



## D1.2 Data Management Plan

Lead Partner:	UNIVIE – University of Vienna
Version:	<b>1.0</b>
Status:	<b>FINAL</b>
Dissemination Level:	<b>PU</b>
Document Link:	

### Deliverable Abstract

**This document describes the project's Data Management Plan.**



## COPYRIGHT NOTICE



This work by Parties of the EOSC-Pillar is licensed under a [Creative Commons Attribution 4.0 International Licence](https://creativecommons.org/licenses/by/4.0/) (CC BY 4.0). The EOSC-Pillar project is co-funded by the European Union Horizon 2020 programme under grant number 857650. It is attributed to Paolo Budroni, Lisa Hönegger and Anita Bodlos and published at Phaidra, Handle: [11353/10.1045211](https://phaidra.univie.ac.at/o:1045211) under <https://phaidra.univie.ac.at/o:1045211>. This version is based on a former version of a DMP Template attributed to Heider, Veronika, Raffetseder, Lena, Sanchez Solis, Barbara & Ulrich, Xenia. (2018). DMP Template for the Social Sciences (Version 1.0). available at Zenodo <https://zenodo.org/record/1291816#.Xba0rppKguU> and based on the Template for Data Management Plans (DMP) at Austrian Research Institutions, Project e-Infrastructures Austria by Blumesberger, Susanne et al. (2016). Cluster C: Template for Data Management Plans (DMP) at Austrian Research Institutions (Word Doc) – English. Handle: [11353/10.459216](https://phaidra.univie.ac.at/o:459216) under <https://phaidra.univie.ac.at/o:459216>.

Version	Effective Date	Description of Document/ Changes
1.0	28/11/2019	First version of DMP – M5 of project

## Contents

1	Introduction .....	6
1.1	This DMP in a nutshell .....	6
1.2	FAIR data principles for an internet of FAIR data and services.....	6
1.3	Notice on the handling of this DMP .....	6
2	Data characteristics .....	7
2.1	Data collection from WPs.....	7
3	Documentation and metadata .....	9
3.1	Metadata standards .....	9
3.2	Documentation .....	9
3.3	Data handling and data quality control .....	10
4	Data availability and storage .....	11
4.1	Data release, data sharing and data access strategy.....	11
4.2	Data storage strategy .....	11
4.3	Data preservation strategy .....	12
4.4	Data interoperability strategy .....	12
5	Legal and Ethical Aspects.....	13
5.1	Legal and Ethical Aspects .....	13

## Administrative Data

Version of DMP, date	1.0, 28 <sup>th</sup> of November, 2019
Project coordinator	Federico Ruggieri, GARR, <a href="mailto:Federico.Ruggieri@garr.it">Federico.Ruggieri@garr.it</a>
Work package leaders	WP1 Fulvio Galeazzi, GARR, <a href="mailto:fulvio.galeazzi@garr.it">fulvio.galeazzi@garr.it</a> WP2 Rob Carrillo, Trust-IT SRL, <a href="mailto:r.carrillo@trust-it-services.com">r.carrillo@trust-it-services.com</a> WP3 Lisa Hönegger, UNIVIE, <a href="mailto:lisa.hoenegger@univie.ac.at">lisa.hoenegger@univie.ac.at</a> WP4 Federica Tanlongo, GARR, <a href="mailto:federica.tanlongo@garr.it">federica.tanlongo@garr.it</a> WP5 Adham Hashibon, Fraunhofer, <a href="mailto:adham.hashibon@iw.fraunhofer.de">adham.hashibon@iw.fraunhofer.de</a> WP6 Frédéric Huynh, IRD, <a href="mailto:frederic.huynh@ird.fr">frederic.huynh@ird.fr</a> WP7 Volker Beckmann, CNRS, <a href="mailto:beckmann@in2p3.fr">beckmann@in2p3.fr</a>
Authors of DMP	Paolo Budroni, University of Vienna, <a href="mailto:paolo.budroni@univie.ac.at">paolo.budroni@univie.ac.at</a> ; Lisa Hönegger, University of Vienna, <a href="mailto:lisa.hoenegger@univie.ac.at">lisa.hoenegger@univie.ac.at</a> ; Anita Bodlos, University of Vienna, <a href="mailto:anita.bodlos@univie.ac.at">anita.bodlos@univie.ac.at</a>
	Acknowledgement: We are grateful to the input provided to this DMP by project partners (work package leaders and reviewers, Lisana Berberi and Lars Kaczmirek).
Data officer and responsible for DMP	Lisa Hönegger, University of Vienna, <a href="mailto:lisa.hoenegger@univie.ac.at">lisa.hoenegger@univie.ac.at</a> Paolo Budroni, University of Vienna, <a href="mailto:paolo.budroni@univie.ac.at">paolo.budroni@univie.ac.at</a>
Project title (Acronym)	EOSC-Pillar; Coordination and Harmonisation of National & Thematic Initiatives to support EOSC
Start and end date of project	Start on 1 <sup>st</sup> July 2019; End on 30 <sup>th</sup> June 2022
Grant number	857650 - H2020

## Executive summary

This DMP can offer guidance to data producers and data stewards involved in the project EOSC-Pillar for all relevant questions concerning data (e.g. the content of data, metadata, software applications) generated during several project activities. This DMP describes the comprehensive lifecycle of data, from collection to storage, preservation, distribution and re-use scenarios.

The guiding principle of data gathering is to gain insights into the environment of the European Open Science Cloud (EOSC) and to support its implementation. Consequently, data collected by EOSC-Pillar may be of special interest to:

1. the EOSC Governance Board
2. the EOSC Executive Board
3. the Working Groups of the Executive Board of EOSC, especially “WG Landscape” and “WG Sustainability”
4. other EOSC supporting projects (“5b projects”: EOSC-Synergy; EOSC-Nordic; NI4OS-Europe, ExPaNDs)
5. other activities and initiatives currently involved in the development of EOSC. Potential addressees are:
  - a) The European Commission
  - b) e-IRG
  - c) ESFRI
  - d) Science Europe
  - e) All European Universities Networks involved in the co-creation processes
  - f) LIBER
  - g) The “Cluster Projects”: ENVRI-FAIR, EOSC Life, ESCAPE, PaNOSC, and SSHOC
  - h) The research community
  - i) Further actors interested in the development of EOSC



# 1 Introduction

## 1.1 This DMP in a nutshell

This Data Management Plan (DMP) is a structured guideline that describes the comprehensive lifecycle of data, from gathering to storage, preservation, distribution and re-use scenarios. Therefore, this DMP can offer guidance to data producers and data stewards involved in the project EOSC-Pillar for all relevant questions concerning data (e.g. the content of data, related metadata or software applications) generated during various project activities. In addition, formulating and following the DMP paves the way for long-term accessibility and subsequent reusability of the digital assets. Following the convention, this DMP is a living document and therefore versioned by definition.

In more detail, this DMP provides a description of what (kind of) data will be collected along the entire lifecycle of the project. Furthermore, we describe how the data will be processed both during the project and after its completion. This description includes statements about the provenance of data, contextual statements, infrastructures used to store and manage data, as well as information regarding the publication, citation, long-term accessibility and, if necessary, deletion of data during or after the research lifecycle. If personal data is processed, we refer to documents handling legal and ethical aspects, including statements on data protection, terms of use, copyright attribution and exploitation rights for further reuse, and licensing.

## 1.2 FAIR data principles for an internet of FAIR data and services

This DMP also contributes to enhancing compliance of data with the FAIR principles in order to improve the findability, accessibility, interoperability, and reusability of the data collected in this project. Therefore, data, related metadata, and contextual tools like software applications and source codes are stored and made available for reuse in suitable repository or archiving systems. Whenever sensible, data are assigned globally unique and persistent identifiers to allow their findability both from humans and computer systems. Data and related services are thus identifiable, accessible, traceable, interoperable, and whenever possible, available and licensed for subsequent re-use (in compliance with intellectual property rights, data protection regulations or other applicable laws and regulations).

## 1.3 Notice on the handling of this DMP

We consider this DMP as an instrument designed to help structure and plan data processing in the EOSC-Pillar project as well as to define related responsibilities. Since it is a versioned document, any version can vary in length and detail following the developments of the (research) focus in the EOSC-Pillar project as well as experiences based on working with different kind of data and different lifecycle-stages. Thus, not all questions addressed in this document may be included in the final version of the DMP. Rather, readers can regard the DMP as a dynamic document, which can and will be updated along the whole lifecycle of the project. In order to keep track of different versions, the version number of each DMP is always included in the administrative section above.

## 2 Data characteristics

The project EOSC-Pillar produces various kinds of data in different stages of the project. All work packages of the project were involved in the first draft of the DMP. Depending on the content and timelines of the individual WPs, information on data collection and management differs in detail at this stage of the project. Information on data processing within different work packages stems from the WP leaders who are also responsible for updates regarding information on data processing (collection, archiving, accessibility, etc.) related to activities within their WPs.

### 2.1 Data collection from WPs

The work packages, their activities relevant to this DMP, the (open source) software used for these endeavors and the related file formats are described in the table below. This table also contains information on the accessibility and reuse options of the described data.

Work Packages (WP) and their activities	(Open source) software	Formats	Accessibility	Timeframe for reuse
<i>WP1 – Management.</i> Two categories of documents are produced: (a) documents for short and midterm preservation and (b) documents for long-term preservation purposes.	LibreOffice, OnlyOffice or Microsoft Office	Different file formats for internal documentation; audio/video documents for midterm-preservation purposes will be produced during virtual or in-person meetings: these will be stored using the MP4 or JPEG/PNG formats. Documents for long-term storage will be in PDF/A format	Project internal/external access. Data for short and midterm storage will be accessible via the project storage system (“EOSC-Pillar repository”), whereas data for long-term accessibility will be stored and available via the repository Phaidra at the University of Vienna.	Short and midterm/long-term
<i>WP2 – The human factor of the EOSC: Dissemination, Outreach and Community building.</i> Data will be collected through various research activities that are still to be determined, mostly based on the outcomes of the data generated in WP3.	STATA, SPSS, MAXQDA, Microsoft Office	.dta (for STATA), .sav (for SPSS), and .tab (for all software used to open delimited files); .pdf/A for reports and documentation	In case there will be further social science datasets collected, they will be available for reuse at the repository AUSSDA (Austrian Social Science Data Archive, University of Vienna, under a doi).	Long-term
<i>WP3 – The “National Initiatives” survey</i> will be available as a dataset. Furthermore, we will publish documentation (accompanying the survey data) and project reports.	LimeSurvey <a href="http://www.limesurvey.org">www.limesurvey.org</a> , STATA, SPSS, Microsoft Office, Notepad++, Latex	.dta (for STATA), .sav (for SPSS), .tab (for all software used to open delimited files), .pdf/A for reports and documentation, and XML for documentation of the questionnaire programming code used in Limesurvey, Textfiles	The resulting dataset and additional documentation will be available for reuse at the repository AUSSDA (Austrian Social Science Data Archive, University of Vienna doi:10.11587/VOSVGK).	Long-term
<i>WP4 – From National Initiatives to Transnational Services</i> consists of documents and reports, technical requirements, data from interviews/consultations with national initiatives and software (the national service registry prototype).	LibreOffice, OnlyOffice or Microsoft Office  gCube (different releases of an open source software)	PDF, PDF/A, XML  Java Code	Tbd  GitHub	Short and midterm/long-term
<i>WP6 – EOSC in Action:</i> use cases and community-driven pilots	Different tools (tbd)	PDF, PDF/A, Textfiles and other formats (tbd)	On-demand	Short and midterm/long-term



## 3 Documentation and metadata

### 3.1 Metadata standards

In order to reach compliance with the FAIR principles, all WPs will carefully draft metadata for publications. At the moment, we plan to use the services of three repositories for publication and storage: AUSSDA and Phaidra (both hosted at the University of Vienna), Zenodo (hosted at CERN in Geneva) and GitHub. These repositories and their metadata standards are described in detail below:

- **AUSSDA** (The Austrian Social Science Data Archive) is based at the University of Vienna and available under <https://data.aussda.at/>. AUSSDA currently supports the export formats DDI, Dublin Core and JSON. Bibliographic entries are available in Endnote XML, RIS, BibTeX formats.
- **Phaidra** (Permanent Hosting, Access, Indexing of Digital Resources and Assets) is a long-term preservation digital archive of the University of Vienna and available under <https://phaidra.univie.ac.at>. Metadata supported by Phaidra are: UW-Metadata, Dublin Core, mods, EXIF-Viewer metadata. Phaidra is able to deliver stored metadata in the format XML.
- **ZENODO** is located at CERN in Geneva, Switzerland and is available under <https://zenodo.org/>. Metadata supported by ZENODO are: BibTeX, CSL, DataCite, Dublin Core, DCAT, JSON, JSON-LD, GeoJSON, MARCXML and Mendeley.
- **GitHub** provides hosting for software development and is available under <https://github.com>. Metadata can be provided as json and DCAT.

### 3.2 Documentation

**WP3:** Data gathered by the “National Initiatives” survey will be published in the repository AUSSDA along with several files for documentation:

- a methods report/codebook, which will outline the data collection and data processing activities. This report will also include a list of targets that were invited to participate in the survey and the non-disclosure agreement for data analysts.
- The questionnaire. This document will also contain the information sheet for legal and ethical aspects for the survey respondents (informed consent) and programmer information to facilitate reuse.
- the code that was used to conduct the survey with the open source software LimeSurvey (in XML format).

All documents are part of a landing page with the same DOI. The terms of use of the repository and specifically for the scientific reuse of the data are available at the following URL:

[https://aussda.at/fileadmin/user\\_upload/p\\_aussda/Documents/AUSSDA\\_Terms\\_of\\_Service\\_en.pdf](https://aussda.at/fileadmin/user_upload/p_aussda/Documents/AUSSDA_Terms_of_Service_en.pdf)

Further documentation and metadata are described in section 2 “Data characteristics” and section 4.2 “Data storage strategy”.



### 3.3 Data handling and data quality control

**WP3:** WP3 will make use of the following services offered by AUSSDA: Data quality checks (e.g. checks of consistency of labels, logical errors in the data), data curation, and version control. Additionally, the repository will perform all necessary transformation of data in relevant formats to ensure the FAIRness and long-term availability of the data and documents. Phaidra has adopted a policy for storage (<https://datamanagement.univie.ac.at/en/about-phaidra/policy-of-phaidra/>) and complies with long-term preservation standards.

We do not intend to delete or block access to data after publication. In case of any updates or changes, these are traceable by means of version control. According to the repositories' policies, it is not foreseen to delete data or metadata that has been published.

**WP6:** To assure data quality, a process (i.e., data flow and the involved communities/groups) is planned to be described. A data quality (DA) dimension will be addressed which consists of accuracy, completeness and consistency or other measures/indicators. Some of the data quality issues can be "missing data", "incorrect data", "irrelevant data" etc. Each use-case community will identify their relevant and specific quantifiable data quality criteria.

## 4 Data availability and storage

### 4.1 Data release, data sharing and data access strategy

Data collected by EOSC-Pillar will be published in the following repositories:

- AUSSDA: <https://www.aussda.at>, with a Creative Commons or Scientific Use licence, with assignment of DOIs.
- Phaidra: <https://phaidra.univie.ac.at/>, with a Creative Commons licence, with assignment of handles.
- Zenodo: <https://zenodo.org/>, with a Creative Commons licence, with assignment of DOIs.

The WP3 “National Initiatives” survey data will be stored and published for scientific reuse in the following repository:

AUSSDA – The Austrian Social Science Data Archive ([www.aussda.at](http://www.aussda.at)) with doi:10.11587/VOSVGK

Related documentation will be licensed under CC BY, metadata will be published as open data under a CC 0 dedication.

### 4.2 Data storage strategy

**WP1:** Concerning midterm preservation purposes: Documents for external use and publication will be made accessible through the project's website and stored in the related file-sharing platform or other open-access platforms (like Zenodo) and/or the Commission's website when applicable. Documents for internal use will be stored in the project's file-sharing platform or GARR's internal platform (based on the software OnlyOffice). The latter option will be applicable, for example, for documents containing administrative or financial information whose accessibility needs to be restricted (to project partners).

**WP2:** The data collected for activities related to communication, outreach, dissemination and stakeholder engagement will consist of contact information including first name, last name, email, organization, stakeholder type and country and will mostly be collected through EOSC-Pillar's website webforms, and webinar and event registrations tools, such as GotoWebinar, Zoom Meetings, and Eventbrite. They are stored first in these data gathering tools. At least monthly, data from these multiple sources are collected, analyzed and stored in the “EOSC-Pillar Repository” under the following URL: <https://repository.eosc-pillar.eu/> This storage solution is a specific file storage system of the project and will be maintained until 2 years after the end of the project, as per H2020 requirements.

**WP3:** During survey data collection, data is stored with LimeSurvey in Germany, according to a GDPR compliant data processing contract between the University of Vienna and LimeSurvey GmbH. After the collection process is finished and during the curation process, data storage, data security and recovery strategies are ensured by the central storage provided by the Computer Centre of the University of Vienna. The Computer Centre of the University of Vienna guarantees automatic backups of the stored data following their standard (virtual and

physical) security procedures. The costs of these factors are covered by the Vienna University Library.

### 4.3 Data preservation strategy

**WP2:** WP2 collects personal data such as contact information including first name, last name, email, organisation, stakeholder type and country through EOSC-Pillar's website webforms, and webinar and event registrations through tools such as GotoWebinar, Zoom Meetings, and Eventbrite. They are stored first in these data gathering tools. At least monthly, data from these multiple sources are collected, analysed and stored in the EOSC-Pillar repository.

**WP3:** WP3 will use the preservation services offered by Phaidra and AUSSDA. Both repositories have standards in place to allow for the long-term preservation of stored data. The repositories do not delete data after the assignment of persistent identifiers (DOIs, handle). Access to data is secured (according to the set access conditions). In certain exceptions, it is possible that the repository blocks access to data (e.g. if a data protection issue has been identified after a data set has been published), however, metadata will remain accessible.

### 4.4 Data interoperability strategy

**WP5 and WP6:** After data is stored in the EOSC-Pillar repository and available for access, it should also be interoperable with other platforms among different researchers, institutions, countries etc.

To allow inter-disciplinary interoperability a specific ontology will be developed and maintained as proposed in WP5.

Datasets collected/generated from use-cases in WP6 will be in compliance with FAIR principles.

Here, a short list of some common tools and mechanisms for FAIR research data providers and consumers will be defined from WP5 deliverables.

Different data formats and schema mapping tools (such as X3ML) will be defined and used to make data use feasible not only by state, but also transnationally.

## 5 Legal and Ethical Aspects

### 5.1 Legal and Ethical Aspects

**WP3:** WP3 has ensured compliance with legal regulations, as well as scientific standards and ethical guidelines during all stages of the “National Initiatives” survey.

To begin with, data collection is based on an information sheet/consent form for respondents including all relevant information on data processing. Only respondents who agreed to the informed consent were allowed to proceed with the survey. Also, the identity of participants to the survey is protected and no negative effects of the participation in the survey is foreseen. Therefore, only project staff who sign a non-disclosure agreement will be granted access to the raw data.

Before publication, the data will be pseudonymized to protect respondents’ privacy according to the GDPR and Austrian national law. In addition, WP3 will choose access conditions offered by AUSSDA for sensitive data: The data set will be available for scientific use, following AUSSDA’s principle to set access conditions to data “as open as possible, but as restricted as necessary”. Data accompanying the survey data (e.g. documents, software, etc.) will be published under licences appropriate for the type of publications and potential necessary restrictions (e.g. tailored, CC and further free licences of use).

All legal aspects related to the storage of data including virtual and physical security procedures are under the authority of AUSSDA and the Computer Center of the University of Vienna.

The party responsible for processing data of the “National Initiatives” survey is the University of Vienna.