



<i>Project Acronym</i>	<i>SoBigData</i>
<i>Project Title</i>	<i>SoBigData Research Infrastructure</i> <i>Social Mining & Big Data Ecosystem</i>
<i>Project Number</i>	<i>654024</i>
<i>Deliverable Title</i>	<i>Legal and Ethical Framework for SoBigData 3</i>
<i>Deliverable No.</i>	<i>D2.4</i>
<i>Delivery Date</i>	<i>October 2018</i>
<i>Authors</i>	<i>David van Putten (TUD), Stefanie Hännold (LUH), Iryna Lishchuck, Nikolaus Forgó, Francesca Pratesi (UNIPi)</i>



DOCUMENT INFORMATION

PROJECT	
Project Acronym	SoBigData
Project Title	SoBigData Research Infrastructure Social Mining & Big Data Ecosystem
Project Start	1st September 2015
Project Duration	48 months
Funding	H2020-INFRAIA-2014-2015
Grant Agreement No.	654024
DOCUMENT	
Deliverable No.	D2.4
Deliverable Title	Legal and Ethical Framework for SoBigData 3
Contractual Delivery Date	30th September 2018
Actual Delivery Date	03 October 2018
Author(s)	David van Putten (TUD), Stefanie Hänold (LUH), Iryna Lishchuck (LUH), Francesca Pratesi (UNIPi)
Editor(s)	Stefanie Hänold (LUH), Beatrice Rapisarda (CNR)
Reviewer(s)	Roberto Trasarti (CNR), Salvatore Ruggieri (UNIPi)
Contributor(s)	
Work Package No.	WP2
Work Package Title	NA1_Legal and Ethical Framework
Work Package Leader	TU Delft
Work Package Participants	CNR, USFD, UNIPi, LUH, ETH Zurich
Dissemination	PU
Nature	Report
Version / Revision	V1.1
Draft / Final	Final
Total No. Pages (including cover)	24
Keywords	MOOC, Operational Ethics Board

DISCLAIMER

SoBigData (654024) is a Research and Innovation Action (RIA) funded by the European Commission under the Horizon 2020 research and innovation programme.

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.

This document contains information on SoBigData core activities, findings and outcomes and it may also contain contributions from distinguished experts who contribute as SoBigData Board members. Any reference to content in this document should clearly indicate the authors, source, organisation and publication date.

The document has been produced with the funding of the European Commission. The content of this publication is the sole responsibility of the SoBigData Consortium and its experts, and it cannot be considered to reflect the views of the European Commission. The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

The European Union (EU) was established in accordance with the Treaty on the European Union (Maastricht). There are currently 27 member states of the European Union. It is based on the European Communities and the member states’ cooperation in the fields of Common Foreign and Security Policy and Justice and Home Affairs. The five main institutions of the European Union are the European Parliament, the Council of Ministers, the European Commission, the Court of Justice, and the Court of Auditors (<http://europa.eu.int/>).

Copyright © The SoBigData Consortium 2015. See <http://project.sobigdata.eu/> for details on the copyright holders.

For more information on the project, its partners and contributors please see <http://project.sobigdata.eu/>. You are permitted to copy and distribute verbatim copies of this document containing this copyright notice, but modifying this document is not allowed. You are permitted to copy this document in whole or in part into other documents if you attach the following reference to the copied elements: “Copyright © The SoBigData Consortium 2015.”

The information contained in this document represents the views of the SoBigData Consortium as of the date they are published. The SoBigData Consortium does not guarantee that any information contained herein is error-free, or up to date. THE SoBigData CONSORTIUM MAKES NO WARRANTIES, EXPRESS, IMPLIED, OR STATUTORY, BY PUBLISHING THIS DOCUMENT.

GLOSSARY

ABBREVIATION	DEFINITION
MOOC	Massive Open Online Course
OEB	Operational Ethics Board
RI	Research Infrastructure

TABLE OF CONTENT

DOCUMENT INFORMATION	2
DISCLAIMER	3
GLOSSARY	4
TABLE OF CONTENT	5
DELIVERABLE SUMMARY	6
EXECUTIVE SUMMARY	7
1 Relevance to SoBigData	8
1.1 Purpose of this document.....	8
1.1 Relevance to project objectives	8
1.2 SOBIGDATA project description	8
1.3 Relation to other workpackages	8
1.4 Structure of the document	9
2 Development of the MOOC.....	10
2.1 The Learning Units.....	11
2.1.1 First Aid for Responsible Data Scientists.....	11
2.1.2 Further Learning Units	18
2.1.3 Experiences with FAIR.....	20
3 The Operational EthicS Board	23
3.1 Composition	23
3.2 Purpose/Tasks	23
3.3 Task example.....	24

DELIVERABLE SUMMARY

This deliverable reports on the advances made in incorporating the legal and ethical framework into the infrastructure and practice of SoBigData. In particular, it shows two major steps that have been made over the past year: the institution of the MOOC, named FAIR, and that of the Operational Ethics Board.

FAIR (First Aid for Responsible data scientists) consists of a series of online training courses on the fundamentals of ethical and legal big data social science. It is aimed at supporting the researchers in conducting their work conscientiously in a time in which the norms concerning social science are subject to major upheavals. This report explains the main features of FAIR, in terms of form and content.

The Operational Ethics Board (OEB) allows for the new possibility of customised, short-term support to researchers applying for access to SoBigData database. It adds a new step in the access request process and it facilitates open discussion on the sometime extremely specific process of protecting the information of data subjects. This report explains its foundation and its mode of operation.

EXECUTIVE SUMMARY

The SoBigData project is aimed at the facilitating open, value-based big data social science. One of the elements of this is a legal and ethical framework, developed in year 1, which sets out legal guidelines and ethical principles that must be upheld in any data science project that is to call itself responsible. The activities of WP2 have been to translate this framework into institutional and practical measures.

Over the past year, two of such measures have been set up. On the one hand, a set of open training materials (a 'MOOC') has been created, intended to enable big data social science researchers to have an overall grasp of the principles that they are expected to abide by. On the other, an Operational Ethics Board (OEB) has been created, which allows for short-term and custom support to researchers applying for access to one more SoBigData database in resolving potential normative issues.

This deliverable outlines the details of these two new institutions: the reasons for their foundation and their mode of operation.

1 RELEVANCE TO SOBIGDATA

SoBigData proposes a research infrastructure (RI) for big data research, which deals with uploading, sharing and conducting research on social science big data. This endeavour puts various legal and ethical issues into play. This document describes the further development of two important elements of the RI with the purpose of ensuring that ethical and legal principles identified in previous work are being pursued: the MOOC named FAIR, a collection of online training materials for social science researchers, as well as the establishment of a new body: the Operational Ethics Board, which will function as a help desk for researchers pursuing research within the SoBigData RI.

1.1 PURPOSE OF THIS DOCUMENT

The purpose of this deliverable is first of all to explain the development of the MOOC, named FAIR, as an instrument that will help users to receive the required knowledge of ethical and legal issues in social data science. As such, this document gives an overview of the structure of FAIR and of its various modules.

Secondly, this document also reports on the Operational Ethics Board, a newly established institution within the RI that supports researchers in pursuing their research in an ethical and legal manner.

1.1 RELEVANCE TO PROJECT OBJECTIVES

SoBigData aims to ensure that social science research is pursued in an ethical and legal manner and that data scientists are given the appropriate information and guidance in order to conduct their research responsibly. The work presented in this deliverable constitutes a prominent key step, as it is as of this year that full course units for FAIR have been developed and tested.

The Operational Ethics Board provides a secondary instrument to achieve this end, supporting researchers in identifying legal and ethical issues and providing valuable advice.

1.2 SOBIGDATA PROJECT DESCRIPTION

“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”.

1.3 RELATION TO OTHER WORKPACKAGES

The work presented in this deliverable will contribute to the aims of WP6 Transnational Access and WP7 Virtual Access, by delivering help to researchers in estimating the legal and ethical issues.

1.4 STRUCTURE OF THE DOCUMENT

This deliverable is organized as follows:

Section 2 reports about the progress of FAIR, describing the overall concept and reporting about the individual learning units, as well as outlining further elements like the quiz parts and future plans.

Section 3 describes the purpose and tasks of the operational ethics board and gives a specific example of its work.

2 DEVELOPMENT OF THE MOOC

One of the main pillars of the SoBigData legal and ethical framework is creating awareness on side of the users. Knowing the legal risks and ethical issues enables data-driven research to involve stronger sense awareness about possible ethical consequences. Only by relying on these premises, it is possible to unlock the immense potential of big data analytics for both innovation and social good. To a large extent, however, legal and ethical issues relating to the use of personal data in social science is still a matter for specialists. Ideally, we would like everybody involved in research based on personal data to be aware and act with ethical and legal considerations in mind. This is why tools are required that spread awareness and information to everyone involved.

This has led to the construction of a specific online course that aims to provide the basis of the ethical issues involving the managing of personal data, especially regarding the access and the use of the SoBigData RI. The course is open to the public (through registration requiring only limited information) at the page <http://fair.sobigdata.eu/moodle/>, and it covers an overview of the main ethical problems, the definition and the obligations of a data controller and intellectual property law. Our plan is to further expand the course offerings, adding for example modules specific on research purpose and a summary of the main anonymization techniques. Each module has an associated quiz composed of 6-10 (depending on the length of the module) multiple-choice questions. The primary purpose of the quiz is to allow for self-testing the knowledge acquired by researchers. Indeed, there are no particularly tricky questions, but a general competence is required. Moreover, for each (wrong) answer, we provided immediate feedback, which explains why the answer is incorrect and makes a link available to the part of the lesson, that debates that question. Glossaries are also available, which explain the relevant ethical/legal terms.

The course is composed of a few lessons. The first one, named simply as the general course, is the core lesson of FAIR: it covers an overview of the main ethical problems, the definition and the obligation of a data controller and the intellectual property law. These topics are developed as three modules within the course. The purpose of the lesson is to show the responsibilities, the possibilities and the boundaries related to the data processing. Moreover, this is also a practical help to all the persons that want to take advantage of some of the SoBigData features, such as accessing the data catalog and applying for the transactional access (TNA). Indeed, this lesson will soon become mandatory for the RI registration and the request of the TNA. Due to this key role within the SoBigData project, we decide to provide the user who completed the lesson with a certification statement.

Other lessons are composed by a single module, due to the specific nature of the topics. At of today, we developed two additional lessons that are currently in the launch phase: a close up about the GDPR in the research applications and some further information about the impact of the GDPR on industries.

At any moment, during the lessons, pop-ups explaining in a few words a key concept are provided, through a glossary, which is accessible also on the home page of the lesson. Moreover, additional material and external links are made available to elaborate some specific aspects. Additional resources are also available in the main page, in order to ensure that users do not miss them.

2.1 THE LEARNING UNITS

2.1.1 FIRST AID FOR RESPONSIBLE DATA SCIENTISTS

2.1.1.1 UNIT 1 – ETHICAL FRAMEWORK FOR DATA SCIENTIST

The first unit is concerned with the fundamental ethical issues associated with big data social science.

Its content is based on the Ethical Framework developed earlier in the SoBigData project. It deals, in the first place, with the different negative effects that may result from ineffective handling of (sensitive) personal information, including direct harms as a result of a loss of privacy (e.g. identity theft), as well as more indirect effects such as discrimination.


ETHICAL CONSIDERATION OPENED UP BY DEALING WITH THIS KIND OF DATA:

- ☐ Citizens might lose control over what is known about them (for further explanations you can watch/read [Glenn Greenwald's conference](#)).
- ☐ The malicious use of data might reveal personal information, thus we have the responsibility to secure the data we use and limit the data we use.
- ☐ The use of data may be unjust when information is used outside of its original context (see also [Nissenbaum, 2004](#)).
- ☐ The use of data may perpetuate inequality, by causing asymmetries in who controls information and deepening societal divides.

Figure 1. Lesson part showing ethical considerations regarding the use of personal data

The unit goes on to suggest a number of solutions to these problems. It explains the importance of several principles that responsible data science should involve, while also emphasising that given the novelty of this field of research, many open issues still exist. That is why it sets out what guidance is available to the researcher in the SoBigData infrastructure. It informs the researcher on where and how to get the necessary information in resolving specific issues, be it through the information repositories created in the previous deliverables, or through the operational ethics board.

ETHICS IN SOBIGDATA

 **Ethics briefs.**

SoBigData provides specialized ethics briefs dealing with specific ethical questions for many data sets or methods. You will find:

- ☐ references to [relevant literature](#)
- ☐ practical examples of ethical problems in a research context
- ☐ contact information to specialists who can help you if you want to discuss open questions.

Integrating ethical consideration into your work starts with a general awareness.

Figure 2. Lesson part reporting about ethics briefs

All this establishes the foundations for the following lessons.

2.1.1.2 UNIT 2 – A JOURNEY THROUGH THE NEW EUROPEAN DATA PROTECTION LAW

This unit of FAIR gives an overview of the most important aspects of the General Data Protection Regulation (GDPR).

At the beginning, it gives an overview of the present legal situation with respect to data protection law as well as a short look back on the legal situation under the Data Protection Directive with an explanation added of what the main differences between the European legal instruments “directive” and “regulation” are.

Secondly, the user receives an explanation what kind of data are to be considered personal data in a legal sense, which shall enable the user to make an own assessment with regard to a specific data set. However, as it can be extremely difficult to make such an assessment, the researcher is also advised to consult the responsible data protection officer/ ethics committee.

Thirdly, the user is acquainted with the concept of the data controller. Again, the advice is given to ask the responsible data protection officer/ethics committee in case of doubts.

Fourthly, the scope of application of the GDPR is explained and the importance of where the establishment of the data controller is situated.

Then, the user is encouraged to get on with the lesson, even if he or she believes himself or herself not to be a data controller (as they might be wrong in their assumption).

In the following, the user is acquainted with the main data protection principles for processing of personal data:

Exploring the European General Data Protection Regulation (GDPR)

Main data protection law principles



You have completed 92% of the lesson
92%

Figure 3. Lesson part about the data protection principles laid down in the GDPR

Thereafter, the user also receives an overview of the obligations that are imposed upon the data controller by the GDPR.

Finally, an explanation follows why the full services of the platform cannot be used by all users throughout the world as not every country provides a level of data protection similar to the EU and a non-restricted transfer of data sets would undermine the protective effect by the GDPR.

The unit ends with a quiz and a case study.

First Aid for Responsible Data Scientists

Home My courses First Aid for Responsible Data Scientists Unit 2- A journey through the new European Data Protection Law Unit Quiz Number 2!

Question 1
Correct
Marked out of 1
Flag question

Which is the main condition for the application of the data protection law?

Select one:

☐ Public importance data processing

☒ Personal data processing ✓

Correct, if your company/research institution process and hold **personal data** of subjects residing in the European Union, regardless of the company's location, you should take in account the rules laid down in the current data protection law.

☐ Every data processing

Question 2
Incorrect
Marked out of 1
Flag question

Who shall be responsible for the principles related to processing personal data?

Select one:

☐ The natural person

☐ The controller

☒ The processor ✗

Your answer is incorrect, the Regulation clearly distinguishes between the obligations placed upon controllers and processors. In the GDPR, Art. 5(2) you can read that the controller shall be responsible for, and be able to demonstrate compliance with, the principles relating to processing of personal data cited in the previous article paragraph ('accountability'). You can also remember these principles reading again this First Lesson section.

Next page

QUIZ NAVIGATION

1 2 3 4 5 6 7 8 9 10

Finish attempt...

Figure 4. Data Protection quiz part

2.1.1.3 UNIT 3 – INTELLECTUAL PROPERTY RIGHTS AND SOCIAL MEDIA CONTENT

This unit provides knowledge about intellectual property rights and social media content.

To begin with, the user receives an explanation what intellectual property is. The use of specific examples shall enable users to receive a better picture of the concept of intellectual property.

Intellectual Property Rights and Social Media Content

The Intellectual Property

WHAT IS INTELLECTUAL PROPERTY?

Intellectual property (indicated with the acronym **IP**) refers to **creations of the mind**, such as **inventions, literary and artistic works, designs and symbols, names and images used in commerce**.

Look [here](#) for more an detailed explanations.

Every original creation in the literary, scientific and/or artistic domain, whatever may be the mode or form of expression, such as:

- ☐ writings,
- ☐ photographic works,
- ☐ melodies,
- ☐ works of applied art,
- ☐ illustrations,
- ☐ maps,
- ☐ plans,
- ☐ sketches
- ☐ three-dimensional works

relative to geography, topography, architecture or science constitute subject matter protected by **copyright**.

Previous Next

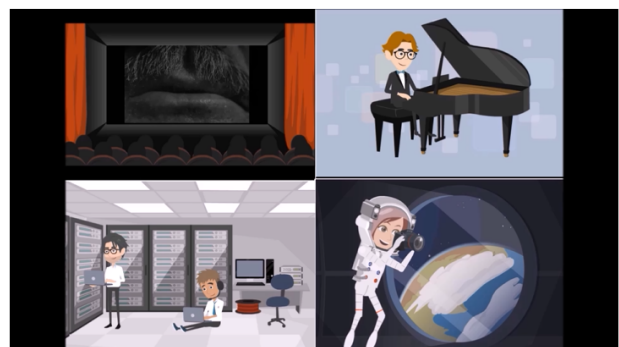


Figure 5. Lesson part explaining the concept of intellectual property

Next it is explained to the user how intellectual property regulations impact research within the SoBigData RI. Research within SoBigData that provides access to social media content often involves copyright regulation. The actions subjected to copyright (e.g. distribution and sharing) are presented.

Intellectual Property Rights and Social Media Content Actions that matter for Copyright

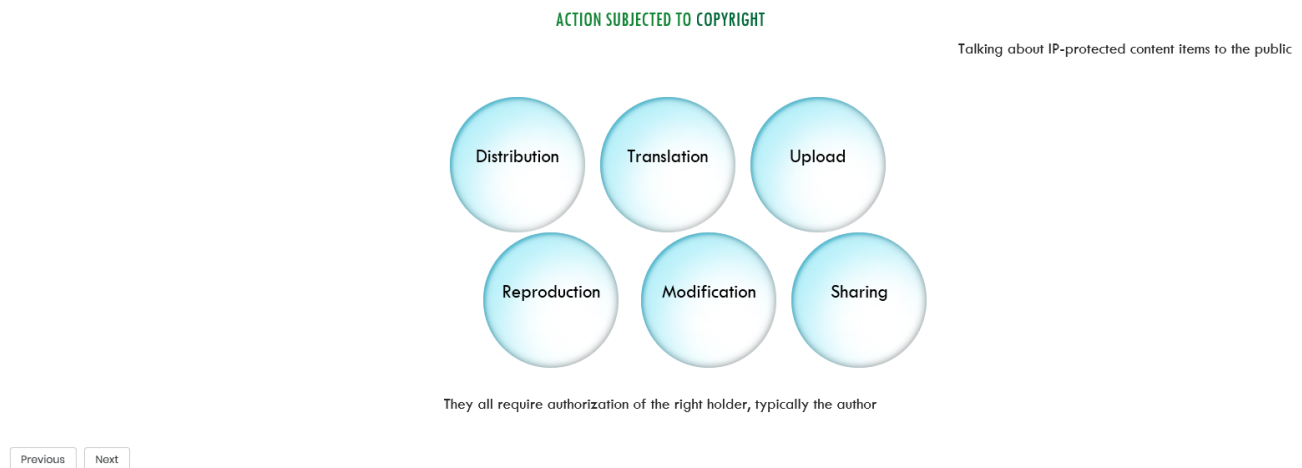
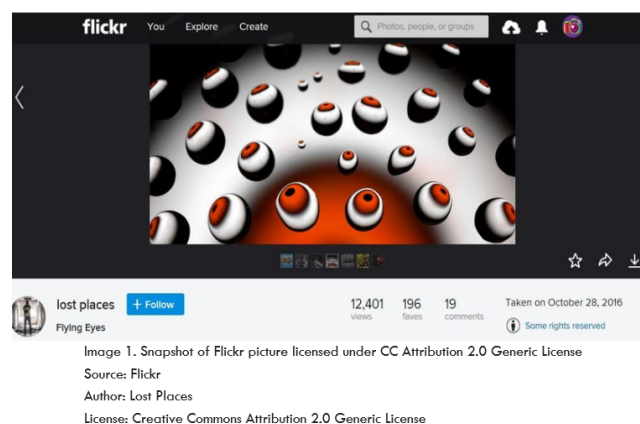


Figure 6. Lesson part explaining actions subject to copyright

Further on, the unit also considers under what terms IP-protected content may be used.

LET'S SEE UNDER WHAT TERMS IP-PROTECTED CONTENT MAY BE USED

For example, an author may upload pictures to Flickr under Creative Commons (CC) share alike license or "all rights reserved" or "all possible uses allowed". The definition of the CC licence is given in the next page.



Previous Next

Figure 7. Lesson part explaining license terms

Intellectual Property Rights and Social Media Content

Possible actions

UNDER CC ATTRIBUTION 2.0 GENERIC LICENSE YOU ARE FREE TO:

SHARE copy and redistribute the material in any medium or format

ADAPT remix, transform, and build upon the material for any purpose, even commercially

UNDER THE FOLLOWING TERMS:

ATTRIBUTION You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

Previous Next

Figure 8. Lesson part explaining what researchers are allowed to under CC Attribution2.0 Generic License

The unit also refers to the meta-data provided by SoBigData RI catalogue, which inform the user which terms govern the use of a specific data set. Users will have to make sure their data use is in compliance with those terms.

Intellectual Property Rights and Social Media Content

SoBigData Rules

HERE YOU FIND WHAT TERMS WE ARE TALKING ABOUT

The access to the SoBigData catalogue works via registration as a D4Science user and acceptance of D4Science Terms available at: <https://services.d4science.org/terms-of-use>.

The D4Science Terms of Use govern in principle the use of D4Science services, such as: Email Service, Workspace Service, Social Service, etc.

However, the use of individual datasets remains governed by the terms of individual datasets.

Section **Copyright** of the D4Science Terms provides:

Copyright

The data and information made available through the D4Science.org Gateway are available under terms described in the **metadata** accompanying each product (e.g. "license" or "constraints" field). Except where otherwise noted (i.e. in case of primary requirement to comply with the data provider's policy) this is the Creative Commons License.

By accessing the data made available by the D4Science Gateway, users agree to the related Terms of Use.

Please also be responsible for the research you make and be respectful of the rights of others parties. By doing so, each user contributes to the community of ethical research.

If you want to do your research in a way that your rights are respected please respect the rights of the others as well.

Previous Next

Figure 9. Lesson part explaining the role of the meta-data

At the end of this unit users are also requested to take part in a quiz to validate their knowledge.

First Aid for Responsible Data Scientists

[Home](#) [My courses](#) [First Aid for Responsible Data Scientists](#) [UNIT 3 – INTELLECTUAL PROPERTY RIGHTS AND SOCIAL MEDIA CONTENT](#) [Unit Quiz number 3!](#)

Question 5
Correct
Mark 100 out of 100
Flag question

What items of social media content can be considered as IP-protected content?

Select one or more:
☒ Music files ✓
☐ Call Data Records
☐ IP-addresses
☒ Posts ✓
☒ Individual symbols, letters, numbers ✓
☒ Photos ✓
☐ Geo-data
☒ Videos ✓

Question 6
Correct
Mark 100 out of 100
Flag question

Under what terms you may use Twitter content as made available via D4Science Infrastructure?

Select one:
☐ D4Science Terms
☒ Twitter Terms ✓
☐ Terms dictated by the researcher providing the data
☐ Metadata

Finish attempt...

QUIZ NAVIGATION

1

2

3

4

5

6

Finish attempt...

Figure 10. Intellectual property quiz part

2.1.1.4 PROVIDING FURTHER RESSOURCES

Aside from the learning units, further resources on ethical and legal aspects are provided. Users will, for example, be able to find ethics briefs for various data sets (for the ethics briefs concept see D2.3). Users will also be able to find various references to articles on specific legal and ethical issues as well as links to the submitted deliverables from WP2 in case they are interested in learning more about legal and ethical concepts and regulations in the context of research with social data.

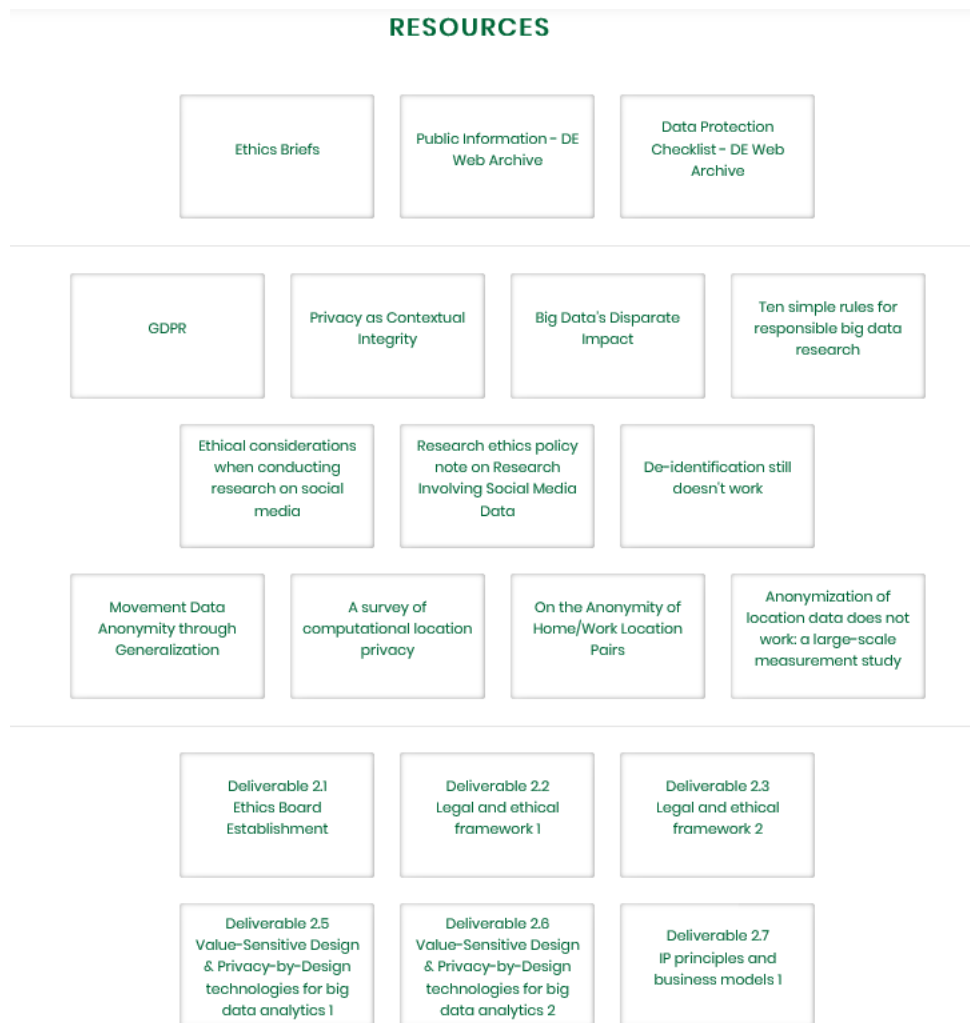


Figure 11. Further resources provided in the MOOC

2.1.2 FURTHER LEARNING UNITS

The second lesson in FAIR is focused on industries. In particular, it lists the legal rights that must be ensured, and the measures that companies shall deploy to be GDPR compliant, e.g. with regard to the right to data portability and the role of a data protection officer.

The impact of GDPR on European Companies
GDPR Compliance for Business

GDPR Compliance for Business

Introduction

As we said, the General Data Protection Regulation came on effect on 25th May 2018.

The regulation affects all the companies collecting and processing personal data of European citizens, and it will grant the European users control over their personal data across Europe.

It is important to point out there will be no distinction between personal data about individuals in their private, public or corporate roles.



Credits: Freepik

Previous
Next

LESSON MENU

- Introduction
- Guarantees for Individual
- Companies measures
- Consent
- Cross border data transfer
- Profiling
- Data portability
- Pseudonymisation
- Mandatory data protection officer requirement
- Cybersecurity and data breach notification
- Codes of conduct and certification
- Consequences for violation
- The End!

Figure 12. Lesson part of GDPR compliance for business

First Aid for GDPR & Data in Business

Home
My courses
First Aid for GDPR & Data in Business
The impact of GDPR on European Companies
GDPR for Business Test

Question 5
Incorrect
Mark 0.00 out of 1.00
Flag question

Which is the role of a DPO?

Select one:

- ☐ The DPO must support the processor in dealing with data breach
- ☐ The DPO must deal with all data protection matters affecting the controller or processor properly and in a timely manner
- ☒ The GDPR does not set out a specific role for DPOs to fulfil **✗ Your answer is incorrect: the GDPR does not establish specific credential to become a DPO, but it list the responsibilities and the role of DPOs. Please, read again the lesson.**

Question 6
Partially correct
Mark 0.30 out of 1.00
Flag question

What does "data breach" mean?

Select one:

- ☒ A data breach is a hacker attack in which sensitive, protected or confidential data has potentially been viewed, stolen or used by an individual unauthorized to do so **✗ Your answer is incorrect: a data breach can occur even without an active attack. Please, read again the lesson.**
- ☐ A data breach is a processing which is not based on data subject's consent or on Union or Member State Law
- ☐ A data breach is an incident in which sensitive, protected or confidential data has potentially been viewed, stolen or used by an individual unauthorized to do so

Previous page
Next page

QUIZ NAVIGATION

1	2	3	4	5	6	7	8
✓	✓	✓	✓	✗	✗		

Finish attempt...

Figure 13. GDPR compliance for business quiz

The third lesson is focused on the research environment, in particular, about the exceptions that the GDPR admits for researchers and public utility. As researchers, we know that we have a privileged role regarding the possibilities of managing personal data. Nevertheless, we also want to point out which are our responsibilities and the safeguards we need to guarantee.

First GDPR Aid for Researchers

Home > My courses > First GDPR Aid for Researchers > Scientific Research in the GDPR

Scientific Research in the GDPR

Scientific Research in the GDPR

Scientific Research

WHAT SCIENTIFIC RESEARCH IS?

The **GDPR** does not define “scientific research” but **Recital 159** states that the definition should be “interpreted in a broad manner”: it includes for example **“technological development and demonstration, fundamental research, applied research and privately funded research”**, as well as “studies conducted in the public interest in the area of public health”.

It is important to point out that the **GDPR** provisions do not distinguish “public interest” from “commercial research”. Therefore, data protection officers and advisory bodies shall answer to this critical interpretative question shortly.

Previous Next

LESSON MENU

- The old directive
- The new Regulation
- Scientific Research
- Purpose of collection
- Legitimate interest
- Special categories
- Safeguards
- The end!

Figure 14. Lesson part First GDPR AID for Researchers

First GDPR Aid for Researchers

Home > Courses > First GDPR Aid for Researchers > Scientific Research in the GDPR > Scientific Quiz > Preview

Scientific Quiz

Question 1
Partially correct
Mark 0.50 out of 1.00
Flag question
Edit question

Does the 95/46 Directive allow processing data for historical, statistic and scientific research?

Select one:

- ☐ Yes, collecting data for historical, scientific and statistic purposes shall not be considered as incompatible provided that Member States shall provide appropriate safeguards
- ☐ No, the Directive does not mention processing data for historical, statistic and scientific research
- ☒ Yes, but data subject must give his/her explicit consent. **The Directive described a specific exception for historical, statistic and scientific research regarding the necessary information and the access to the data. The key point, in the Directive, was the obligation of providing appropriate safeguards.**

Question 2
Correct
Mark 1.00 out of 1.00
Flag question
Edit question

Does the GDPR define "scientific research"?

Select one:

- ☐ No, but Recital 159 states that it must be "considered in a strict manner"
- ☒ No, but Recital 159 states the definition should be "interpreted in a broad manner" ✓

QUIZ NAVIGATION

1 2 3 4 5 6

Finish attempt ...

Start a new preview

Figure 15. First GDPR AID for Researchers quiz part

UNIPi is planning to add further learning modules to FAIR in the future.

2.1.3 EXPERIENCES WITH FAIR

At the current date, 160 users registered to FAIR, even if some of them do not complete the course.

The test on the Ethics Unit was attempted 109 times, from 80 different users. The following figure shows the grade of the last attempt (the used evaluation criteria) for each user. As one can see, only one

participant withdrew from the quiz, while the others received a very good average judgment (in every quiz, we fixed a threshold to 60% in order to consider the quiz passed).

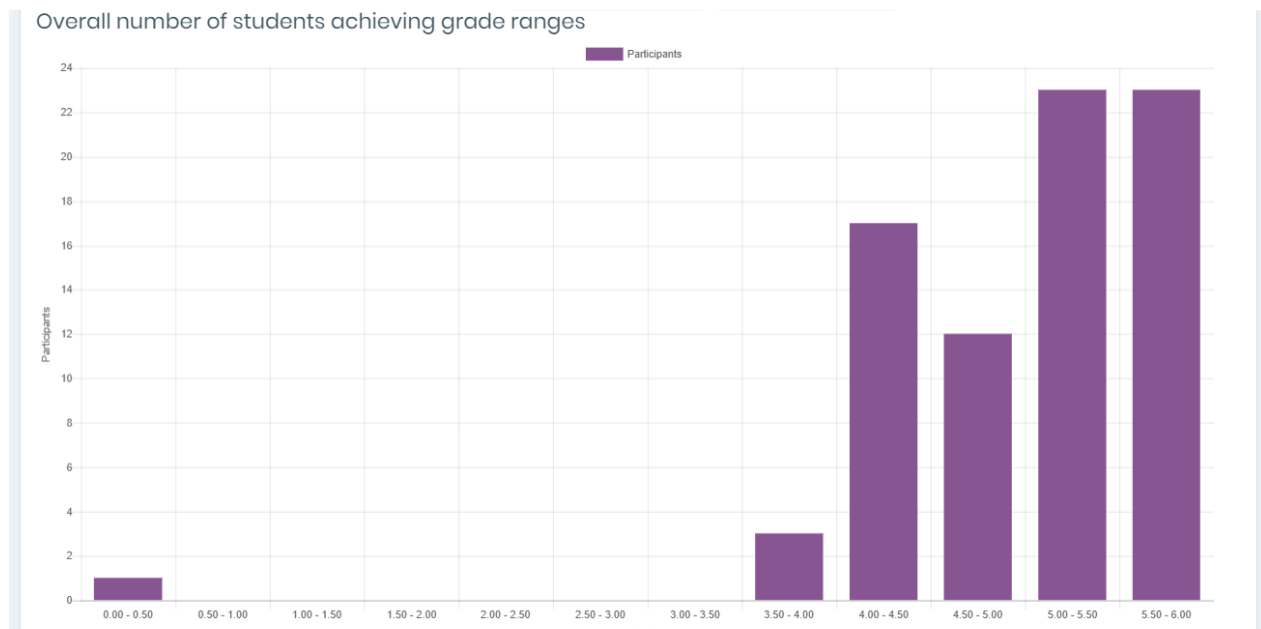


Figure 16. Ethics Quiz Statistics

The Data Protection test was attempted 101 times, from 71 different participants. Here, the grade distribution is wider due to the higher number of questions (10 instead of 6), but the overall behavior is similar.

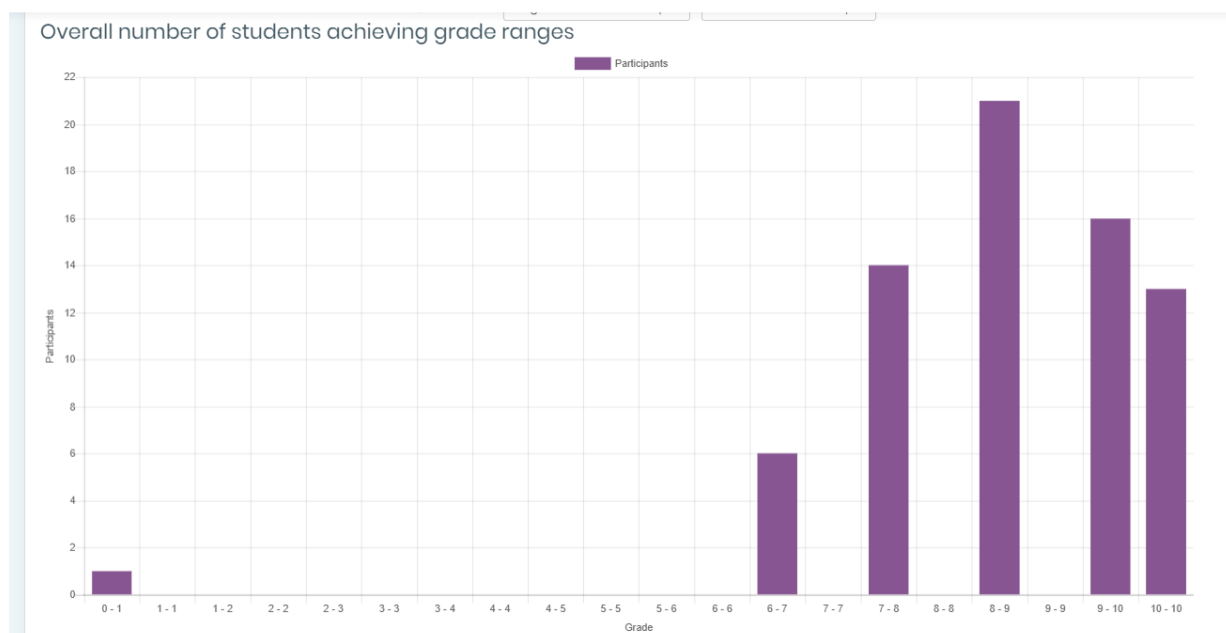


Figure 17. Data Protection Quiz Statistics

Regarding the Intellectual Property Unit, we have 102 attempts from 66 participants, with a general grade distribution, which is similar to the previous ones.

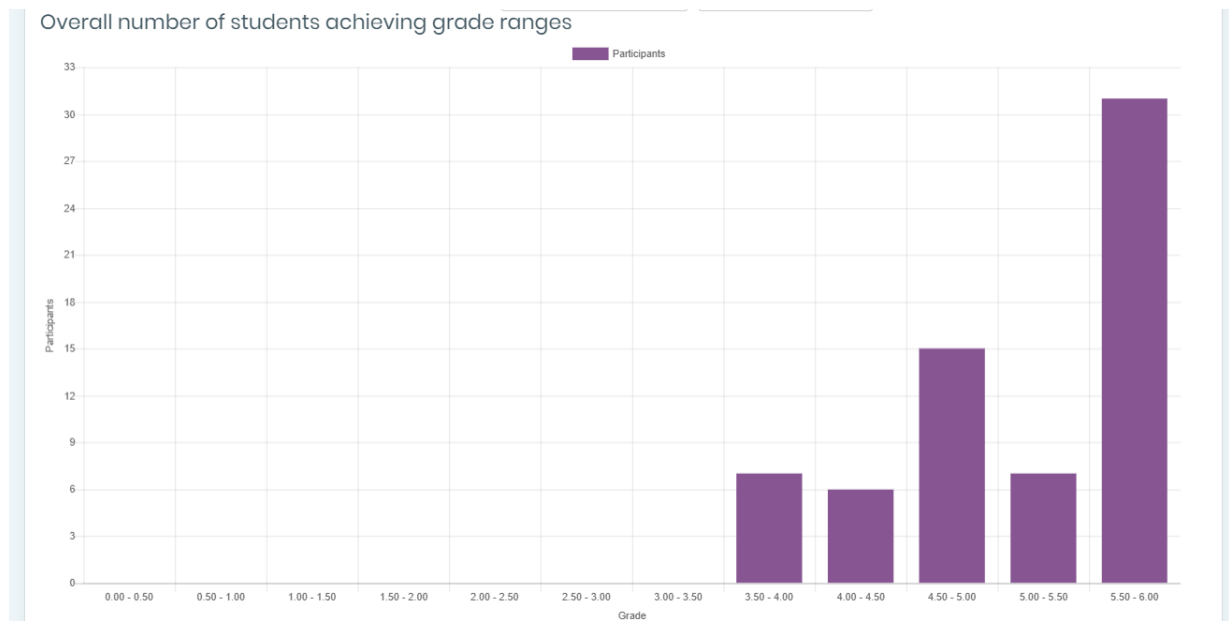


Figure 18. Intellectual Property Quiz Statistics

In general, we saw that the participation remains quite stable if the participants are motivated enough. Indeed, a significant number of users only register on the platform, maybe just for the curiosity due to the dissemination we have done. The real interested users, i.e., who started to test themselves, in the vast majority of cases, will complete the whole course.

3 THE OPERATIONAL ETHICS BOARD

3.1 COMPOSITION

The operational ethics board consists of an interdisciplinary group of researchers, knowledgeable in data science (Francesca Pratesi), law (Francesca Donati) and ethics (David van Putten). The need for an operational ethics board arose from the consideration that the ethics board, consisting of senior researchers, can focus on critical issues, leaving the day-to-day ethical support to the operational ethics board.

Francesca Pratesi is a post-doc at the Computer Science Department of the University of Pisa and a member of the KDD-LAB, a joint research group with the Information Science and Technology Institute of the National Research Council in Pisa. She received her MS degree and Ph.D. degree in Computer Science from the University of Pisa in 2013 and 2017. Her research is in privacy-aware data mining, data privacy, spatio-temporal data analysis and privacy risk evaluation.

Maria Francesca Donati is an Italian lawyer specialized in data protection. In 2017 she attended a Master in "Internet Ecosystem: Governance and Rights" at the University of Pisa. Currently, she offers data protection legal advice to corporations, in particular for GDPR compliance.

David van Putten is a PhD student at the Erasmus University in Rotterdam. Over the past two years, he had been affiliated with the TU Delft in contributing to the Legal and Ethical Framework of SoBigData. His research is in digital ethics and philosophy of economics.

3.2 PURPOSE/TASKS

The aim of the OEB is thus to allow for short-term assessment of the ethical problems faced by the researchers applying to SoBigData datasets. The OEB reviews incoming access requests on the basis of a number of possible problems and, if necessary, engages in a dialogue with the researcher on the best way to tackle the raised ethical issues. As such, the OEB works in tandem with the ethics boards of individual institutions. The OEB can be directly contacted via e-mail, addressed to info.ethicsboard@sobigdata.eu. The sender should specify his/her institution, the problem he/she wants to address, and the ethical issues he/she faced, possibly along with a proposal of a solution. The answer will be provided as advice to point out the primary ethical concerns about the specific case or approval of the mode of operation of the writer.

Moreover, the OEB is an integral part of the process of the transactional access. Previously, the WP 6 asked for a waiver that certifies the scope and the methods of the applicants from an ethical point of view. Recently, it was decided to transfer this control internally to the SoBigData project. Thus, now, each application is reviewed from a technical and scientific point of view and, then, a request for ethical approval is sent to the OEB.

This approach has several advantages: firstly, we are making the ethical guidelines more uniform, since they are coordinated by the OEB; secondly, we are making it cleared the process, since the hosting institutions know the persons who certified the applicants; thirdly, we are speeding up the process, since the applicants

usually have only a vague idea of the presence of ethics boards within their own institutions, these boards are often busy and not easy to contact.

3.3 TASK EXAMPLE

A researcher may have the idea of analysing consumer behaviour. Through the SoBigData infrastructure, (s)he applies for access to a dataset containing consumer data, a log of transactions collected by a supermarket for example. Several risks might exist. The transactions could reveal personal or sensitive information about the consumers in question. (e.g. through inferences made on the basis of what they are purchasing.) The owner of the dataset might impose certain restrictions. The GDPR or national legislations might contain a clause affecting the research in question. And so on.

The task of the OEB is to review the application and deliver feedback. If the risks are severe enough, the ethics board might advise against conducting the research in question. In some cases, the issues may be resolved with a remedial action. The researcher might need to sign a non-disclosure agreement, or apply an algorithm that anonymises or sanitises the datasets. In other cases, the research may proceed without any noticeable issues.

It is thus the task of the OEB to advise the researcher on possible problems *and* possible solutions. A subsidiary activity of the OEB is therefore to establish a database of case information, which facilitates this process.