



# PROTEUS

Scalable online machine learning for predictive analytics and real-time interactive visualization

687691

---

## D2.4 Proteus Data Management and Ethics Plan

---

**Lead Author: Rachel Finn, Graham Hesketh, Giuseppe Maio & Kush Wadhwa, Trilateral Research**

**Reviewers: Hamid Bouchachia (BU) & Marta Garcia Barzana (AMIII)**

Deliverable nature:	Report (R)
Dissemination level: (Confidentiality)	Consortium (CO)
Contractual delivery date:	30/11/2018
Actual delivery date:	30/11/2018
Version:	1.5
Total number of pages:	46
Keywords:	Data management, exploitation, research ethics, open access

**Abstract**

This deliverable represents the final iteration of the PROTEUS data management and ethics plan, a comprehensive document that described how the data used by and generated within the project was managed, and how it will be managed after the close of the project. This report includes a consideration of the legal, ethical and policy issues relevant to PROTEUS research and the management of project data and how the project has met and will meet these obligations.

## Executive summary

This deliverable represents the final iteration of the PROTEUS data management and ethics plan, a comprehensive document that described how the data used by and generated within the project was managed, and how it will be managed after the close of the project.

Data has always been an integral part of the scientific research process and is an object of scientific interest both as an object of investigation and as a potential asset for those within and outside the research process. Data management assists researchers in outlining how data needs to be protected and making informed decisions about exploiting research data or sharing them to allow others to exploit them.

Particular policies, legal frameworks and ethical guidance relevant to the management of research data frame these issues. As such, this document begins by providing foundational and contextual information related to the project, the relevant policies and legal frameworks and the initial plans for ensuring that the collection and use of data within the project conforms to the issues within this larger context. Specifically, the document examines the European Commission's Open Data Research Pilot, intellectual property rights relevant to PROTEUS, data protection standards and practices and guidance on ethical research processes.

The project used and generated three types of data

- a) Raw, numeric data from the company's databases that formed the basis for the tools developed by the PROTEUS project.
- b) Data derived from the algorithms that was used to validate the algorithms and fed into the visualisation tools to enable visualisation of the results of the analytics.
- c) Evaluations, opinions and insights by company employees, who volunteered to evaluate the utility of the tools in the operational context of ArcelorMittal.

The first data set has always raised significant intellectual property issues as it has commercial sensitivity for ArcelorMittal. The second and third data sets were generated by the project and raised both intellectual property issues for AMIII and data protection and ethical research issues via the PROTEUS evaluation data. Specifically, the usability evaluation exercises were conducted with a volunteer who was also an ArcelorMittal employee, and thus, issues around voluntary participation, informed consent and data protection were relevant here.

The deliverable progresses by describing how the project has met all of our legal, ethical and policy obligations surrounding PROTEUS data. In doing so, the project has agreed the following principles:

- Data owned by ArcelorMittal was shared with consortium members, although consortium members can only access this data through the project coordinator
- PROTEUS partners agree to respect the commercial confidence of the data provided by AMIII
- Human participants in PROTEUS research activities were under no obligation to participate and their involvement will be strictly voluntary
- PROTEUS partners have respected and will respect data protection and ethical research principles related to the following:
  - Participant confidentiality and anonymisation
  - Informed consent
  - Data minimisation
  - Purpose limitation
  - Transparency
  - Rights of access, correction and erasure
- Personal data that can identify the research participants will be stored by AMIII securely
- The coordinator will hold all commercially sensitive data will manage access to this data for PROTEUS partners

The project has achieved compliance with European and Spanish regulations, and registration with the Spanish Data Protection Authority was undertaken by AMIII as part of their normal business activities. Any personal data relating to AMIII employees stayed within the confines of AMIII, who processed the data using their employment contract with participants as the legal basis for processing. Furthermore, the pilot activities were

restricted to one key informant, meaning there was no systematic collection and processing of personal data. Nevertheless, the participant received the PROTEUS information sheet and signed informed consent forms (included in Annex A). An independent ethical and data protection expert, Ricard Martinez, has reviewed these materials and added comments and suggestions to ensure the materials and procedures are as robust as possible (see Annex B for his evaluation and Annex C for his short CV).

While much of the information in this document is based on previous versions of the data management plan, some key issues have emerged in this final version, specifically that the consortium has agreed that project data can only be used after the close of the project for dissemination purposes and that the data will not be made openly accessible after the close of the project due to the intellectual property issues associated with the data and the small size of the evaluation data.

## Document Information

<b>IST Project Number</b>	687691	<b>Acronym</b>	PROTEUS
<b>Full Title</b>	Scalable online machine learning for predictive analytics and real-time interactive visualization		
<b>Project URL</b>	http://proteus-bigdata.com		
<b>EU Project Officer</b>	Martina EYDNER		

<b>Deliverable</b>	<b>Number</b>	D2.4	<b>Title</b>	PROTEUS data management and ethics plan
<b>Work Package</b>	<b>Number</b>	WP2	<b>Title</b>	Industrial case: Requirements, challenges, validation and demonstration

<b>Date of Delivery</b>	<b>Contractual</b>	M36	<b>Actual</b>	M36
<b>Status</b>	version 1.5		final <input type="checkbox"/>	
<b>Nature</b>	report <input checked="" type="checkbox"/> demonstrator <input type="checkbox"/> other <input type="checkbox"/>			
<b>Dissemination level</b>	public <input type="checkbox"/> restricted <input checked="" type="checkbox"/>			

<b>Authors (Partner)</b>	Rachel Finn, Graham Hesketh, Giuseppe Maio, Kush Wadhwa (Trilateral Research)			
<b>Responsible Author</b>	<b>Name</b>	Rachel Finn	<b>E-mail</b>	Rachel.finn@trilateralresearch.com
	<b>Partner</b>	Trilateral Research	<b>Phone</b>	+44 (0)207 559 3550

<b>Abstract (for dissemination)</b>	This deliverable represents the final iteration of the PROTEUS data management plan, a comprehensive document that described how the data used by and generated within the project was managed, and how it will be managed after the close of the project. This plan includes a consideration of the legal, ethical and policy issues relevant to the management of PROTEUS data and how the project will meet these obligations. It also outlines plans for further developing this strategy to guide partners in exploiting PROTEUS data, including providing open access to the data to allow other stakeholders to re-use it.
<b>Keywords</b>	Data management, exploitation, research ethics, open access

<b>Version Log</b>			
<b>Issue Date</b>	<b>Rev. No.</b>	<b>Author</b>	<b>Change</b>
10/5/16	0.3	Rachel Finn	Draft circulated to project partners
19/5/16	1.0	Rachel Finn	Changes from reviewers incorporated
28/4/17	1.1	Tally Hatzakis, Rachel Finn	Updates to the document given project progress over Y1.
15/5/17	1.2	Ricard Martinez	Comments on the draft
29/5/17	1.3	Tally Hatzakis, Rachel Finn & Kush Wadhwa	Revision to the document following Ricard Martinez' review

20/11/18	1.4	Rachel Finn, Graham Hesketh, Giuseppe Maio, Kush Wadhwa	Updates to the document given project progress to completion.
29/11/18	1.5	Rachel Finn, Graham Hesketh, Giuseppe Maio, Kush Wadhwa	Updates to the document after internal review by PROTEUS partners.

## Table of Contents

Executive summary .....	3
Document Information .....	5
Table of Contents .....	7
Abbreviations .....	8
1 Introduction.....	9
2 Project description .....	11
3 Relevant policies.....	12
3.1 EC open research data pilot.....	12
3.2 Intellectual property rights.....	13
3.2.1 Trade secrets.....	13
3.2.2 Copyright and database rights .....	13
3.3 Data protection law.....	14
3.3.1 General Data Protection Regulation (679/2018).....	14
3.3.2 Spanish data protection law and regulatory requirements.....	15
3.4 Ethical research guidance .....	15
4 Data set description.....	17
4.1 Data collection and characteristics .....	17
4.1.1 Personal data collection.....	20
5 Ethical and legal issues .....	21
5.1 Participant selection and informed consent.....	21
5.2 Personal data protection.....	22
5.3 Intellectual property rights.....	23
6 Data governance.....	24
6.1 Access.....	24
6.2 Storage and processing .....	24
6.3 Sharing.....	24
7 Standards and metadata.....	25
8 Data exploitation.....	26
9 Long-term archiving and preservation (including open access).....	27
10 Conclusion.....	28
References.....	29
Annex A – PROTEUS informed consent and information sheets .....	30
PROTEUS Information sheet .....	30
PROTEUS Informed consent sheet .....	32
Annex B – Ricard Martinez Review report.....	34
1. Executive summary .....	35

2. Introduction .....	36
3. General Issues .....	37
3.1 The Right to data protection a European Fundamental right.....	37
3.2 A risk analysis under General Data Protection Regulation (GDPR).....	37
4. Ethical and Legal Issues .....	39
4.1 Excluded processes.....	39
4.2 The processing of personal data in the framework of labour relations. ....	39
4.3 Pseudonymization and anonymization. ....	41
4.4 Some considerations about safety. ....	43
5.1 RECOMMENDATIONS .....	44
Annex C - Ricard Martinez CV .....	45
Annex D – List of changes after Ricard Martinez review .....	46

## Abbreviations

**AEPD:** Agencia Española de Protección de Datos

**AMIII:** ArcelorMittal

**BSA:** British Sociological Association

**DPIA:** Data Protection Impact Assessment

**EC:** European Commission

**GDPR:** General Data Protection Regulation

**ICT:** Information and Communication Technologies

**IPR:** Intellectual Property Rights

**ORD Pilot:** Open Research Data Pilot



# 1 Introduction

Data has always been an integral part of the scientific research process and is an object of scientific interest both as an object of investigation and as a potential asset for those within and outside the research process. Scientists have long been engaging in the protection of their data assets from unauthorised access and sharing and the dissemination of their data assets for use by others. However, disciplinary differences and traditions have often governed whether data was kept under lock and key or shared more widely with other scientists and stakeholders (Wessels, et al., 2014). However, recent policy advances in the field of open access have encouraged scientists and other researchers to consider managing this asset more explicitly, including assessing whether the data needs to be protected or whether it can be shared more widely to enable others to exploit already existing data for scientific and other gains. Rather than relying on tradition, all researchers are being encouraged to make an informed decision about exploiting research data or sharing them to allow others to exploit them.

In this context, research data has a specific focus and meaning. According to the European Commission, who funds this research, “research data” refers to “information, in particular facts or numbers), collected to be examined and considered and as a basis for reasoning, discussion, or calculation (EC, 2016, p. 3).” This may include data as varied as statistics, interview recordings or notes from field observations (ibid.).

This document will assist the consortium to identify:

- a) what data needs to be protected, and
- b) what data might be sharable outside of the consortium.

Data management plans are an important tool to enable this consideration. Data management plans encourage researchers to consider a range of different legal and ethical issues that govern whether data needs to be protected and the extent to which it can be made openly accessible or shared in other ways. Legal and ethical issues to be considered include open access policies, intellectual property rights, data protection legislation and research ethics. Data management plans enable researchers to consider these and to outline their compliance with each of these requirements. In addition, proper data management also encourages researchers to explicate how their data is being collected, processed, stored and exploited. According to Jones (2011), such planning helps researchers to check they have the necessary support for their research and enables them to make sound decisions with a clear understanding of the different options.

The PROTEUS project developed a predictive analytics toolset to help companies identify defects early in the manufacturing process to reduce economic and environmental waste; hence improving companies’ efficiency and competitiveness, as well as environmental friendliness. The work has been based on the steel coil manufacturing process of ArcelorMittal, a large steel manufacturing company. To this end, the project used and generated three types of data

- d) Raw, numeric data from the company’s databases that formed the basis for the tools developed by the PROTEUS project.
- e) Data derived from the algorithms that was used to validate the algorithms and fed into the visualisation tools to enable visualisation of the results of the analytics.
- f) Evaluations, opinions and insights by company employees, who volunteered to evaluate the utility of the tools in the operational context of ArcelorMittal.

The project has considered sensitivities in the management of each of these data types, using this data management plan. First, the project needed to consider intellectual property rights, including trade secrets, as the data upon which we relied to build the PROTEUS toolset is owned by ArcelorMittal and is commercially sensitive. Second, the project generated new forms of data and insights through big data analytics tools, the ownership of which is outlined in the Consortium agreement and has been revisited within this document as the data and the project matured. Third, PROTEUS undertook evaluative research with company employees, in the process of which it collected some personal data. Hence the project has been guided by the EU research ethics and data protection framework to respect participants’ rights alongside sensitivities raised by doing research with employees of a specific company. Finally, the project sits within the European Commission’s Horizon 2020 Open Data Research Pilot, and consequently, the project did consider the extent to which the data from this publicly funded research can be made open access. However, the consortium concluded this was not possible given the intellectual property issues surrounding the data.

As such, this data management plan has formed a guidance document for the project. It assisted the project to navigate intellectual property and data protection issues and govern whether and when the consortium can provide open access to research data. The data management plan has also provided a template that will help the partners utilise and protect the data effectively during and at the end of the project. Finally, this document provided essential guidance on ensuring adequate protections for those participating in research, including providing personal data.

While these procedures were being co-managed by the project coordinator(s), Treelogic and finally Bournemouth University, and Trilateral Research, providing such protections is the responsibility of every partner in the PROTEUS project. This document provided guidance to all partners to help them effectively manage PROTEUS data. It has been continually updated as the project progresses. This final version is a further revision of the version produced in M18 of the project, prior to the evaluation. The UK's Data Curation Centre (2016) has the most comprehensive guidance on data management practices and the creation of a data management plan, and this document relies heavily on their advice and good practice guidance.

## 2 Project description

The PROTEUS project developed a data analytics system for steel company ArcelorMittal to help the company identify defects in the steel coils they manufacture earlier in the manufacturing process. At present, these defects are discovered after manufacturing the coils, when they are inspected for quality and thickness, and the post-production discovery of defective coils has a significant economic impact. Being able to detect defective coils during the manufacturing process will enable earlier intervention, and ultimately, fewer defects and less economic and environmental waste.

PROTEUS tested whether it is possible to build a predictive analytics tools using ArcelorMittal production data, and which variables are relevant for identifying manufacturing defects. The predictive analytics system relied on new scalable online machine learning techniques. These were developed by the project and tested in the ArcelorMittal industrial environment. The project also developed visualization methods and tools to help manufacturers identify defects and respond to alerts.

The PROTEUS project used the large volume and variety of data collected during the manufacturing process to build the predictive platform. The project used sensor data measuring approximately 7440 variables across the production process. These were combined with data related to the eventual thickness and quality of the steel coil. The processing of these data was essential to build the scalable online machine learning platform that will assist ArcelorMittal in augmenting their efficiency and competitiveness by identifying defective steel coils earlier in the manufacturing process.

### 3 Relevant policies

As mentioned in the introduction, there are a number of legal requirements and policies relevant to the management of PROTEUS data. First, the project falls within the European Commission's Open Data Research Pilot, which means that the project is obligated to consider providing open access to the research data utilised and generated by the project. In addition, intellectual property rights are relevant, as much of the data utilised within the project are owned by ArcelorMittal, are vital to the company's production process and directly relate to the company's competitive advantage. Finally, data protection laws in Europe, and in Spain where the research primarily occurred, are relevant as the research involved human participants and thus collection of personal data, although it was a very limited amount. Hence, research ethics associated with human participants are also of paramount importance and good practice in this area governed how the project managed the participation of these voluntary research subjects.

#### 3.1 EC open research data pilot

The PROTEUS project is funded by the European Commission and responded to the ICT-16-2015 call on Big data - research. As it is part of the "Introduction to Leadership in Enabling and Industrial Technologies (LEIT): Information and communication technologies (ICT)" work programme, PROTEUS falls under the auspices of the European Commission's Open Research Data Pilot (ORD Pilot).

The ORD Pilot is an initiative by the European Commission (EC) to provide greater access to scientific information. The ORD Pilot specifically refers to H2020 projects and is "designed to improve and maximise access to and reuse of research data generated by [those] projects" (European Commission, 2016, p.7). The Commission feels that providing open access to the data generated by publicly funded research will boost innovation, prevent duplication and provide transparency to members of the public (ibid.). As such, the H2020 programme within the EC is moving in concert with other public funding bodies. Finally, a provision to provide open access to the scientific publications emanating from projects is also part of the Grant Agreement signed by projects and the EC in H2020.

In addition, the Commission also recommends the use of data management plans to guide projects in how they should manage the research data used and generated by the project. This includes identifying any intellectual property issues, privacy or data protection issues or commercial sensitivity that might require data security restrictions that will prevent unauthorised sharing (OpenAIRE, 2016).

As part of a project's participation in the ORD Pilot, they must do three things (EC, 2016):

1. Evaluate the sensitivity of their data, including issues related to intellectual property rights, privacy and data protection and commercial sensitivity. This information is used to decide how the data should be collected, stored, processed and preserved.
2. Where possible, they should deposit their research data in an appropriate repository to ensure that it is adequately preserved.
3. Take measures to enable third parties to access and re-use this data, free of charge. This may include providing information about any tools necessary to take advantage of the data, including specialised software or algorithms. The EC recommends the use of Creative Commons or other open licenses to manage the re-use of data after it has been deposited and opened.

While the EC is committed to the ethical benefits of providing open access to research data used and generated by the project, they stress that the decision to provide open access must be taken after any intellectual property, security, privacy or other legal issues have been evaluated. As such, they offer an "opt-out" if the data used or produced by the project includes issues around confidentiality, personal data protection or security issues (EC, 2016). In addition, projects may also opt out if they plan to exploit the data produced within the project themselves.

PROTEUS partners chose at the outset of the project to participate in the pilot in order to maximise access to and reuse of research data and insights that can boost innovation and derived social benefits yet respect commercial sensitivities and ethical considerations that arise as part of its work. However, as part of the data management process, PROTEUS partners discovered that all of the data provided by ArcelorMittal was

commercially sensitive and could not be made open access. The following sections describes these IP issues as well as the additional issues related to privacy and data protection and ethical research practice that will further guide our decisions about managing PROTEUS data.

## **3.2 Intellectual property rights**

Using data owned by ArcelorMittal inevitably raised intellectual property rights. Intellectual property rights protect works by individuals that are the result of creativity, innovation, skill and specialist effort (Korn and Oppenheim, 2011). This section examines the intellectual property rights relevant to PROTEUS, including rights around trade secrets, for ArcelorMittal data, and copyright and database rights, for project-generated data. This information will be used to govern the extent to which this data can be made openly accessible after the close of the project.

### **3.2.1 Trade secrets**

Trade secrets are a specific form of intellectual property rights that cover commercial data. In order to qualify for protection as a trade secret, the information must be:

- secret (i.e. it is not generally known among, or readily accessible to, circles that normally deal with the kind of information in question).
- have commercial value because it is a secret.
- subject to reasonable steps by the rightful holder of the information to keep it secret (e.g., through confidentiality agreements) (World Intellectual Property Organization, no date).

Trade secrets can be protected indefinitely, and they are not subject to any registration or other formal requirements. Furthermore, they must be relevant within business and provide some form of competitive advantage to the holder.

The data held by ArcelorMittal and provided to PROTEUS are subject to protection as trade secrets. They relate specifically to the core manufacturing process of the specific products. This information is not known to other organisations and forms an important part of ArcelorMittal's competitive advantage. The data is protected within ArcelorMittal, and the company has consulted with their legal team to approve the limited sharing of this data with the PROTEUS consortium.

### **3.2.2 Copyright and database rights**

Copyright is an automatically bestowed intellectual property right that protects the author or creator of a work and determines where, how and when the work may be made publicly available as well as how it may be used. While copyright is most often associated with scientific publications in the research arena, copyright might also be relevant to data where the creation of a data set represents a significant intellectual investment (de Vries, 2012). In Europe, the 1996 Database Directive protects the intellectual investment utilised to create data sets. While factual information is not subject to intellectual property legislation, building a collection of data is protected within this framework. The Directive prevents third parties from publishing, distributing and copying research data where the owners have claimed protection. These rights can be utilised to allow researchers to dictate how their data is utilised to allow them to exploit the data themselves, and to control how the data is utilised even where open access is provided. Previous research in this area has found a potential conflict between such database rights and open access mandates (Finn, et al., 2014). Nevertheless, the push by research funders and public bodies to support the exploitation of data by researchers themselves gives significant license to researchers to enable them to choose between exploitation expectations and open access expectations or to meet both in incremental steps (ibid.). Intellectual property owners can exercise these rights through licences, and the data management plan will assist partners to identify appropriate licenses given the constraints and necessary protections identified within this document.

This legal framework has allowed PROTEUS to meaningfully consider how the consortium might exploit the data produced by the project and the extent to which PROTEUS can provide open access to the derived data resulting from the project. In the project Consortium Agreement, partners have already agreed that each party owns the results, including the data, produced by them. Furthermore, in the case of joint ownership, each party has the right to exploit the results as they see fit. However, as the project developed it became clear that all of

the data given to the partners and produced within the project was commercially sensitive and as such could not be made openly accessible to other researchers.

### 3.3 Data protection law

#### 3.3.1 General Data Protection Regulation (679/2018)

In addition to the EC's open research data pilot, the project was also guided by policies on the protection of personal data both in Europe and in Spain, where the research with human subjects' element of the project were situated. Rights to privacy, as contained within the European Convention of Human Rights focus on "respect for private and family life, home and communications", which are less relevant for the research carried out by the PROTEUS project.

In Europe, the 2018 General Data Protection Regulation (GDPR) governs the protection of personal data and sets out specific principles for the protection of personal data when data is being collected or processed. The personal data collected by PROTEUS was very low risk from the perspective of data protection; nevertheless, the following data protection principles are relevant for PROTEUS.

- **Fair, lawful and transparent processing** – Information must be processed in accordance with a stated and specific lawful basis, including consent, the performance of a contract or a legal obligation on the part of the organisation processing the data (as well as other possibilities). PROTEUS relied on both consent and performance of a contract in relation to the collection, storage and processing of personal data.
- **Consent** – Where consent is relied upon as the lawful basis of processing, it must be freely given, and organisations must be able to demonstrate that consent was given. The informed consent sheets create a record that consent was given.
- **Purpose limitation** – Personal data can only be collected for an explicit, legitimate purpose that was specified at the time of collection. Data in PROTEUS was only used to evaluate the toolset.
- **Data minimisation** – Data collected must be limited to and relevant for the data required to achieve the purpose specified. The only personal data collected, processed and stored by PROTEUS is the information on the informed consent sheet.
- **Storage limitation** – Data cannot be kept for longer than necessary to achieve the specified purpose. Personal data collected by PROTEUS will only be stored for one year after the close of the project, unless further queries are raised by the EC requiring the data to be stored for a longer period of time.
- **Data security** – Data must be protected by technical and procedural measures against loss, damage or theft. AMIII has agreed to keep the consent forms stored securely.
- **Accountability** – Organisations collecting, storing and processing personal data must ensure that they can demonstrate their compliance with the principles above. The informed consent sheets and this document comprise a record of data processing and a data retention schedule.

In addition, the following requirements relevant for PROTEUS are also outlined in the GDPR.

- **Right to be forgotten** – Under Art. 17 data subjects have a right to obtain erasure from the data controller without undue delay. This means that PROTEUS research participants will have the right to have the record of their participation in the research deleted.
- **Data protection by design and default** – Data controllers must include appropriate provisions for anonymisation, pseudonymisation and data minimisation for all areas and stakeholders involved have been considered [IT (Chief Information-Security Officer, Legal (Legal Officer or Data Protection Officer), Business and Partners in the Project). Full anonymisation of AMIII employee data is a PROTEUS priority. However, we may need to prove the "real or effective" participation in the project for the administrative justification of the Project and guarantee the scientific accuracy to eventual publications and the ability to re-identify user participants may be required. As such, AMIII will keep the information about the employee participant, and partners will only have access to an anonymized version of the evaluation data.
- **Notification of data breaches** – Authorities should be notified of any data breaches within 72 hours of their occurrence.

- Processing personal data for research purposes – AMIII will ensure that data processing will respect the principles of Article 30 of GDPR. This document comprises a Record of processing activities and ensures the confidentiality, integrity, availability and resilience of processing systems and services. In addition, data security safeguards (as per article 32 of GDPR) have been put in place to secure data from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed.
- Transparency and communication with data subjects – The consortium has ensured that all research participants were fully informed prior to data collection and as part of the process of informed consent inherent in its methodology about the contact details of the data controller, the purpose of their data processing, the recipients of processed data, forms and modalities of their communication. Moreover, the consortium agrees that that information processing logs are kept accessible. Specifically, that this document will be stored by all partners to provide data subjects with clear and timely accounts of the trail of their data processing on demand as per the requirement of article 12 of GDPR.

PROTEUS has taken each of these obligations into consideration when processing personal data to ensure the project complies with the General Data Protection Regulation.

### 3.3.2 Spanish data protection law and regulatory requirements

During the course of the PROTEUS project the National Spanish Data Protection Regime was replaced with the General Data Protection Regulation, covering the whole of Europe. Nevertheless, the project complies with all requirements to register contained within Spanish Law, and both Treelogic and ArcelorMittal were already registered with the General Data Protection Registry. This was accomplished via the Agencia Española de Protección de Datos' (AEPD, Spanish data protection authority) website. The registration describes:

- The purpose of the data file
- The categories of personal data contained within the file
- Any intended data disclosures
- The security measures applied to the data
- Any intended international transfers (Practical Law, 2015)

Registration was accomplished via this website:

[https://www.agpd.es/portalwebAGPD/canalresponsable/inscripcion\\_ficheros/index-iden-idphp.php](https://www.agpd.es/portalwebAGPD/canalresponsable/inscripcion_ficheros/index-iden-idphp.php)

As demonstrated below, many of these requirements overlap with requirements associated with ethical research practice with human volunteers. As such, PROTEUS was able to meet each of these obligations via our informed consent mechanisms.

## 3.4 Ethical research guidance

Good practice in research with human volunteers has been the subject of much study in relation to social science as well as other disciplines. Disciplinary societies, such as the British Sociological Association, have published good practice guidance in relation to social science research with human volunteers, and PROTEUS has followed these guidelines very closely. The research with human volunteers that was carried out within PROTEUS is most closely related to social science in that we are interested in how people interact with a specific technology product within an employment context. As such, PROTEUS integrated aspects of sociology, science and technology studies and organisational studies, each of which fall under a larger interdisciplinary social science umbrella. Given this interdisciplinary social science framework, the consortium has selected the ethical research guidance produced by the British Sociological Association (BSA) to steer our research activities with human volunteers (BSA, 2002).

The BSA standards include a number of ethical research principles relevant to PROTEUS, and many of these overlap with data protection requirements. Specifically, the BSA ethical principles include an obligation to consider data protection law, human rights law and other relevant legal frameworks. Thus, data protection laws and other legal requirements are brought to the forefront of social research work.

These requirements include principles that link with data minimisation principles as well as obligations surrounding confidentiality, storage, security, sharing and rights of access, correction and erasure. With respect to data minimisation, BSA ethical standards explain that researchers should:

- consider whether specific types of data should be recorded, including sensitive data.
- keep personal information confidential,
- anonymise and pseudonymise the data as much as possible, and
- store data securely in an anticipation of threats to anonymity and confidentiality.

The standards also recognise that full anonymity cannot be absolutely guaranteed, and that participants should be given information to enable them to understand the limits of this anonymity. With respect to data storage and sharing, the guidance explains that researchers should provide information about

- how data will be stored,
- whether it will be shared with other researchers, and
- how it might possibly be used by those researchers.

In some cases, including providing open access to anonymised data, additional consent should be obtained. Finally, participants may be given rights to see copies of notes, transcripts and other research materials, and researchers should explain whether participants will be granted rights to alter or correct these. In each of these cases, the obligations surrounding research ethics and data protection are overlapping and PROTEUS has met the requirements of both by relying upon the more stringent obligations.

In addition, research ethics and data protection also include principles of transparency and informed consent. The creation of an information sheet and informed consent form are central to the processes of both transparency and informed consent. Research ethics guidance asserts the following:

Participation in sociological research should be based on the freely given informed consent of those studied. This implies a responsibility on the sociologist to explain in appropriate detail, and in terms meaningful to participants, what the research is about, who is undertaking and financing it, why it is being undertaken and how it is to be disseminated and used. (BSA, 2002, p. 3)

The informed consent form has accomplished the transparent provision of information about the research, the funding, the purpose of the project and how the results will be disseminated and used. It also includes information about how the data will be used by the project. Furthermore, providing this information has enabled the project to obtain informed consent from participants, given that the information sheet is a transparency mechanism.

Finally, the BSA guidance discusses the other professional and moral responsibilities that researchers have in relation to research participants. The guidance requires researchers to consider their responsibilities to:

- Protect the interests of those involved in the research.
- Ensure, as far as possible, that the research process may be disturbing to participants and whether it may produce any unintended effects
- Ensure that participants are aware that they have a right to refuse to participate

In addition, researchers have a responsibility over the ways in which their data may be utilised and how their findings might be disseminated. These responsibilities are particularly relevant as PROTEUS research participants are/were employees of one of the partner organisations - AMIII. Thus, their employer had, and will continue to have, have access to the research data held about them, and research participants will be in an unequal power relationship with the researchers themselves. As such, issues around confidentiality, data security and data governance are particularly important in PROTEUS.

The following sections describe the PROTEUS data governance considerations that have enabled the project to meet each of these policy, legal and ethical requirements. Specifically, the following chapters describe each of the data sets in more detail, examine our responses to the ethical and legal issues and look at issues related to data handling including access storage and sharing. Finally, the last chapters consider whether and how the consortium can exploit the project data as well as the extent to which PROTEUS data can be opened via open access provisions.



## 4 Data set description

This chapter outlines the specific characteristics of the three types of data that were collected and processed by the project. The three data sets fall within the following categories:

1. ArcelorMittal data: ArcelorMittal data provided to the project
2. PROTEUS system data: Data generated by the project via the running of the algorithms and worked on by the visualisation tools.
3. PROTEUS evaluation data: User testing data of the visualisation aspects of the tool by ArcelorMittal employees

The first data set consists of real, anonymised, production data provided to the project by ArcelorMittal (AMIII), which has significant commercial sensitivity and thus is covered by intellectual property rights. This data was used to develop and test the PROTEUS predictive analytics tools. The second includes derived data from the project allowing the data visualisation tools to function on the algorithmic outputs produced within the project. Although these two datasets are differentiated here, it is also possible to view them as one dataset that was used by different partners to achieve different goals: e.g., algorithm validation, data visualisation, etc. The third data set came from the piloting and assessing of the PROTEUS system. ArcelorMittal employees have assessed the visualisation dashboard in relation to its utility and functionality. In the sub-sections that follow we outline the methods for data collection that were used for each of the three PROTEUS data categories, their characteristics and associated sensitivities.

### 4.1 Data collection and characteristics

This section provides a description of the characteristics of the different types of data that were collected by the project. This includes information about where the data originated, the stakeholders for whom it was useful and the uses to which it was put within the project. It also includes information about the scale and volume of the data, its format and the extent to which the data might be interoperable with other data sets.

#### ArcelorMittal data

The ArcelorMittal data originates from the AMIII steel coil production process. Deliverable 2.1: *Scenario analysis and objectives description* gives a comprehensive overview of the steel production processes and a description of the data set relevant to PROTEUS. PROTEUS partners used the data generated from the steel coil production process to identify variables relevant to and develop and test algorithms for detecting potential defects in the coils. As such, the data is useful to all of the consortium partners. However, because the production process has commercial sensitivity, the data that was provided by AMIII might also be of interest to customers and competitors. Because of this sensitivity, the data provided by AMII to the project is anonymised – any data relating to specific customers has been removed and the variables are mostly unlabelled to protect the sensitivity of the production process.

The ArcelorMittal data is made up of two datasets: the coil production data set and the flatness data set. These data sets comprise complex, large and high velocity data. In particular, the coil production data consists on different measures that are made available during the production process at different moments in time:

**Time-series:** these data are produced and available in real time during the coil production. It consists of **53 different variables** that are measured and made available in **real time** by different sensors deployed in the actual production environment. These variables present a **variety of formats**, combining both 1D and 2D information, and are produced at **different rates** (depending on the specific variable) **ranging from 50 milliseconds to around 1 second period**.

**HSM:** historical and final production coil data (known as HMS data). It provides more than **7,000 variables** as aggregated information per coil at the end of the production process, thus this is **available only once the production has finalised for one particular coil**.

**The flatness dataset:** measures of the flatness of the resulting coil. It consists of **3 variables** (2 of them as 1D flatness, and an additional 2D flatness measure) following a format that is compliant with the one used in *time-series* and synchronised with it by the spatial information (x, y). This information is measured and **available**

**only after the coil has been fully produced, and after a certain delay** due to the production environment setup and infrastructure. As previously introduced and also in previous deliverable D2.1, this information **was one of the main targets for the real-time learning and predictive processes** enabled by PROTEUS, since the capability to predict flatness in real time (without the need to wait until the process has finalised to get the actual measures) is key to improve business operations.

The data generated by the sensors is both qualitative and quantitative, and stores approximately 7440 variables relevant to the production process. The data dates from 2010 and thus there are approximately 840,000 records for each variable. The types of measurements included are temperature, vibration intensity, tension in the rollers, speed of the plate when entering the roller and surface flatness. With respect to surface flatness, the generation rate for variables is approximately 1 per second, this requires significant computing capacity to manage this high-velocity data set. In addition, while the process data set to be shared includes historical records for each variable, these data are also generated in real time during the production process at a rate of between 32 and 500 milliseconds, depending on the specific measure. The initial data controller, Treelogic, used historical data from both data sets to emulate real-time data that can be used to test the algorithms. This data was eventually transferred to Bournemouth University, DFKI and TRI to enable further project activities. All partners have agreed to keep the data fully confidential and to only use the data for the purposes of developing and evaluating elements within the PROTEUS project.

As noted above, this raw data has been shared in an anonymised form, so any sensitive information is excluded. This analysed data set is comprised of both quantitative and qualitative data that are stored across 42 different tables. All of the tables share the same key variable, the coil identifier, which allows AMIII to join and relate information for each coil across multiple tables. The size of each of the 42 data tables is approximately 300-700MB. This data provided some context to the initial variables that are thought to be relevant to PROTEUS.

### **PROTEUS derived data**

The second set of data includes data produced within the PROTEUS project that allows the different tools developed to be integrated and cross functional. This data consists of output data from the algorithms, including simple-moments, sax-results and lasso-results, that allow the data visualisation tools to display the results of the analytics. These data are the output of the use cases of AMIII, as described in D2.11. Specifically, the data are aligned with the use cases as follows:

- simple-moments: average and deviation, classical statistical functions, the only difference is that they're calculated in streaming in this case.
- Aggregate approximation-results: summarisation of the coil data, allowing tracking the evolution of the variable over consecutive windows using the average metric. This was done using symbolic approximation (SAX)
- Prediction-results: A regression model for prediction of flatness in real-time. A number of algorithms (e.g. online Lasso) have been developed and tested on the data.

The data is streaming data, produced in real-time during the process. This means their characteristics (e.g., size) are as follows:

- simple-moments: these data are produced once (one for average and another one for deviation) for every input from the variable, in streaming. i.e. if a coil produces 120 sensor measures for a variable, then simple-moments produces 120 values for average and 120 values for deviation, once for every sensor value, iteratively.
- sax-results: aggregation is done for a given WindowSize, so this data is produced once every WindowSize measures of the input variable. This means that if a coil produces 120 sensor measures for a variable, SAX produces  $120/\text{WindowSize}$  measures. This can be configured in the algorithm. In the current setup,  $\text{WindowSize} = 5$ , so for the example the size is  $120/5 = 24$ .
- lasso-results: once for every 'x' value for any input from all the variables, in streaming. There is no fixed amount, because this depends on the coil and how many different 'x' values there are in all the different variables. It is as many values as different 'x' the coil has.

In terms of format, the data is recorded and presented as follows in order to enable PROTEIC to transform it into a visualisation.

--

MomentsResult1D

```
{
"coilld": "",
"varld": "",
"x": "",
"mean": "",
"variance": "",
"counter": ""
}
```

MomentsResult2D

```
{
"coilld": "",
"varld": "",
"x": "",
"y": "",
"mean": "",
"variance": "",
"counter": ""
}
```

SAXResult

```
{
"coilld": "",
"varName": "",
"x1": "",
"x2": "",
"classld": "",
"similarity": ""
}
```

LassoResult

```
{
"coilld": "",
"varld": "",
"x": "",
"label": ""
}
```

This data is not useful beyond the project itself, as it is a transformation of raw AMIII data into data useful for the visualisation tools. In that respect, it is specific to PROTEUS and commercially protected like the raw data provided by AMIII. It is also not likely to be interoperable with other data sets, although the algorithms could work on different raw data and produce similar results useful for the PROTEIC visualisation tools.

### **PROTEUS user evaluation data**

Data generated during the evaluation phase of the research involves human subjects and included the collection of personal data in order to ensure informed consent and provide evidence of data collection if required by the EC. The purpose of the evaluation exercise was to assess whether the PROTEUS solution is fit for purpose

from the point of view of industrial operators. The anonymised research participant was presented with a series of visualisations generated by PROTEUS predictive analytics system and requested to evaluate these visualisations with respect to their utility, reliability and validity for identifying defects. They were also asked to consider what features they would like and how well this might be integrated into their current operations workflow. Given that the participant was asked to comment on and critique current and planned workflow processes to be initiated by their employer, the research may have put him or her in a potentially difficult situation, because essentially, they were asked to indirectly critique their current workflow and work process and critique the new system. As their employer is a participant in the project and driving the project, there was a low risk that the employee may experience adverse consequences from stating their opinions. As such, the consortium planned a number of protections to ensure that research participants are not adversely affected by their participation and that their participation was fully voluntary.

Potential research participants included employees of ArcelorMittal with, at least some, responsibility for quality control. Evaluation data was collected via one-to-one consultation exercises. The research process initially targeted participants from a small group of HSM experts within the organisation. In the end, only one participant was approached to provide feedback on the system, given that person's level of expertise and their positioning within the organisation. Given that the person was relatively senior, there was little chance that he or she would have felt undue pressure to participate or that he or she was in a position to experience adverse effects as a result of being asked to critique current workflows and processes.

The data itself consisted of anonymised, written notes compiled by another AMIII employee. This fully anonymised data has been converted into a word document and shared with the consortium. The Word document is stored on the project's data repository in a secure cyber-location that meets existing standards on data security and data protection. As per EC requirements governing H2020 projects set out in the Grant Agreement, the anonymised research data will be stored for no less than 5 years after the close of the project.

#### **4.1.1 Personal data collection**

The project sought to collect as little personal data as possible from human participants. Nevertheless, requirements around ethical research practice necessitate the collection of some personal data to manage informed consent processes and to re-connect with participants who may have questions about PROTEUS or who may wish to exercise their rights of access, correction and erasure. The personal information was limited to the names and contact details of the volunteer as filled out on their informed consent sheets. After the informed consent process, data from the participant was fully anonymised as described in Chapter 5, and the personal information contained within their informed consent forms is being held securely by their employer as described in Chapter 6.

## 5 Ethical and legal issues

The types of data described above raise specific issues related to intellectual property, data protection and research ethics that the project has had to manage appropriately. Where relevant, Spanish employment law is considered alongside the European law, as Spain is the primary research location where data collection and processing activities are taking place. The following sections outline how PROTEUS managed each of the relevant legal requirements and describes the agreed data governance processes around access storage and sharing. Consequently, this section makes consistent reference to the material to be discussed in Chapter 6: Data Governance. This chapter addresses the management of informed consent, personal data protection, and intellectual property rights.

### 5.1 Participant selection and informed consent

Issues related to research ethics can largely be addressed through the management of informed consent when the research is being conducted with healthy, adult volunteers. However, as the participants are employees of a partner organisation, there are some risks that those approached to participate may have felt pressured to do so. This risk was managed through the participant selection and informed consent processes.

**Participant selection** is crucial in ensuring that the research activities are targeted to individuals who can answer the questions posed, and add value to the project, without undue burden on AMIII and their employees. Since legal issues meant that the PROTEUS platform could not be integrated directly into existing AMIII systems to test the toolset live, the partners decided to reduce the size of the trials to limit the evaluation to employees with expertise in the Hot Strip Mill process and the existing quality control mechanisms. This significantly reduced the pool of potential participants, and in the end one key informant was selected for the evaluation. This informant was a relatively senior person at AMIII and, as noted above, was unlikely to experience any negative impacts resulting from his participation in the research or to feel pressure to participate against his or her will. Nevertheless, the informant was put through all of the appropriate informed consent processes to ensure that he or she was participating on a voluntary basis and fully understood the implications of participating in the research.

**In relation to informed consent** the consortium provided the participant with adequate information about the purpose of the research, the data that would be collected, the research funders, the ways in which the data would be utilised and who would benefit from the research to ensure that participants understood the potential implications of their participation. An **information sheet** provided this information in appropriate detail and in language that is meaningful to participant(s). (See Annex A for the final version of the information sheet). The information sheet sets out information about:

- how data would be anonymised if they chose to participate
- how data would be stored and shared with other researchers if they chose to participate
- how participants may access the data they provided after the consultation
- whether they can make corrections after the consultation
- how they can request their data be removed, and
- where they can go if they have any questions, comments or complaints.

In addition, the information sheet explains any unintended effects that may result from the research. Combining each of these pieces of information enabled potential participants to evaluate whether they would like to participate in the research and whether they might experience any unintended or adverse effects.

The consortium also took steps to ensure that potential participants are well informed and that they felt empowered to volunteer or decline participation as they saw fit. Given that this research was carried out with employees of one of the partner organisations, ensuring voluntary participation is paramount and required that the information sheet make it clear that their participation is purely voluntary. Given that the project does not involve many sensitive topics, the consortium agreed that this was sufficient to ensure voluntary participation. The final informed consent form is also included in Annex A of the deliverable.

Moreover, given that the participant is also an employee of ArcelorMittal, the consortium also complied with legislation, guidelines and recommendations regarding *jurisprudence* in accordance with the Spanish Data Protection Agency and the Working Party<sup>1</sup> to ensure that employee rights prevail as per the law, and guarantee that such rights cannot be circumvented by the employer under article 20 of the workers' statute<sup>2</sup>. To this end, the consortium made clear that the research activity does not form part of the ordinary work experience or that the conditions of realisation do not affect their job. It explicitly highlighted the freedom of choice of the participant, the absence of adverse consequences and the possibility to withdraw from the Project at any point by approaching the research staff. However, given the limited size of the evaluation, the consortium decided against informing and consulting the relevant Works council about the project.

## 5.2 Personal data protection

However, seeking informed consent did raise issues around the protection of personal data, as personal data, including names and contact information was needed to record informed consent. The consortium has managed and will continue to manage this using the following steps:

1. The personal data supplied by participants was only collected by and stored by Arcelor Mittal's members of the PROTEUS research team.
2. This personal information has not been and will not be shared with the project partners.
3. Any enquiries about participants' personal information will be fed through the coordinator who will liaise with the AMIII PROTEUS team directly to manage any potential conflicts between AMIII and their employee(s)
4. Participants have been given the right to review their data and make any corrections or erasures should they have any concerns.

The project has avoided the collection of data that is not necessary for the purposes of the research (*purpose limitation and data minimisation principles*) and limited the personal data collected to names and contact details only. All other data has been anonymised and storage has been limited to a small team of individuals within AMIII. As noted above, both consent and the employment contract are the legal basis for storing the names of those who have participated in the research. In addition, the information sheet enables the project to meet requirements around transparency and provide a mechanism through which participants can exercise their rights of access, correction and erasure. The information sheet also assists the project in meeting requirements around data retention, as the information sheet as well as this document set out how long the data will be stored and with whom it may be shared.

In addition to these, the project also met Spanish data protection requirements. AMIII has registered with the AEPD as a processor of personal data given their employment functions. In addition, should a data breach occur, the coordinator will inform both the AEPD and research participants about the breach and provide advice on any consequences.

Thus, the overlapping requirements around ethical research practice and the protection of personal data was met simultaneously using both the information sheet and informed consent forms for the PROTEUS research. While the amount of personal data that was collected by the project is relatively minimal, the project has used the data protection principles to guide the collection of all data about human participants, whether personal or

---

<sup>1</sup> SSTC 292/2000, 241/2012, 29/2013, 170/2013, 39/2016. Bărbulescu V. Romania (Application no. 61496/08) Working document on the surveillance of electronic communications in the workplace. (29 May 2002). Working Party Article 29. Opinion 15/2011 on the definition of consent. Working document on the surveillance of electronic communications in the workplace. (29 May 2002). Spanish Data Protection Agency. Data Protection in Labour Relationships.

<sup>2</sup> Article 20. Management and control of the labour activity.(...)

<sup>3</sup> The employer may adopt the measures he believes more convenient for supervision and control to verify the compliance by the worker of his duties and obligations, keeping in its adoption and application the consideration of his dignity and taking into account, where appropriate, the actual capacity of workers with disabilities.

not, to ensure that we meet ethical research requirements. Attention to both ensures participants receive the maximum level of protection and consideration during and after the project's completion.

### **5.3 Intellectual property rights**

As noted above, the ArcelorMittal corporate data (including the raw data supplied by AMIII and the derived data from PROTEUS) is subject to intellectual property protections, and the consortium has taken the following specific steps to address these. First, as noted above, the data has been anonymised so that any sensitive data associated with AMIII customers is removed. Second, the data governance procedures around access, storage and sharing, discussed in Chapter 6 below, ensure that consortium members respect AMIII's intellectual property rights. Finally, each of the partners have agreed to only use the data for the purposes of the PROTEUS project and the development and testing of PROTEUS algorithms and software. This has been agreed via the PROTEUS Consortium Agreement, a legally binding document that governs the project and partnership arrangements.

The Consortium Agreement and this document also guides the intellectual property rights claimed by the consortium with respect to PROTEUS toolset data. The Consortium Agreement outlines how the data has been and will be used and shared within the consortium. Partners also agree that the data will not be re-used after the close of the project. Underpinning this will be the agreement, contained within the Consortium Agreement, that each partner owns the intellectual property, including data, which they create. Nevertheless, the Consortium Agreement also provides for joint or multiple ownership, and in these cases, relevant partners will agree on the license to be used. Consideration of these intellectual property rights has also governed our decision not to make any of the project data openly accessible after the close of the project.

## **6 Data governance**

This section outlines the rules for the governance of PROTEUS data. The information here relies upon information provided by consortium partners, as well as governance mechanisms agreed in the Consortium Agreement. Each of the sub-chapters that follow outline the agreed rules about how data was and will be accessed, stored and shared.

### **6.1 Access**

With respect to the ArcelorMittal data provided to PROTEUS and the data derived from PROTEUS activities, the partners have agreed that the project coordinator is responsible for managing access to the data. AMIII agreed to allow access to the required historical and simulated real-time data in an anonymised format, provided that data is only accessed by consortium partners and only for project activities. Partners have agreed not to seek access to raw, non-anonymised, data. Anonymised AMIII data has been provided directly to the coordinator, who is storing the data in their existing ICT infrastructure and enabled access by consortium partners. Any partner with a user password was able to access the data to develop or test their algorithm or software. Partners maintained access only rights during testing but were not able to store such data. Unauthorised access to this data was and will be prevented via BU's existing data and information security mechanisms, which meet existing information security standards.

Access to PROTEUS derived data and PROTEUS evaluation data was restricted to the consortium during the course of the project. As a starting point, the algorithms and system evaluation data were held by the partners participating in system development and evaluation and shared among the consortium for the purposes of carrying out the work described in the Description of Action. Access to the PROTEUS data after the close of the project will be restricted to activities required to further disseminate PROTEUS results, e.g., via demonstrations of the SOLMA library or PROTEIC tools.

### **6.2 Storage and processing**

With respect to storage, each of the three types of data was stored differently and will be maintained differently, as required by EC rules, once the project is completed. BU will store the ArcelorMittal provided data and PROTEUS derived data in their existing infrastructure. All of this data will be backed-up in the cluster, automatically, so that the data can be recovered in the event of an incident. The security of this data was maintained via the security policies and mechanisms that BU already has in place for protecting their sensitive data. Finally, AMII will store the personal data associated with the informed consent within a separate, but equally secure, storage space that is not accessible to project partners to protect the personal data of those participating in the project.

After the close of the project, BU will continue to store this data for as long as required given EC rules on H2020 funding as outlined in the Grant Agreement.

### **6.3 Sharing**

PROTEUS data will not be shared outside the consortium.



## 7 Standards and metadata

This section considers the standards that PROTEUS used to represent data generated by the project, and the additional standards around data security, etc. that might be useful to govern the data used within and generated by the project. In addition, this section also includes a consideration of the metadata for the PROTEUS data set.

This includes a consideration and evaluation of existing standards and our reasoning for selecting specific standards. The consortium has agreed to avoid proprietary data formats as far as possible to make it easy for the consortium to carry out the required activities, including dissemination after the close of the project if required.

Given the proprietary nature and commercial sensitivity of utilised data and the requirements of the project for a realistic simulation environment on which the PROTEUS prototype could test, a concerted effort between ArcelorMittal and BU was undertaken to jointly curate the data and ensure that there is a clear and accurate understanding of what the data categories represent amongst data partners.

In addition to these, PROTEUS also consider standards around other issues that could govern the storage and representation of PROTEUS data. This included data security standards such as ISO 27001.

Finally, the project also considered effective metadata for describing the PROTEUS system and evaluation data. However, the consortium decided that since the data will only be used by existing partners for the purpose of executing the PROTEUS project and would not be shared outside the consortium or used for any other purpose after the close of the project, specific metadata was not required.

## **8 Data exploitation**

This section describes the exploitation parameters for PROTEUS data by project partners and external stakeholders for additional commercial or research purposes. The consortium has agreed that the PROTEUS data will not be shared outside the consortium or used by project partners for any purpose beyond disseminating the results of the project. This is because of the commercial sensitivity of the raw data and the derived data and the small size of the evaluate data. As such, there is no plan to exploit the datasets after the close of the project.

## **9 Long-term archiving and preservation (including open access)**

PROTEUS partners will not store the AMIII raw data, PROTEUS derived data or evaluation data beyond what is required for EC reporting. As such, the data will only be stored by the coordinator BU for the length of time required to fulfil EC reporting requirements and potential audits. Given the commercial sensitivity of the data and the small size of the evaluation data, no data will be made open access after the close of the project.

## 10 Conclusion

This deliverable represents the final iteration of the PROTEUS data management plan. It provides foundational and contextual information related to the project, the relevant policies and legal frameworks and the project partners' plans for ensuring that the collection and use of data within the project conforms to the issues within this larger context. Specifically, the document examines the European Commission's Open Data Research Pilot, intellectual property issues relevant to PROTEUS, data protection standards and practices and guidance on ethical research processes. The deliverable progresses by describing the data used and generated by PROTEUS and outlining how the project met all of our legal, ethical and policy obligations surrounding PROTEUS data. In doing so, the project has agreed the following principles:

- Data owned by ArcelorMittal will be shared with consortium members, although consortium members can only access this data through the project coordinator
- PROTEUS partners agree to respect the commercial confidence of the data provided by AMIII
- Human participants in PROTEUS research activities were under no obligation to participate and their involvement will be strictly voluntary
- PROTEUS partners have respected and will respect data protection and ethical research principles related to the following:
  - Participant confidentiality and anonymisation
  - Informed consent
  - Data minimisation
  - Purpose limitation
  - Transparency
  - Rights of access, correction and erasure
- Personal data that can identify the research participants will be stored by AMIII securely
- The coordinator will hold all commercially sensitive data will manage access to this data for PROTEUS partners

In addition to meeting European regulations, the project also met legal regulations set by the Spanish government as the data and many of the research activities will be located in Spain. This includes the registration of AMIII as a data controller with the Spanish data protection authority and the provision of specific information required by Spanish data protection legislation. Final versions of the PROTEUS information sheet and informed consent forms are included below to demonstrate how we met these obligations.

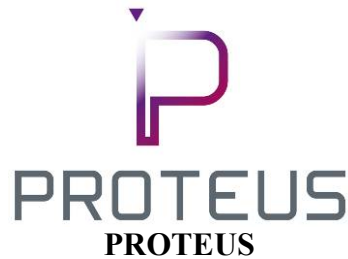
Finally, the project partners have agreed that no data will be exploited by the project partners or made openly accessible after the close of the project, given the commercial sensitivity of the AMIII data and the PROTEUS derived data as well as the small size of the evaluation data set. Nevertheless, partners may use the data to further disseminate PROTEUS results, including through demonstrations of the system.

## References

- [1] Wessels, Bridgette, Rachel L. Finn, Peter Linde, Paolo Mazzetti, Stefano Nativi, Susan Riley, Rod Smallwood, Mark J. Taylor, Victoria Tsoukala, Kush Wadhwa and Sally Wyatt, “Issues in the development of open access to research data”, *Prometheus: Critical Studies in Innovation*, Vol. 32, Issue 1, 2014, pp. 49-66.
- [2] European Commission, *Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020*, Version 2.1, 15 February 2016. [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)
- [3] Jones, Sarah. (2011). ‘How to Develop a Data Management and Sharing Plan’. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides>
- [4] Data Curation Centre, “Data management plans”, 2016. <http://www.dcc.ac.uk/resources/data-management-plans>
- [5] OpenAIRE, “What is the Open Research Data Pilot?”, 22 Feb 2016. <https://www.openaire.eu/opendatapilot>
- [6] Korn, Naomi, and Charles Oppenheim, *Licensing Open Data: A Practical Guide*, June 2011 version 2.0. [http://discovery.ac.uk/files/pdf/Licensing\\_Open\\_Data\\_A\\_Practical\\_Guide.pdf](http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf)
- [7] World Intellectual Property Organization, “How are trade secrets protected?”, no date. [http://www.wipo.int/sme/en/ip\\_business/trade\\_secrets/protection.htm](http://www.wipo.int/sme/en/ip_business/trade_secrets/protection.htm)
- [8] Marc de Vries, *Open Data and Liability*, European Public Sector Information Platform Topic Report No. 2012 / 13, December 2012.
- [9] Finn, Rachel, Kush Wadhwa, Mark Taylor, Thordis Sveinsdottir, Merel Noorman and Jeroen Sondervan, *Legal and ethical issues in open access and data dissemination and preservation*, RECODE project Deliverable 3.1, 30 April 2014.
- [10] European Parliament and the Council, Directive 95/46/EC of 24.10.1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ L 281, 23.11.1995.
- [11] Article 29 Data Protection Working Party, Opinion 4/2007 on the concept of personal data, Brussels, 20 June 2007, [http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136\\_en.pdf](http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf) (“A29WP Opinion 4/2007”).
- [12] Practical Law, Data protection in Spain: overview, 1 Dec 2015. <http://uk.practicallaw.com/1-520-8264#>
- [13] Linklaters, “Data Protected: Spain”, July 2015. <https://clientsites.linklaters.com/Clients/dataprotected/Pages/Spain.aspx>
- [14] British Sociological Association, *Statement of ethical practice for the British Sociological Association*, British Sociological Association, Durham UK, March 2002.
- [15] Apache Kafka, “7.1 Security Overview”, 2016. <https://kafka.apache.org/090/security.html>
- [16] UK Data Service, “Data management planning for ESRC researchers”, 2016. <https://www.ukdataservice.ac.uk/manage-data/plan/dmp-esrc>

## Annex A – PROTEUS informed consent and information sheets

### PROTEUS Information sheet



#### Scalable online machine learning for predictive analytics and real-time interactive visualization

15 October 2018

The PROTEUS research project<sup>3</sup> conducted by Marta Garcia of Arcelor Mittal is funded by the European Union<sup>4</sup>. The purpose of the research is to evaluate a new system for identifying defective steel coils sooner in the production process. The evaluation exercise is based on the tool developed within the project that will provide alerts and information about the quality of the steel coils. This research will help ArcelorMittal and other large companies make better use of the data they have about the manufacturing process and cut down on waste and defective products. I will contribute to this research by evaluating the usefulness and user-friendliness of the toolset. My role is to help the researchers test how easy the system is to use from operatives' point of view and tell them whether I understand the alerts and information I see, so they can assess and where necessary refine the design of the system to make it useful and user-friendly.

This research will involve a consultation lasting 1-2 hours where I will be invited to give feedback about the tool, discuss my understanding of what I see and provide feedback about how this might be integrated into the manufacturing process. Participation to this project is *not* an employment obligation and it is entirely in my discretion to participate. Participating in this research is entirely voluntarily and I am free to leave at any time, without any impacts on my employment. I am also aware that I am free to refuse to answer any questions that I feel are commercially or institutionally sensitive or relate to topics that I do not wish to discuss. I understand that I have the right to ask questions and receive understandable answers before making any decision.

I understand that I will only be asked to provide professional, not personal, information, and the record of my involvement in the research will be kept confidential. I have been informed that everything I say will be anonymous. The data from this consultation will be recorded via paper notes and I understand that I can request a copy of the notes to review, if I wish. I understand that I am also allowed to delete or make any changes to the notes if I feel the information I provided could be improved or clarified. I understand that a record about my personal involvement in the research will be kept in a secure location at AMIII, solely for the purpose of double checking the accuracy of data and for proving to funding and other authorities that the evaluation exercise was indeed carried out. I understand that the data I provide will be anonymised and the record of my participation will be kept in a file separate from the research data to ensure my anonymity. All other parties and researchers within PROTEUS will work with information in an anonymised format. I understand that the data will be kept for one year after the PROTEUS project ends.

---

<sup>3</sup> <http://proteus-bigdata.com/>

<sup>4</sup> Grant agreement number 687691

I understand that this research conforms to European Commission guidelines and that it has been approved by the Ethics Committee in the Research Executive Agency managing Horizon 2020 projects. Finally, I have been given the contact details of the research team and I have been informed that I am free to contact Hamid Bouchachia (Project Coordinator) about any queries relating to my data or the project itself.

*Project Coordinator Hamid Bouchachia*  
Bournemouth University, UK  
+44 (0) 1202 962401  
[abouchachia@bournemouth.ac.uk](mailto:abouchachia@bournemouth.ac.uk)





*information provided to the participants. A copy of the signed and dated consent form should be placed in the project's main record at AMIII, which must be kept in a secure location.*

**Annex B – Ricard Martinez Review report****PROTEUS**

**Scalable online machine learning for predictive analytics and real-time  
interactive visualization**

**687691**

---

## **D2.3 Annex: Proteus Data Management and Ethics Plan data Protection Review**

---

**Lead Author: Ricard Martínez**

Deliverable nature:	<Report (R)>
Dissemination level: (Confidentiality)	
Contractual delivery date:	31/05/2017
Actual delivery date:	15/05/2017
Version:	First Draft
Total number of pages:	16
Keywords:	Data management, exploitation, research ethics, open access

## 1. Executive summary

This deliverable contains the analysis of the first iteration of the PROTEUS data management plan. According with the Document D2.2 [Proteus Data Management and Ethics Plan] PROTEUS will use and generate three specific data sets that fall within the following categories:

1. ArcelorMittal data provided to the project (ArcelorMittal data)
2. Derived data about the functioning of the scalable online machine learning tools, including data from the benchmarking and technical evaluation process (PROTEUS toolset data)
3. Data from the evaluation of the visualisation aspects of the tool (PROTEUS evaluation data)

As the document states the project has agreed the following principles:

- Data owned by ArcelorMittal will be shared with consortium members, although consortium members can only access this data through the project coordinator
- PROTEUS partners agree to respect the commercial confidence of the data provided by AMIII
- Human participants in PROTEUS research activities are under no obligation to participate and their involvement will be strictly voluntary
- PROTEUS partners will respect data protection and ethical research principles related to the following:
  - Participant confidentiality and anonymization
  - Informed consent
  - Data minimisation
  - Purpose limitation
  - Transparency
  - Rights of access, correction and erasure
- The coordinator will hold all commercially sensitive data and personal data and will manage access to this data for PROTEUS partners

In addition to meeting European regulations, the project will also meet legal regulations set by the Spanish government since data and many of the research activities will be located in Spain. This includes the registration of the project coordinator as a data controller with the Spanish data protection authority and the provision of specific information required by Spanish data protection legislation.

The object of this report consists on the reviewing of the document.

## 2. Introduction

This report refers exclusively to the compliance of the legal framework in the protection of personal data. This therefore excludes any opinion regarding intellectual property and other legal aspects concerning the PROTEUS Project.

For this purpose, the following references have been taken into account

### **A.-Legal Framework.**

- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance).
- Organic Law 15/1999, 13 December, of protection of Personal data. (Consolidated text. Last updated: 5 march 2011).
- Royal Decree 1720 / 2007, of 21 December, approving the regulations implementing the organic law 15/1999, 13 December, of protection of personal data. (Consolidated text. Last updated: 8 march 2012).
- Royal Legislative Decree 2/2015 of 23 October approving the consolidated text of the Statute of Workers updates and derogates Royal Legislative Decree 1/1995 of 24 March approving the consolidated text of the Statute of Workers.

### **B.- Case Law**

- *Court of Justice of the European Union.*
  - Volker und Markus Schecke (C-92/09).
  - Hartmut Eifert (C-93/09) y Land Hessen.
  - Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González (C-131/12).
  - Digital Rights Ireland (C-293/12).
  - Commission v Hungary (C-288/12).
- *Spanish Constitutional Court.*
  - STC 292/2000
  - STC 241/2012.
  - STC 29/2013.
  - STC 170/2013.
  - STC 39/2016.
- *European Court of Human Rights.*
  - Case of Bărbulescu V. Romania (Application no. 61496/08)

### **C.-Reports and Guidelines**

- Working document on the surveillance of electronic communications in the workplace. (29 May 2002).
- Working Party Article 29. Opinion 15/2011 on the definition of consent.
- Working Party Article 29. Opinion 05/2014 on Anonymization Techniques.
- Spanish Data Protection Agency. Data Protection in Labour Relationships.

- Spanish Data Protection Agency. Guidance and guarantees in the procedures of anonymization of personal data.
- Spanish Data Protection Agency. Code of good practices in data protection for Big Data projects.

### 3. General Issues

#### **3.1 The Right to data protection a European Fundamental right.**

An appropriate approach to data protection requires that the Court of Justice of the European Union has consolidated in its various judgments the fundamental right to data protection. In this regard, the consequences of three recent judgments (Costeja, Digital Right Ireland and European Commission V. Hungary) and the judgment of the European Court of Human Rights in the Bărbulescu case are briefly outlined.

The following conclusions can be drawn from these judgments relevant to the PROTEUS Project:

- Any processing of personal data affects a fundamental right. It therefore receives the highest consideration in the Constitutional Orders of Member State.
- The fundamental right is based on respect for the principles of transparency and consent, and the guarantee of rights of access, rectification, cancellation and opposition to the processing.
- Data protection also requires guaranteeing what has been inserted in the doctrine in the so-called fair information principles or fairness principles: purpose limitation principle, proportionality and data de minimisation principles, retention principle and security principle.
- The identifiability in the internet of the search engines can have an intense repercussion on the data released in the networks.
- The publication of information in open data environments must respond to the principle of minimizing the impact on data protection.
- Data Protection Authorities play a fundamental role as guarantors of the fundamental right to data protection. A very direct consequence of this is that their guidelines, reports or recommendations must be considered as Soft Law.
- In the context of labour relations, strict compliance with data protection principles must be ensured. In this sense the duty of transparency is fundamental to eliminate in its case the expectation of privacy.

On the other hand, it should be taken into account that consent should not be required in those cases where the processing is part of the labour relationship.

#### **3.2 A risk analysis under General Data Protection Regulation (GDPR).**

The General Data Protection Regulation provides a new approach based on accountability and risk analysis as well as new obligations that will be fully implemented as from May 25, 2018.

Nevertheless it is necessary to point out that the Spanish Agency of Data Protection has distinguished in its website between 'entry into force' and 'full implementation' pointing out that the Regulation shall enter into force on the twentieth day following its publication in the Official Journal of the European Union (Article 99).

The Deliverable subject of this report includes different references to elements of the GDPR in section "3.3.2 3.3.2 European General Data Protection Regulation". In addition an explicit commitment is included: "PROTEUS will take each of these obligations into consideration when processing personal data to ensure the project complies with both current legislation and forthcoming legislation".

However, there are at least three recommendations that this expert believes should be followed. The Project should consider the inclusion of specific policies of GDPR compliance in the following areas:

- *Protection of data from the design*

In this matter the Deliverable suggests that this principle has been adequately considered. This is demonstrated by the design of consent processes, or references to pseudonymization.

However, it would be convenient to document:

- a. If all areas or stakeholders involved have been considered [IT (Chief Information-Security Officer, Legal (Legal Officer or Data Protection Officer), Business and Partners in the Project).
  - b. If any specific methodology has been followed, such as the Privacy by Design Guidelines of the Ontario<sup>5</sup> Data Protection Authority or Guidelines issued in the matter by the above mentioned Spanish Data Protection Agency, or other authorities<sup>6</sup>.
  - c. If the opening of a Record of processing activities (article 30 GDPR) has been considered.
  - d. Whether the information will be extended to the data in accordance with the duty of Transparency provided for in Articles 12 to 14 of the GDPR.
- From the reading of the document it is deduced that PROTEUS complies with the risk-based approach. However, there is no specific section where the methodology developed is confirmed or where the identified risks and the measures necessary to be eliminated or mitigated are listed.

---

<sup>5</sup> See for example Privacy by Design (2011), Privacy by Design Solutions for Biometric One-to-Many Identification Systems (2014) Big Data and Innovation – Setting the Record Straight: De-Identification Does Work (2014). Available at <https://www.ipc.on.ca/about-us/guidance-documents/?yr=&topic=privacy>.

<sup>6</sup> ICO. Big data, artificial intelligence, machine learning and data protection. Available at <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>.

## 4. Ethical and Legal Issues

Following the structure of Deliverable, sections 4 and 5 in relation to Annex A, this section will discuss the following issues:

- Excluded processes
- Informed consent. The processing of personal data in the framework of labour relations.
- The impact of pseudonymization and anonymization on PROTEUS and its impact on the project.
- Some considerations regarding safety

### 4.1 Excluded processes.

According to the documentation the processes in the datasets Arcelor Mittal data and PROTEUS predictive analytics system data do not include personal data. The expert has not had access of such data. However from the descriptions contained in the Deliverable, the absence of data is clearly deduced. The data obtained refer exclusively to technical data related to industrial processes and to results obtained through Big Data Analytics Tools.

### 4.2 The processing of personal data in the framework of labour relations.

The fact that the data subject to processing belong to Arcelor Mittal workers implies taking into account a series of specific forecasts. In this respect the Deliverable states:

#### 5.1 Informed Consent

(...)

The consortium will take steps not only to ensure that potential participants are well informed but that they also feel empowered to volunteer or decline participation as they see fit. Given that this research will be carried out with employees of one of the partner's organisations, ensuring voluntary participation is paramount and will require a few additional steps. First, following good practice, the information sheet will advise participants that their participation is purely voluntary, and partner AMIII has confirmed this. In addition, non-AMIII personnel will undertake recruitment and advise participants that their participation is voluntary. Finally, during the research activity itself, those conducting the research will invite participants to re-consider their participation and to excuse themselves from the research activity. Given that the project does not involve many sensitive topics, this should be sufficient to ensure voluntary participation, but the consortium remains dedicated to addressing any further concerns raised by employees in the course of the project. Hence, the project will remain vigilant about potential conflicts and will carry out a rolling risk assessment to ensure voluntary participation. To date, an informed consent form has been drafted for future reference and appended in Annex A herein

It is considered that the regulatory compliance policies foreseen here are correct although they can be supplemented by some issues consolidated in the legislation, jurisprudence and guidelines and recommendations of the Spanish Data Protection Agency and the Working Party<sup>7</sup>.

---

<sup>7</sup> SSTC 292/2000, 241/2012, 29/2013, 170/2013, 39/2016.

Bărbulescu V. Romania (Application no. 61496/08)

Working document on the surveillance of electronic communications in the workplace. (29 May 2002).

Working Party Article 29. Opinion 15/2011 on the definition of consent. Working document on the surveillance of electronic communications in the workplace. (29 May 2002).

Spanish Data Protection Agency. Data Protection in Labour Relationships.

The jurisprudence of the Constitutional Court and the Supreme Court in Spain, as well as of the European Court of Human Rights, has repeatedly pointed out that within the framework of the labour relationship, the worker's right to privacy, prevails<sup>8</sup>. This right may be limited in different cases, either by the nature of the work performed or because the employer exercises the control for which he/she is empowered by article 20 of the workers' statute<sup>9</sup>. To this concern there are two basic principles:

- When the processing of a worker's personal data in the context of labour relations is directly related to the job benefit, it is not feasible to base such processing on consent.

In this sense, for a better understanding of the voluntary nature of the processing, elements such as making it clear that the survey does not form part of the ordinary work experience or that the conditions of realization do not affect the job, could be highlighted. The specific references to the freedom of choice of the worker, the absence of adverse consequences and the possible withdrawal of the Project when it is decided should be noted as an extremely positive element of the Deliverable and Annex. And also, to the existence of a contact person who, as Data Protection Officer, plays a role of guarantor of their rights.

In any case it is advisable to describe in as much detail as possible the conditions under which the processing will be carried out in order to establish whether it is in fact a voluntary activity or if it should be considered part of the contractual relationship. In this sense, the statement in Annex A is sufficiently clear, although it could be supplemented by a brief description of the project or a document describing the project in general and the voluntary nature of the participation. The textual structure of the informed consent of Annex A would probably be more understandable or friendly if, instead of introducing the explanatory details following phrases like "I consent", "I understand", there would be a brief introduction of the Project and the consequences of participation of the worker.

- The duty of transparency plays a key role in the

- INFORMATION ON THE PROCESSING OF PERSONAL DATA. MODES:

The duty of disclosure of Article 5 of the LOPD is an essential part of the right to data protection. The essential nature of this data applies both to the collection of personal data for processing requiring consent, and to cases in which consent is not required.

Article 5 of Organic Law 15/1999 establishes a duty generally imposed on data controllers, so that in principle affected parties must be notified of the processing of their data, both in cases where processing has the consent of the data subject, and for cases in which processing has been enabled due to other causes admissible under Article 6 of this Law." (Report 60/2004)

In the first case, this is because the consent, in addition to being given freely, specifically and in advance, must also be communicated and therefore the failure to inform invalidates the declaration of consent by

---

<sup>8</sup> Thus the Statute of Workers establishes:

Article 18. Inviolability of the worker.

Records may only be made on the worker, in their lockers and private effects, when they are necessary for the protection of the corporate assets and of the other workers of the company, in the workplace and in working hours. In doing so, dignity and privacy of the worker will be respected and a legal representative of the workers or, in the absence of the work center, of another worker of the company, will be available whenever possible.

<sup>9</sup> Article 20. Management and control of the labour activity.

(...)

3. The employer may adopt the measures he believes more convenient for supervision and control to verify the compliance by the worker of his duties and obligations, keeping in its adoption and application the consideration of his dignity and taking into account, where appropriate, the actual capacity of workers with disabilities.



the data subject or interested party. In addition, the content of the information defined under Article 5 of the LOPD constitutes a guarantee of data subjects' rights, since it enables them to know in respect of whom these rights may be exercised. (...)

In the area of personnel management, organisations must be particularly careful, both in the procedure they choose and the time at which they carry out the capture of personal data.

As regards compliance with the principle of transparency, it should be emphasized that the GDPR leaves a broad field to national legislation.

#### Article 88

##### Processing in the context of employment

1. Member States may, by law or by collective agreements, provide for more specific rules to ensure the protection of the rights and freedoms in respect of the processing of employees' personal data in the employment context, in particular for the purposes of the recruitment, the performance of the contract of employment, including discharge of obligations laid down by law or by collective agreements, management, planning and organisation of work, equality and diversity in the workplace, health and safety at work, protection of employer's or customer's property and for the purposes of the exercise and enjoyment, on an individual or collective basis, of rights and benefits related to employment, and for the purpose of the termination of the employment relationship.

2. Those rules shall include suitable and specific measures to safeguard the data subject's human dignity, legitimate interests and fundamental rights, with particular regard to the transparency of processing, the transfer of personal data within a group of undertakings, or a group of enterprises engaged in a joint economic activity and monitoring systems at the work place.

Precisely for this reason the provisions on data protection must be supplemented by those of the Workers' Statute. Article 63 of this regulation refers to the works council as "representative and associated of all workers in the company or work center to defend their interests, being constituted in each work center which census is of fifty or more workers. Article 64 gives this Committee information and consultation rights. The works council will have the right to be informed and consulted by the employer on those issues that may affect the workers. Therefore, although the rule does not specifically foresee a duty of information in this regard, since it is an issue related to fundamental rights, it is advisable to inform the Works council about the Project if it exists. It should be noted that in accordance with article 65.2 of the statute, the members of the committee may be required to «observe the duty of discretion with respect to information that, in the legitimate and objective interest of the company or the work center, has been expressly communicated on a reserved basis».

### **4.3 Pseudonymization and anonymization.**

Throughout the document, policies of pseudonymization and/or anonymization are proposed. In this regard, it should be noted that both the Working Party and the Spanish Data Protection Agency are based on a strict interpretation based on the "26th Whereas" of Directive 95/46 /EC of the European Parliament and of the Council of 24 October 1995 On the protection of individuals with regard to the processing of personal data and on the free movement of such data

(26) Whereas the principles of protection must apply to any information concerning an identified or identifiable person; whereas, to determine whether a person is identifiable, account should be taken of all the means likely **reasonably to be used either by the controller or by any other person to identify** the said person; whereas the principles of protection shall not apply to data rendered anonymous in such a way that the data subject is no longer identifiable; whereas codes of conduct within the meaning of Article 27 may be a useful instrument for providing guidance as to the ways in which data may be

rendered anonymous and retained in a form in which identification of the data subject is no longer possible;

This implies that the conditions of anonymization require that "no third party" can reidentify. In the Project, pseudonymization is designated as the default policy instead of anonymization. For this, the identity of the participants and the responses or evaluations derived from the interviews and from viewing the results are separated into two data sets.

In health research, and in general when some risk can be derived, or a right must be guaranteed for the participants, research methodologies advise or require re-identification. In the case of PROTEUS, it appears that there is probably a need to prove the "real or effective" participation in the investigation for the administrative justification of the Project, in order to guarantee scientific accuracy to eventual publications, or to guarantee the value of data in content release in open data environments. If these were reasons for pseudonymization, it would be appropriate to underline them explicitly and explain why a complete anonymization is not done.

On the other hand, section 5.2 describes how a simple method for anonymization/pseudonymization has been used by assigning a numerical value to each participant. In addition, special cases are considered:

4. Participants with particular identifying features or experiences may be managed by mixing these with other participants' characteristics (e.g., switching places of birth) to make each participant less identifiable. Where necessary, some identifying features may be removed from the data if it cannot be anonymised.

The methodology is fully shared although the example "switching places of birth" is not adequately understood. Probably, from the point of view of third parties there are more sensitive elements that allow reidentification as the unique characteristics of the workplace. Also, ignoring if the questionnaires or battery of questions to be used are different depending on the job profile of the worker, an opinion cannot be given on whether there could be some type of reidentificación from the answers.

On the other hand, if we take into account the Spanish sector legislation there are two laws from which we can extract complementary ideas useful for the project. We refer to:

- Law 37/2007, of November 16, on reuse of public sector information.
- Law 19/2013, of December 9, on transparency, access to public information and good governance.

Law 37/2007 allows conditions to be imposed on the re-use of anonymized information in Open data contexts:

Article 8. Conditions for reuse.

Reusing the information of the Administrations and the public sector bodies referred to in article 2 of this Law may be subject, among others, to the following general conditions:

(F) Where information, even if provided in a decoupled form, contains sufficient elements that would allow identification of stakeholders in the re-use process, a ban on reversing the dissociation procedure by adding new data obtained from other sources.

Both this rule and the Transparency Law attribute legal responsibility to those who obtain data from a public source process them, when they are personal, or perform some type of reidentification.

Therefore, it would be advisable to include in the regulatory framework among the parties in the consortium and in the terms and conditions of use of the open data spaces a clause adding a prohibition or express commitment of non-reidentification both for the Internal use of the data and in eventual publications.

#### ***4.4 Some considerations about safety.***

According to the description of the data in the Project, the applicable security level will be the basic one provided in Article 81 of Royal Decree 1720/2007 of 21 December, which approves the Regulation for the development of Organic Law 15/1999, of 13 December, on the protection of personal data. As stated in the document, the organization applies the ISO / IEC 27001: 2013 standard capable of meeting the high standard of the same regulation.

On the other hand, security information regarding encryption of communications is completed. It should be recalled that the Working Party, and now Articles 32 and 34 GDPR consider encrypting data itself as a high quality measure. In this sense, article 34 excludes the duty of notification of security incidents to the data subject when the data had been encrypted.

If personal data are kept in local systems there is no consideration to be made. However it should be remembered that for the processing in Cloud there is the security standard ISO / IEC 27017: 2015, Information technology, Security techniques, Code of practice for information security controls based on ISO / IEC 27002 for cloud services.

## 5.1 RECOMMENDATIONS

Taking into account the above mentioned considerations, the following recommendations should be followed:

- To document, if applicable, the methodologies of data protection from design and by default and give an impact analysis on data protection, and their results.
- Consider opening an internal Registry of the processes.
- Verify if the information to data subjects complies with the provisions of art. on transparency provided for in the GDPR.
- Describe in as much detail as possible the conditions under which the processing will be carried out in order to establish whether it is in fact a voluntary activity or if it should be considered part of the contractual relationship.
- Inform about the processing of personal data in the framework of the Project to the works council
- Explicitly document the reasons for the pseudonymization and explain in such a case why a complete anonymization is not performed.
- Include in the regulatory framework between the parties in the consortium and in the terms and conditions of use of the open data spaces that a clause could be established to add a prohibition or express commitment of non-reidentification both for the internal use of the data and for any publications.
- It should be remembered that for the processing in Cloud there is the security standard ISO / IEC 27017: 2015, Information technology, Security techniques, Code of practice for information security controls based on ISO / IEC 27002 for cloud services.

## **Annex C - Ricard Martinez CV**

Data Protection Officer of the University of Valencia and Director of the Microsoft- University of Valencia Privacy & Digital Transformation Institutional Chair.

Former Head of Area of Transparency and Open Government (Diputación de Valencia). Former President of the Spanish Privacy Professionals Association (May 2016). Data Protection Officer (University of Valencia- November 2015). Head of the Studies Area of the Spanish Data Protection Authority (2007-2011). Assistant Professor in Constitutional Law, and law and Information Technologies at Open Catalonia University (UOC). Collaborator in various master's degrees at U. Valencia, U. Polytechnic of Valencia, U. Carlos III. Director of the Official Master of Data protection Law at the International University of la Rioja. Part Time Researcher at U. Jaume I (General Data Protection Regulation Impact (DER2015-63635-R)). Ph. D. in Law (with distinction), University of Valencia. Dissertation about Data protection Right, Technology and Freedoms. Member of the Academic Council of Fide. Publications.

<https://dialnet.unirioja.es/servlet/autor?codigo=173672>

## Annex D – List of changes after Ricard Martinez review

Following the review and recommendations of our dedicated Spanish data expert, Richard Martinez, the following amendments have taken place in D2.3.

	In response to the following recommendations from Richard Martinez' report:		Changes in the following sections of the final D2.3 report were made:
p.7	Section 3.2	p.14	<p><b>Section 3.3.2</b>            “Data protection by design and default” was amended to include all stakeholders and clarify the rationale for pseudonymisation            “Processing personal data for research purposes” was amended to include requirement of article 30, 32 of GDPR.            A new bullet point on Transparency and communication with data subjects was added to include GDPR requirements outlined in Articles 12-14 of GDPR</p>
P11-12	Section 4.3	p.15	<p><b>Section 3.3.3</b>            Reference to Law 37/2007 (article 8) regarding re-identification was introduced at the end of the section</p>
p.9		p.22	<p><b>Section 5.1</b>            Reference to principles of jurisprudence as per the Spanish law and prior consultation with the Works council about the project was introduced at the end of the section.</p>
p.12	Section 4.4	p.24	<p><b>Section 6.2</b>            Reference to ISO standards with respect to cloud security processing of personal data was added.</p>
		Annex A	<p>Changes in the information sheet were made to make more explicit and prominent recommendation regarding jurisprudence and transparency and explicitly declare a Data Protection Officer</p>