



**European Research Council**  
Executive Agency

Established by the European Commission



**Horizon 2020**  
**European Union Funding**  
**for Research & Innovation**

**European Research Council (ERC)**

**ERC Data Management Plan**

## ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)



European Research Council

Established by the European Commission

1

Project Acronym	Project Number
ISEProD	835201

**DISCLAIMER.** Please note that the ERC Data Management Plan is not a part of the Ethics Review. It is the responsibility of the Principal Investigator to inform the ERCEA Ethics Team of any ethics issues/concerns regarding the collection, processing, sharing and storage of data in relation to the project. The Principal investigator can also be asked to submit an Ethics Data Management Plan (Ethics DMP).

Deliverable title	D1.1 Data Management Plan
Due date (in months)	6 <sup>1</sup>
Lead Beneficiary	Luiss Libera Università Internazionale degli Studi Sociali Guido Carli (Luiss University)
Type	ORDP
Dissemination Level	Public
Related WP	WP1 - Participation in the Open Research Data Pilot
Author	Fabiano Schivardi (Luiss)
Internal Reviewers	Anna Elisa D'Agostino (Luiss), Roberta Pellicano (Luiss), Chiara Annulli (Luiss)
Contact(s)	<a href="mailto:fschivardi@luiss.it">fschivardi@luiss.it</a>
Funder	European Research Council
Project Title	Industrial Structure and the European Productivity Growth Divergence
Project Acronym	ISEProD
Grant Agreement Number	835201
Start Date	2020-01-01
Project Duration (months)	60
Max ERC Funding	1 944 581 €
Approved by	Fabiano Schivardi (Luiss)
Version	1.0
Date Last Update	12 August 2020

---

<sup>1</sup> Delayed due to COVID19. Given the COVID-19 situation, the due date for the deliverable has been postponed for a maximum of 60 days as agreed with the Project Officer upon the HI's request of June, 22<sup>nd</sup>, 2020.

Version History			
Version	Date	Author(s)	Notes
0.1	03 August 2020	Fabiano Schivardi	Draft created
0.2	06 August 2020	Anna Elisa D'Agostino (Luiss), Roberta Pellicano (Luiss), Chiara Annulli (Luiss)	Internal Review, administrative information added.
0.3	07 August 2020	Fabiano Schivardi (Luiss)	Document reviewed
0.4	7 August 2020	Roberta Pellicano (Luiss)	Institutional details and policies links added, document formatted with ERC template
0.5	11 August 2020	Chiara Annulli (Luiss)	Document reviewed, typos removed.
1.0	12 August 2020	Fabiano Schivardi (Luiss)	First version release approved

# INDEX

1		2
<b>1</b>	<b>ISEPROD APPROACH TO DATA MANAGEMENT</b>	<b>6</b>
<b>1.1</b>	<b>Project Overview</b>	<b>6</b>
1.1.1	ISEProD Abstract	6
<b>1.2</b>	<b>Data Management Scope</b>	<b>6</b>
1.2.1	The role of data collection in ISEProD	6
<b>1.3</b>	<b>Overall Structure of the data management in ISEProd</b>	<b>6</b>
1.3.1	Data Management Reference Documents and Framework	7
<b>1.4</b>	<b>Data management during the project</b>	<b>8</b>
1.4.1	Roles and Responsibilities	8
<b>1.5</b>	<b>Long-term data management</b>	<b>8</b>
1.5.1	Standard citation	8
1.5.2	Licenses	9
<b>1.6</b>	<b>Open Access to Publications and other types of research results</b>	<b>9</b>
<b>1.7</b>	<b>Allocation of resources</b>	<b>10</b>
1.7.1	Estimated Costs	10
1.7.2	Potential value of long-term data preservation	10
<b>2</b>	<b>DATA COLLECTION AND GENERATION</b>	<b>10</b>
<b>2.1</b>	<b>Data Types</b>	<b>10</b>
2.1.1	Secondary data	11
2.1.2	Primary data	14
<b>2.2</b>	<b>Anonymisation</b>	<b>15</b>
<b>3</b>	<b>DATA STORAGE AND BACKUP</b>	<b>16</b>
<b>4</b>	<b>DATA DOCUMENTATION</b>	<b>16</b>
4.1.1	Project Folder Structure	18
4.1.2	File naming convention	18
<b>5</b>	<b>DATA ACCESS</b>	<b>18</b>
<b>6</b>	<b>DATA SHARING AND REUSE</b>	<b>19</b>
<b>7</b>	<b>DATA PRESERVATION AND ARCHIVING</b>	<b>19</b>
<b>ETHICAL ISSUES</b>		<b>20</b>
<b>8</b>	<b>REVISION OF THE DATA MANAGEMENT PLAN</b>	<b>20</b>

# 1 ISEProD approach to Data Management

## 1.1 Project Overview

### 1.1.1 ISEProD Abstract

The last two decades have witnessed substantial progress in the measurement of productivity. However, the understanding of the deep determinants of productivity growth is still limited. This is important in the face of both the global productivity slowdown emerged since mid 2000s and the large productivity growth divide in Continental Europe between the North and the South. ISEProD PI hypothesizes that ownership, control and finance (in short “industrial structure”, IS) play a key role and that their importance, due to technological change, has grown over time. The PI plans to merge insights and state-of-the-art techniques from industrial organization and corporate finance to assess the causes and consequences of different ISs. Methodologically, ISEProD will use a combination of unique data sources (credit register, matched employer- employee data, firm input-output relationships), quasi-experimental design (changes in the banking regulation, executives mortality shocks in local labor markets) as well as structural techniques (selection models and Bayesian learning models) to address the fundamental identification issues that plague the literature. The PI also plans a major data collection effort in four European countries to study succession in family firms. The project will pursue four specific objectives in four deeply integrated work packages:

- a) The role of firm ownership and control on skill acquisition and risk taking;
- b) Skills supply: the effects of the education of entrepreneurs and managers;
- c) Industrial structure, input/output relationships and IT adoption;
- d) Access to finance with asymmetric information and imperfectly competitive financial markets.

The ambition is to generate a discontinuous change in our understanding of the relationship between corporate governance/finance and productivity growth. On the way, ISEProD expects to contribute to the key debate for the future of the European Union on the diverging productivity performance of European economies.

## 1.2 Data Management Scope

### 1.2.1 The role of data collection in ISEProD

The project has a strong empirical orientation, so the data will have a fundamental role in its development. In fact, all the theoretical hypotheses considered to explain the sources of differences in productivity growth between Southern and Northern Europe will be put to test empirically. The data will therefore both determine which hypotheses are supported and their relative importance in determining growth differences.

## 1.3 Overall Structure of the data management in ISEProD

Research Data Management is an important part of the research activity, focusing on the development of an internal project strategy that covers all the data lifecycle.

Efficiently managing data within ISEProD will guarantee that the data is properly handled during acquisition and processing, and assures that data will be archived and disseminated to gain larger impact and assure no information is lost.

The Data Management Plan is one of the ISEProD deliverables and consists in a living document to be delivered in its first version at the beginning of the project activities and later updated. The plan also aims at ensuring consistency in the resource and budgetary planning for data management related costs. The ISEProD DMP also includes details on how the Open Access to publications will be carried out.

ISEProD DMP will describe the types of data that will be handled to achieve project objectives, and information about how the data will be handled both during and after the project.

According to the guidelines provided by the ERC under Horizon 2020, the project data must be FAIR (Findable, Accessible, Interoperable and Reusable) as much as it is possible, except when there are substantial reasons to keep the data confidential or otherwise protected. Among other aspects, the DMP also describes the possible levels of data openness within the ISEProD project, the differences in their handling and compliance with the Intellectual Property Rights (IPR).

In ISEProD, participation to the Open Research Data pilot is managed by WP1 - Participation in the Open Research Data Pilot, which will run for the overall project life.

The aims of WP1 are:

- the development and necessary update of ISEProD Data Management Plan;
- the deposit of the published version or final peer-reviewed manuscript accepted for publication of ISEProD scientific publications in a research data repository,
- ensuring open access to the deposited publications within twelve months of publication;
- the deposit of the research data that validate the results of the deposited scientific publications
- the deposit of the digital research data generated by the project in a research data repository so that the research community can exploit them for further research;
- clarifying to the research community which tools they have to rely on to use the raw data in order to validate ISEProD Research.

#### 1.3.1 Data Management Reference Documents and Framework

- 1) This Data Management Plan was prepared to comply with the Grant Agreement Art. 29.2 and 29.3.
- 2) This Data Management Plan is compliant with the “European Research Council (ERC) Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020” ([https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf))
- 3) This Data Management Plan (DMP) has been prepared by taking into account the ERC template attached to the “Guidelines on Data Management in Horizon 2020” ([https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management\\_en.htm](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm)).
- 4) The project will comply with the requirements of 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)
- 5) Type of data, storage, recruitment process, confidentiality, ownership, management of intellectual property and access: The Grant Agreement is to be referred to for these aspects, particularly Articles 18, 23a (Management of intellectual property), 25 (Access right to background), 26 (Ownership of results), 27 (Protection of results – Visibility of EU Funding), 28 (Exploitation of results); 29 (Dissemination of results – Open Access – Visibility of EU Funding), 30 (Transfer and licensing of results), 31 (Access Rights to results), 32 (Working Conditions for the PI and his/her team), 36 (Confidentiality), 39 (Processing of personal data) and 52 (Communication between Parties) and “Annex I – Description of Work” of the Grant Agreement.

The Grant Agreement was signed on 05/11/2019 and was set into force on 01/01/2020. The procedures that will be implemented for data collection, storage, access, sharing policies, protection, retention and destruction will be according to the requirements of the national legislation (Italy) and in line with the EU standards.

- 6) The Principal Investigator and its Host Institution of the project will ensure that EU standards are followed. Regarding the issue of informed consent for all survey procedures, all participants will be provided with a Participant Information Sheet and Consent Form to provide informed consent. The default position for all data relating to personal information will be anonymous.
- 7) An ethical approach will be adopted and maintained throughout the fieldwork process. The Principal Investigator and its Host Institution will assure that the EU standards regarding ethics and Data Management are fulfilled. The Principal Investigator will proceed with the survey according to the provisions of the national legislation that are adjusted according to the respective EU Directives for Data Management and ethics.
- 8) This document was drafted according to the related institutional policies and regulations of the Host Institution: Luiss university policy on open access to scientific publications (<https://www.luiss.edu/sites/www.luiss.it/files/POLICY-OA-LUISS-ENG-2016.pdf>) and Luiss regulations for the processing of sensitive and judicial data pursuant to legislative decree no. 196/2003 ([https://www.luiss.edu/sites/www.luiss.it/files/luiss\\_regulations\\_processing\\_sensitive\\_judicial\\_data.pdf](https://www.luiss.edu/sites/www.luiss.it/files/luiss_regulations_processing_sensitive_judicial_data.pdf))

## 1.4 Data management during the project

### 1.4.1 Roles and Responsibilities

The PI is responsible of ensuring the quality of the DMP and for its implementation during the project lifetime.

The PI will be advised by Luiss Data Protection Officer (Mr. Francesco Flego) and supported by the Research office and the Luiss University Guido Carli Committee for Research.

In this context the PI is responsible for:

- uploading all data related to their case study into the designated data repository;
- verify that sufficient metadata is provided;
- ensure that data structure, naming, etc. comply to the DMP;
- manage data privacy and ethical issues.

### 1.5 Long-term data management

After the project end, all data that do not contain personal information (including necessary documentation, metadata, code, consent form, software, etc.), or that do not have specific restrictions, or that are not otherwise protected will be stored and made publicly available in the Luiss institutional repository under appropriate license. A standard citation will be made available for users to ensure author(s) will be appropriately cited and get credit for their work.

Luiss institutional repository (<http://iris.luiss.it/>) is conceived to ensure long term preservation of its contents. The possible access granted by the platform include:

- Open access
- Closed access
- Restricted access
- Embargoed access

Luiss repository is compliant with OpenAIRE guidelines and registered in OpenAIRE.

The detailed Luiss institutional repository user guide is available at:

<https://biblioteca.luiss.it/sites/biblioteca.luiss.it/files/IRIS-LUISS-GUIDE-1.4-eng.pdf>

The Institutional policy on Open Access to scientific publication is available at:

<https://www.luiss.edu/sites/www.luiss.it/files/POLICY-OA-LUISS-ENG-2016.pdf>

#### 1.5.1 Standard citation

When the data will be deposited and access will be granted for reuse, a standard citation will be provided to the users for copy and paste purposes.



This simple yet important element will allow the users to cite the data directly to give the data producer appropriate credit, enabling easier access to the data for repurposing or reuse and research results verification.

Based on the [Force 11 Joint Declaration of Data Citation Principles](#), the current [DataCite Metadata Schema](#) and the [DMPTool service guidelines](#) v2.2.2 of the [University of California Curation Center](#), the ISEProD dataset standard citation includes the following elements

- **Author(s)**
- **Title**
- **Publication year** when the dataset was released (may be different from the Access date)
- **Publisher** – the data center, archive, or repository
- **Identifier** – a unique public identifier (e.g., the Luiss repository handle or DOI)

**Standard citation:**

<<Author(s)>> (<<Publication year>>): <<Title>>. <<Publisher>>, <<Identifier>>

**Example citation**

Schivardi F (2020): ISEProD Sample Dataset Title. Luiss institutional repository, <https://doi.org/10.xxxx/yyyy>.

For datasets deposited in the Luiss institutional repository, the field “note” will be used to provide standard dataset citation.

### 1.5.2 Licenses

Following the “European Research Council (ERC) Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020”, the default license for datasets will be:

- CC-0 or Public domain
- CC-BY

Both cases should be accompanied by a suggestion for citation (see corresponding paragraph). This choice is also in line with Fact Sheet on Creative Commons and Open Science<sup>2</sup>.

## 1.6 Open Access to Publications and other types of research results

ISEProD will comply with ERC Open Access mandate on scientific literature by depositing a copy of each published articles derived by the project activities that will undergo a peer review process. All deposited article will be assigned an open access right with a maximum embargo period of 12 months.

Green open access will be the preferred route, although the PI reserved a specific budget to publish in gold open access journals (see paragraph 1.7.1).

The corresponding author of the paper published is responsible for depositing the open access version of the paper that is compliant with the publisher copyright and the ERC mandates, and the PI should ensure application to the rules specified in this paragraph, that apply to all research literature that undergo peer review and that result from ISEProD activities. All articles should contain the reference to the project, possibly in the “funding” or “acknowledgment” sections. Suggested form is the following<sup>3</sup>:

*“This [article/result/equipment/video is part of] ISEProD project (Industrial Structure and the European Productivity Growth Divergence) that has received funding from the European Research Council (ERC) under the European Horizon 2020 research and innovation programme (Grant agreement No. 835201)”*

---

<sup>2</sup> ‘Fact Sheet on Creative Commons and Open Science’, Creative Commons UK, DOI:10.5281/zenodo.840652, CC BY 4.0, <https://creativecommons.org/licenses/by/4.0/>

<sup>3</sup> See a similar generic text available in all official EU language at the following link: <https://erc.europa.eu/sites/default/files/content/pages/pdf/Language-versions-of-Acknowledgement.pdf>

Prior to submission, the corresponding author should:

- 1) Check the journal policy on Sherpa Romeo (<https://v2.sherpa.ac.uk/romeo/>) and on the journal and/or editor website;
- 2) In case the embargo period and or other conditions applied by the publisher do not comply with ERC Open Access mandate, the corresponding author should contact the journal editor and possibly contract conditions that comply with the ERC mandate. This may include the [https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-oa-guide-model-for-publishing-a\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-oa-guide-model-for-publishing-a_en.pdf) or finding a different venue for publication.
- 3) The corresponding author must keep a copy of the pre-print (original manuscript submitted to the editor) and of the post-print (accepted manuscript) in the corresponding publication folder (see section 4) for depositing purposes.
- 4) As soon as possible after paper acceptance, at latest at the time of publication, the corresponding author should deposit in compliance with ERC open access mandate and publisher open access policy, the version of the article that is permitted for open access in the Luiss repository, include reference to the ISEProD grant in the metadata, and set the proper access or embargo rights to the file containing the machine readable version of the article

For corresponding authors not affiliated to Luiss, the above should be guaranteed by the PI, who should deposit a copy of the manuscript in Luiss repository.

## 1.7 Allocation of resources

### 1.7.1 Estimated Costs

The budget related to compliance with the Open Access mandated of ERC within ISEProD project is allocated in two ways

**Cost to make data management FAIR** (findable, accessible, interoperable and reusable):

budget 5,000 euros to prepare the documentation and to make sure that the data can safely be distributed, addressing any possible confidentiality concern. All empirical work will be clearly documented and codes distributed. Given that this will mostly be performed by the RAs, the cost is already included under the corresponding voice above.

**Journal open access costs.**

An overall budget of 17500 euros will be allocated for publications in Gold OA Journals (average cost of 2,500 euros for each paper for submission fees and open access corresponding to at least seven papers).

### 1.7.2 Potential value of long-term data preservation

ISEProD data will be preserved for future reuse by the scientific community. The data will most likely be of interest to researches in the industry dynamics, productivity and corporate governance literature. I expect other researchers to use them to address other questions in these fields. Subject to privacy issues, the data could be combined with other datasets and also be updated in the future with follow-up surveys.

## 2 Data Collection and Generation

### 2.1 Data Types

The Principal Investigator identified 6 different types of data sources, including primary and secondary data. Each data type is accompanied here by a table providing the following information:

<b>Dataset type</b>	Taken from the list provided in paragraphs 2.1.1 and 2.1.2
<b>Origin</b>	Primary or secondary
<b>Source</b>	Each type of data has a specific Source (INPS, Bank of Italy, etc...). If data has been generated or reused, the original source(s) of the dataset must be indicated. That can be also a paper, report, public dataset, etc. ideally, a

	DOI, PID or URL can be added to the original source.
<b>Expected size</b>	To be estimated in GB
<b>Format</b>	Eg: txt, csv, pdf, etc...
<b>Documentation</b>	The documentation can be provided in different forms and will be stored in the corresponding data subfolder (description text files, codes developed to analyse the data, tools used to process the data, methods, algorithms, related informed consent, anonymization tool/algorithm, ...)
<b>Access right</b>	May be one of the following: <ul style="list-style-type: none"> <li>- Open Access</li> <li>- Embargoed access (in this case: motivate embargo, define embargo rules)</li> <li>- Restricted access (in this case: motivate restricted access, define who and how can access the data)</li> <li>- Closed Access (in this case: motivate closed access)</li> </ul>
<b>License</b>	For secondary datasets, the license will be provided by the License has to be provided only for Open Access or Embargoed access rights. In this cases, only CC-0, CC-BY or public domain should be used.
<b>Related vocabulary(ies)</b>	If relevant, ant domain specific vocabulary that will be used in relation to this dataset should be listed
<b>Data Quality Assurance Procedures</b>	

For each data type, the description of the documentation that will be stored together with each dataset belonging to that type is provided in the corresponding data subfolder (see paragraph 4). For some of the secondary datasets, it may not be possible to store local copies; in this case, only the metadata.txt file will be kept by the PI in accordance to the DMP strategy.

### 2.1.1 Secondary data

All the 5 types of secondary data sources, the datasets are owned and managed by public institutions within the EU, which are in charge of ensuring that all the national and union-wide privacy requirements are satisfied. In particular, in all cases the data will not be delivered to the Principal investigator directly but will be accessed on a server of the owing institution. The data are always anonymized and only aggregate results (descriptive tables, regression coefficients, data for graphs) can be downloaded, after a check from the owning institution that makes sure that they pass the anonymity test.

All data collected within the research will be subject to specific security measures All data will be processed with confidentiality and by electronic instruments which need authentication, according to the suitable standards to guarantee an adequate security level.

Within all the ISEProD activities that involve secondary data, only strictly necessary data for the realization of the project objectives will be used. All the secondary data are accessed at the providing institutions and will not be given directly to the PI. Details are provided in the tables below.

The secondary data types are the following:

### **Danish Administrative Tax Records.**

These data are accessible through Copenhagen Business School.

<b>Dataset type</b>	<b>Danish Administrative Tax Records</b>
<b>Origin</b>	Secondary
<b>Source</b>	Statistics Denmark. <a href="https://www.dst.dk/en">https://www.dst.dk/en</a>
<b>Expected size</b>	Unknown: Statistics Denmark grant access to different datasets.
<b>Format</b>	Stata.
<b>Documentation</b>	Data description and programs will be provided.
<b>Access right</b>	Restricted access. The data can be accessed working with researchers located in Denmark at Copenhagen Business School that provided the PI with an institutional email. Through this and an authentication procedure the PI can access remotely the data. Nothing can be exported directly. All output are inspected by Statistics Denmark prior to being sent to the authors.
<b>License</b>	Will be provided with exact access procedure
<b>Related vocabulary(ies)</b>	No domain specific vocabulary be used in relation to this dataset
<b>Data Quality Assurance Procedures</b>	These data are managed by a public institution which is responsible for the quality of the data.

#### **Social security records of the Italian National Institute for Social Security (INPS).**

These data are only accessible on site at the Social Security Administration. Access to the INPS data was granted via a public call for research proposals for the Visitinps<sup>4</sup> program in 2018.

<b>Dataset type</b>	<b>Italian social security records</b>
<b>Origin</b>	Secondary
<b>Source</b>	Italian National Institute for Social Security (INPS)
<b>Expected size</b>	Several gigabytes.
<b>Format</b>	Stata.
<b>Documentation</b>	Data description and programs will be provided.
<b>Access right</b>	Restricted access. Data can be accessed by applying to the periodic call for projects of the Visitinps program (see <a href="https://www.inps.it/NuovoportaleINPS/default.aspx?itemDir=53372">https://www.inps.it/NuovoportaleINPS/default.aspx?itemDir=53372</a> ). Access is given only on site at INPS premises. Nothing can be exported directly. All output are inspected by INPS prior to being sent to the authors.
<b>License</b>	Will be provided with exact access procedure
<b>Related vocabulary(ies)</b>	No domain specific vocabulary be used in relation to this dataset
<b>Data Quality Assurance Procedures</b>	These data are managed by a public institution which is responsible for the quality of the data.

**Data from INAPP, the Italian public policy institute that performs research on the labor market.**

<sup>4</sup> <https://www.inps.it/nuovoportaleinps/default.aspx?itemDir=47212>

The rules for accessing the data are similar to those of the INPS. INAPP is responsible for performing the match with the data from INPS.

<b>Dataset type</b>	<b>Italian labor market</b>
<b>Origin</b>	Secondary
<b>Source</b>	Italian public policy institute (INAPP, <a href="https://www.inapp.org/">https://www.inapp.org/</a> )
<b>Expected size</b>	Unknown at the moment.
<b>Format</b>	Stata.
<b>Documentation</b>	Data description and programs will be provided.
<b>Access right</b>	Restricted access. INAPP requires external researchers to work together with an INAPP researcher, who maintains exclusive access rights to the data.
<b>License</b>	Will be provided with exact access procedure
<b>Related vocabulary(ies)</b>	No domain specific vocabulary be used in relation to this dataset
<b>Data Quality Assurance Procedures</b>	These data are managed by a public institution which is responsible for the quality of the data.

**Data on IT adoption and input/output relations between firms is based on data “owned” by the Italian tax authority (the Department of Fiscal Affairs at the Treasury).**

The rules are the same as above. Differently from INPS and INAPP, the Department of Fiscal Affairs has not yet set up a codified procedure for accessing the data. The PI is in the process of discussing with them. The project is to set up a system similar to INPS: on-site access to researchers, with a set of rules that ensures that all confidentiality and privacy issues are taken care of. Due to the heavy involvement of the Department of Fiscal Affairs in the policy measures enacted to contrast the Covid crisis, the procedures to allow researchers access to the data have been suspended till further notice. It is therefore not possible to provide any precise information at the moment.

<b>Dataset type</b>	<b>Italian tax authority data</b>
<b>Origin</b>	Secondary
<b>Source</b>	Department of Fiscal Affairs at the Italian Treasury.
<b>Expected size</b>	--
<b>Format</b>	--
<b>Documentation</b>	--
<b>Access right</b>	--
<b>License</b>	--
<b>Related vocabulary(ies)</b>	--
<b>Data Quality Assurance Procedures</b>	These data are managed by a public institution which is responsible for the quality of the data.

**Data of firm-bank relationships managed by the Bank of Italy.**

In this case, access to the data is only possible to Bank of Italy employees, that are therefore directly involved in the projects. As for the other cases, the data are accessed only on site, and the Bank of Italy has strict procedures to make sure that privacy and confidentiality constraints are met.

Preliminary activities on the data have already been realized and a working paper has been published<sup>5</sup>.

<b>Dataset type</b>	<b>Firm-bank relationships</b>
<b>Origin</b>	Secondary
<b>Source</b>	Bank of Italy.
<b>Expected size</b>	50 GB.
<b>Format</b>	Stata.
<b>Documentation</b>	Data description and programs will be provided.
<b>Access right</b>	Restricted access. The Bank of Italy requires external researchers to work together with a bank of Italy researcher, who maintains exclusive access rights to the data.
<b>License</b>	--
<b>Related vocabulary(ies)</b>	No domain specific vocabulary be used in relation to this dataset
<b>Data Quality Assurance Procedures</b>	These data are managed by a public institution which is responsible for the quality of the data.

### 2.1.2 Primary data

#### **Survey on family firm successions.**

The only project activity that entails a direct collection of data is that on succession in family firms. Firms that went through a recent succession at the helm of the firm will be identified through the consultation of public sources. Some countries, such as Italy, France and Nordic countries, have public registries on board members and owners. An alternative is to consult firms' web sites, either manually or through a web scraping algorithm. The data collection will be based on a direct (either in person or via phone) interview. Recruitment will be based on a direct consent of the participant. Personal data collected (related to the succession in family firms) will be treated by qualified staff, formed and authorized with suitable act containing the instructions to guarantee to data subjects the right levels of security (authentication, limitation and differentiation of accesses...). To maximize take-up, national association of entrepreneurs will be involved, that will introduce the initiative and support it.

In particular, the survey/interviews will be design to collect information about the number, the gender and the educational attainments of the entrepreneur's children. This information is essential to estimate how internal succession vs. hiring an external manager responds to differences in the "supply" of managerial skills within the family.

For the project activities that imply a treatment of personal data, the principle of the minimization of the data will be respected. For this reason, only data which are useful to develop the study will be collected and only the treatments necessary to pursue the purposes adequately indicated in the privacy notice will be carried out, in compliance with article 13 of Reg. EU/2016/679. The informed consent for data sharing and long term preservation will be included in the survey and interviews dealing with personal data on **family firm successions data**.

<b>Dataset type</b>	<b>Family firm successions data</b>
<b>Origin</b>	Primary
<b>Source</b>	Data collected within ISEProd through interviews/survey/desk search
<b>Expected size</b>	Unknow.

<sup>5</sup> <http://docenti.luiss.it/schivardi/files/2012/11/BankNetworks.pdf>

<b>Format</b>	Stata.
<b>Documentation</b>	Data description and programs will be provided.
<b>Access right</b>	<p>In general, we plan to use Embargoed access, that is, to grant access once we will have finished our working paper. We haven't yet defined clearly the access rules. Those will depend on the capacity to ensure anonymity, which cannot be determined ex-ante. The plan is to have a public database where anonymity is fully preserved, at the cost of having to aggregate some information (for examples, for each firm, reporting the sales class rather than the exact sales figure).</p> <p>For the secondary datasets, when provided to ISEProD after proper anonymization or in aggregated fashion, the data will be made openly available (open Access) through the Luiss repository when the data is needed to validate the results presented in scientific publications, including associated metadata. In this case, the dataset used by ISEProD will be deposited at the time of publication submission and access to the datasets should be restricted/embargoed until the article is published. This way the dataset could be provided with a proper DOI and cited directly in the article at the time of the submission to the journal editor.</p> <p>For the other raw secondary datasets (eg. Non-anonymised dataset or full raw data deriving from an extraction which will not be used entirely in a publication, etc), the default should be: deposit the dataset and related documentation and decide the access on single dataset basis (could also remain restricted until project ends or even after if restriction motivations still apply).</p>
<b>License</b>	To be determined depending on the specific dataset. CC-BY and CC0 will be preferred.
<b>Related vocabulary(ies)</b>	No domain specific vocabulary be used in relation to this dataset
<b>Data Quality Assurance Procedures</b>	We will compare our survey data with public sources or other comparable datasets to ensure data quality.

## 2.2 Anonymisation

ISEProD will handle data that include personal information. For the managing and sharing purposes, the data collected by ISEProD (**Survey on family firm successions**) will have to be anonymized.

Depending on the structure of the data, that is still under definition, an appropriate algorithm will be applied to the dataset.

The use of anonymising tools such as AMNESIA<sup>6</sup> will be considered.

With regard to the personal data, only anonymised data is allowed to be placed on the selected repository. The raw confidential data will be securely stored by the PI in accordance to the Luiss Institutional provision and policies.

The only data that will be directly managed by the PI is the self-collected one on succession in family firms. Potentially, these data could be used for other projects by other researchers.

Moreover, it might be important to be able to conduct follow up studies. To satisfy these two requirements, ISEProD plans to store the original, non-anonymized data on a secure storing infrastructure. Pseudonymization techniques will be applied to the dataset that will eventually be shared. The pseudonymization will entail:

- a. The removal of all personal information (individual and company names etc);
- b. The treatment of the characteristics that might allow to reconstruct the identity of the subjects, such as firm characteristics. For example, from the exact value of the sales data, one might be able to identify the firm and, through this, the owner. All the information that allows to “reverse-engineer” identity will be properly treated, for example by substituting sales with sales classes. The exact procedure can only be determined with the actual data in hand.

When shared in Open Access, anonymization will be ensured before depositing the data into Luiss repository.

### 3 Data Storage and Backup

All data and documentation resulting from the project activity will be subject to Luiss storage and backup policies.

Files are stored in cloud environment based on Microsoft Azure. The disaster recovery solution is based on the backup solutions offered by Microsoft Azure cloud services. On a daily basis a backup is made for all data stored on Microsoft Azure cloud environment. The backups of data are securely kept and always ready to be used in case of fault or incidents affecting the IT Applications. The existing contract between Luiss and CRUI regarding Microsoft Azure cloud services does not make DR multi region solutions available, because the only accessible region is Amsterdam.

Microsoft designed Azure with industry-leading security controls, compliance tools, and privacy policies to safeguard the data in the cloud, including the categories of personal data identified by the GDPR. These also help to comply with other important global and regional privacy standards such as ISO/IEC 27018, EU-U.S. Privacy Shield, EU Model Clauses, HIPAA/HITECH, and HITRUST. Access to the tools used to conduct the research is regulated providing authentication with 8-feature passwords. Information obtained in the context of the study are preserved inside the university premises, for which limited and controlled access is provided.

Subjects who will process personal and anonymized data will be identified and appointed with a specific act that binds them to respect the obligation of confidentiality and all the principles established by the law regarding the protection of personal data.

### 4 Data Documentation

All primary data from ISEProD will be stored locally and will be subject to the storage and backup conditions and policies highlighted in section 3.

For each dataset, three subproduct will be created:

**Raw data:** unprocessed data with full personal information (none-anonymized data)

**Anonymized data:** raw data in which any personal identifiers are removed

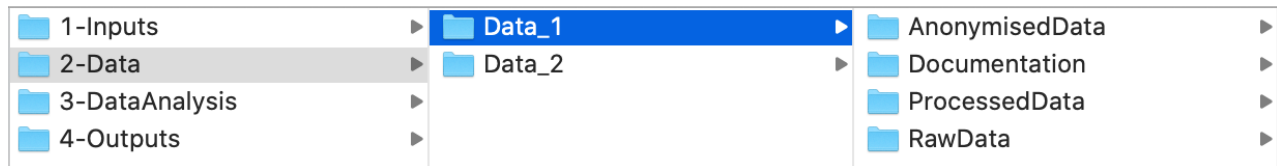
---

<sup>6</sup> <https://amnesia.openaire.eu/>



**Processed data:** e.g. recodes, transformations, selections, or enrichment of the anonymized data, which are all captured and described in metadata files

For each subproduct, a separated folder will be created, generated by a single specific data folder. The dataset related documentation as described in paragraph 2.1 will be stored in documentation folder under the related dataset folder.



Each dataset used in the ISEProD project will be listed in the DMP according to the following data description (general metadata) template. The metadata will be stored in a metadata.txt in each folder of a separate dataset. Metadata.txt file is updated any time it is needed and at least when the dataset is deposited in the designated data repository. The structure of the metadata.txt is the following:

Field	Description
<b>Dataset Title</b>	Name of the dataset. Dataset title must always begin with the task code it is associated with and end with the relevant case study and keyflow. e.g Dataset title should reflect its origin (RAW/ANONYM/AGGR)
<b>Creator</b>	Might be a person, a group of people or an institution
<b>Description</b>	A short description of the dataset (should include information on how the data was generated, what parts does it consist of, quickly introduce software or experiment setup)
<b>Keywords</b>	Keywords that will help indexing, tagging and finding the data once it is made publicly available.
<b>Data Purpose</b>	The purpose of the data collection/generation and its relation to the objectives of the project. The concrete task, milestone or deliverable should also be mentioned if relevant.
<b>Language</b>	Use the <a href="#">ISO 639 code</a> standard for representing names of languages
<b>Data Collection Start Date</b>	Indicates the Start of Data Collection
<b>Data Collection Last Update</b>	Indicates the End of Data Collection/latest update
<b>Data Type</b>	Identifies the data type referring to the categories identified in paragraph 2.1 of ISEProD DMP
<b>Data formats</b>	Text / numbers / images / 3D models / code / audio files / video files / reports / surveys / maps / scientific articles .txt; .csv; .shp; .wkt; .wmv; .zolca; .pdf; .stadat; .sav,...
<b>Data Origin</b>	Primary or secondary, raw data/anonymized/aggregated/processed....
<b>Data Source</b>	If data has been generated or reused, the original source(s) of the dataset must be indicated. That can be an email correspondence, paper, report, public dataset, etc. ideally, a DOI, PID or URL can be added to the original source.
<b>Software</b>	If viewing or editing the data requires a specific software, it must be indicated which. If multiple are available, one is sufficient with preference to the Open Source or Free ones.

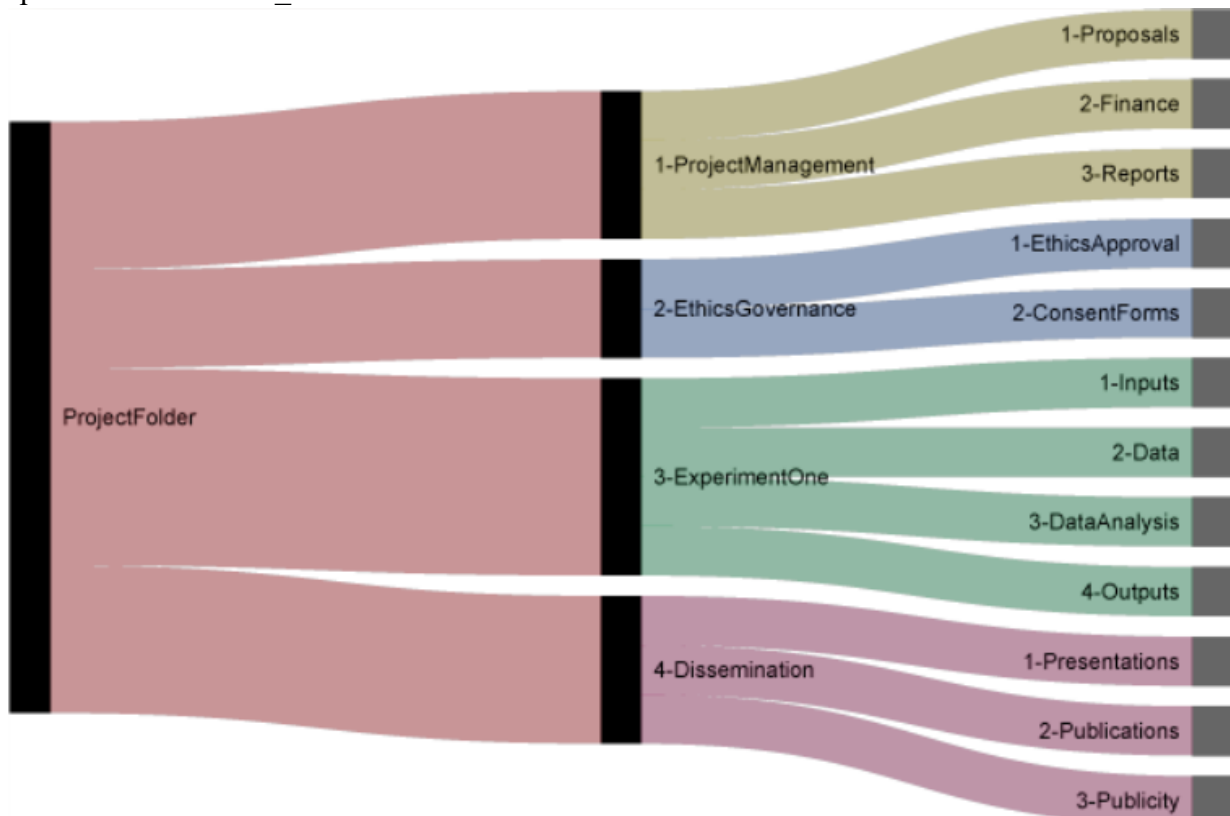
<b>Data Size</b>	Estimated (known) data size
<b>Stability</b>	Fixed (never change after being collected, generated) / Growing (new data may be added, but old data is never changed, deleted) / Revisable (new data may be added, old data may be changed, deleted)
<b>Confidentiality</b>	Public / Internal /Restricted (indicate restrictions, e.g. embargo period) / Confidential
<b>Archiving</b>	Long term or project timeframe (Data should be preserved for long-term archiving / It is sufficient to keep the data only until the end of the project.)

The metadata text file will be deposited together with the dataset in the Luiss institutional repository as separated attached files. The metadata.txt should be granted open access permission even if the access to the dataset file(s) remains embargoed/closed/restricted.

When depositing the dataset, the PI will also decide which related documentation needs to be deposited in order to adhere to the FAIR principles, in order to allow for a better data reuse and understanding.

#### 4.1.1 Project Folder Structure

The ISEProD folder structure will follow guidelines defined by Nikola Vukovic in [http://nikola.me/folder\\_structure.html](http://nikola.me/folder_structure.html)



#### 4.1.2 File naming convention

The strategy used in ISEProD project for file naming follows the TILS Document Naming Convention ([https://www.data.cam.ac.uk/files/gdl\\_tilsdocnaming\\_v1\\_20090612.pdf](https://www.data.cam.ac.uk/files/gdl_tilsdocnaming_v1_20090612.pdf))

## 5 Data Access

ISEProD envisages four levels of confidentiality for the project data:

- Public - the data can be made public at any time during or after the project;

- Internal - the data is shared only between ISEProD team members, cannot be made public during or after the project;
- Restricted - access is restricted due to a period of embargo or if only certain members of the ISEProD team can access it;
- Confidential - the data is only accessible by a certain group of people defined by the PI

The level of confidentiality is assigned by the PI and are indicated it in the corresponding metadata field.

During the project public data will be made accessible through the designated repository.

## 6 Data Sharing and Reuse

Data interoperability and reusability is ensured by applying standardized metadata, as well as internationally accepted preferred formats for the deposited data files. These elements are also described in the tables provided in paragraph 2. Non-proprietary or open standard file formats will be used to ensure accessibility and reuse:

- Spreadsheets: Comma Separated Values (.csv) or Tab Separated Values (.tsv)
- Text: plain text (.txt), or if formatting is needed, PDF/A (.pdf)
- Presentations: PDF/A (.pdf)
- Images: TIFF (.tif, .tiff), or PNG (.png)
- Videos: MPEG-4 (.mp4)
- Maps: ESRI shapefiles (.shp, .shx, .dbf)

Dataset that have been tagged as Public will be made available on the designated data repository immediately after upload. If certain restrictions apply, the datasets will be released following the restrictions (including embargo or restricted access rules). In principle all data will be reusable by third parties unless special restrictions apply which will be indicated in the accompanying metadata.txt file and or in the related documentation.

## 7 Data Preservation and Archiving

A good data management strategy includes the selection of material intended for long term preservation. ISEProD PI will carefully select what needs to be preserved and what can be disposed. The general strategy for long term data archiving in ISEProD project is the following:

### **Data needs to be preserved if:**

- The original data is collected or generated within ISEProD activities (primary data);
- The data is reused from other data sources that do not have a persistent identifier and run a risk to become inaccessible in the future.

If the data is generated using other data sources, that have been reused within ISEProD, then the decision about archiving needs to be based on the ease of replicating the dataset: if it is possible to completely replicate exactly the same dataset, then only the metadata and data generation process (e.g. in a form of code) is sufficient to preserve. Preservation has to be compliant with the license assigned to the secondary dataset.

If the data is reused from other data sources, that have a persistent identifier, then it is sufficient to preserve the metadata of such datasets.

## Ethical issues

The PI already submitted the ETHICS REPORT to the ERC. The WP2: Ethics Requirements is the 'ethics requirements' that the project must comply with will be covered by the deliverables listed below:

### HUMANS

H - Requirement No. 2-

Copies of ethics approvals for the research with humans.

### PROTECTION OF PERSONAL DATA

D2.2 POPD - Requirement No. 9

Detailed information on the informed consent procedures in regard to data processing

D2.3: POPD - Requirement No. 10

Templates of the informed consent forms and information sheets (in language and terms intelligible to the participants);

D2.4: POPD - Requirement No. 13

The applicant must evaluate the ethics risks related to the data processing activities of the project. This includes also an opinion by the DPO if data protection impact assessment should be conducted under art.35 General Data Protection Regulation 2016/679. If not, the DPO must provide a justification. The risk evaluation and the opinion must be provided to the ERC Executive Agency.

D2.5: POPD - Requirement No. 14

A copy of the Data Processing Agreement with the two data collection companies must be provided.

The due date for the delivery of the deliverables above is fixed on Month 12 – December 2020

Luiss Privacy Policy is published on Luiss Website. The information published at <https://www.luiss.edu/privacy-policy> describes how Luiss treats data gathered through the use of [www.luiss.edu](http://www.luiss.edu) and [www.luiss.it](http://www.luiss.it).

Luiss University Regulation on Personal Data Protection was issued by D.R. Nr. 203 of October 27th, 2015 and Luiss University Regulation on the Luiss University Guido Carli Committee for Research was updated by the D.R. N. 76 of May, 12th, 2020. This last Committee is in charge of the evaluation and approval of the ethical issues in all research proposals upon individual request of the researchers at Luiss University.

## 8 Revision of the Data Management Plan

ISEProD DMP will be revised on yearly basis to evaluate its efficiency and efficacy. The DMP will also be updated: at least in conjunction with each project revisions and audit or any time the PI will consider it necessary (for example when a new dataset will be created/collected, whenever a dataset is deposited/updated, or it changes its access right. The evaluation will be guided by the following questions:

- Do data sufficiently preserve the anonymity?
- Is the anonymization occurred correctly and issued with unique identifier?
- Is the confidential and sensitive data being stored safely in the database?
- Is file identification occurred correctly and meaningful?
- Is the file identifier understandable for outsiders?
- Is the plan still consistent with what is being done in ISEProD description of action?
- Is the instantiated database consistent with the specifications in this document?

In any case, it must be mentioned that any anonymized / pseudonymised data that cannot be cross-linked and contains no further identifiable variables will be regarded as free to be used as 'regular' data.