19-07-2022

Deliverable D2.3
Data Management Plan

Deliverable D2.3

Contractual Date:	01-10-2020
Actual Date:	19-07-2022
Grant Agreement No.:	951886
Work Package	WP2
Task Item:	Task 2.3
Nature of Milestone:	R (Report)
Dissemination Level:	PU (Public)
Lead Partner:	PTB
Document ID:	CLONETS-M22-001
Authors:	Schnatz, Harald (PTB)

Abstract

With this document we put in place a Data Management Plan to ensure that all envisioned users profit from a common data platform in an appropriate way.

Table of Contents

Executive Summary

With this document we put in place a Data Management Plan to ensure that all envisioned users profit from a common data platform in an appropriate way.

With this we:

1. identify where the data platform is to be hosted to be easily accessible for every member

2. categorise the data processed in the CLONETS network

3. set up rules for the access to data

4. and assure the security of the data processed

This Data Management Plan (DMP) describes the data management for all of the data sets that will be collected, processed or generated by the project. The DMP will cover the following aspects:

1. the handling of the data and the rules for sharing data between different groups of users;

2. specification of the data that will be collected, processed or generated;

3. the methodology and standards (including data security and ethical aspects) that will be applied;

4. plans for data curation and preservation.

5. definition of a standard, structured data format (e.g. JSON, XML, ASCII or TIFF) and naming rules for easy data exchange, and reliable use.

6. support of the principles of fair access to scientific data - Findable, Accessible, Interoperable and Reusable (FAIR).

The selection of data to be openly accessible will be made on a case-by-case basis and agreed by the consortium. This includes ethical aspects and data security such as for the protection of IP for any project outputs that are considered to be commercially exploitable.

1   Data Management Plan

Task 2.3 Data Management (lead PTB)

This document provides a Data Management Plan to ensure that all partners and future users profit from a common data platform in an appropriate way.

In particular this document:
1.   identifies where the data platform is hosted to be easily accessible for every member
2.   categorises the data processed in the CLONETS-DS network
3.   defines rules of access to data
4.   and assures the security of the data processed

The Data Management Plan (DMP) applies to all data sets that will be collected, processed or generated by the project. The DMP covers the following aspects:

1.   the handling of the data and the rules for sharing data between different groups of users;

2.   specification of the data that will be collected, processed or generated;

3.   the methodology and standards (including data security and ethical aspects) that will be applied;

4.   plans for data curation and preservation.

5.   definition of a standard, structured data format (e.g. JSON, XML, ASCII or TIFF) and naming rules for easy data exchange, and reliable use.

The developed methods of accessing and processing data use the latest best practice analysis on data management. The data collection method considers the network structure developed within the framework of the project. Within this document the consortium develops guidelines for scientific network users in order to support the principles of fair access to scientific data - Findable, Accessible, Interoperable and Reusable (FAIR).

For internal use the consortium agrees to upload all files and communication to the platform provided by PSNC. Access to this platform is restricted to members of CLONETS and the members of the external advisory board.

The structure of the platform is organized as folders and subfolders dedicated to specific topics such as contract documentation, organisation, work packages, deliverables, promotion, etc.

The consortium has agreed that the selection of data to be openly accessible will be made on a case by case basis. This includes ethical aspects and data security such as for the protection of IP for any project outputs that are considered to be commercially exploitable. In such cases, it may be necessary to withhold all or some of the data generated. This will be decided by the relevant partner(s) and managed by the DMP, the Consortium Agreement and if appropriate the project's exploitation plan.

This document has had regular updates at each project meeting.

1.1 Data Summary

Questions	Answers
What is the purpose of the data collection / generation?	Data collected and generated will be used to create a comprehensive design study, paving the way towards establishing a pan-European time and frequency reference system as a European Research Infrastructure, which serves the European science community.
What is its relation to the objectives of the project?	**Survey responses (WP1)** will be used to design an architecture for a time and frequency reference system at the technical level (WP2); to design a sustainable governance of the reference system (WP3); to support end-user uptake (WP4); and to support claims and references made in conference presentations and peer-reviewed publications (WP4). The architecture designs, site information and service requirements collected and generated will support the drafting of an engineering plan which achieves interoperability at the European level (WP2).
What types and formats of data will the project generate/collect?	**Raw data** will be generated by instruments in their native format, which may be proprietary. Where possible, open formats such as CSV, XLSX and HDF will be preferred. **Processed data** may be generated from raw data for the purpose of sharing and archiving and will be in an open format such as txt, dat, CSV, XLSX and HDF. **Numerical simulation data** will be generated in an open format such as CSV, XLSX and HDF. **Survey responses** will be stored in an open text or spreadsheet format such as DOCX, XLSX or PDF. Online survey results may additionally be stored in another platform-dependent format.
Will you re-use any existing data and how?	Re-use of existing data is not currently foreseen.
What is the origin of the data?	**Survey results** will be collected from targeted organisations in the form of questionnaires, including online.
What is the expected size of the data *(if known)*?	The total amount of data generated over the duration of the project is not expected to exceed 1 TB (equivalent to 1 GB per day).
Outline who might find it useful *('data utility')*?	ESFRI - listed RI and other European users of time and frequency standards, who wish to develop use models for the different groups of users. All identified users in WP1 and businesses interested in commercialisation of the architecture designs and sustainable governance structures developed in this project.

	Extended studies beyond CLONETS-DS and implementation of CLONETS-DS will require the reuse the data from this project.

1.2 Findable, Accessible, Interoperable and Reusable (FAIR) Data

1.2.1 Making Data Findable, Including Provisions for Metadata

Questions	Answers
Are the data produced and/or used in the project discoverable with metadata?	Data produced by the project and intended for sharing openly or within the consortium will be discoverable with metadata. Data referred to in published papers will be available via referenced links.
Are the data identifiable and locatable by means of a standard identification mechanism *(e.g. persistent and unique identifiers such as Digital Object Identifiers)*?	Datasets released for open access will be assigned a DOI. The project's open access peer-reviewed publications will each have a DOI.
What naming conventions will you follow?	The conventions to be used will include key information such as project name, dataset name, laboratory name, time and date.
Will search keywords be provided that optimise possibilities for re-use?	We use the keywords whenever we publish conference proceedings, publications and datasets.
Will you provide clear version numbers?	Datasets will be allocated version numbers while they are being worked on / generated, except for raw data which will not be modified once generated. The implementation of versioning will depend on the storage location.
What metadata will be created? *If metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.*	Metadata will be created for all datasets intended for sharing (openly or within the consortium), with a minimum of: Description, Creator/Owner, Access, Persistent Identifiers

1.2.2 Making Data Openly Accessible

Questions	Answers
Which data produced and/or used in the project will be made openly	Data associated with published papers and with the validation and characterisation of open hard- and software produced by the project will be made openly available by default, unless there is a specific

available as the default? *If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.*	limitation on that data, such as IP or commercial restrictions. Restricted background IP has been declared within the Consortium Agreement. WP4 aims at making data openly available, subject to legal and ethical constraints. The accessibility of datasets generated by collaborators will be defined when setting up collaboration agreements.
How will the data be made accessible *(e.g. by deposition in a repository)*?	Once processing, quality control, organisation, analysis and publication are complete, the data will be deposited in a public repository such as **Zenodo**.
What methods or software tools are needed to access the data?	Data will be in standard formats readable by common office and analytic software, which is commercially or otherwise publicly available. This software may include Microsoft Office, Matlab, Python, Mathematica, Origin and Adobe Reader (depending on the type of data).
Is documentation about the software required in order to access the data included?	Where the data format is not obvious from the file extension (such as .txt or .xlsx), information will be provided explaining the format and indicating suitable software to access it. This information may be provided in the form of a text file stored alongside the data, or in another suitable way depending on the repository.
Is it possible to include the relevant software *(e.g. in open source code)*?	This will not be necessary, because the data generated in the project is expected to be accessible and usable with common office and analytic software, including existing open source software.
Where will the data and associated metadata, documentation and code be deposited? *Preference should be given to certified repositories that support open access where possible.*	At the outset of the project, we envisage storing the data in the open access, secure public repository **Zenodo**. However, this will be reviewed as the project progresses and if there is any good reason to do so, certain data may be stored in other repositories supporting open access.
Have you explored appropriate arrangements with the identified repository?	Arrangements will be as per standard procedure with the public repository.
If there are restrictions on use, how will access be provided?	No restrictions on the use of the published data are envisaged, but users will be required to acknowledge the consortium and the source of the data in any resulting presentations or publications.
Is there a need for a data access committee?	The consortium will act as the data management and access committee, deciding which data should be openly accessible on a case by case basis. Ethical aspects and data security, including intellectual property requirements, will be considered. If necessary, some or all of a potential publication's data will be withheld.

Are there well described conditions for access *(i.e. a machine readable license)*?	Data will be accessible to the public without restrictions. An appropriate standard license such as **Creative Commons Attribution (CC-BY)** will be used.
How will the identity of the person accessing the data be ascertained?	User identification or registration is not expected to be required as a pre-requisite for access.

1.2.3   Making Data Interoperable

Questions	Answers
Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (*i.e. adhering to standards for formats, that are as far as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)*?	The data produced in the project will be interoperable as the datasets will adhere to standardised open formats and will contain metadata.
What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?	Datasets will have metadata defined to match at least minimum standards for the data repository, and for access and discovery. These will be defined as part of the project activities.
Will you be using standard vocabularies for all of the data types present in your data set, to allow inter-disciplinary interoperability?	For all datasets: the data, metadata and documentation are compliant to disciplinary standards, open file formats and use controlled vocabularies and the standard metadata scheme for easy interoperability and re-use.
If it is essential to use uncommon, or generate project specific, ontologies or vocabularies, will you provide mappings to more commonly used ontologies?	This is not currently envisaged.

1.2.4 Increase Data Re-Use (Through Clarifying Licenses)

Questions	Answers
How will the data be licensed to permit the widest re-use possible?	We anticipate licensing data made openly available under a **Creative Commons license**. The precise form of license, e.g. Creative Commons Attribution 4.0 (CC BY 4.0) or Creative Commons Attribution and ShareAlike 4.0 (CC BY-SA 4.0), will be determined on a case by case basis.
When will the data be made available for re-use? *If an embargo is required to allow time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.*	Data associated with papers published in a peer-reviewed open-access journal will be made available from the date of article publication. WP4 aims for publishing data at monthly intervals. For other data made available openly, the decision will be taken on a case-by-case basis.
Are the data produced and / or used in the project useable by third parties, in particular after the end of the project? *If the re-use of some data is restricted, explain why.*	Any data made openly available will remain so indefinitely, subject to availability of the repository, and may be used with standard software.
How long will the data remain re-usable?	Any data made openly available will remain reusable for the lifetime of the repository, which is expected to be a minimum of 20 years.
Are data quality assurance processes described?	Data quality assurance processes will be followed, according to the quality systems in place within each partner organisation, and these will be described within the project. Data quality will be assured through repeated and comparison measurements, through the metrological characterisation of measurement setups and through the validation of the data collected. Other quality assurance processes will include the documented testing of software and the peer-review of publications based on the data.

1.3 Allocation of Resources

Questions	Answers
What are the estimated costs for making data Findable, Accessible,	The chosen repository, **Zenodo**, is free of charge.

Interoperable and Reusable (FAIR) in your project?	There will be costs associated with the resources required to prepare and lodge datasets on the repository, and these are expected to be absorbed as part of the overall project costs.
How will these costs be covered? *Note that costs related to open access to research data are eligible in EMPIR (if compliant with the Grant Agreement conditions).*	The costs for making the data FAIR are included in the project budget.
Who will be responsible for data management in your project?	The WP2 lead (Harald Schnatz, PTB) is responsible for coordinating updates to the data management plan. However, partners in the consortium will be responsible for managing data generated within their own organisations, including organising data backup and storage, data archiving and depositing the data within the repository.
What are the costs and potential value of the long-term preservation of the data *(also state who decides on what data will be kept and for how long)*?	No costs beyond costs of preparing and depositing the dataset are expected. The data has long-term value for subsequent use in publications, projects or by individuals. The Consortium will collectively decide on what data should be kept and for how long.

1.4   Data Security

Questions	Answers
What provisions are in place for data security (including data recovery as well as secure storage and the transfer of sensitive data)?	*Data recovery and secure storage* All partners are either accredited to, or work in compliance with, the ISO 17025 standard on the "General requirements for the competence of testing and calibration laboratories". The partners will store data on their organisations' networks, or cloud-based file storage, which have backup and cybersecurity measures in place. Data deposited in the **Zenodo** public repository will provide additional security as it has multiple replicas in a distributed file system which is backed up on a nightly basis. *Transfer of sensitive data* No personal or other sensitive data is expected to be generated.
Is the data safely stored in certified repositories for long term preservation and curation?	Yes. The data will be safely stored in the **Zenodo** open access repository.

1.5 Ethical Aspects

Questions	Answers
Are there any ethical or legal issues that could impact on data sharing? *You can also discuss this in the context of the outcomes of the ethics review and if relevant, include references to ethics report(s) and the ethics section in the Annex 1.*	WP5 will assess any legal and ethical questions that may arise in the context of this activity.
Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?	The consortium has no plans to share data with identifiable personal information. Surveys conducted within this project will be addressed at organisations, not individuals and the results will be shared only in anonymised / aggregated form.

1.6 Other

Questions	Answers
Do you use other national / funder / sectorial / departmental procedures for data management? If yes, which ones?	Data management will be compliant with European laws about data security and the protection of privacy (e.g. GDPR).

References

[Zenodo]   **https://www.zenodo.org/**

Glossary

DOI	Digital object identifier
DMP	Data Management Plan
EMPIR	European Metrology Programme for Innovation and Research
ESFRI	European Strategy Forum on Research Infrastructures
FAIR	Findable, Accessible, Interoperable and Reusable
GDRP	General Data Protection Regulation
PSNC	Poznan Supercomputing and Networking Center
PTB	Physikalisch-Technische Bundesanstalt