be provided on the informed consent procedures that will be implemented.*	See section 5.1.3 (Human participants) All participating users will have given signed consent, following EU requirements and using a standardized form in the appropriate language.
*3. The applicant has to further clarify how passive actors are involved for the purpose of data collection, and what **informed consent** procedures will be implemented for that case.*	These issues will be considered by the **legal and ethical framework** proposed by the Ethics Board, it will be revised after year 2 of the project, in order to cover necessary modifications in the development of the project or changed legislation. As stated in Task 2.2 a key legal and ethical issue will be the question under which provisions and codes of conduct collected datasets can be (re)used for secondary research purposes. Guidelines will also take into account existing national codes; for example, the Italian code of conduct stipulates that in certain cases information to users can be transmitted to through newspaper or other media instead of personal communication.

	Basic principles are however already presented in section 5.1.2 (Data Protection) and are inspired to Article 29 Working Party on *purpose limitation*[3] and the Data Protection Directive "*either by a valid informed consent or a legal exception, such as "processing for scientific purposes"*.
Category PROTECTION OF PERSONAL DATA	
1. *The applicants must provide a policy for dealing with re-identifiable data. Moreover, the applicants must provide a policy for dealing with findings which are unrelated to the proposed work.*	One major duty of the Ethics Board is to design and update the **ethical and legal framework** to be adopted in SoBigData. Basic principles of the policy for dealing with re-identifiable data are established in section 5.1.2 (Data Protection).
2. *Copies of ethical approvals for the collection of personal data by the competent University Data Protection Officer / National Data Protection authority must be submitted.*	It is part of the planned procedures that will be implemented by the ethics board as established in Task 2.1.
3. *Justification must be given in case of collection and/or processing of personal sensitive data.*	The collection of personal sensitive data is not an objective in SoBigData. If needed, SoBigData will provide the appropriate justification and will put in place the appropriate security procedures.
4. *Detailed information must be provided on the procedures that will be implemented for data collection, storage, protection, retention and destruction and confirmation that they comply with national and EU legislation.*	Details will be provided as part of Task 8.1 (Data Management and Integration Plan) following the guidelines established by the Ethics Board in the Ethics and Legal framework (T2.2) so that they will be guaranteed to be compliant with the Data Protection Directive.
5. *Detailed information must be provided on the informed consent procedures that will be implemented.*	See point 3 of Category "Humans".
6. *The applicant must explicitly confirm that the existing data are publicly available.*	Dataset that are publicly available under different access provisions will be curated and managed as established in task8.1 (Data Management and Integration Plan).
7. *In case of data not publicly available, relevant authorisations must be provided.*	Dataset not publicly available will have appropriate authorization and appropriate access procedures to be used by researchers on site as established in task8.1 (Data Management and Integration Plan).
8. *In the course of the project, it is possible that the consortium encounters new ethics issues. Any such new issues and how they have been addressed are to be documented to contribute to the understanding of ethics in big data research.*	The Ethics Board installed and managed in work package NA1 proactively addresses ethical issues as part of the research activities in order to tackle current and new ethical issues. The Ethics Board will provide guidance to researchers as to how to address issues concerning research ethics and data.
Category NON-EU COUNTRIES	
1. *The applicants must provide confirmation whether the proposed work requires local/national ethics clearance in those countries where the research will take place.*	We do not foresee such cases, but if it happens the Ethics Board will take care of these cases.
2. *The applicant must confirm that the ethical standards and guidelines of Horizon2020 will be*	Yes, we confirm.

[3] See Article 29 Working Party (2013). Opinion 03/2013 on purpose limitation. Available at http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp203_en.pdf

rigorously applied, regardless of the country in which the research is carried out.	
3. The applicant must provide details on the material which will be imported to/exported from EU and provide the adequate authorisations.	These details concerning data exchange with Swiss and non-EU countries will be inserted in the (Data Management and Integration Plan).

5.2 Security

We do not identify issues of security related to the current proposal, other than the above mentioned provisions for secure data management of potentially sensitive or otherwise restricted access to data.

Bibliographic references

- **[Andrienko 2011]** - N. Andrienko, G.Andrienko. Spatial Generalization and Aggregation of Massive Movement Data. IEEE Transactions on Visualization and Computer Graphics, 2011, v.17 (2), pp.205-219
- **[Barabasi 1999]** - A. L. Barabasi, R. Albert, Emergence of Scaling in Random Networks, Science 286, 509 (1999).
- **[Barabasi 2008]** - M. C. Gonzalez, C. A. Hidalgo, A. L. Barabasi, Understanding human mobility patterns, Nature 454, (2008) 779-782.
- **[Berlingerio 2010]** - M. Berlingerio, M. Coscia, F. Giannotti, A. Monreale, D. Pedreschi, As Time Goes by: Discovering Eras in Evolving Social Networks, PAKDD 2010.
- **[Berlingerio 2012]** - M. Berlingerio, M. Coscia, F. Giannotti, A. Monreale, D. Pedreschi, Multidimensional Networks: Foundations of Structural Analysis, WWW Journal. 2012.
- **[Bontcheva 2014]** - K. Bontcheva, I. Roberts, L. Derczynski, D. Rout. The GATE Crowdsourcing Plugin: Crowdsourcing Annotated Corpora Made Easy. Proceedings of the meeting of the European chapter of the Association for Computation Linguistics (EACL) 2014.
- **[Bosch 2013]** - Harald Bosch and Dennis Thom and Florian Heimerl and Edwin Puttmann and Steffen Koch and Robert Kruger and Michael Worner and Thomas Ertl. ScatterBlogs2: Real-Time Monitoring of Microblog Messages through User-Guided Filtering. IEEE Transactions on Visualization and Computer Graphics 2013 (vol 19, num 12)
- **[Bringmann 2010]** - B.Bringmann, M. Berlingerio, F. Bonchi, A. Gionis, Learning and Predicting the Evolution of Social Networks, IEEE Intelligent Systems (EXPERT), 2010.
- **[Brockmann 2006]** - D. Brockmann, L. Hufnagel, T. Geisel, The scaling laws of human travel, Nature 439, 2006.
- **[Buckingham-Shum 2012]** - Simon Buckingham Shum et al., Democratising Big Data, Complexity Modelling and Collective Intelligence. Eur. Phys. J. Special Topics 214, 2012.
- **[Caldarelli 2007]** - G. Caldarelli, Scale free networks, Oxford University Press.
- **[Ceccarelli 2013]** - Diego Ceccarelli, Claudio Lucchese, Salvatore Orlando, Raffaele Perego, Salvatore Trani: Learning relatedness measures for entity linking. CIKM 2013: 139-148
- **[Cornolti 2013]** - M. Cornolti, P. Ferragina and M. Ciaramita. A Framework for Benchmarking Entity-Annotation Systems. ACM WWW 2013.
- **[Coscia 2011]** - M. Coscia, F. Giannotti, D. Pedreschi, Statistical Analysis and Data Mining 4(5), 512-546 (2011).
- **[Easley 2010]** - D. Easley and J. Kleinberg, Networks, Crowds, and Markets: Reasoning About a Highly Connected World, Cambridge University Press, 2010.
- **[Esuli 2010]** - A. Esuli and F. Sebastiani, Machines that learn how to code open-ended survey data. International Journal of Market Research, 52(6), (2010) 775-800.
- **[Fekete 2012]** - J.D. Fekete and S. Claudio. Managing Data for Visual Analytics: Opportunities and Challenges. IEEE Data Eng. Bull. Journal 2012

- **[Ferrari 2012]** - L. Ferrari, M. Mamei, Classification and prediction of whereabouts patterns from reality mining dataset, Pervasive and Mobile Computing, 2012.
- **[Fortunato 2010]** - S. Fortunato, Physics Report 486 (3-5), 75-174 (2010).
- **[Ghani 2013]** - Ghani, Sohaib and Kwon, Bum chul and Lee, Seungyoon and Yi, Ji Soo and Elmqvist, Niklas. Visual Analytics for Multimodal Social Network Analysis: A Design Study with Social Scientists. IEEE Transactions of Visualization and Computer Graphics 2013 (vol. 19, num. 2)
- **[Giannotti 2007]** - F. Giannotti, M. Nanni, F. Pinelli, D. Pedreschi, Trajectory pattern mining,Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2007, 330-339.
- **[Giannotti 2008]** - F. Giannotti and D. Pedreschi, Mobility, Data Mining and Privacy, Springer, 2008.
- **[Giannotti 2011]** - F. Giannotti, M. Nanni, D. Pedreschi, F. Pinelli, C. Renso, S. Rinzivillo, R. Trasarti, Unveiling the complexity of human mobility by querying and mining massive trajectory data. The VLDB Journal 20(5), (2011) 695-719.
- **[Giannotti 2012]** - F. Giannotti, D. Pedreschi, A. Pentland, P. Lukowicz, D. Kossmann, J. Crowley and D. Helbing. A planetary nervous system for social mining and collective awareness. Eur. Phys. J. Special Topics 214, 49-75 (2012). http://dx.doi.org/10.1140/epjst/e2012-01688-9
- **[Gurtner 2014]** - G. Gurtner, S. Vitali, M. Cipolla, F. Lillo, R.N. Mantegna, S. Micciché, S. Pozzi, Multi-scale analysis of the European airspace using network community detection, PLoS ONE 9(5): e94414. doi:10.1371/journal.pone.0094414 (2014)
- **[Hastie 2009]** – T. Hastie, R. Tibshirani, and J. Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition. Springer Series in Statistics, 2009.
- **[Helbing 2009]** - D. Helbing and W. Yu, The outbreak of cooperation among success-driven individuals under noisy conditions, PNAS 106 (8), (2009) 3680-3685.
- **[Holme 2011]** - P. Holme, J. Saramaki, Temporal Networks. Physics Reports 519(3) eprint arXiv:1108.1780, 2011
- **[Hoven 2012]** - J. van den Hoven, D. Helbing, D. Pedreschi, J. Domingo-Ferrer, F. Giannotti and M. Christen. FuturICT – The road towards ethical ICT. Eur. Phys. J. Special Topics 214, 153-181 (2012). http://dx.doi.org/10.1140/epjst/e2012-01691-2
- **[Howe 2006]** - J. Howe, 2006. The rise of crowdsourcing. Wired 14 (6).
- **[Jiang 2012]** - S. Jiang, J. Ferreira and M. C. Gonzalez, Clustering daily patterns of human activities in the city. Data Mining and Knowledge Discovery 25, 2012.
- **[Jianxi 2012]** - Gao Jianxi, Buldyrev Sergey V., Stanley H. Eugene, Havlin S, Networks formed from interdependent networks, Nature Physics 8, 40-48 (2012).
- **[Kashima 2009]** - H. Kashima, T. Kato, Yoshihiro Yamanishi, M. Sugiyama, and K. Tsuda, Link propagation: A fast semi-supervised learning algorithm for link prediction, In SIAM, 2009.
- **[Keeling 2005]** - M. J. Keeling, K. T. D. Eames, J. Royal, Networks and epidemic models, Soc. Interface, 2005.
- **[Keim 2010]** - DA Keim, J Kohlhammer, G Ellis, F Mansmann. Mastering the Information Age: Solving Problems with Visual Analytics, Eurographics Association. 2010

- **[Kempe 2013]** - D. Kempe, J. Kleinberg, and E. Tardos, Maximizing the spread of influence through a social network, in Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '03), ACM, New York, NY, USA, 137-146.
- **[Kleinberg 2000]** - J. Kleinberg, Navigation in a Small World, Nature 406 (2000), 845.
- **[Kumar 2006]** - Pang-Ning Tan, Michael Steinbach, Vipin Kumar. Introduction to Data Mining. Addison Wesley, 2006.
- **[Küngas 2010]** - P. Küngas and M. Dumas. Redundancy detection in service-oriented systems. In Proc. of WWW'2010, pp. 581-590
- **[Lamirel 2011]** - J.-C. Lamirel, R. Mall, S. Al Shehabi, G. Safi. A New Label Maximization Based Incremental Neural Clustering Approach: Application to Text Clustering. In Proc. of WSOM'2011, pp. 257-266
- **[Lazer 2009]** - D. Lazer, A. Pentland et al, Life in the network: the coming age of computational social science. Science 323, (2009) 721-723.
- **[Leskovec 2005]** - J. Leskovec, J. Kleinberg, and C. Faloutsos, Graphs over time: densification laws, shrinking diameters and possible explanations, in Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining (KDD '05), ACM, New York, NY, USA, 177-187, 2005.
- **[Leskovec 2010]** - J. Leskovec, D. Huttenlocher, and J. Kleinberg, Predicting positive and negative links in online social networks, In WWW, 2010
- **[Liben-Nowell 2003]** - D. Liben-Nowell, J. Kleinberg, The link prediction problem for social networks, In CIKM, 2003.
- **[Lucchese 2011]** - C. Lucchese, S. Orlando, R. Perego, F. Silvestri, and G. Tolomei. Identifying task-based sessions in search-engines query logs. WSDM 2011, 277-286, ACM.
- **[Mokarizadeh 2012]** - S. Mokarizadeh, P. Küngas, M. Matskin, M. Crasso, M. Campo, A. Zunino. Information Diffusion in Web Services Networks. In Proc. of ICWS 2012, pp. 488-495
- **[Monreale 2009]** - A. Monreale, F. Pinelli, R. Trasarti, F. Giannotti, WhereNext: a location predictor on trajectory pattern mining,Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2009, 637-646.
- **[Moussad 2011]** - M. Moussad, D. Helbing, and G. Theraulaz, How simple rules determine pedestrian behavior and crowd disasters. Proceedings of the National Academy of Sciences of the USA (PNAS) 108(17), (2011) 6884-6888.
- **[Mucha, 2010]** - P. J. Mucha, T. Richardson, K. Macon, M. A. Porter, and J.-P. Onnela, Community structure in time-dependent, multiscale, and multiplex networks, Science 328, 876, 2010.
- **[Newman 2010]** - M. E. J. Newman, Networks: An Introduction, Oxford University Press, 2010.
- **[Pang 2008]** - B. Pang and L. Lee, Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1/2), (2008) 1-135
- **[Papadakis 2012]** - Papadakis, G.; Ioannou, E.; Palpanas, T.; Niederée, C.; Nejdl, W.: A Blocking Framework for Entity Resolution in Highly Heterogeneous Information Spaces. Transactions on Knowledge and Data Engineering, 2012

- **[Parent 2013]** - C. Parent, S. Spaccapietra, C. Renso, G. Andrienko, N. Andrienko, V. Bogorny, M. Damiani, A. Gkoulalas-Divanis, J. Macedo, N. Pelekis, Y. Theodoridis, Z. Yan, Semantic Trajectories Modeling and Analysis, ACM Computing Surveys (2013).
- **[Pastor 2001]** - R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, Phys. Rev. Lett. 86, (2001).
- **[Pentland 2012]** - A.Pentland. Society's Nervous System: Building Effective Government, Energy, and Public Health Systems. IEEE Computer 45(1), 31-38 (2012).
- **[Song 2010]** - C. Song, T. Koren, P. Wang, A. L. Barabasi, Modelling the scaling properties of human mobility, Nature Physics, 2010
- **[Tang 2009]** - L. Tang and H. Liu, Relational learning via latent social dimensions, In KDD 2009.
- **[Trasarti 2011]** - R. Trasarti, F. Pinelli, M. Nanni, F. Giannotti, Mining mobility user profiles for car pooling,Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2011, 1190-1198.
- **[Tumminello 2010]** - M. Tumminello, F. Lillo, R. N. Mantegna, Correlation, hierarchies, and networks in financial markets. Journal of Economic Behavior and Organization 75, 40-58. (2010)
- **[Tumminello 2011]** - M. Tumminello, S. Miccichè, F. Lillo, J. Piilo, and R.N. Mantegna, Statistically validated networks in bipartite complex systems, PLoS ONE 6(3): e17994. doi:10.1371/journal.pone.0017994 (2011).
- **[Wang 2011]** - D. Wang, D. Pedreschi, C. Song, F. Giannotti, A. L. Barabasi, Human Mobility, Social Ties, and Link Prediction,Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2011, 1100-1108.
- **[Watts 1998]** - D. J. Watts and S. H. Strogatz, Collective dynamics of ‘small-world' networks, Nature 393, 440 (1998).
- **[Zimmermann 2002]** - A. Zimmermann, S. Schonfelder, G Rindsfuser, T. Haupt, Observing the rhythms of daily life: a six-week travel diary. Transportation (2002) 29(2):95-124.

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

ESTIMATED BUDGET FOR THE ACTION (page 1 of 2)

	Estimated eligible[1] costs (per budget category)									EU contribution			Additional information		
	A. Direct personnel costs				B. Direct costs of subcontracting	[C. Direct costs of fin. support]	D. Other direct costs	E. Indirect costs[2]	F. Special unit costs	Total costs	Reimbursement rate %	Maximum EU contribution[3]	Maximum grant amount[4]	Information for indirect costs	Information for auditors	Other information:
	A.1 Employees (or equivalent) A.2 Natural persons under direct contract A.3 Seconded persons [A.6 Personnel for providing access to research infrastructure]		A.4 SME owners without salary A.5 Beneficiaries that are natural persons without salary				D.1 Travel D.2 Equipment D.3 Other goods and services D.4 Costs of large research infrastructure		F. 1 "Costs for providing trans-national access to research infrastructure" **					Estimated costs of in-kind contributions not used on premises	Declaration of costs under Point D.4	Estimated costs of beneficiaries/ linked third parties not receiving EU funding
Form of costs[6]	Actual	Unit[7]	Unit[8]		Actual	Actual	Actual	Flat-rate[9]	Unit[11]							
								25%								
	(a)	Total (b)	No hours	Total (c)	(d)	(e)	(f)	(g)=0,25x ((a)+(b)+ (c)+(f) +[(h1)+(h2)]- (m))	Total (h1)	(i)= (a)+(b)+(c)+ (d)+(e)+(f)+ (g)+(h1)+(h2)+(h3)	(j)	(k)	(l)	(m)	Yes/No	
1. CNR	811000.00	0.00			0.00	0.00	57000.00	217000.00		1085000.00	100.00	1085000.00	1085000.00	0.00	No	
2. USFD	521900.00	0.00			0.00	0.00	267000.00	197225.00		986125.00	100.00	986125.00	986125.00	0.00	No	
3. UNIPI	229000.00	0.00			0.00	0.00	59000.00	72000.00		360000.00	100.00	360000.00	360000.00	0.00	No	
4. Fraunhofer	0.00	437000.00			0.00	0.00	56500.00	123375.00		616875.00	100.00	616875.00	616875.00	0.00	No	
5. UT	147000.00	0.00			0.00	0.00	35000.00	45500.00		227500.00	100.00	227500.00	227500.00	0.00	No	
6. IMT LUCCA	108000.00	0.00			0.00	0.00	32000.00	35000.00		175000.00	100.00	175000.00	175000.00	0.00	No	
7. LUH	267600.00	0.00			0.00	0.00	58000.00	81400.00		407000.00	100.00	407000.00	407000.00	0.00	No	
8. KCL	255441.00	0.00			0.00	0.00	46559.00	75500.00		377500.00	100.00	377500.00	377500.00	0.00	No	
9. SNS	108000.00	0.00			0.00	0.00	32000.00	35000.00		175000.00	100.00	175000.00	175000.00	0.00	No	
10. AALTO UNIVERSITY	0.00	288144.00			0.00	0.00	51856.00	85000.00		425000.00	100.00	425000.00	425000.00	0.00	No	
11. ETH Zurich[13]										917500.00						917500.00
12. TU Delft	116000.00	0.00			0.00	0.00	16000.00	33000.00		165000.00	100.00	165000.00	165000.00	0.00	No	
Total consortium	3149941.00	725144.00			0.00	0.00	858915.00	1183500.00		5917500.00		5000000.00	5000000.00	0.00		917500.00

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

ESTIMATED BUDGET FOR THE ACTION (page 2 of 2)

(1) See Article 6 for the eligibility conditions
(2) The indirect costs covered by the operating grant (received under any EU or Euratom funding programme; see Article 6.5.(b)) are ineligible under the GA. Therefore, a beneficiary that receives an operating grant during the action's duration cannot declare indirect costs for the year(s)/reporting period(s) covered by the operating grant (see Article 6.2.E).
(3) This is the theoretical amount of EU contribution that the system calculates automatically (by multiplying all the budgeted costs by the reimbursement rate). This theoretical amount is capped by the 'maximum grant amount' (that the Commission/Agency decided to grant for the action) (see Article 5.1).
(4) The 'maximum grant amount' is the maximum grant amount decided by the Commission/Agency. It normally corresponds to the requested grant, but may be lower.
(5) Depending on its type, this specific cost category will or will not cover indirect costs. Specific unit costs that include indirect costs are: costs for energy efficiency measures in buildings, access costs for providing trans-national access to research infrastructure and costs for clinical studies.
(6) See Article 5 for the forms of costs
(7) Unit : hours worked on the action; costs per unit (hourly rate) : calculated according to beneficiary's usual accounting practice
(8) See Annex 2a 'Additional information on the estimated budget' for the details (costs per hour (hourly rate)).
(9) Flat rate : 25% of eligible direct costs, from which are excluded: direct costs of subcontracting, costs of in-kind contributions not used on premises, direct costs of financial support, and unit costs declared under budget category F if they include indirect costs
(10) See Annex 2a 'Additional information on the estimated budget' for the details (units, costs per unit).
(11) See Annex 2a 'Additional information on the estimated budget' for the details (units, costs per unit, estimated number of units, etc)
(12) Only specific unit costs that do not include indirect costs
(13) See Article 9 for beneficiaries not receiving EU funding
(14) Only for linked third parties that receive EU funding

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

THE UNIVERSITY OF SHEFFIELD (USFD), RC000667, established in FIRTH COURT WESTERN BANK, SHEFFIELD S10 2TN, United Kingdom, GB648238808, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary ('2')

in Grant Agreement No 654024 ('the Agreement')

between CONSIGLIO NAZIONALE DELLE RICERCHE **and** *the European Union ('the EU'), represented by the European Commission ('the Commission'),*

for the action entitled 'SoBigData Research Infrastructure (SoBigData)'.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement the grant in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

Deborah LODGE with ECAS id nlodgdeb signed in the Participant Portal on 25/06/2015 at 09:24:44 (transaction id SigId-136-n41lvaG60gycXySRWcdFtzwwh3aSn1pPP6AXXby7WxzX7XwzTPTa1Qu VmgmoE1[V5m0VEuWeNlgbPUszYkTOa-PHsUMVSXYC#y5EW6BzsR-x0x/2cWszRt3G9loO1Vqu2JEDaY1ZH0CXSjKWv5uP). Timestamp by third party at Thu Jun 25 10:24:49 CEST 2015

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

UNIVERSITA DI PISA (UNIPI), 125096/CF80003670504, established in LUNGARNO PACINOTTI 43/44, PISA 56126, Italy, IT00286820501, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary (‘3’)

in Grant Agreement No 654024 (‘the Agreement’)

between CONSIGLIO NAZIONALE DELLE RICERCHE **and** *the European Union (‘the EU’), represented by the European Commission ('the Commission'),*

for the action entitled ‘SoBigData Research Infrastructure (SoBigData)’.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement the grant in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

Franco TURINI with ECAS id murifra signed in the Participant Portal on 25/05/2015 at 13:28:20 (transaction id SigId-526-1JZVRpqzQNyU+rj8zfHA7+SQzIe8Szz0Ni+3tDMAzwe7ST0B+uwPL77qOgPKFl+QLOL6VLmBZApZcPaomh=m-PHgUMVS0YCHy5EV6BzsR-hR7EtuZdWQDozYzML8SP5m4RmNWygnLAYS(hocbps0)
Timestamp by third party at
Thu Jun 25 14:28:32 CEST 2015

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

FRAUNHOFER GESELLSCHAFT ZUR FORDERUNG DER ANGEWANDTEN FORSCHUNG EV (Fraunhofer) EV, VR4461, established in HANSASTRASSE 27C, MUENCHEN 80686, Germany, DE129515865, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary (‘4’)

in Grant Agreement No 654024 (‘the Agreement’)

between CONSIGLIO NAZIONALE DELLE RICERCHE **and** *the European Union (‘the EU’), represented by the European Commission ('the Commission'),*

for the action entitled ‘SoBigData Research Infrastructure (SoBigData)’.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement the grant in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

Sabine MAYER with ECAS id nmayersa signed in the Participant Portal on 26/06/2015 at 16:12:04 (transaction id SigId-645-zWLzSasvKvyrHYQmPvhZvdFCuoDkgF7zdSO7rLjezGzzu1Gphjtkl-Tq9OrzzikGoaKLcPYpxWiszb4uj-9u7m-Jj71zxYb8yrQLoJPs9TM_-WJDuGq6VlcWWM3Ja43avFKdkoDo5ox2dEsqVyWLXs4)
Timestamp by third party at
Fri Jun 26 17:12:16 CEST 2015

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

TARTU ULIKOOL (UT), 74001073, established in ULIKOOLI 18, TARTU 50090, Estonia, EE100030417, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary (‘5’)

in Grant Agreement No 654024 (‘the Agreement’)

between CONSIGLIO NAZIONALE DELLE RICERCHE **and** *the European Union (‘the EU’), represented by the European Commission ('the Commission'),*

for the action entitled ‘SoBigData Research Infrastructure (SoBigData)’.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement the grant in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

Marco KIRM with ECAS id rikrmima signed in the Participant Portal on 25/06/2015 at 09:40:51 (transaction id SigId-156-YxUGn93ywqJUjxlrrLigNDcZywV8SPLQ1pa82h20EpeSzH42uyw9N23FswcfzehwcE5Liv3KV7IlIzYvbAb0-PHslUMVSXYCiHy5EV6BzsR-z4cIAMP1HR9941B1SZP5f7zwrA31N5X2G4qyXPEMmOm).
Timestamp by third party at
Thu Jun 25 10:40:56 CEST 2015

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

SCUOLA IMT (ISTITUZIONI, MERCATI, TECNOLOGIE) ALTI STUDI DI LUCCA (IMT LUCCA), CF92037570469, established in PIAZZA SAN PONZIANO 6, LUCCA 55100, Italy, IT02235840465, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary (‘6’)

in Grant Agreement No 654024 (‘the Agreement’)

between CONSIGLIO NAZIONALE DELLE RICERCHE **and** *the European Union (‘the EU’), represented by the European Commission ('the Commission'),*

for the action entitled ‘SoBigData Research Infrastructure (SoBigData)’.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement the grant in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

Alberto BEMPORAD with ECAS id nbermpaa signed in the Participant Portal on 29/05/2015 at 13:32:50 (transaction id: Sigid-536-ZbCHjABCfiEfcFBcm95EzNZzQuemOKzKCnT9zSb4Jw9w2aWKfDMcYx wmpyzbsCmxoKlCfDeaIgAw2P8GGzzWm-PHaiUMVSKYC2DedozTzwBeG-CVmcX0yzaQ9lA@vkNtHu1mSoTqq@Ezzu.NfP9kbJ). Timestamp by third party at Mon Jun 29 14:33:02 CEST 2015

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

GOTTFRIED WILHELM LEIBNIZ UNIVERSITAET HANNOVER (LUH), established in Welfengarten 1, HANNOVER 30167, Germany, DE811245527, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary (‘7’)

in Grant Agreement No 654024 (‘the Agreement’)

between CONSIGLIO NAZIONALE DELLE RICERCHE **and** *the European Union (‘the EU’), represented by the European Commission ('the Commission'),*

for the action entitled ‘SoBigData Research Infrastructure (SoBigData)’.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement the grant in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

Henning HOWIND with ECAS id nhowind signed in the Participant Portal on 13/07/2015 at 09:23:10 [transaction id Sig3c-281-Yd2zJdqpVPUxBH8GT-JINX5LzDqS2NOZHCE7zNWtc8aGALarGyDcZ7Z mNgnPzzrawLZMGBDWezNpWWu6U18RzgDD-Jj71zxYbByroOrPG7NJ0a-mlauvaW5YJO6rEPa1BK73DfNuC5vS1-Ly02LrzRWFQ]. Timestamp by third party at Mon Jul 13 10:23:23 CEST 2015

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

KING'S COLLEGE LONDON (KCL), RC000297, established in Strand, LONDON WC2R 2LS, United Kingdom, GB627403551, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary (‘8’)

in Grant Agreement No 654024 (‘the Agreement’)

between CONSIGLIO NAZIONALE DELLE RICERCHE **and** *the European Union (‘the EU’), represented by the European Commission ('the Commission'),*

for the action entitled ‘SoBigData Research Infrastructure (SoBigData)’.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement the grant in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

Daniel WALKER with ECAS id nwalkeda signed in the Participant Portal on 23/07/2015 at 09:53:04 (transaction id SigId-5313-fQjBch6wwzMa2SohOaUl6trPbzbt1oFQr91fLuqlMt4NCzGZOchrtsmEZSIDBMHEcxFS9Czp8e8Yelze6aFjazLm-Jj71zxYb8yroDrPG7Nu0a-55zquynRzJN5bzm+46LobIlqzMQzzhbd85CJgN6CrmYcG).
Timestamp by third party at
Thu Jul 23 10:53:19 CEST 2015

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

SCUOLA NORMALE SUPERIORE (SNS), CF80005050507, established in PIAZZA DEI CAVALIERI 7, PISA 56126, Italy, IT00420000507, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary (‘9’)

in Grant Agreement No 654024 (‘the Agreement’)

between CONSIGLIO NAZIONALE DELLE RICERCHE **and** *the European Union (‘the EU’), represented by the European Commission ('the Commission'),*

for the action entitled ‘SoBigData Research Infrastructure (SoBigData)’.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement the grant in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

Fabio BELTRAM with ECAS id nbeltfa signed in the Participant Portal on 25/06/2015 at 11:11:31 [transaction id SigId-294-JPGd9HDaTzT8mB6QmT101d8YHoTzSOCNGEINMzoAwc0ptrMeFxwYrzzo@encGZ8e=prXq4xzpRzJMUC8oC0-PHsIJMVSXYCHy5EW6BrsR-o2DehOon29trypzrVJT=R7vGZxBKzTYBMqgq3rxSPujJ). Timestamp by third party at Thu Jun 25 12:11:37 CEST 2015

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

AALTO-KORKEAKOULUSAATIO (AALTO UNIVERSITY) FI1, 22283574, established in OTAKAARI 1, ESPOO 02150, Finland, FI22283574, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary (‘10’)

in Grant Agreement No 654024 (‘the Agreement’)

between CONSIGLIO NAZIONALE DELLE RICERCHE **and** *the European Union (‘the EU’), represented by the European Commission ('the Commission'),*

for the action entitled ‘SoBigData Research Infrastructure (SoBigData)’.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement the grant in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

Tuija PULKKINEN with ECAS id npulktu signed in the Participant Portal on 26/06/2015 at 10:05:06 (transaction id SigId-256-ykw3dhgCEichUbSFpsdsOEM0SSxMDA2zTrPAL0axLztPLzJYhqEV6vJQEqd3mqWlMg1dQnWVbzTUC0eM2q4Nts-J!71zxYb8yrQLqJPe9TMu-g4yvCzgzY5qlBUPKpT3nIs2Ygfe!4JibBad1agorcCu5G!). Timestamp by third party at Fri Jun 26 11:05:20 CEST 2015

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

EIDGENOESSISCHE TECHNISCHE HOCHSCHULE ZUERICH (ETH Zurich), GESETZSR414110, established in Raemistrasse 101, ZUERICH 8092, Switzerland, CHE115203630MWST, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary (‘11’)

in Grant Agreement No 654024 (‘the Agreement’)

between CONSIGLIO NAZIONALE DELLE RICERCHE **and** *the European Union (‘the EU’), represented by the European Commission ('the Commission'),*

for the action entitled ‘SoBigData Research Infrastructure (SoBigData)’.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement the grant in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

Keller AGATHA with ECAS id nagathke signed in the Participant Portal on 10/07/2015 at 09:45:36 (transaction id SigId-161-XgSYITRGHwq4DRM75x6FBRcc05Dz=Ds,i8e7Yn5g7pR!DCCXYFRWYXzRjw6HAxjmvL4xC1zKzd0J9VCCpm-PHslUMVSXYO13xcLZM8U,W-XvwdcPR!6rpa1zWPHM0T1,i02tw6sb6W6zxe82,HOmy0). Timestamp by third party at Fri Jul 10 19:45:48 CEST 2015

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

ANNEX 3

ACCESSION FORM FOR BENEFICIARIES

TECHNISCHE UNIVERSITEIT DELFT (TU Delft), 27364265, established in STEVINWEG 1, DELFT 2628 CN, Netherlands, NL001569569B01, ('the beneficiary'), represented for the purpose of signing this Accession Form by the undersigned,

hereby agrees

to become beneficiary (‘12’)

in Grant Agreement No 654024 (‘the Agreement’)

between CONSIGLIO NAZIONALE DELLE RICERCHE **and** *the European Union (‘the EU’), represented by the European Commission ('the Commission'),*

for the action entitled ‘SoBigData Research Infrastructure (SoBigData)’.

and mandates

the coordinator to submit and sign in its name and on its behalf any **amendments** to the Agreement, in accordance with Article 55.

By signing this Accession Form, the beneficiary accepts the grant and agrees to implement the grant in accordance with the Agreement, with all the obligations and conditions it sets out.

SIGNATURE

For the beneficiary

Jeroen VAN DEN HOVEN with ECAS id rhovenje signed in the Participant Portal on 26/06/2015 at 09:41:12 [transaction id SigId-174-CF0_QT0xlYCoJmWxpjQHgpx1o17Heqv6H3EUMExMI7PYBp7yzS5pOxgM=8Nzh1AWNFEIEGtPVc9JtZVlw4x-Jj71zaYb0yrrQLgJPe9TMu-JJDeLjqgyFGH5=cWsHxf3I7VBs4qCVZt1hp_WWJ8]. Timestamp by third party at Fri Jun 26 09:41:25 CEST 2015

ⓘ print format A4 landscape

MODEL ANNEX 4 FOR H2020 GENERAL MGA — MULTI

FINANCIAL STATEMENT FOR *[BENEFICIARY [name]/ LINKED THIRD PARTY [name]]* FOR REPORTING PERIOD [reporting period]

	Eligible[1] costs (per budget category)												Receipts	EU contribution			Additional information
	A. Direct personnel costs				B. Direct costs of subcontracting	*[C. Direct costs of fin. support]*	D. Other direct costs		E. Indirect costs[2]	*[F. Costs of …]*		Total costs	Receipts	Reimbursement rate %	Maximum EU contribution[3]	Requested EU contribution	Information for indirect costs :
	A.1 Employees (or equivalent) A.2 Natural persons under direct contract A.3 Seconded persons *[A.6 Personnel for providing access to research infrastructure]*		A.4 SME owners without salary A.5 Beneficiaries that are natural persons without salary				D.1 Travel D.2 Equipment D.3 Other goods and services	*[D.4 Costs of large research infrastructure]*		*[F.1 Costs of …]*			Receipts of the action, to be reported in the last reporting period, according to Article 5.3.3				Costs of in-kind contributions not used on premises
Form of costs[4]	Actual	Unit	Unit		Actual	Actual	Actual	Actual	Flat-rate[5] 25%	Unit	Unit						
	a	Total b	No hours	Total c	d	*[e]*	f	*[g]*	h=0,25 x (a+b+ c+f+*[g]* + *[i1]*[6] +*[i2]*[6] - o)	No units	Total *[i1]*	Total *[i2]*	j = a+b+c+d+*[e]* +f +*[g]* +h+*[i1]* +*[i2]*	k	l	m	n	o
[short name beneficiary/linked third party]																	

The beneficiary/linked third party hereby confirms that:
The information provided is complete, reliable and true.
The costs declared are eligible (see Article 6).
The costs can be substantiated by adequate records and supporting documentation that will be produced upon request or in the context of checks, reviews, audits and investigations (see Articles 17, 18 and 22).
For the last reporting period: that all the receipts have been declared (see Article 5.3.3).

ⓘ Please declare all eligible costs, even if they exceed the amounts indicated in the estimated budget (see Annex 2). Only amounts that were declared in your individual financial statements can be taken into account lateron, in order to replace other costs that are found to be ineligible.

[1] See Article 6 for the eligibility conditions

[2] The indirect costs claimed must be free of any amounts covered by an operating grant (received under any EU or Euratom funding programme; see Article 6.2.E). If you have received an operating grant during this reporting period, you cannot claim any indirect costs.

[3] This is the *theoretical* amount of EU contribution that the system calculates automatically (by multiplying the reimbursement rate by the total costs declared). The amount you request (in the column 'requested EU contribution') may have to be less (e.g. if you and the other beneficiaries are above budget, if the 90% limit (see Article 21) is reached, etc).

[4] See Article 5 for the form of costs

[5] Flat rate : 25% of eligible direct costs, from which are excluded: direct costs of subcontracting, costs of in-kind contributions not used on premises, direct costs of financial support, and unit costs declared under budget category F if they include indirect costs (see Article 6.2.E)

[6] Only specific unit costs that do not include indirect costs

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

ANNEX 5

MODEL FOR THE CERTIFICATE ON THE FINANCIAL STATEMENTS

- **For options [*in italics in square brackets*]: choose the applicable option. Options not chosen should be deleted.**
- **For fields in [grey in square brackets]: enter the appropriate data**

TABLE OF CONTENTS

TERMS OF REFERENCE FOR AN INDEPENDENT REPORT OF FACTUAL FINDINGS ON COSTS DECLARED UNDER A GRANT AGREEMENT FINANCED UNDER THE HORIZON 2020 RESEARCH FRAMEWORK PROGRAMME..... 2

INDEPENDENT REPORT OF FACTUAL FINDINGS ON COSTS DECLARED UNDER A GRANT AGREEMENT FINANCED UNDER THE HORIZON 2020 RESEARCH FRAMEWORK PROGRAMME 7

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

Terms of Reference for an Independent Report of Factual Findings on costs declared under a Grant Agreement financed under the Horizon 2020 Research and Innovation Framework Programme

This document sets out the '**Terms of Reference (ToR)**' under which

[OPTION 1: [insert name of the beneficiary] ('the Beneficiary')] [OPTION 2: [insert name of the linked third party] ('the Linked Third Party'), third party linked to the Beneficiary [insert name of the beneficiary] ('the Beneficiary')]

agrees to engage

[**insert legal name of the auditor**] ('the Auditor')

to produce an independent report of factual findings ('the Report') concerning the Financial Statement(s)[1] drawn up by the *[Beneficiary] [Linked Third Party]* for the Horizon 2020 grant agreement [insert number of the grant agreement, title of the action, acronym and duration from/to] ('the Agreement'), and

to issue a Certificate on the Financial Statements' ('CFS') referred to in Article 20.4 of the Agreement based on the compulsory reporting template stipulated by the Commission.

The Agreement has been concluded under the Horizon 2020 Research and Innovation Framework Programme (H2020) between the Beneficiary and *[OPTION 1: the European Union, represented by the European Commission ('the Commission')][OPTION 2: the European Atomic Energy Community (Euratom,) represented by the European Commission ('the Commission')][OPTION 3: the [Research Executive Agency (REA)] [European Research Council Executive Agency (ERCEA)] [Innovation and Networks Executive Agency (INEA)] [Executive Agency for Small and Medium-sized Enterprises (EASME)] ('the Agency'), under the powers delegated by the European Commission ('the Commission').]*

[1] By which costs under the Agreement are declared (see template 'Model Financial Statements' in Annex 4 to the Grant Agreement).

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

The *[Commission] [Agency]* is mentioned as a signatory of the Agreement with the Beneficiary only. The *[European Union][Euratom][Agency]* is not a party to this engagement.

1.1 Subject of the engagement

The coordinator must submit to the *[Commission][Agency]* the final report within 60 days following the end of the last reporting period which should include, amongst other documents, a CFS for each beneficiary and for each linked third party that requests a total contribution of EUR 325 000 or more, as reimbursement of actual costs and unit costs calculated on the basis of its usual cost accounting practices (see Article 20.4 of the Agreement). The CFS must cover all reporting periods of the beneficiary or linked third party indicated above.

The Beneficiary must submit to the coordinator the CFS for itself and for its linked third party(ies), if the CFS must be included in the final report according to Article 20.4 of the Agreement..

The CFS is composed of two separate documents:

- The Terms of Reference ('the ToR') to be signed by the *[Beneficiary] [Linked Third Party]* and the Auditor;
- The Auditor's Independent Report of Factual Findings ('the Report') to be issued on the Auditor's letterhead, dated, stamped and signed by the Auditor (or the competent public officer) which includes the agreed-upon procedures ('the Procedures') to be performed by the Auditor, and the standard factual findings ('the Findings') to be confirmed by the Auditor.

If the CFS must be included in the final report according to Article 20.4 of the Agreement, the request for payment of the balance relating to the Agreement cannot be made without the CFS. However, the payment for reimbursement of costs covered by the CFS does not preclude the *[Commission,][Agency,]* the European Anti-Fraud Office and the European Court of Auditors from carrying out checks, reviews, audits and investigations in accordance with Article 22 of the Agreement.

1.2 Responsibilities

The *[Beneficiary] [Linked Third Party]:*

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

- must draw up the Financial Statement(s) for the action financed by the Agreement in compliance with the obligations under the Agreement. The Financial Statement(s) must be drawn up according to the *[Beneficiary's] [Linked Third Party's]* accounting and book-keeping system and the underlying accounts and records;
- must send the Financial Statement(s) to the Auditor;
- is responsible and liable for the accuracy of the Financial Statement(s);
- is responsible for the completeness and accuracy of the information provided to enable the Auditor to carry out the Procedures. It must provide the Auditor with a written representation letter supporting these statements. The written representation letter must state the period covered by the statements and must be dated;
- accepts that the Auditor cannot carry out the Procedures unless it is given full access to the *[Beneficiary's] [Linked Third Party's]* staff and accounting as well as any other relevant records and documentation.

The Auditor:

- *[Option 1 by default:* is qualified to carry out statutory audits of accounting documents in accordance with Directive 2006/43/EC of the European Parliament and of the Council of 17 May 2006 on statutory audits of annual accounts and consolidated accounts, amending Council Directives 78/660/EEC and 83/349/EEC and repealing Council Directive 84/253/EEC or similar national regulations*]*.
- *[Option 2 if the Beneficiary or Linked Third Party has an independent Public Officer:* is a competent and independent Public Officer for which the relevant national authorities have established the legal capacity to audit the Beneficiary*].*
- *[Option 3 if the Beneficiary or Linked Third Party is an international organisation:* is an [internal] [external] auditor in accordance with the internal financial regulations and procedures of the international organisation*].*

The Auditor:

- must be independent from the Beneficiary *[and the Linked Third Party]*, in particular, it must not have been involved in preparing the *[Beneficiary's] [Linked Third Party's]* Financial Statement(s);
- must plan work so that the Procedures may be carried out and the Findings may be assessed;
- must adhere to the Procedures laid down and the compulsory report format;
- must carry out the engagement in accordance with this ToR;
- must document matters which are important to support the Report;
- must base its Report on the evidence gathered;
- must submit the Report to the *[Beneficiary] [Linked Third Party]*.

The Commission sets out the Procedures to be carried out by the Auditor. The Auditor is not responsible for their suitability or pertinence. As this engagement is not an assurance engagement, the Auditor does not provide an audit opinion or a statement of assurance.

1.3 Applicable Standards

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

The Auditor must comply with these Terms of Reference and with[2]:

- the International Standard on Related Services ('ISRS') 4400 *Engagements to perform Agreed-upon Procedures regarding Financial Information* as issued by the International Auditing and Assurance Standards Board (IAASB);
- the *Code of Ethics for Professional Accountants* issued by the International Ethics Standards Board for Accountants (IESBA). Although ISRS 4400 states that independence is not a requirement for engagements to carry out agreed-upon procedures, the *[Commission][Agency]* requires that the Auditor also complies with the Code's independence requirements.

The Auditor's Report must state that there is no conflict of interests in establishing this Report between the Auditor and the Beneficiary *[and the Linked Third Party]*, and must specify - if the service is invoiced - the total fee paid to the Auditor for providing the Report.

1.4 Reporting

The Report must be written in the language of the Agreement (see Article 20.7).

Under Article 22 of the Agreement, the [*Commission*] [*Agency*], the European Anti-Fraud Office and the Court of Auditors have the right to audit any work that is carried out under the action and for which costs are declared from [*the European Union*] [*Euratom*] budget. This includes work related to this engagement. The Auditor must provide access to all working papers (e.g. recalculation of hourly rates, verification of the time declared for the action) related to this assignment if the *[Commission] [Agency]*, the European Anti-Fraud Office or the European Court of Auditors requests them.

1.5 Timing

The Report must be provided by *[*dd Month yyyy*]*.

[2] Supreme Audit Institutions applying INTOSAI-standards may carry out the Procedures according to the corresponding International Standards of Supreme Audit Institutions and code of ethics issued by INTOSAI instead of the International Standard on Related Services ('ISRS') 4400 and the Code of Ethics for Professional Accountants issued by the IAASB and the IESBA.

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

1.6 Other terms

[*The [Beneficiary] [Linked Third Party] and the Auditor can use this section to agree other specific terms, such as the Auditor's fees, liability, applicable law, etc. Those specific terms must not contradict the terms specified above.*]

[legal name of the Auditor]	[legal name of the *[Beneficiary][Linked Third Party]*]
[name & function of authorised representative]	[name & function of authorised representative]
[dd Month yyyy]	[dd Month yyyy]
Signature of the Auditor	Signature of the *[Beneficiary][Linked Third Party]*

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

Independent Report of Factual Findings on costs declared under Horizon 2020 Research and Innovation Framework Programme

(To be printed on the Auditor's letterhead)

To

[name of contact person(s)], [Position]

[[*Beneficiary's*] [*Linked Third Party's*] name]

[Address]

[dd Month yyyy]

Dear [Name of contact person(s)],

As agreed under the terms of reference dated [dd Month yyyy]

with *[OPTION 1: [insert name of the beneficiary] ('the Beneficiary')] [OPTION 2: [insert name of the linked third party] ('the Linked Third Party'), third party linked to the Beneficiary [insert name of the beneficiary] ('the Beneficiary')],*

we

[name of the auditor] ('the Auditor'),

established at

[full address/city/state/province/country],

represented by

[name and function of an authorised representative],

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

have carried out the procedures agreed with you regarding the costs declared in the Financial Statement(s)[3] of the *[Beneficiary] [Linked Third Party]* concerning the grant agreement

[insert grant agreement reference: number, title of the action and acronym] ('the Agreement'),

with a total cost declared of

[total amount] EUR,

and a total of actual costs and 'direct personnel costs declared as unit costs calculated in accordance with the [*Beneficiary's*] [*Linked Third Party's*] usual cost accounting practices' declared of

[sum of total actual costs and total direct personnel costs declared as unit costs calculated in accordance with the *[Beneficiary's] [Linked Third Party's]* usual cost accounting practices] EUR

and **hereby provide our Independent Report of Factual Findings ('the Report')** using the compulsory report format agreed with you.

The Report

Our engagement was carried out in accordance with the terms of reference ('the ToR') appended to this Report. The Report includes the agreed-upon procedures ('the Procedures') carried out and the standard factual findings ('the Findings') examined.

The Procedures were carried out solely to assist the [*Commission*] [*Agency*] in evaluating whether the [*Beneficiary's*] [*Linked Third Party's*] costs in the accompanying Financial Statement(s) were declared in accordance with the Agreement. The *[Commission] [Agency]* draws its own conclusions from the Report and any additional information it may require.

[3] By which the Beneficiary declares costs under the Agreement (see template 'Model Financial Statement' in Annex 4 to the Agreement).

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

The scope of the Procedures was defined by the Commission. Therefore, the Auditor is not responsible for their suitability or pertinence. Since the Procedures carried out constitute neither an audit nor a review made in accordance with International Standards on Auditing or International Standards on Review Engagements, the Auditor does not give a statement of assurance on the Financial Statements.

Had the Auditor carried out additional procedures or an audit of the *[Beneficiary's] [Linked Third Party's]* Financial Statements in accordance with International Standards on Auditing or International Standards on Review Engagements, other matters might have come to its attention and would have been included in the Report.

Not applicable Findings

We examined the Financial Statement(s) stated above and considered the following Findings not applicable:

Explanation (to be removed from the Report):

*If a Finding was not applicable, it must be marked as '**N.A.**' ('Not applicable') in the corresponding row on the right-hand column of the table and means that the Finding did not have to be corroborated by the Auditor and the related Procedure(s) did not have to be carried out.*

The reasons of the non-application of a certain Finding must be obvious i.e.

i) if no cost was declared under a certain category then the related Finding(s) and Procedure(s) are not applicable;

ii) if the condition set to apply certain Procedure(s) are not met the related Finding(s) and those Procedure(s) are not applicable. For instance, for 'beneficiaries with accounts established in a currency other than euro' the Procedure and Finding related to 'beneficiaries with accounts established in euro' are not applicable. Similarly, if no additional remuneration is paid, the related Finding(s) and Procedure(s) for additional remuneration are not applicable.

List here all Findings considered not applicable for the present engagement and explain the reasons of the non-applicability.

….

Exceptions

Apart from the exceptions listed below, the *[Beneficiary] [Linked Third Party]* provided the Auditor all the documentation and accounting information needed by the Auditor to carry out the requested Procedures and evaluate the Findings.

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

Explanation (to be removed from the Report):

- *If the Auditor was not able to successfully complete a procedure requested, it must be marked as '**E**' ('Exception') in the corresponding row on the right-hand column of the table. The reason such as the inability to reconcile key information or the unavailability of data that prevents the Auditor from carrying out the Procedure must be indicated below.*
- *If the Auditor cannot corroborate a standard finding after having carried out the corresponding procedure, it must also be marked as '**E**' ('Exception') and, where possible, the reasons why the Finding was not fulfilled and its possible impact must be explained here below.*

List here any exceptions and add any information on the cause and possible consequences of each exception, if known. If the exception is quantifiable, include the corresponding amount.

….

Example (to be removed from the Report):

1. *The Beneficiary was unable to substantiate the Finding number 1 on … because ….*
2. *Finding number 30 was not fulfilled because the methodology used by the Beneficiary to calculate unit costs was different from the one approved by the Commission. The differences were as follows: …*
3. *After carrying out the agreed procedures to confirm the Finding number 31, the Auditor found a difference of _____ EUR. The difference can be explained by …*

Further Remarks

In addition to reporting on the results of the specific procedures carried out, the Auditor would like to make the following general remarks:

Example (to be removed from the Report):

1. *Regarding Finding number 8 the conditions for additional remuneration were considered as fulfilled because …*
2. *In order to be able to confirm the Finding number 15 we carried out the following additional procedures: ….*

Use of this Report

This Report may be used only for the purpose described in the above objective. It was prepared solely for the confidential use of the *[Beneficiary] [Linked Third Party]* and the [*Commission*] [*Agency*], and only to be submitted to the [*Commission*] [*Agency*] in connection with the requirements set out in Article 20.4 of the Agreement. The Report may not be used by the *[Beneficiary] [Linked Third Party]* or by the [*Commission*] [*Agency*] for any other purpose, nor may it

Associated with document Ref. Ares(2015)2187417 - 26/05/2015

be distributed to any other parties. The [*Commission*] [*Agency*] may only disclose the Report to authorised parties, in particular to the European Anti-Fraud Office (OLAF) and the European Court of Auditors.

This Report relates only to the Financial Statement(s) submitted to the *[Commission] [Agency]* by the *[Beneficiary] [Linked Third Party]* for the Agreement. Therefore, it does not extend to any other of the [*Beneficiary's*] [*Linked Third Party's*] Financial Statement(s).

There was no conflict of interest[4] between the Auditor and the Beneficiary *[and Linked Third Party]* in establishing this Report. The total fee paid to the Auditor for providing the Report was EUR _____ (including EUR_____ of deductible VAT).

We look forward to discussing our Report with you and would be pleased to provide any further information or assistance.

[legal name of the Auditor]

[name and function of an authorised representative]

[dd Month yyyy]

Signature of the Auditor

[4] A conflict of interest arises when the Auditor's objectivity to establish the certificate is compromised in fact or in appearance when the Auditor for instance:
- was involved in the preparation of the Financial Statements;
- stands to benefit directly should the certificate be accepted;
- has a close relationship with any person representing the beneficiary;
- is a director, trustee or partner of the beneficiary; or
- is in any other situation that compromises his or her independence or ability to establish the certificate impartially.

Agreed-upon procedures to be performed and standard factual findings to be confirmed by the Auditor

The European Commission reserves the right to i) provide the auditor with additional guidance regarding the procedures to be followed or the facts to be ascertained and the way in which to present them (this may include sample coverage and findings) or to ii) change the procedures, by notifying the Beneficiary in writing. The procedures carried out by the auditor to confirm the standard factual finding are listed in the table below.

If this certificate relates to a Linked Third Party, any reference here below to 'the Beneficiary' is to be considered as a reference to 'the Linked Third Party'.

The 'result' column has three different options: 'C', 'E' and 'N.A.':

- 'C' stands for 'confirmed' and means that the auditor can confirm the 'standard factual finding' and, therefore, there is no exception to be reported.
- 'E' stands for 'exception' and means that the Auditor carried out the procedures but cannot confirm the 'standard factual finding', or that the Auditor was not able to carry out a specific procedure (e.g. because it was impossible to reconcile key information or data were unavailable),
- 'N.A.' stands for 'not applicable' and means that the Finding did not have to be examined by the Auditor and the related Procedure(s) did not have to be carried out. The reasons of the non-application of a certain Finding must be obvious i.e. i) if no cost was declared under a certain category then the related Finding(s) and Procedure(s) are not applicable; ii) if the condition set to apply certain Procedure(s) are not met then the related Finding(s) and Procedure(s) are not applicable. For instance, for 'beneficiaries with accounts established in a currency other than the euro' the Procedure related to 'beneficiaries with accounts established in euro' is not applicable. Similarly, if no additional remuneration is paid, the related Finding(s) and Procedure(s) for additional remuneration are not applicable.

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
A	**ACTUAL PERSONNEL COSTS AND UNIT COSTS CALCULATED BY THE BENEFICIARY IN ACCORDANCE WITH ITS USUAL COST ACCOUNTING PRACTICE**		

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	The Auditor draws a sample of persons whose costs were declared in the Financial Statement(s) to carry out the procedures indicated in the consecutive points of this section A. (*The sample should be selected randomly so that it is representative. Full coverage is required if there are fewer than 10 people (including employees, natural persons working under a direct contract and personnel seconded by a third party), otherwise the sample should have a minimum of 10 people, or 10% of the total, whichever number is the highest)* The Auditor sampled _____ people out of the total of _____ people.		
A.1	**PERSONNEL COSTS** <u>For the persons included in the sample and working under an employment contract or equivalent act (general procedures for individual actual personnel costs and personnel costs declared as unit costs)</u> To confirm standard factual findings 1-5 listed in the next column, the Auditor reviewed following information/documents provided by the Beneficiary: ○ a list of the persons included in the sample indicating the period(s) during which they worked for the action, their position (classification or category) and type of contract; ○ the payslips of the employees included in the sample; ○ reconciliation of the personnel costs declared in the Financial Statement(s) with the accounting system (project accounting and general ledger) and payroll system; ○ information concerning the employment status and employment conditions of personnel included in the sample, in particular their employment contracts or equivalent;	1) The employees were i) directly hired by the Beneficiary in accordance with its national legislation, ii) under the Beneficiary's sole technical supervision and responsibility and iii) remunerated in accordance with the Beneficiary's usual practices.	
		2) Personnel costs were recorded in the Beneficiary's accounts/payroll system.	
		3) Costs were adequately supported and reconciled with the accounts and payroll	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	○ the Beneficiary's usual policy regarding payroll matters (e.g. salary policy, overtime policy, variable pay); ○ applicable national law on taxes, labour and social security and ○ any other document that supports the personnel costs declared. The Auditor also verified the eligibility of all components of the retribution (see Article 6 GA) and recalculated the personnel costs for employees included in the sample.	records.	
		4) Personnel costs did not contain any ineligible elements.	
		5) There were no discrepancies between the personnel costs charged to the action and the costs recalculated by the Auditor.	
	Further procedures if 'additional remuneration' is paid To confirm standard factual findings 6-9 listed in the next column, the Auditor: ○ reviewed relevant documents provided by the Beneficiary (legal form, legal/statutory obligations, the Beneficiary's usual policy on additional remuneration, criteria used for its calculation…); ○ recalculated the amount of additional remuneration eligible for the action based on the supporting documents received (full-time or part-time work, exclusive or non-exclusive dedication to the action, etc.) to arrive at the applicable FTE/year and pro-rata rate (see data collected in the course of carrying out the procedures under A.2 'Productive hours' and A.4 'Time recording system').	6) The Beneficiary paying "additional remuneration" was a non-profit legal entity.	
		7) The amount of additional remuneration paid corresponded to the Beneficiary's usual remuneration practices and was consistently paid whenever the same kind of work or expertise was required.	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	IF ANY PART OF THE REMUNERATION PAID TO THE EMPLOYEE IS NOT MANDATORY ACCORDING TO THE NATIONAL LAW OR THE EMPLOYMENT CONTRACT ("ADDITIONAL REMUNERATION") AND IS ELIGIBLE UNDER THE PROVISIONS OF ARTICLE 6.2.A.1, THIS CAN BE CHARGED AS ELIGIBLE COST TO THE ACTION UP TO THE FOLLOWING AMOUNT: *(A) IF THE PERSON WORKS FULL TIME AND EXCLUSIVELY ON THE ACTION DURING THE FULL YEAR: UP TO EUR 8 000/YEAR;* *(B) IF THE PERSON WORKS EXCLUSIVELY ON THE ACTION BUT NOT FULL-TIME OR NOT FOR THE FULL YEAR: UP TO THE CORRESPONDING PRO-RATA AMOUNT OF EUR 8 000, OR* *(C) IF THE PERSON DOES NOT WORK EXCLUSIVELY ON THE ACTION: UP TO A PRO-RATA AMOUNT CALCULATED IN ACCORDANCE TO ARTICLE 6.2.A.1.*	8) The criteria used to calculate the additional remuneration were objective and generally applied by the Beneficiary regardless of the source of funding used.	
		9) The amount of additional remuneration included in the personnel costs charged to the action was capped at EUR 8,000 per FTE/year (up to the equivalent pro-rata amount if the person did not work on the action full-time during the year or did not work exclusively on the action).	
	Additional procedures in case "unit costs calculated by the Beneficiary in accordance with its usual cost accounting practices" is applied: Apart from carrying out the procedures indicated above to confirm standard factual findings 1-5 and, if applicable, also 6-9, the Auditor carried out following procedures to confirm standard factual findings 10-13 listed in the next column:	10) The personnel costs included in the Financial Statement were calculated in accordance with the Beneficiary's usual cost accounting practice. This methodology was consistently used in all H2020 actions.	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	○ obtained a description of the Beneficiary's usual cost accounting practice to calculate unit costs;. ○ reviewed whether the Beneficiary's usual cost accounting practice was applied for the Financial Statements subject of the present CFS; ○ verified the employees included in the sample were charged under the correct category (in accordance with the criteria used by the Beneficiary to establish personnel categories) by reviewing the contract/HR-record or analytical accounting records; ○ verified that there is no difference between the total amount of personnel costs used in calculating the cost per unit and the total amount of personnel costs recorded in the statutory accounts; ○ verified whether actual personnel costs were adjusted on the basis of budgeted or estimated elements and, if so, verified whether those elements used are actually relevant for the calculation, objective and supported by documents.	11) The employees were charged under the correct category.	
		12) Total personnel costs used in calculating the unit costs were consistent with the expenses recorded in the statutory accounts.	
		13) Any estimated or budgeted element used by the Beneficiary in its unit-cost calculation were relevant for calculating personnel costs and corresponded to objective and verifiable information.	
	<u>For natural persons included in the sample and working with the Beneficiary under a direct contract other than an employment contract, such as consultants (no subcontractors).</u> To confirm standard factual findings 14-18 listed in the next column the Auditor reviewed following information/documents provided by the Beneficiary: ○ the contracts, especially the cost, contract duration, work description, place of work, ownership of the results and reporting obligations to the Beneficiary;	14) The natural persons reported to the Beneficiary (worked under the Beneficiary's instructions).	
		15) They worked on the Beneficiary's premises (unless otherwise agreed with the Beneficiary).	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	○ the employment conditions of staff in the same category to compare costs and; ○ any other document that supports the costs declared and its registration (e.g. invoices, accounting records, etc.).	16) The results of work carried out belong to the Beneficiary.	
		17) Their costs were not significantly different from those for staff who performed similar tasks under an employment contract with the Beneficiary.	
		18) The costs were supported by audit evidence and registered in the accounts.	
	<u>For personnel seconded by a third party and included in the sample (not subcontractors)</u> To confirm standard factual findings 19-22 listed in the next column, the Auditor reviewed following information/documents provided by the Beneficiary: ○ their secondment contract(s) notably regarding costs, duration, work description, place of work and ownership of the results; ○ if there is reimbursement by the Beneficiary to the third party for the resource made available (in-kind contribution against payment): any documentation that supports the costs declared (e.g. contract, invoice, bank payment, and proof of registration in its accounting/payroll, etc.) and reconciliation of the Financial Statement(s) with the accounting system (project accounting and general ledger) as well as any proof that the amount invoiced by the third party did not include any profit;	19) Seconded personnel reported to the Beneficiary and worked on the Beneficiary's premises (unless otherwise agreed with the Beneficiary).	
		20) The results of work carried out belong to the Beneficiary.	
		If personnel is seconded against payment: 21) The costs declared were supported with documentation and recorded in the	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	○ if there is no reimbursement by the Beneficiary to the third party for the resource made available (in-kind contribution free of charge): a proof of the actual cost borne by the Third Party for the resource made available free of charge to the Beneficiary such as a statement of costs incurred by the Third Party and proof of the registration in the Third Party's accounting/payroll; ○ any other document that supports the costs declared (e.g. invoices, etc.).	Beneficiary's accounts. The third party did not include any profit.	
		If personnel is seconded free of charge: 22) The costs declared did not exceed the third party's cost as recorded in the accounts of the third party and were supported with documentation.	
A.2	**PRODUCTIVE HOURS** To confirm standard factual findings 23-28 listed in the next column, the Auditor reviewed relevant documents, especially national legislation, labour agreements and contracts and time records of the persons included in the sample, to verify that: ○ the annual productive hours applied were calculated in accordance with one of the methods described below, ○ the full-time equivalent (FTEs) ratios for employees not working full-time were correctly calculated.	23) The Beneficiary applied method [*choose one option and delete the others*] [**A**: 1720 hours] [**B**: the 'total number of hours worked'] [**C**: 'annual productive hours' used correspond to usual accounting practices]	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	If the Beneficiary applied method B, the auditor verified that the correctness in which the total number of hours worked was calculated and that the contracts specified the annual workable hours. If the Beneficiary applied method C, the auditor verified that the 'annual productive hours' applied when calculating the hourly rate were equivalent to at least 90 % of the 'standard annual workable hours'. The Auditor can only do this if the calculation of the standard annual workable hours can be supported by records, such as national legislation, labour agreements, and contracts. *BENEFICIARY'S PRODUCTIVE HOURS' FOR PERSONS WORKING FULL TIME SHALL BE ONE OF THE FOLLOWING METHODS:* ***A.*** *1720 ANNUAL PRODUCTIVE HOURS (PRO-RATA FOR PERSONS NOT WORKING FULL-TIME)* ***B.*** *THE TOTAL NUMBER OF HOURS WORKED BY THE PERSON FOR THE BENEFICIARY IN THE YEAR (THIS METHOD IS ALSO REFERRED TO AS 'TOTAL NUMBER OF HOURS WORKED' IN THE NEXT COLUMN). THE CALCULATION OF THE TOTAL NUMBER OF HOURS WORKED WAS DONE AS FOLLOWS: ANNUAL WORKABLE HOURS OF THE PERSON ACCORDING TO THE EMPLOYMENT CONTRACT, APPLICABLE LABOUR AGREEMENT OR NATIONAL LAW PLUS OVERTIME WORKED MINUS ABSENCES (SUCH AS SICK LEAVE OR SPECIAL LEAVE).*	24) Productive hours were calculated annually.	
		25) For employees not working full-time the full-time equivalent (FTE) ratio was correctly applied.	
		If the Beneficiary applied method B. 26) The calculation of the number of 'annual workable hours', overtime and absences was verifiable based on the documents provided by the Beneficiary.	
		If the Beneficiary applied method C. 27) The calculation of the number of 'standard annual workable hours' was verifiable based on the documents provided by the Beneficiary.	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	C. THE STANDARD NUMBER OF ANNUAL HOURS GENERALLY APPLIED BY THE BENEFICIARY FOR ITS PERSONNEL IN ACCORDANCE WITH ITS USUAL COST ACCOUNTING PRACTICES (THIS METHOD IS ALSO REFERRED TO AS 'TOTAL ANNUAL PRODUCTIVE HOURS' IN THE NEXT COLUMN). THIS NUMBER MUST BE AT LEAST 90% OF THE STANDARD ANNUAL WORKABLE HOURS. *'ANNUAL WORKABLE HOURS' MEANS THE PERIOD DURING WHICH THE PERSONNEL MUST BE WORKING, AT THE EMPLOYER'S DISPOSAL AND CARRYING OUT HIS/HER ACTIVITY OR DUTIES UNDER THE EMPLOYMENT CONTRACT, APPLICABLE COLLECTIVE LABOUR AGREEMENT OR NATIONAL WORKING TIME LEGISLATION.*	28) The 'annual productive hours' used for calculating the hourly rate were consistent with the usual cost accounting practices of the Beneficiary and were equivalent to at least 90 % of the 'annual workable hours'.	
A.3	**HOURLY PERSONNEL RATES** <u>I) For unit costs calculated in accordance to the Beneficiary's usual cost accounting practice (unit costs):</u> If the Beneficiary has a "Certificate on Methodology to calculate unit costs " (CoMUC) approved by the Commission, the Beneficiary provides the Auditor with a description of the approved methodology and the Commission's letter of acceptance. The Auditor verified that the Beneficiary has indeed used the methodology approved. If so, no further verification is necessary. If the Beneficiary does not have a "Certificate on Methodology" (CoMUC) approved by the	29) The Beneficiary applied [*choose one option and delete the other*]: [Option I: "Unit costs (hourly rates) were calculated in accordance with the Beneficiary's usual cost accounting practices"] [Option II: Individual hourly rates were applied]	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	Commission, or if the methodology approved was not applied, then the Auditor: ○ reviewed the documentation provided by the Beneficiary, including manuals and internal guidelines that explain how to calculate hourly rates; ○ recalculated the unit costs (hourly rates) of staff included in the sample following the results of the procedures carried out in A.1 and A.2. II) For individual hourly rates: The Auditor: ○ reviewed the documentation provided by the Beneficiary, including manuals and internal guidelines that explain how to calculate hourly rates; ○ recalculated the hourly rates of staff included in the sample following the results of the procedures carried out in A.1 and A.2. *"UNIT COSTS CALCULATED BY THE BENEFICIARY IN ACCORDANCE WITH ITS USUAL COST ACCOUNTING PRACTICES":* *IT IS CALCULATED BY DIVIDING THE TOTAL AMOUNT OF PERSONNEL COSTS OF THE CATEGORY TO WHICH THE EMPLOYEE BELONGS VERIFIED IN LINE WITH PROCEDURE A.1 BY THE NUMBER OF FTE AND THE ANNUAL TOTAL PRODUCTIVE HOURS OF THE SAME CATEGORY CALCULATED BY THE BENEFICIARY IN ACCORDANCE WITH PROCEDURE A.2.* *HOURLY RATE FOR INDIVIDUAL ACTUAL PERSONAL COSTS:* *IT IS CALCULATED BY DIVIDING THE TOTAL AMOUNT OF PERSONNEL COSTS OF AN EMPLOYEE VERIFIED IN LINE WITH*	*For option I concerning unit costs and if the Beneficiary applies the methodology approved by the Commission (CoMUC):* 30) The Beneficiary used the Commission-approved methodology to calculate hourly rates. It corresponded to the organisation's usual cost accounting practices and was applied consistently for all activities irrespective of the source of funding.	
		For option I concerning unit costs and if the Beneficiary applies a methodology not approved by the Commission: 31) The unit costs re-calculated by the Auditor were the same as the rates applied by the Beneficiary.	
		For option II concerning individual hourly rates:	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	PROCEDURE A.1 BY THE NUMBER OF ANNUAL PRODUCTIVE HOURS VERIFIED IN LINE WITH PROCEDURE A.2.	32) The individual rates re-calculated by the Auditor were the same as the rates applied by the Beneficiary.	
A.4	**TIME RECORDING SYSTEM** To verify that the time recording system ensures the fulfilment of all minimum requirements and that the hours declared for the action were correct, accurate and properly authorised and supported by documentation, the Auditor made the following checks for the persons included in the sample that declare time as worked for the action on the basis of time records: o description of the time recording system provided by the Beneficiary (registration, authorisation, processing in the HR-system); o its actual implementation; o time records were signed at least monthly by the employees (on paper or electronically) and authorised by the project manager or another manager; o the hours declared were worked within the project period; o there were no hours declared as worked for the action if HR-records showed absence due to holidays or sickness (further cross-checks with travels are carried out in B.1 below) ;	33) All persons recorded their time dedicated to the action on a **daily/ weekly/ monthly** basis using a **paper/computer-based** system. (*delete the answers that are not applicable)*	
		34) Their time-records were authorised at least monthly by the project manager or other superior.	
		35) Hours declared were worked within the project period and were consistent with the presences/absences recorded in HR-records.	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	○ the hours charged to the action matched those in the time recording system. *ONLY THE HOURS WORKED ON THE ACTION CAN BE CHARGED. ALL WORKING TIME TO BE CHARGED SHOULD BE RECORDED THROUGHOUT THE DURATION OF THE PROJECT, ADEQUATELY SUPPORTED BY EVIDENCE OF THEIR REALITY AND RELIABILITY (SEE SPECIFIC PROVISIONS BELOW FOR PERSONS WORKING EXCLUSIVELY FOR THE ACTION WITHOUT TIME RECORDS).*	36) There were no discrepancies between the number of hours charged to the action and the number of hours recorded.	
	<u>If the persons are working exclusively for the action and without time records</u> For the persons selected that worked exclusively for the action without time records, the Auditor verified evidence available demonstrating that they were in reality exclusively dedicated to the action and that the Beneficiary signed a declaration confirming that they have worked exclusively for the action.	37) The exclusive dedication is supported by a declaration signed by the Beneficiary's and by any other evidence gathered.	
B	**COSTS OF SUBCONTRACTING**		
B.1	**The Auditor obtained the detail/breakdown of subcontracting costs and sampled _____ cost items selected randomly** (*full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is highest*). To confirm standard factual findings 38-42 listed in the next column, the Auditor reviewed the	38) The use of claimed subcontracting costs was foreseen in Annex 1 and costs were declared in the Financial Statements under the subcontracting category.	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	following for the items included in the sample: o the use of subcontractors was foreseen in Annex 1; o subcontracting costs were declared in the subcontracting category of the Financial Statement; o supporting documents on the selection and award procedure were followed; o the Beneficiary ensured best value for money (key elements to appreciate the respect of this principle are the award of the subcontract to the bid offering best price-quality ratio, under conditions of transparency and equal treatment. In case an existing framework contract was used the Beneficiary ensured it was established on the basis of the principle of best value for money under conditions of transparency and equal treatment). In particular, i. if the Beneficiary acted as a contracting authority within the meaning of Directive 2004/18/EC or of Directive 2004/17/EC, the Auditor verified that the applicable national law on public procurement was followed and that the subcontracting complied with the Terms and Conditions of the Agreement. ii. if the Beneficiary did not fall under the above-mentioned category the Auditor verified that the Beneficiary followed their usual procurement rules and respected the Terms and Conditions of the Agreement..	39) There were documents of requests to different providers, different offers and assessment of the offers before selection of the provider in line with internal procedures and procurement rules. Subcontracts were awarded in accordance with the principle of best value for money. *(When different offers were not collected the Auditor explains the reasons provided by the Beneficiary under the caption "Exceptions" of the Report. The Commission will analyse this information to evaluate whether these costs might be accepted as eligible)*	
	For the items included in the sample the Auditor also verified that: o the subcontracts were not awarded to other Beneficiaries in the consortium;	40) The subcontracts were not awarded to other Beneficiaries of the consortium.	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	○ there were signed agreements between the Beneficiary and the subcontractor; ○ there was evidence that the services were provided by subcontractor;	41) All subcontracts were supported by signed agreements between the Beneficiary and the subcontractor.	
		42) There was evidence that the services were provided by the subcontractors.	
C	**COSTS OF PROVIDING FINANCIAL SUPPORT TO THIRD PARTIES**		
C.1	**The Auditor obtained the detail/breakdown of the costs of providing financial support to third parties and sampled _____ cost items selected randomly** (*full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is highest*). The Auditor verified that the following minimum conditions were met: a) the maximum amount of financial support for each third party did not exceed EUR 60 000, unless explicitly mentioned in Annex 1; b) the financial support to third parties was agreed in Annex 1 of the Agreement and the other provisions on financial support to third parties included in Annex 1 were	43) All minimum conditions were met	

Ref	Procedures	Standard factual finding	Result (C / E / N.A.)
	respected.		

D	OTHER ACTUAL DIRECT COSTS		
D.1	**COSTS OF TRAVEL AND RELATED SUBSISTENCE ALLOWANCES** **The Auditor sampled _____ cost items selected randomly** (*full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is the highest*). The Auditor inspected the sample and verified that: ○ travel and subsistence costs were consistent with the Beneficiary's usual policy for travel. In this context, the Beneficiary provided evidence of its normal policy for travel costs (e.g. use of first class tickets, reimbursement by the Beneficiary on the basis of actual costs, a lump sum or per diem) to enable the Auditor to compare the travel costs charged with this policy; ○ travel costs are correctly identified and allocated to the action (e.g. trips are directly linked to the action) by reviewing relevant supporting documents such as minutes of meetings, workshops or conferences, their registration in the correct project account, their consistency with time records or with the dates/duration of the workshop/conference; ○ no ineligible costs or excessive or reckless expenditure was declared.	44) Costs were incurred, approved and reimbursed in line with the Beneficiary's usual policy for travels.	
		45) There was a link between the trip and the action.	
		46) The supporting documents were consistent with each other regarding subject of the trip, dates, duration and reconciled with time records and accounting.	
		47) No ineligible costs or excessive or reckless expenditure was declared.	
D.2	**DEPRECIATION COSTS FOR EQUIPMENT, INFRASTRUCTURE OR OTHER ASSETS** **The Auditor sampled _____ cost items selected randomly** (*full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is the highest*). For “equipment, infrastructure or other assets” [from now on called “asset(s)”] selected in the	48) Procurement rules, principles and guides were followed.	
		49) There was a link between the grant agreement and the asset charged to the action.	

	sample the Auditor verified that: ○ the assets were acquired in conformity with the Beneficiary's internal guidelines and procedures; ○ they were correctly allocated to the action (with supporting documents such as delivery note invoice or any other proof demonstrating the link to the action) ○ they were entered in the accounting system; ○ the extent to which the assets were used for the action (as a percentage) was supported by reliable documentation (e.g. usage overview table); The Auditor recalculated the depreciation costs and verified that they were in line with the applicable rules in the Beneficiary's country and with the Beneficiary's usual accounting policy (e.g. depreciation calculated on the acquisition value). The Auditor verified that no ineligible costs such as deductible VAT, exchange rate losses, excessive or reckless expenditure were declared (see Article 6.5 GA).	50) The asset charged to the action was traceable to the accounting records and the underlying documents. 51) The depreciation method used to charge the asset to the action was in line with the applicable rules of the Beneficiary's country and the Beneficiary's usual accounting policy. 52) The amount charged corresponded to the actual usage for the action. 53) No ineligible costs or excessive or reckless expenditure were declared.	
D.3	**COSTS OF OTHER GOODS AND SERVICES** **The Auditor sampled _____ cost items selected randomly** (*full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is highest*). For the purchase of goods, works or services included in the sample the Auditor verified that: ○ the contracts did not cover tasks described in Annex 1;	54) Contracts for works or services did not cover tasks described in Annex 1. 55) Costs were allocated to the correct action and the goods were not placed in the inventory of durable equipment.	

	○ they were correctly identified, allocated to the proper action, entered in the accounting system (traceable to underlying documents such as purchase orders, invoices and accounting); ○ the goods were not placed in the inventory of durable equipment; ○ the costs charged to the action were accounted in line with the Beneficiary's usual accounting practices; ○ no ineligible costs or excessive or reckless expenditure were declared (see Article 6 GA). In addition, the Auditor verified that these goods and services were acquired in conformity with the Beneficiary's internal guidelines and procedures, in particular: ○ if Beneficiary acted as a contracting authority within the meaning of Directive 2004/18/EC or of Directive 2004/17/EC, the Auditor verified that the applicable national law on public procurement was followed and that the procurement contract complied with the Terms and Conditions of the Agreement. ○ if the Beneficiary did not fall into the category above, the Auditor verified that the Beneficiary followed their usual procurement rules and respected the Terms and Conditions of the Agreement. For the items included in the sample the Auditor also verified that: ○ the Beneficiary ensured best value for money (key elements to appreciate the respect of this principle are the award of the contract to the bid offering best price-quality ratio, under conditions of transparency and equal treatment. In case an existing framework contract was used the Auditor also verified that the Beneficiary ensured it was established on the basis of the principle of best value for money under conditions of transparency and equal treatment); *SUCH GOODS AND SERVICES INCLUDE, FOR INSTANCE, CONSUMABLES AND SUPPLIES, DISSEMINATION (INCLUDING OPEN ACCESS), PROTECTION OF RESULTS, SPECIFIC EVALUATION OF THE ACTION IF IT IS REQUIRED BY THE*	56) The costs were charged in line with the Beneficiary's accounting policy and were adequately supported. 57) No ineligible costs or excessive or reckless expenditure were declared. For internal invoices/charges only the cost element was charged, without any mark-ups. 58) Procurement rules, principles and guides were followed. There were documents of requests to different providers, different offers and assessment of the offers before selection of the provider in line with internal procedures and procurement rules. The purchases were made in accordance with the principle of best value for money. *(When different offers were not collected the Auditor explains the reasons provided by the Beneficiary under the*	

	AGREEMENT, CERTIFICATES ON THE FINANCIAL STATEMENTS IF THEY ARE REQUIRED BY THE AGREEMENT AND CERTIFICATES ON THE METHODOLOGY, TRANSLATIONS, REPRODUCTION.	*caption "Exceptions" of the Report. The Commission will analyse this information to evaluate whether these costs might be accepted as eligible)*	
D.4	**AGGREGATED CAPITALISED AND OPERATING COSTS OF RESEARCH INFRASTRUCTURE** The Auditor ensured the existence of a positive ex-ante assessment (issued by the EC Services) of the cost accounting methodology of the Beneficiary allowing it to apply the guidelines on direct costing for large research infrastructures in Horizon 2020. ***In the cases that a positive ex-ante assessment has been issued*** *(see the standard factual findings 59-60 on the next column)****,*** The Auditor ensured that the beneficiary has applied consistently the methodology that is explained and approved in the positive ex ante assessment; ***In the cases that a positive ex-ante assessment has NOT been issued*** *(see the standard factual findings 61 on the next column),* The Auditor verified that no costs of Large Research Infrastructure have been charged as direct costs in any costs category;	59) The costs declared as direct costs for Large Research Infrastructures (in the appropriate line of the Financial Statement) comply with the methodology described in the positive ex-ante assessment report. 60) Any difference between the methodology applied and the one positively assessed was extensively described and adjusted accordingly. 61) The direct costs declared were free from any indirect costs items related to the Large Research Infrastructure.	

	In the cases that a draft ex-ante assessment report has been issued with recommendation for further changes *(see the standard factual findings 61 on the next column),* • The Auditor followed the same procedure as above (when a positive ex-ante assessment has NOT yet been issued) and paid particular attention (testing reinforced) to the cost items for which the draft ex-ante assessment either rejected the inclusion as direct costs for Large Research Infrastructures or issued recommendations.		
E	**USE OF EXCHANGE RATES**		
E.1	a) For Beneficiaries with accounts established in a currency other than euros **The Auditor sampled _____ cost items selected randomly and verified that the exchange rates used for converting other currencies into euros were in accordance with the following rules established in the Agreement** (*full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is highest)*: *COSTS INCURRED IN ANOTHER CURRENCY SHALL BE CONVERTED INTO EURO AT THE AVERAGE OF THE DAILY EXCHANGE RATES PUBLISHED IN THE C SERIES OF OFFICIAL JOURNAL OF THE EUROPEAN UNION (https://www.ecb.int/stats/exchange/eurofxref/html/index.en.html), DETERMINED OVER THE CORRESPONDING REPORTING PERIOD.* *IF NO DAILY EURO EXCHANGE RATE IS PUBLISHED IN THE OFFICIAL JOURNAL OF THE EUROPEAN UNION FOR THE CURRENCY IN QUESTION, CONVERSION SHALL BE MADE AT THE AVERAGE OF THE MONTHLY ACCOUNTING RATES ESTABLISHED BY THE COMMISSION AND PUBLISHED ON ITS WEBSITE (http://ec.europa.eu/budget/contracts_grants/info_contracts/inforeuro/inforeuro_en.cfm),*	62) The exchange rates used to convert other currencies into Euros were in accordance with the rules established of the Grant Agreement and there was no difference in the final figures.	

	DETERMINED OVER THE CORRESPONDING REPORTING PERIOD.		
	b) For Beneficiaries with accounts established in euros **The Auditor sampled _____ cost items selected randomly and verified that the exchange rates used for converting other currencies into euros were in accordance with the following rules established in the Agreement** (*full coverage is required if there are fewer than 10 items, otherwise the sample should have a minimum of 10 item, or 10% of the total, whichever number is highest)*: *COSTS INCURRED IN ANOTHER CURRENCY SHALL BE CONVERTED INTO EURO BY APPLYING THE BENEFICIARY'S USUAL ACCOUNTING PRACTICES.*	63) The Beneficiary applied its usual accounting practices.	

[legal name of the audit firm]

[name and function of an authorised representative]

[dd Month yyyy]

<Signature of the Auditor>

 Associated with document Ref. Ares(2015)2187417 - 26/05/2015

ANNEX 6

MODEL FOR THE CERTIFICATE ON THE METHODOLOGY

- **For options [*in italics in square brackets*]: choose the applicable option. Options not chosen should be deleted.**
- **For fields in [grey in square brackets]: enter the appropriate data.**

TABLE OF CONTENTS

TERMS OF REFERENCE FOR AN AUDIT ENGAGEMENT FOR A METHODOLOGY CERTIFICATE IN CONNECTION WITH ONE OR MORE GRANT AGREEMENTS FINANCED UNDER THE HORIZON 2020 RESEARCH AND INNOVATION FRAMEWORK PROGRAMME……2

INDEPENDENT REPORT OF FACTUAL FINDINGS ON THE METHODOLOGY CONCERNING GRANT AGREEMENTS FINANCED UNDER THE HORIZON 2020 RESEARCH AND INNOVATION FRAMEWORK PROGRAMME …… 7

 Associated with document Ref. Ares(2015)2187417 - 26/05/2015

Terms of reference for an audit engagement for a methodology certificate in connection with one or more grant agreements financed under the Horizon 2020 Research and Innovation Framework Programme

This document sets out the '**Terms of Reference (ToR)**' under which

[OPTION 1: [insert name of the beneficiary] ('the Beneficiary')] [OPTION 2: [insert name of the linked third party] ('the Linked Third Party'), third party linked to the Beneficiary [insert name of the beneficiary] ('the Beneficiary')]

agrees to engage

[**insert legal name of the auditor**] ('the Auditor')

to produce an independent report of factual findings ('the Report') concerning the *[Beneficiary's] [Linked Third Party's]* usual accounting practices for calculating and claiming direct personnel costs declared as unit costs ('the Methodology') in connection with grant agreements financed under the Horizon 2020 Research and Innovation Framework Programme.

The procedures to be carried out for the assessment of the methodology will be based on the grant agreement(s) detailed below:

[title and number of the grant agreement(s)] ('the Agreement(s)')

The Agreement(s) has(have) been concluded between the Beneficiary and *[OPTION 1: the European Union, represented by the European Commission ('the Commission')][OPTION 2: the European Atomic Energy Community (Euratom,) represented by the European Commission ('the Commission')][OPTION 3: the [Research Executive Agency (REA)] [European Research Council Executive Agency (ERCEA)] [Innovation and Networks Executive Agency (INEA)] [Executive Agency for Small and Medium-sized Enterprises (EASME)] ('the Agency'), under the powers delegated by the European Commission ('the Commission').]*.

 Associated with document Ref. Ares(2015)2187417 - 26/05/2015

The *[Commission] [Agency]* is mentioned as a signatory of the Agreement with the Beneficiary only. The *[European Union] [Euratom] [Agency]* is not a party to this engagement.

1.1 Subject of the engagement

According to Article 18.1.2 of the Agreement, beneficiaries *[and linked third parties]* that declare direct personnel costs as unit costs calculated in accordance with their usual cost accounting practices may submit to the *[Commission] [Agency]*, for approval, a certificate on the methodology ('CoMUC') stating that there are adequate records and documentation to prove that their cost accounting practices used comply with the conditions set out in Point A of Article 6.2.

The subject of this engagement is the CoMUC which is composed of two separate documents:

- the Terms of Reference ('the ToR') to be signed by the *[Beneficiary] [Linked Third Party]* and the Auditor;

- the Auditor's Independent Report of Factual Findings ('the Report') issued on the Auditor's letterhead, dated, stamped and signed by the Auditor which includes; the standard statements ('the Statements') evaluated and signed by the *[Beneficiary] [Linked Third Party]*, the agreed-upon procedures ('the Procedures') performed by the Auditor and the standard factual findings ('the Findings') assessed by the Auditor. The Statements, Procedures and Findings are summarised in the table that forms part of the Report.

The information provided through the Statements, the Procedures and the Findings will enable the Commission to draw conclusions regarding the existence of the *[Beneficiary's] [Linked Third Party's]* usual cost accounting practice and its suitability to ensure that direct personnel costs claimed on that basis comply with the provisions of the Agreement. The Commission draws its own conclusions from the Report and any additional information it may require.

1.2 Responsibilities

The parties to this agreement are the *[Beneficiary] [Linked Third Party]* and the Auditor.

The *[Beneficiary] [Linked Third Party]:*

 Associated with document Ref. Ares(2015)2187417 - 26/05/2015

- is responsible for preparing financial statements for the Agreement(s) ('the Financial Statements') in compliance with those Agreements;
- is responsible for providing the Financial Statement(s) to the Auditor and enabling the Auditor to reconcile them with the *[Beneficiary's] [Linked Third Party's]* accounting and bookkeeping system and the underlying accounts and records. The Financial Statement(s) will be used as a basis for the procedures which the Auditor will carry out under this ToR;
- is responsible for its Methodology and liable for the accuracy of the Financial Statement(s);
- is responsible for endorsing or refuting the Statements indicated under the heading 'Statements to be made by the Beneficiary/ Linked Third Party' in the first column of the table that forms part of the Report;
- must provide the Auditor with a signed and dated representation letter;
- accepts that the ability of the Auditor to carry out the Procedures effectively depends upon the *[Beneficiary] [Linked Third Party]* providing full and free access to the *[Beneficiary's] [Linked Third Party's]* staff and to its accounting and other relevant records.

The Auditor:

- *[Option 1 by default: is qualified to carry out statutory audits of accounting documents in accordance with Directive 2006/43/EC of the European Parliament and of the Council of 17 May 2006 on statutory audits of annual accounts and consolidated accounts, amending Council Directives 78/660/EEC and 83/349/EEC and repealing Council Directive 84/253/EEC or similar national regulations]*.
- *[Option 2 if the Beneficiary or Linked Third Party has an independent Public Officer: is a competent and independent Public Officer for which the relevant national authorities have established the legal capacity to audit the Beneficiary]*.
- *[Option 3 if the Beneficiary or Linked Third Party is an international organisation: is an [internal] [external] auditor in accordance with the internal financial regulations and procedures of the international organisation]*.

The Auditor:

- must be independent from the Beneficiary *[and the Linked Third Party]*, in particular, it must not have been involved in preparing the Beneficiary's *[and Linked Third Party's]* Financial Statement(s);
- must plan work so that the Procedures may be carried out and the Findings may be assessed;
- must adhere to the Procedures laid down and the compulsory report format;
- must carry out the engagement in accordance with these ToR;
- must document matters which are important to support the Report;
- must base its Report on the evidence gathered;
- must submit the Report to the *[Beneficiary] [Linked Third Party]*.

The Commission sets out the Procedures to be carried out and the Findings to be endorsed by the Auditor. The Auditor is not responsible for their suitability or pertinence. As this engagement is not an assurance engagement the Auditor does not provide an audit opinion or a statement of assurance.

1.3 Applicable Standards

The Auditor must comply with these Terms of Reference and with[1]:

- the International Standard on Related Services ('ISRS') 4400 *Engagements to perform Agreed-upon Procedures regarding Financial Information* as issued by the International Auditing and Assurance Standards Board (IAASB);
- the *Code of Ethics for Professional Accountants* issued by the International Ethics Standards Board for Accountants (IESBA). Although ISRS 4400 states that independence is not a requirement for engagements to carry out agreed-upon procedures, the Commission requires that the Auditor also complies with the Code's independence requirements.

The Auditor's Report must state that there was no conflict of interests in establishing this Report between the Auditor and the Beneficiary *[and the Linked Third Party]* that could have a bearing on the Report, and must specify – if the service is invoiced - the total fee paid to the Auditor for providing the Report.

1.4 Reporting

The Report must be written in the language of the Agreement (see Article 20.7 of the Agreement).

Under Article 22 of the Agreement, the Commission, *[the Agency]*, the European Anti-Fraud Office and the Court of Auditors have the right to audit any work that is carried out under the action and for which costs are claimed from [*the European Union*] [*Euratom*] budget. This includes work related to this engagement. The Auditor must provide access to all working papers related to this assignment if the Commission*, [the Agency]*, the European Anti-Fraud Office or the European Court of Auditors requests them.

1.5 Timing

The Report must be provided by [dd Month yyyy].

[1] Supreme Audit Institutions applying INTOSAI-standards may carry out the Procedures according to the corresponding International Standards of Supreme Audit Institutions and code of ethics issued by INTOSAI instead of the International Standard on Related Services ('ISRS') 4400 and the Code of Ethics for Professional Accountants issued by the IAASB and the IESBA.

 Associated with document Ref. Ares(2015)2187417 - 26/05/2015

1.6 Other Terms

[The [Beneficiary] [Linked Third Party] and the Auditor can use this section to agree other specific terms, such as the Auditor's fees, liability, applicable law, etc. Those specific terms must not contradict the terms specified above.]

[legal name of the Auditor]	[legal name of the [Beneficiary] [Linked Third Party]]
[name & title of authorised representative]	[name & title of authorised representative]
[dd Month yyyy]	[dd Month yyyy]
Signature of the Auditor Signature	Signature of the *[Beneficiary] [Linked Third Party]*

 Associated with document Ref. Ares(2015)2187417 - 26/05/2015

Independent report of factual findings on the methodology concerning grant agreements financed under the Horizon 2020 Research and Innovation Framework Programme

(To be printed on letterhead paper of the auditor)

To

[name of contact person(s)], [Position]

[[Beneficiary's] [Linked Third Party's] name]

[Address]

[dd Month yyyy]

Dear [Name of contact person(s)],

As agreed under the terms of reference dated [dd Month yyyy]

with *[OPTION 1: [insert name of the beneficiary] ('the Beneficiary')] [OPTION 2: [insert name of the linked third party] ('the Linked Third Party'), third party linked to the Beneficiary [insert name of the beneficiary] ('the Beneficiary')]*,

we

[name of the auditor] ('the Auditor'),

established at

[full address/city/state/province/country],

represented by

[name and function of an authorised representative],

 Associated with document Ref. Ares(2015)2187417 - 26/05/2015

have carried out the agreed-upon procedures ('the Procedures') and provide hereby our Independent Report of Factual Findings ('the Report'), concerning the *[Beneficiary's] [Linked Third Party's]* usual accounting practices for calculating and declaring direct personnel costs declared as unit costs ('the Methodology').

You requested certain procedures to be carried out in connection with the grant(s)

[title and number of the grant agreement(s)] ('the Agreement(s)').

The Report

Our engagement was carried out in accordance with the terms of reference ('the ToR') appended to this Report. The Report includes: the standard statements ('the Statements') made by the *[Beneficiary] [Linked Third Party]*, the agreed-upon procedures ('the Procedures') carried out and the standard factual findings ('the Findings') confirmed by us.

The engagement involved carrying out the Procedures and assessing the Findings and the documentation requested appended to this Report, the results of which the Commission uses to draw conclusions regarding the acceptability of the Methodology applied by the *[Beneficiary] [Linked Third Party]*.

The Report covers the methodology used from [dd Month yyyy]. In the event that the *[Beneficiary] [Linked Third Party]* changes this methodology, the Report will not be applicable to any Financial Statement[2] submitted thereafter.

The scope of the Procedures and the definition of the standard statements and findings were determined solely by the Commission. Therefore, the Auditor is not responsible for their suitability or pertinence.

Since the Procedures carried out constitute neither an audit nor a review made in accordance with International Standards on Auditing or International Standards on Review Engagements, we do not

[2] Financial Statement in this context refers solely to Annex 4 of the Agreement by which the Beneficiary declares costs under the Agreement.

give a statement of assurance on the costs declared on the basis of the *[Beneficiary's] [Linked Third Party's]* Methodology. Had we carried out additional procedures or had we performed an audit or review in accordance with these standards, other matters might have come to its attention and would have been included in the Report.

Exceptions

Apart from the exceptions listed below, the *[Beneficiary] [Linked Third Party]* agreed with the standard Statements and provided the Auditor all the documentation and accounting information needed by the Auditor to carry out the requested Procedures and corroborate the standard Findings.

List here any exception and add any information on the cause and possible consequences of each exception, if known. If the exception is quantifiable, also indicate the corresponding amount.

…..

Explanation of possible exceptions in the form of examples (to be removed from the Report):

i. the [Beneficiary] [Linked Third Party] did not agree with the standard Statement number … because…;

ii. the Auditor could not carry out the procedure … established because …. (e.g. due to the inability to reconcile key information or the unavailability or inconsistency of data);

iii. the Auditor could not confirm or corroborate the standard Finding number … because ….

Remarks

We would like to add the following remarks relevant for the proper understanding of the Methodology applied by the [Beneficiary] [Linked Third Party] or the results reported:

Example (to be removed from the Report):

Regarding the methodology applied to calculate hourly rates …

Regarding standard Finding 15 it has to be noted that …

The [Beneficiary] [Linked Third Party] explained the deviation from the benchmark statement XXIV concerning time recording for personnel with no exclusive dedication to the action in the following manner: …

Annexes

 Associated with document Ref. Ares(2015)2187417 - 26/05/2015

Please provide the following documents to the auditor and annex them to the report when submitting this CoMUC to the Commission:

1. Brief description of the methodology for calculating personnel costs, productive hours and hourly rates;
2. Brief description of the time recording system in place;
3. An example of the time records used by the [*Beneficiary] [Linked Third Party*];
4. Description of any budgeted or estimated elements applied, together with an explanation as to why they are relevant for calculating the personnel costs and how they are based on objective and verifiable information;
5. A summary sheet with the hourly rate for direct personnel declared by the [*Beneficiary*] [*Linked Third Party*] and recalculated by the Auditor for each staff member included in the sample (the names do not need to be reported);
6. A comparative table summarising for each person selected in the sample a) the time claimed by the [*Beneficiary*] [*Linked Third Party*] in the Financial Statement(s) and b) the time according to the time record verified by the Auditor;
7. A copy of the letter of representation provided to the Auditor.

Use of this Report

This Report has been drawn up solely for the purpose given under Point 1.1 Reasons for the engagement.

The Report:

- is confidential and is intended to be submitted to the Commission by the [*Beneficiary*] [*Linked Third Party*] in connection with Article 18.1.2 of the Agreement;
- may not be used by the [*Beneficiary*] [*Linked Third Party*] or by the Commission for any other purpose, nor distributed to any other parties;
- may be disclosed by the Commission only to authorised parties, in particular the European Anti-Fraud Office (OLAF) and the European Court of Auditors.
- relates only to the usual cost accounting practices specified above and does not constitute a report on the Financial Statements of the [*Beneficiary*] [*Linked Third Party*].

No conflict of interest[3] exists between the Auditor and the Beneficiary *[and the Linked Third Party*] that could have a bearing on the Report. The total fee paid to the Auditor for producing the Report was EUR _____ (including EUR _____ of deductible VAT).

[3] A conflict of interest arises when the Auditor's objectivity to establish the certificate is compromised in fact or in appearance when the Auditor for instance:

- was involved in the preparation of the Financial Statements;

 Associated with document Ref. Ares(2015)2187417 - 26/05/2015

We look forward to discussing our Report with you and would be pleased to provide any further information or assistance which may be required.

Yours sincerely

[legal name of the Auditor]

[name and title of the authorised representative]

[dd Month yyyy]

Signature of the Auditor

- stands to benefit directly should the certificate be accepted;

- has a close relationship with any person representing the beneficiary;

- is a director, trustee or partner of the beneficiary; or

- is in any other situation that compromises his or her independence or ability to establish the certificate impartially.

 Associated with document Ref. Ares(2015)2187417 - 26/05/2015

Statements to be made by the Beneficiary/Linked Third Party ('the Statements') and Procedures to be carried out by the Auditor ('the Procedures') and standard factual findings ('the Findings') to be confirmed by the Auditor

The Commission reserves the right to provide the auditor with guidance regarding the Statements to be made, the Procedures to be carried out or the Findings to be ascertained and the way in which to present them. The Commission reserves the right to vary the Statements, Procedures or Findings by written notification to the Beneficiary/Linked Third Party to adapt the procedures to changes in the grant agreement(s) or to any other circumstances.

If this methodology certificate relates to the Linked Third Party's usual accounting practices for calculating and claiming direct personnel costs declared as unit costs any reference here below to 'the Beneficiary' is to be considered as a reference to 'the Linked Third Party'.

Please explain any discrepancies in the body of the Report.	
Statements to be made by Beneficiary	**Procedures to be carried out and Findings to be confirmed by the Auditor**
A. Use of the Methodology I. The cost accounting practice described below has been in use since *[*dd Month yyyy*]*. II. The next planned alteration to the methodology used by the Beneficiary will be from [dd Month yyyy*]*.	**Procedure:** ✓ The Auditor checked these dates against the documentation the Beneficiary has provided. **Factual finding:** 1. The dates provided by the Beneficiary were consistent with the documentation.
B. Description of the Methodology III. The methodology to calculate unit costs is being used in a consistent manner and is reflected in the relevant procedures. *[Please describe the methodology your entity uses to calculate personnel costs, productive hours and hourly rates, present your description to the Auditor and annex it to this certificate]* *[If the statement of section "B. Description of the methodology" cannot be endorsed by the Beneficiary or there is no written methodology to calculate unit costs it should be listed here below and reported as exception by the Auditor in the main Report of*	**Procedure:** ✓ The Auditor reviewed the description, the relevant manuals and/or internal guidance documents describing the methodology. **Factual finding:** 2. The brief description was consistent with the relevant manuals, internal guidance and/or other documentary evidence the Auditor has reviewed. 3. The methodology was generally applied by the Beneficiary as part of its usual costs accounting practices.

Please explain any discrepancies in the body of the Report.	
Statements to be made by Beneficiary	**Procedures to be carried out and Findings to be confirmed by the Auditor**
Factual Findings: - *…]*	
C. **Personnel costs** <u>General</u> IV. The unit costs (hourly rates) are limited to salaries including during parental leave, social security contributions, taxes and other costs included in the remuneration required under national law and the employment contract or equivalent appointing act; V. Employees are hired directly by the Beneficiary in accordance with national law, and work under its sole supervision and responsibility; VI. The Beneficiary remunerates its employees in accordance with its usual practices. This means that personnel costs are charged in line with the Beneficiary's usual payroll policy (e.g. salary policy, overtime policy, variable pay) and no special conditions exist for employees assigned to tasks relating to the European Union or Euratom, unless explicitly provided for in the grant agreement(s); VII. The Beneficiary allocates its employees to the relevant group/category/cost centre for the purpose of the unit cost calculation in line with the usual cost accounting practice; VIII. Personnel costs are based on the payroll system and accounting system. IX. Any exceptional adjustments of actual personnel costs resulted from relevant budgeted or estimated elements and were based on objective and verifiable information. *[Please describe the 'budgeted or estimated elements' and their relevance to personnel costs, and explain how they were reasonable and based on objective and verifiable information, present your explanation to the Auditor and annex it to this certificate].* X. Personnel costs claimed do not contain any of the following ineligible costs: costs related to return on capital; debt and debt service charges; provisions for future losses	**Procedure:** *The Auditor draws a sample of employees to carry out the procedures indicated in this section C and the following sections D to F.* *[The Auditor has drawn a random sample of 10 full-time equivalents made up of employees assigned to the action(s). If fewer than 10 full-time equivalents are assigned to the action(s), the Auditor has selected a sample of 10 full-time equivalents consisting of all employees assigned to the action(s), complemented by other employees irrespective of their assignments.].* For this sample: ✓ the Auditor reviewed all documents relating to personnel costs such as employment contracts, payslips, payroll policy (e.g. salary policy, overtime policy, variable pay policy), accounting and payroll records, applicable national tax , labour and social security law and any other documents corroborating the personnel costs claimed; ✓ in particular, the Auditor reviewed the employment contracts of the employees in the sample to verify that: i. they were employed directly by the Beneficiary in accordance with applicable national legislation; ii. they were working under the sole technical supervision and responsibility of the latter; iii. they were remunerated in accordance with the Beneficiary's usual practices; iv. they were allocated to the correct group/category/cost centre for the purposes of calculating the unit cost in line with the Beneficiary's usual cost accounting practices; ✓ the Auditor verified that any ineligible items or any costs claimed under other costs categories or costs covered by other types of grant or by other grants financed from the European Union budget have not been taken

 Associated with document Ref. Ares(2015)2187417 - 26/05/2015

Please explain any discrepancies in the body of the Report.	
Statements to be made by Beneficiary	**Procedures to be carried out and Findings to be confirmed by the Auditor**
or debts; interest owed; doubtful debts; currency exchange losses; bank costs charged by the Beneficiary's bank for transfers from the Commission/Agency; excessive or reckless expenditure; deductible VAT or costs incurred during suspension of the implementation of the action. XI. Personnel costs were not declared under another EU or Euratom grant (including grants awarded by a Member State and financed by the EU budget and grants awarded by bodies other than the Commission/Agency for the purpose of implementing the EU budget). <u>If additional remuneration as referred to in the grant agreement(s) is paid</u> XII. The Beneficiary is a non-profit legal entity; XIII. The additional remuneration is part of the beneficiary's usual remuneration practices and paid consistently whenever the relevant work or expertise is required; XIV. The criteria used to calculate the additional remuneration are objective and generally applied regardless of the source of funding; XV. The additional remuneration included in the personnel costs used to calculate the hourly rates for the grant agreement(s) is capped at EUR 8 000 per full-time equivalent (reduced proportionately if the employee is not assigned exclusively to the action). *[If certain statement(s) of section "C. Personnel costs" cannot be endorsed by the Beneficiary they should be listed here below and reported as exception by the Auditor in the main Report of*	into account when calculating the personnel costs; ✓ the Auditor numerically reconciled the total amount of personnel costs used to calculate the unit cost with the total amount of personnel costs recorded in the statutory accounts and the payroll system. ✓ to the extent that actual personnel costs were adjusted on the basis of budgeted or estimated elements, the Auditor carefully examined those elements and checked the information source to confirm that they correspond to objective and verifiable information; ✓ if additional remuneration has been claimed, the Auditor verified that the Beneficiary was a non-profit legal entity, that the amount was capped at EUR 8 000 per full-time equivalent and that it was reduced proportionately for employees not assigned exclusively to the action(s). ✓ the Auditor recalculated the personnel costs for the employees in the sample. **Factual finding:** 4. All the components of the remuneration that have been claimed as personnel costs are supported by underlying documentation. 5. The employees in the sample were employed directly by the Beneficiary in accordance with applicable national law and were working under its sole supervision and responsibility. 6. Their employment contracts were in line with the Beneficiary's usual policy; 7. Personnel costs were duly documented and consisted solely of salaries, social security contributions (pension contributions, health insurance, unemployment fund contributions, etc.), taxes and other statutory costs included in the remuneration (holiday pay, thirteenth month's pay, etc.); 8. The totals used to calculate the personnel unit costs are consistent with those registered in the payroll and accounting records; 9. To the extent that actual personnel costs were adjusted on the basis of budgeted or estimated elements, those elements were

 Associated with document Ref. Ares(2015)2187417 - 26/05/2015

Please explain any discrepancies in the body of the Report.	
Statements to be made by Beneficiary	**Procedures to be carried out and Findings to be confirmed by the Auditor**
Factual Findings: - *…]*	relevant for calculating the personnel costs and correspond to objective and verifiable information. The budgeted or estimated elements used are: — (indicate the elements and their values). 10. Personnel costs contained no ineligible elements; 11. Specific conditions for eligibility were fulfilled when additional remuneration was paid: a) the Beneficiary is registered in the grant agreements as a non-profit legal entity; b) it was paid according to objective criteria generally applied regardless of the source of funding used and c) remuneration was capped at EUR 8000 per full-time equivalent (or up to up to the equivalent pro-rata amount if the person did not work on the action full-time during the year or did not work exclusively on the action).
D. Productive hours XVI. The number of productive hours per full-time employee applied is *[delete as appropriate]*: A. 1720 productive hours per year for a person working full-time (corresponding pro-rata for persons not working full time). B. the total number of hours worked in the year by a person for the Beneficiary C. the standard number of annual hours generally applied by the beneficiary for its personnel in accordance with its usual cost accounting practices. This number must be at least 90% of the standard annual workable hours. <u>If method B is applied</u> XVII. The calculation of the total number of hours worked was done as follows: annual workable hours of the person according to the employment contract, applicable labour agreement or national law plus overtime worked minus absences (such as sick leave and special leave). XVIII. 'Annual workable hours' are hours	**Procedure (same sample basis as for Section C: Personnel costs):** ✓ The Auditor verified that the number of productive hours applied is in accordance with method A, B or C. ✓ The Auditor checked that the number of productive hours per full-time employee is correct and that it is reduced proportionately for employees not exclusively assigned to the action(s). ✓ If method B is applied the Auditor verified i) the manner in which the total number of hours worked was done and ii) that the contract specified the annual workable hours by inspecting all the relevant documents, national legislation, labour agreements and contracts. ✓ If method C is applied the Auditor reviewed the manner in which the standard number of working hours per year has been calculated by inspecting all the relevant documents, national legislation, labour agreements and contracts and verified that the number of productive hours per year used for these calculations was at least 90% of the standard number of working hours per year.

 Associated with document Ref. Ares(2015)2187417 - 26/05/2015

Please explain any discrepancies in the body of the Report.	
Statements to be made by Beneficiary	**Procedures to be carried out and Findings to be confirmed by the Auditor**
during which the personnel must be working, at the employer's disposal and carrying out his/her activity or duties under the employment contract, applicable collective labour agreement or national working time legislation. XIX. The contract (applicable collective labour agreement or national working time legislation) do specify the working time enabling to calculate the annual workable hours. <u>If method C is applied</u> XX. The standard number of productive hours per year is that of a full-time equivalent; for employees not assigned exclusively to the action(s) this number is reduced proportionately. XXI. The number of productive hours per year on which the hourly rate is based i) corresponds to the Beneficiary's usual accounting practices; ii) is at least 90% of the standard number of workable (working) hours per year. XXII. Standard workable (working) hours are hours during which personnel are at the Beneficiary's disposal preforming the duties described in the relevant employment contract, collective labour agreement or national labour legislation. The number of standard annual workable (working) hours that the Beneficiary claims is supported by labour contracts, national legislation and other documentary evidence. *[If certain statement(s) of section "D. Productive hours" cannot be endorsed by the Beneficiary they should be listed here below and reported as exception by the Auditor:* - *…]*	**Factual finding:** <u>General</u> 12. The Beneficiary applied a number of productive hours consistent with method A, B or C detailed in the left-hand column. 13. The number of productive hours per year per full-time employee was accurate and was proportionately reduced for employees not working full-time or exclusively for the action. <u>If method B is applied</u> 14. The number of 'annual workable hours', overtime and absences was verifiable based on the documents provided by the Beneficiary and the calculation of the total number of hours worked was accurate. 15. The contract specified the working time enabling to calculate the annual workable hours. <u>If method C is applied</u> 16. The calculation of the number of productive hours per year corresponded to the usual costs accounting practice of the Beneficiary. 17. The calculation of the standard number of workable (working) hours per year was corroborated by the documents presented by the Beneficiary. 18. The number of productive hours per year used for the calculation of the hourly rate was at least 90% of the number of workable (working) hours per year.
E. Hourly rates The hourly rates are correct because: XXIII. Hourly rates are correctly calculated since they result from dividing annual personnel	**Procedure** ✓ The Auditor has obtained a list of all personnel rates calculated by the Beneficiary in accordance with the methodology used. ✓ The Auditor has obtained a list of all the relevant employees, based on which the

Please explain any discrepancies in the body of the Report.	
Statements to be made by Beneficiary	**Procedures to be carried out and Findings to be confirmed by the Auditor**
costs by the productive hours of a given year and group (e.g. staff category or department or cost centre depending on the methodology applied) and they are in line with the statements made in section C. and D. above. *[If the statement of section 'E. Hourly rates' cannot be endorsed by the Beneficiary they should be listed here below and reported as exception by the Auditor:* - *…]*	personnel rate(s) are calculated. For 10 full-time equivalent employees selected at random (same sample basis as Section C: Personnel costs): ✓ The Auditor recalculated the hourly rates. ✓ The Auditor verified that the methodology applied corresponds to the usual accounting practices of the organisation and is applied consistently for all activities of the organisation on the basis of objective criteria irrespective of the source of funding. **Factual finding:** 19. No differences arose from the recalculation of the hourly rate for the employees included in the sample.
F. Time recording XXIV. Time recording is in place for all persons with no exclusive dedication to one Horizon 2020 action. At least all hours worked in connection with the grant agreement(s) are registered on a **daily/weekly/monthly** basis *[delete as appropriate]* using a **paper/computer-based system** *[delete as appropriate]*; XXV. For persons exclusively assigned to one Horizon 2020 activity the Beneficiary has either signed a declaration to that effect or has put arrangements in place to record their working time; XXVI. Records of time worked have been signed by the person concerned (on paper or electronically) and approved by the action manager or line manager at least monthly; XXVII. Measures are in place to prevent staff from: i. recording the same hours twice, ii. recording working hours during absence periods (e.g. holidays, sick leave), iii. recording more than the number of productive hours per year used to calculate the hourly rates, and	**Procedure** ✓ The Auditor reviewed the brief description, all relevant manuals and/or internal guidance describing the methodology used to record time. The Auditor reviewed the time records of the random sample of 10 full-time equivalents referred to under Section C: Personnel costs, and verified in particular: ✓ that time records were available for all persons with not exclusive assignment to the action; ✓ that time records were available for persons working exclusively for a Horizon 2020 action, or, alternatively, that a declaration signed by the Beneficiary was available for them certifying that they were working exclusively for a Horizon 2020 action; ✓ that time records were signed and approved in due time and that all minimum requirements were fulfilled; ✓ that the persons worked for the action in the periods claimed; ✓ that no more hours were claimed than the productive hours used to calculate the hourly

 Associated with document Ref. Ares(2015)2187417 - 26/05/2015

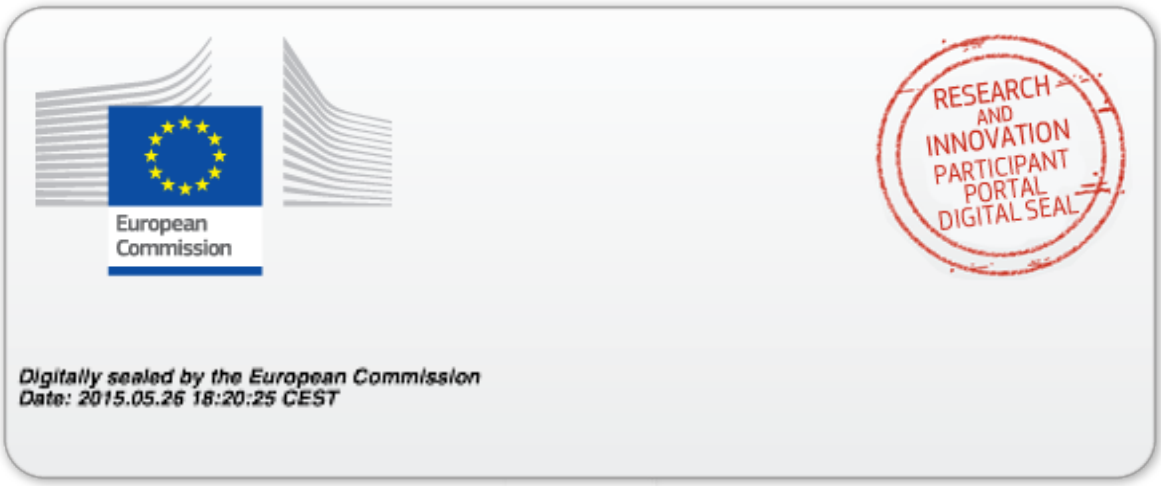
Please explain any discrepancies in the body of the Report.	
Statements to be made by Beneficiary	**Procedures to be carried out and Findings to be confirmed by the Auditor**
iv. recording hours worked outside the action period. XXVIII. No working time was recorded outside the action period; XXIX. No more hours were claimed than the productive hours used to calculate the hourly personnel rates. *[Please provide a brief description of the time recording system in place together with the measures applied to ensure its reliability to the Auditor and annex it to the present certificate[4]].* *[If certain statement(s) of section "F. Time recording" cannot be endorsed by the Beneficiary they should be listed here below and reported as exception by the Auditor:* *- …]*	personnel rates; ✓ that internal controls were in place to prevent that time is recorded twice, during absences for holidays or sick leave; that more hours are claimed per person per year for Horizon 2020 actions than the number of productive hours per year used to calculate the hourly rates; that working time is recorded outside the action period; ✓ the Auditor cross-checked the information with human-resources records to verify consistency and to ensure that the internal controls have been effective. In addition, the Auditor has verified that no more hours were charged to Horizon 2020 actions per person per year than the number of productive hours per year used to calculate the hourly rates, and verified that no time worked outside the action period was charged to the action. **Factual finding:** 20. The brief description, manuals and/or internal guidance on time recording provided by the Beneficiary were consistent with management reports/records and other documents reviewed and were generally applied by the Beneficiary to produce the financial statements. 21. For the random sample time was recorded or, in the case of employees working exclusively for the action, either a signed declaration or time records were available; 22. For the random sample the time records were signed by the employee and the action manager/line manager, at least monthly. 23. Working time claimed for the action occurred in the periods claimed; 24. No more hours were claimed than the number productive hours used to calculate the hourly

[4] The description of the time recording system must state among others information on the content of the time records, its coverage (full or action time-recording, for all personnel or only for personnel involved in H2020 actions), its degree of detail (whether there is a reference to the particular tasks accomplished), its form, periodicity of the time registration and authorisation (paper or a computer-based system; on a daily, weekly or monthly basis; signed and countersigned by whom), controls applied to prevent double-charging of time or ensure consistency with HR-records such as absences and travels as well as it information flow up to its use for the preparation of the Financial Statements.

 Associated with document Ref. Ares(2015)2187417 - 26/05/2015

Please explain any discrepancies in the body of the Report.	
Statements to be made by Beneficiary	**Procedures to be carried out and Findings to be confirmed by the Auditor**
	personnel rates; 25. There is proof that the Beneficiary has checked that working time has not been claimed twice, that it is consistent with absence records and the number of productive hours per year, and that no working time has been claimed outside the action period. 26. Working time claimed is consistent with that on record at the human-resources department.

[official name of the** [Beneficiary] [Linked Third Party]**]	***[official name of the Auditor]***
[name and title of authorised representative]	***[name and title of authorised representative]***
[dd Month yyyy]	***[dd Month yyyy]***
<Signature of the** [Beneficiary] [Linked Third Party]**>	***<Signature** of the Auditor**>***



This document is digitally sealed. The digital sealing mechanism uniquely binds the document to the modules of the Participant Portal of the European Commission, to the transaction for which it was generated and ensures its integrity and authenticity.

Any attempt to modify the content will lead to a breach of the electronic seal, which can be verified at any time by clicking on the digital seal validation symbol.