



D1.5

2nd Updated Data Management Plan

Work Package	1
Lead partner	UHEL
Status	Final
Deliverable type	Report
Dissemination level	Public
Due date	31-01-2021
Submission date	01-03-2021

Deliverable abstract

The Data Management Plan (DMP) contains information on the kind of data generated in the project, suggested policies, standards, and sustainability activities related to the created data services, including software solutions for the services. This is done in close collaboration with the work packages on FAIR Policies and common technical solutions, to ensure the management of knowledge in the project is maintained following the FAIR principles.

The DMP covers data, documents and solutions created within the project (i.e. not the data provided by the RIs in their services, which is part of their service policies). It describes the data (including metadata) and software standards, availability, curation and preservation methods, and is based on the Digital Curation Centre model for DMPs.

The first update of the DMP details the tools and systems which are used to maintain personal data and digital assets generated in the project. The second update includes additional data types and corresponding tools for storage.

DELIVERY SLIP

	Name	Partner Organization	Date
Main Author	Ari Asmi	UHEL	
Contributing Authors	Ulrich Bundke Katrin Seemeyer Alex Vermeulen Zhiming Zhao	FZJ FZJ ULUND UvA	
Reviewer(s)	Andreas Petzold	FZJ	
Approver	Andreas Petzold	FZJ	01 March 2021

DELIVERY LOG

Issue	Date	Comment	Author
V 1.0	26 Feb 2021	Final version for submission	Ari Asmi

DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the Project Manager at manager@envri-fair.eu.

GLOSSARY

A relevant project glossary is included in Appendix A. The latest version of the master list of the glossary is available at <https://doi.org/10.5281/zenodo.4471374>.

PROJECT SUMMARY

ENVRI-FAIR is the connection of the ESFRI Cluster of Environmental Research Infrastructures (ENVRI) to the European Open Science Cloud (EOSC). Participating research infrastructures (RI) of the environmental domain cover the subdomains Atmosphere, Marine, Solid Earth and Biodiversity / Terrestrial Ecosystems and thus the Earth system in its full complexity.

The overarching goal is that at the end of the proposed project, all participating RIs have built a set of FAIR data services which enhances the efficiency and productivity of researchers, supports innovation, enables data- and knowledge-based decisions and connects the ENVRI Cluster to the EOSC.

This goal is reached by: (1) well defined community policies and standards on all steps of the data life cycle, aligned with the wider European policies, as well as with international developments; (2) each participating RI will have sustainable, transparent and auditable data services, for each step of data life cycle, compliant to the FAIR principles. (3) the focus of the proposed work is put on the implementation of prototypes for testing pre-production services at each RI; the catalogue of prepared services is defined for each RI independently, depending on the maturity of the involved RIs; (4) the complete set of thematic data services and tools provided by the ENVRI cluster is exposed under the EOSC catalogue of services.

TABLE OF CONTENTS

D1.5 - 2nd Updated Data Management Plan	4
1 Preamble	4
1.1 Rationale	4
1.2 Data Management Team	4
2 Data Management Plan – 2 nd Update	5
2.1 Data Type Summary.....	5
2.2 Data Management Platforms and Tools	5
2.2.1 Redmine.....	5
2.2.2 Zenodo.....	6
2.2.3 GitLab/GitHub.....	6
2.2.4 MediaWiki.....	7
2.2.5 Open semantic search	7
2.2.6 OntoWiki	7
2.3 Management Plan by Data Type	8
2.3.1 Personal contact data, email lists	8
2.3.2 Project documents.....	8
2.3.3 Surveys, questionnaires	8
2.3.4 Further digital objects created via the demonstrators defined in the work packages	8
2.3.5 Geophysical data sets collected via the RIs during the project	9
2.3.6 Software.....	9
2.3.7 Documentation.....	9
2.3.8 Websites.....	9
2.4 Ethical Aspects.....	12
2.5 Other.....	12
3 Appendix.....	13
3.1 List of Abbreviations and Acronyms	13

D1.5 - 2nd Updated Data Management Plan

1 Preamble

1.1 Rationale

The Data Management Plan (DMP) contains information on the kind of data generated in the project, the suggested policies, standards, and sustainability activities related to the created data services, including software solutions for the services. This is done in close collaboration with the work packages on FAIR Policies and common technical solutions, to ensure the management of knowledge in the project is maintained following the FAIR principles.

The DMP covers data, documents and solutions created within the project and thus excludes the data provided by the RIs in their services, that is dealt with through the individual RI service policies). This DMP describes data (including metadata) and software standards, availability, curation and preservation methods, and is based on the Digital Curation Centre model for DMPs accessible at <https://dmponline.dcc.ac.uk/>.

The initial DMP ([ENVRI-FAIR Deliverable D1.3](#)) focused on the tools and systems which are used to maintain personal data of involved project participants. The first updated DMP ([ENVRI-FAIR Deliverable D1.4](#)) and further subsequent annual updates of the DMP include the description of data and metadata, software standards, availability, curation and preservation methods used for the data, documents and solutions created within the project.

1.2 Data Management Team

The Data Management Team (DMT) is responsible for providing ENVRI-FAIR with a Data Management Plans (DMP) during the project period and oversees the DMP revisions.

The DMT consists of the Co-coordinator (chair), Project Management Team (PMT) representative, responsible of drafting the DMP text, and WP 5 and 7 leaders. The General Assembly (GA) confirms the membership of the DMT Team during annual GA meeting.

DMT Members confirmed by the GA:

Ari Asmi (co-coordinator)	ari.asmi@helsinki.fi
Katrin Seemeyer (PMT representative)	k.seemeyer@fz-juelich.de
Alex Vermeulen (WP 5 lead)	alex.vermeulen@icos-ri.eu
Zhiming Zhao (WP 7 lead)	z.zhao@uva.nl
Ulrich Bundke (ENVRI-FAIR data manager)	u.bundke@fz-juelich.de

The first update of the DMP is produced as a deliverable of WP1 and subsequent annual updates will be also delivered via WP1.

2 Data Management Plan – 2nd Update

2.1 Data Type Summary

In the course of the project conduction, the following types of data assets will be created:

- Personal contact data, email lists
 - These are generally related to overall project management (i.e., information on project participants), but also include such information on the mailing lists related to other ENVRI community members, communication WP partners, etc. The form of these are typically databases, and spreadsheet documents
- Project documents, i.e., deliverables, milestone reports. official or internal project documents, typically documents as PDF files
- Survey, questionnaire result documents. Typically, spreadsheets of answers collected during the project runtime.
- Digital objects created via the demonstrators defined in the work packages. There are a variety of these depending on the type of demonstrator.
 - Geophysical data sets collected via the RIs during the project. These are additional datasets which are on top of the typically derived data products of the RI (i.e., normal RI products are not part of the data types covered by this DMP)
 - Software products e.g., ENVRI hub and related software implementations
 - ENVRI Knowledge Base
- Documentation of software, services and service use examples, in the ENVRI Knowledge Base or other repositories. These can be RDF triples, documents (PDF, etc), or example scripts, workflows, notebooks etc.
- Websites
 - Training website (includes training material and videos)
 - ENVRI website content
 - ENVRI Knowledge Base content
 - ENVRI wiki content and wiki structure.

2.2 Data Management Platforms and Tools

2.2.1 Redmine

In its initial phase the project data collection and management were exclusively related to personal data of project participants and of members of the ENVRI community. Participation in the project is granted by invitation or on personal request.

ENVRI-FAIR uses the open-source Redmine environment (www.redmine.org) for internal communication, storage of project-related documents and project management. The system is a password-protected and secured environment for project data storage with access control; see below for details.

2.2.1.1 Allocation of Resources (Redmine)

The Redmine environment (www.redmine.org) is a free and open source, web-based project management and issue tracking tool. It allows users to manage multiple projects and associated subprojects. It features per project wikis, forums, time tracking, and flexible, role-based access control. It includes a calendar and Gantt charts to aid visual representation of projects and their deadlines. Redmine includes a detailed version control system and a repository browser and a document comparison tool (diff viewer).

The implemented Document Management System Features (DMSF) include a directory structure, document versioning and revision history, email notifications for directories and/or documents, document locking, configurable document approval workflow, document access auditing and document content full text search.

Project participants use the internal ENVRI-FAIR Redmine environment for communication and sharing of documents. This ensures that communication and related documents are stored in the Redmine archive

and can be accessed throughout the lifetime of the project. Redmine includes version control of documents and keeps the previous versions so that continuing version control is in place. Notably, the control of each individual section is limited to only those project participants which need the access. Access control lies with the ENVRI-FAIR data manager (see below).

The ENVRI-FAIR Redmine system is hosted at Forschungszentrum Juelich GmbH and administrated by the ENVRI-FAIR data manager Ulrich Bundke, email u.bundke@fz-juelich.de. The ENVRI-FAIR data manager is responsible for the implementation of the DMP and related data management activities. The Data Management Team is responsible for the review and of the DMP and oversees the DMP revisions.

Hardware resources required for the delivery of the DMP are provided in-kind by Forschungszentrum Juelich GmbH. Staff resources are covered by project funding.

The ENVRI-FAIR Redmine system will be kept online minimum one year after the project. This period can be extended on request by the ENVRI community to ensure the continuation of the ENVRI-FAIR work beyond the project lifetime.

2.2.1.2 Data Security (Redmine) and Retention times

The ENVRI-FAIR Redmine environment is hosted at a local Linux cluster of the Institute of Energy and Climate Research 8 of Forschungszentrum Juelich GmbH which is coordinating the project. The Linux cluster is backed up daily by the IT Services of Forschungszentrum Juelich GmbH.

Retention times of data assets are planned as follows:

- Personal contact data of the project participants will be maintained and kept accessible in the ENVRI-FAIR Redmine environment during the lifetime of the project.
- Personal information contained in surveys is destroyed at the end of the lifetime of the project, preferably immediately after the conclusions of the study are finalised.
- Project internal data assets like internal reports will be maintained and kept accessible in the ENVRI-FAIR Redmine environment past the lifetime of the project; the exact retention time will be specified during the duration of the project.
- Openly accessible data assets will be kept accessible via the Zenodo Environmental Research Infrastructures Community ENVRI and/or via the ENVRI community webpage past the lifetime of the project.

2.2.2 Zenodo

Zenodo is used as the main system for storing documents for public re-use and sharing. ENVRI has created its own Zenodo Community page¹ which contains among others all public deliverables of the ENVRI-FAIR project. The Community is curated by the ENVRI project coordination team.

2.2.2.1 Allocation of Resources

CERN, OpenAire (and, by association, European Commission) support the Zenodo system and it is free of charge to use for the ENVRI-FAIR project.

2.2.2.2 Data Security

Access to ENVRI-FAIR products in Zenodo is free of charge and anonymous. The longevity of the system is guaranteed for at least 20 years by the Zenodo Longevity policy².

2.2.3 GitLab/GitHub

Public repositories like GitLab (<https://about.gitlab.com/>) or GitHub (<https://github.com/>) are used by the developers for controlling versions. Mature versions of the software or documents from the GitLab/GitHub are published in Zenodo.

¹ <https://zenodo.org/communities/envri/?page=1&size=20>

² <https://about.zenodo.org/policies/>

2.2.3.1 Allocation of Resources

The public GitLab/GitHub repositories are free of charge services offered by gitlab.com and github.com. The GitHub is hosted by github.com whereas Gitlab is hosted by the individual participating partners.

2.2.3.2 Data Security

The public GitLab/GitHub repositories are operated by the third party and under certain conditions services are offered free of charge. The third party operates the servers and takes care for central data security in addition to the local copy of the project stored on the working server of the developers. Final products are published and stored long time in Zenodo.

2.2.4 MediaWiki

MediaWiki is an open-source platform for publicly editable Wikipages, used for example by Wikipedia. This platform provides necessary editability and structure for the ENVRI community pages. The software is optimised to efficiently handle large projects, which can have terabytes of content and hundreds of thousands of hits per second. Because Wikipedia is one of the world's largest websites, achieving scalability through multiple layers of caching and database replication has been a major concern for developers. Another major aspect of MediaWiki is its internationalisation; its interface is available in more than 300 languages.

2.2.4.1 Allocation of Resources

The ENVRI Wikipages instance of MediaWiki is hosted on the same server as the homepage. MediaWiki is written in the PHP programming language and stores all text content into a MySQL database operated and maintained by the host. The Mediawiki will be available for the whole e project. The long-term location for this service is under consideration for the time after the project end.

2.2.4.2 Data Security

User login and editorial functionality is limited to selected project members. The public editorial system is deactivated. The MediaWiki Database is backed up regularly as part of the database server systems of the host.

2.2.5 Open semantic search

Open Semantic Search (opensemanticsearch.org) is an open-source framework for enabling online information semantic search. The open semantic search instance has been customised and further developed as the ENVRI Knowledge Base (ENVRI KB) search engine.

2.2.5.1 Allocation of Resources

The ENVRI Knowledge Base is currently hosted on a dedicated virtual server (search.envri.eu). The open semantic search will be available for the minimum of the project runtime. The long-term location beyond the project lifetime for this service is under consideration.

2.2.5.2 Data Security

The database of the search engine will be regularly backed up as part of fast recovery systems of the ENVRI KB.

2.2.6 OntoWiki

OntoWiki is an open-source framework for managing RDF content of the ENVRI Knowledge Base and provides a user interface for querying, search and updating the RDF content. The RDF content of the Knowledge Base includes the description of the research infrastructures using ENVRI reference ontology

(open information linking for ENVRI (OIL-e), semantic description of the software tools and services, and descriptions of the FAIRness assessment results.

2.2.6.1 Allocation of Resources

The OntoWiki instance is currently hosted on a dedicated server (ontowiki.envri.eu). OntoWiki will be available for the minimum of the project runtime. The long-term location beyond the project lifetime for this service is under consideration. OntoWiki provides the Sparql endpoint for public access (sparql.envri.eu).

2.2.6.2 Data Security

The database of the OntoWiki is backed up regularly via the version control system.

2.3 Management Plan by Data Type

This section provides short descriptions of the locations of each data type and the data management structures supported in the project. It is also supported by the table describing the FAIR aspects of each type of data.

2.3.1 Personal contact data, email lists

For the registration to Redmine users must provide their first name, last name, and an email address. Lists of participants to general project meetings and to work package meetings are collected on site and are kept as internal documents in the respective work package storages of the ENVRI-FAIR Redmine environment.

2.3.2 Project documents

Official project documents are prepared using the ENVRI-FAIR document preparation policy, which consider the deliverable format and the associated workflow including project review. After acceptance of the documents as project products, the prepared document is finalised by the Project Management team (i.e., final formatting etc.), and created into a PDF document. These documents are then submitted to the European Commission, and for public deliverables and reports, copies are also made available via the project webpages and Zenodo repository. Each document is minted a DOI number by the Zenodo repository and is made visible for metadata search via that platform.

2.3.3 Surveys, questionnaires

Each survey follows the ENVRI general policies for survey preparation, already prepared in the previous ENVRIplus project. This mechanism will provide the necessary ethical guidelines, among the GDPR regulations.

Survey results are initially stored by the survey creators, using different platforms which fulfil necessary privacy concerns. Results are then interpreted and used in the project, particularly in project deliverables. After usage of the survey results, they are discarded and erased by the initiator of the survey. There is no central repository for raw survey data.

2.3.4 Further digital objects created via the demonstrators defined in the work packages

Through the project demonstrators and test applications further data objects may be offered or generated and shared according to the FAIR principles. These data objects and their metadata will be curated by the project participants through their dedicated data portals and thus their data lifecycle is governed by the DMPs of the relevant Research Infrastructures.

2.3.5 Geophysical data sets collected via the RIs during the project

Geophysical data sets will be stored at the responsible research infrastructures with a reference link archived in the ENVRI-FAIR Redmine system, or in case of no RI able to host them (e.g., due to their special nature), open access is provided in the Zenodo Environmental Research Infrastructures Community ENVRI at <https://zenodo.org/communities/envri/Software>.

2.3.6 Software

Software developed for ENVRI-FAIR purposes includes software and use cases within the ENVRI- Hub concept, and common software solutions for FAIRness developed in the subdomains. During the project runtime, software source versions are stored in a central ENVRI Gitlab instance, hosted by the project management. The long-term preservation plan for these software codes is being developed, but one fallback option is to store fixed versions in Zenodo as a backup.

Runtime versions used in the ENVRI-Hub demonstrator will depend on the long-term preservation of these services.

2.3.7 Documentation

Documentations, particularly connected to the ENVRI-Hub, are a crucial part of the project initiatives. They are developed as part of the ENVRI Knowledge Base, in the use cases of the subdomains, as a side product of the technical and policy development. In some cases, the information is also contained within the official documentation above (see 2.3.2).

Overall storage of some of this information is done in the Knowledge Base itself, but the central location for most of the documentation is intended to be part of the websites (particularly the ENVRI Wiki, see below). Overall, the ENVRI website also collects links (and copies) to all (public) deliverables of all ENVRI projects.

2.3.8 Websites

The ENVRI websites have been kept running with additional institutional funding for those periods where there were no cluster projects. However, the sustainability of the whole ENVRI website content (inc. upkeep of the Wiki pages) is currently being considered.

Type of data	Tool	Findable	Accessible	Interoperable	Re-usable	Retention time
Personal contact data and mailing lists	Redmine	Will be store, no further information or other metadata is stored	Not openly accessible due to privacy protection	Not intended	n/a	Lifetime of the project
Project documents (deliverables, milestones)	Redmine, Zenodo	Will be maintained and kept accessible in the ENVRI-FAIR Redmine environment	Open-access project documents will be made accessible via the Zenodo Environmental Research Infrastructures Community ENVRI at https://zenodo.org/communities/envri/ and/or via the webpage https://envri.eu/home-envri-fair/ .	Not intended, however, e.g. the deliverable format is controlled via template, and the Zenodo requires some level of metadata (including author names).	Zenodo publication makes the documents re-usable for other purposes.	Lifetime of the project or for the documents in Zenodo about 20 years (see 2.2.2.2)
Surveys, questionnaires	(individual survey creators in work packages)	Results from internal surveys of target groups are stored in the respective work package storages in the protected ENVRI-FAIR Redmine environment.	Not intended, however the final results of the surveys (e.g. conclusions or non-personalized answers), can be embedded in project documentation (see 2.3.3). Questions of the surveys can be submitted to, e.g., Zenodo if not part of the general documentation.	Not intended. However, the overall results are treated as project documentation (see 2.3.3).	n/a	After usage of the survey results, they are discarded and erased
Digital objects, geophysical data sets	Redmine, Zenodo,	Project-related digital objects will be stored in the respective ENVRI-FAIR Redmine environment, and in case of open access in the Zenodo Environmental Research Infrastructures Community ENVRI at https://zenodo.org/communities/envri/ .	Digital objects, geophysical data sets and software considered for open access will be made available at the Zenodo Environmental Research Infrastructures Community ENVRI at https://zenodo.org/communities/envri/ or via GitLab, respectively (see also software)	The interoperability of digital objects, geophysical data sets and software will be specified in the next DMP update.	Digital objects, geophysical data sets and software intended for re-use will be made openly accessible via the repositories at Zenodo and GitLab, together with the respective metadata.	

Type of data	Tool	Findable	Accessible	Interoperable	Re-usable	Retention time
		Geophysical data sets will be stored at the responsible research infrastructures with a reference link archived in the ENVRI-FAIR Redmine system, or in case of no RI able to host them (e.g., due to their special nature), open access in the Zenodo Environmental Research Infrastructures Community ENVRI at https://zenodo.org/communities/envri/				
Software	GitLab, GitHub	Software developed in the course of the project will be stored and maintained at GitLab resources hosted by Forschungszentrum Juelich GmbH. The preparation of the GitLab repository is ongoing at the time of submission of the document.				
Documentation	ENVRI KB					
Websites, e.g. Training website, ENVRI website Wiki Pages	Wordpress,Media Wiki	Project related webpages and Wikipages of public interest will be stored on https://envri.eu . The project intern search engine https://search.envri.eu fosters the content and makes the content findable using free text keywords in combination with different categories.	Content is openly accessible		MediaWiki Pages makes the content reusable for other purpose	Long-time storage is under consideration

2.4 Ethical Aspects

Based on the self-assessment presented in the ENVRI-FAIR document [D12.1](#) “POPD Requirement No. 1”³, ENVRI-FAIR involves processing of personal data only on the level of collecting contact information of participants to meetings and workshops, and within a survey of target groups’ needs; participants will be provided in writing with details of what personal information will be processed and with other relevant information on processing of their personal data as required by General Data Protection Regulation (Regulation EU 2016/679).

Forschungszentrum Juelich GmbH ensures the required ENVRI-FAIR related file keeping of

- detailed information on the procedures for data collection, storage, protection, retention, destruction, and confirmation that they comply with national and EU legislation;
- detailed information on the informed consent procedures in regard to the collection, storage, and protection of personal data;
- templates of the informed consent forms and information sheets.

2.5 Other

Further details on the procedures for data management used in ENVRI-FAIR are described in

- ENVRI-FAIR [D12.1](#) POPD Requirement No. 1
- Declaration of Conformity with 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 the Free Movement of such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation); issued by FZ Juelich Data Protection Officer in July 2018 (Appendix to ENVRI-FAIR D12.1)
- ENVRI-FAIR Deliverable [D1.3](#) Initial Data Management Plan
- ENVRI-FAIR Deliverable [D1.4](#) 1st updated Data Management Plan

³ The self-assessment was prepared according to the Ethics Self-Assessment Guidance: Horizon 2020 Guidance — How to complete your ethics self-assessment: V6.1 – 04.02.2019

3 Appendix

3.1 List of Abbreviations and Acronyms

DMP	Data Management Plan
DMT	Data Management Team
DOI	Digital Object Identifier
DMSF	Document Management System Features
EC	European Commission - is the executive body of the European Union responsible for proposing legislation, implementing decisions, upholding the EU treaties and managing the day-to-day business of the EU
ENVRI	Environmental Research Infrastructures
ENVRI KB	ENVRI Knowledge Base
EOSC	European Open Science Cloud
ESFRI	European Strategy Forum on Research Infrastructures
FAIR	Findability, Accessibility, Interoperability, and Reusability of digital assets
GA	(1) Grant Agreement - Contract between Coordinator and Commission (2) General Assembly - GA is the ultimate decision-making body of the consortium
GDPR	General Data Protection Regulation
OIL-e	open information linking for ENVRI
PMT	Project Management Team
POPD	Protection of Personal Data
RDF	Resource Description Framework
RI	Research Infrastructure
WP	Work Package