



Reliability, Resilience and Defense technology for the grid

D1.2 – Data Management Plan

Date: 30/03/2023



**Funded by
the European Union**

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. Horizon Europe Grant agreement N° 101075714.

Deliverable details

Title	WP	Version
Data Management Plan	1	1.0

Contractual delivery date	Actual delivery date	Delivery type*
31/03/2023	31/03/2023	public

*Delivery type: R: Document, report; DEM: Demonstrator, pilot, prototype; DEC: Websites, patent fillings, videos, etc; OTHER; ETHICS: Ethics requirement; ORDP: Open Research Data Pilot.

Author(s)	Organisation
Ugo Stecchi	ETRA
Aleksandra Krkoleva,	UKIM
Petar Krstevski	UKIM
Aris Dimeas	ICCS

Version	Date	Person	Action	Status*	Dissemination**
V0.1	07/03/2023	Ugo Stecchi	First draft of the document	draft	CO
V0.2	09/03/2023	Lucas Pons	Documents improvements	Preliminary version	CO
V0.3	09/03/2023	Ugo Stecchi	Document consolidation	Version ready for peer review	CO
V0.4	22/03/2023	Aleksandra Krkoleva, Petar Krstevski	Peer review	Reviewed version	CO
V0.5	24/03/2023	Aris Dimeas	Peer review	Reviewed version	CO
V1.0	29/03/2023	Ugo Stecchi	Final version	Changes accepted	PU

*Status: Draft, Final, Approved, Submitted (to European Commission).

**Dissemination Level: PU: Public; CO: Confidential, only for members of the consortium (including the Commission Services)

Executive Summary

The Data Management Plan (DMP) is a living document that aims at providing an analysis of the main elements of the data management policy that will be used by the R²D² Consortium regarding the project research data.

This document describes initially the methodological framework taking into account the Horizon Europe Programme Guide regarding Open Research Data with a twofold objective: on the one hand, to clearly define how all public data assets issued by the project consortium will be further disseminated to a wider audience and on the other hand, to outline the methodology which can make the research data generated in the context of R²D², findable, accessible, interoperable and reusable (FAIR principles).

It is allowed that this document may evolve during the development of the project, in case significant changes occur, in order to keep an updated version of the guidelines and recommendations for making the research data Findable, Accessible, Interoperable and Reusable (FAIR) and therefore contribute to knowledge discovery and innovation.

The current version of the document is the first iteration in which the envisioned data management strategy is presented. It makes a first effort to plan the definition of the types of research data that will be generated or collected during the project, the standards that will be used, how the research data will be preserved and what parts of the datasets will be shared for verification or reuse.

Keywords

Data Management, FAIR, Privacy, Security, Ethics

Copyright Statement

The work described in this document has been conducted within the R²D² project. This document reflects only the R²D² Consortium view and the European Union is not responsible for any use that may be made of the information it contains.

This document and its content are the property of the R²D² Consortium. All rights relevant to this document are determined by the applicable laws. Access to this document does not grant any right or license on the document or its contents. This document or its contents are not to be used or treated in any manner inconsistent with the rights or interests of the R²D² Consortium or the Partners detriment and are not to be disclosed externally without prior written consent from the R²D² Partners.

Each R²D² Partner may use this document in conformity with the R²D² Consortium Grant Agreement provisions.

1 Index

1	Index	4
1.1	list of tables.....	5
1.2	list of Figures.....	5
2	INTRODUCTION	6
2.1	Purpose of the Document	6
2.2	Scope of the Document.....	6
2.3	Structure of the Document.....	6
3	DATA SUMMARY.....	7
3.1	Purpose of the data management and relation to the project.....	7
3.2	Data set types, formats and standards	7
3.3	Increase data Re-use.....	8
4	FAIR DATA.....	10
4.1	making data findable, including provisions of metadata	10
4.2	making data accessible	11
4.2.1	Open research data repository	12
4.2.2	Open Access publications.....	12
4.3	making data interoperable	13
4.4	PROJECT'S GUIDELINES FOR data re-use	14
5	OTHER RESEARCH OUTPUTS.....	15
5.1	R ² D ² Website and public deliverables	15
5.2	open source code	15
6	ALLOCATION OF RESOURCES	17
6.1	Roles and Responsibilities	17
7	DATA SECURITY.....	19
8	ETHICS.....	20
9	CONCLUSIONS.....	21
10	References and Acronyms	22
10.1	References.....	22
10.2	Acronyms	23



1.1 LIST OF TABLES

Table 1: Template of the table for the project’s data mapping	8
Table 2: Dublin Core Metadata Element Set - 4 Interoperability Levels	11
Table 3: Dublin Core Metadata Element Set – 15 Elements Overview	11
Table 4: Acronyms	23

1.2 LIST OF FIGURES

Figure 1: Project website under construction, with “library” section where to store public deliverables	15
---	----

2 INTRODUCTION

2.1 PURPOSE OF THE DOCUMENT

The Data Management Plan (DMP) is a living document, and it is expected to evolve during the lifespan of the project. As the implementation of the project progresses and significant changes occur, the DMP will be updated with information at a finer level of granularity.

This first version of the DMP aims to outline how the R²D² project will try to make the research data FAIR and therefore contribute to knowledge discovery and innovation. Although this first version submitted by month 6 of the project does not provide very detailed information on the specific datasets to be collected, generated and processed during the project, the objective is to lay the foundations for creating an effective data management strategy covering the complete research data life cycle.

2.2 SCOPE OF THE DOCUMENT

The focus of this deliverable is to cover the entire life cycle of research data. This data management plan, which will be updated on an ongoing basis, will serve as a guide by the R²D² consortium on how to manage the research data during and after the end of the project. The current version is the first iteration presenting the planned data management strategy it provides an initial definition of the types of research data that will be generated or collected during the project, the standards that will be used, how the research data will be stored and what portions of the datasets will be shared for review or reuse. The next versions of the DMP will get into more detail and describe the datasets to be produced by the project, the specific conditions that are attached to them and the practical data management procedures to be implemented under the R²D² project.

2.3 STRUCTURE OF THE DOCUMENT

This deliverable follows the template provided by the services of the European Commission (EC) [1] on Data Management Plans, adapted to follow R²D² document procedures –e.g. section 1 “Introduction” is mandatory in all the project documents.

In the next section, the data summary is resuming information about datasets that will be used in the project, including possible re-use of data. Then, the document focuses on making R²D² research data FAIR. The summary of the European Commission guidelines can be found in [2].

Respecting the guidelines provided by the Commission in art. 17 of the Model Grant Agreement [3], this document will be updated in correspondence of the end of the review periods or differently, when required, in case significant changes may arise.

3 DATA SUMMARY

3.1 PURPOSE OF THE DATA MANAGEMENT AND RELATION TO THE PROJECT

The goal of R²D² project is to improve the overall resilience of the power system, through the improvement of the security (both digital cyber-security of IT infrastructure and operational security of operational technologies (OT)) and reliability of the systems at high, medium and low voltage level.

The R²D² DMP is in compliance with European recommendations for Data Management Plans [1], which ensures that data is FAIR by providing clear procedures and regularly updating the project's document. The DMP analyses the critical aspects of data management policies that the R²D² Consortium needs to consider for project research data.

Beneficiaries must manage the digital research while the project data is generated in the action ('data') according to FAIR principles as stated in annex 5 of the Model Grant Agreement [3], following the philosophy of:

- *“as open as possible as closed as necessary, unless providing open access would in particular:*
- *be against the beneficiary's legitimate interests, including regarding commercial exploitation, or*
- *be contrary to any other constraints, in particular the EU competitive interests or the beneficiary's obligations under this Agreement; if open access is not provided (to some or all data), this must be justified in the DMP”.*

The DMP serves as an analytical document for the primary components of the data management policy that will be employed by the Consortium for project research data. It encompasses the consortium's data management policies, procedures, and systems, which will be implemented and integrated into research practices, and regularly assessed throughout the research cycle.

While the project aims to make data openly available, any restrictions imposed by partner institutions on data they contribute (as specified in their Background stated in the R²D² Consortium Agreement) will not be overridden. Additionally, the project will adopt and maintain an ethical approach throughout the fieldwork process. Responsible partners will ensure that the project adheres to EU standards concerning ethics and data management, as stated in D1.1 Project Management Plan [4].

3.2 DATA SET TYPES, FORMATS AND STANDARDS

Throughout the duration of the project, R²D² will gather (and also generate) a variety of data. All of it must be stored in a way such that it is easily accessible by both humans and software, as appropriate. Broadly, the data falls into two categories:

- Organizational data – e.g. data relevant to the implementation of the Innovation action.
- Technical and scientific data – this includes raw and processed experimental data, scientific analyses/publications as well as software code and algorithms.

The detailed definition of the different types, formats and standards of the data to be collected, processed and/or generated during the project will be done in the framework of Work Package 2 (WP2), more specifically in task 2.2 “System Architecture Definition” and presented in the associated deliverables D2.1 and D2.3 “Requirements and Detailed architecture design”. As indicated in the R²D² Description of Action (DoA), these deliverables will be respectively submitted in month 9 and in month 16. They will provide a description of the open, secure and flexible architecture including details of the types and formats of data to be handled.

In this context, in month 6 (the due month for the current deliverable) the project is not yet in the position to provide a detailed list of the datasets that will be used in the project, and neither the format nor specific standards to be used. As defined in the DoA and the structure of the project, it was not expected/planned to have the details of this information by month 6 and therefore it cannot be included in the current version of the DMP. The next version of the DMP will include the final decisions on the datasets to be used.

The consortium will work on the definition and refinement of all types of research data to be handled during and after the end of the project, the methodologies and standards to be applied and how it will be shared, curated and preserved.

The starting point of task 2.2 mentioned before and of T2.4 “Pilot Sites Survey and Recognition” will be to list the types of data to be handled in the framework of R²D² and to associate to each type of data the specific details, which can define the sensitivity of data in terms of: a) privacy; b) data protection due to privacy or other reasons, such as commercial interest or critical infrastructure data. As a result of the activities from the overmentioned tasks, the next version of the DMP will list the dataset to be used in the project providing at least the information reported in the template table below:

Table 1: Template of the table for the project’s data mapping

#	Dataset Name	Diss. Level	Stored in	Relevant WP	Format	Relevant Standard

In any case, as already stated in deliverable D1.1-Project Management Plan [4], the project is not supposed to employ private data related to individual consumption/production of energy. R²D² will exclusively manage aggregated and/or anonymised data sets. In addition, the Project Management Plan identifies the need to apply security measures when collecting and storing data from partners. Considering the early phase when D1.1 was edited, it is recognised the need to apply technical measures to secure the data transfer from field devices to the repository, through secure transfer protocols and tools. Similarly, data encryption mechanisms will be applied to secure data storage as well.

3.3 INCREASE DATA RE-USE

Along the project’s lifetime, existing research findings, publications, and other pertinent information that is accessible will be examined. The primary purpose of this analysis will be



to conduct internal project assessments, and the relevant information will be included in the appropriate project deliverables with attribution to the original sources.

Results available from other research activities, publications, and further relevant available information will be analysed. This information will be mainly used for internal project studies and will be provided in respective project deliverables with appropriate references to the origins of the gathered information.

Moreover, R²D² will use and exploit some of the preliminary solutions and technologies developed in various H2020 projects: SYNERGY [6], TRINITY [7], CROSSBOW [8], WISEGRID [9], PHOENIX [10] and X-FLEX [11] projects. These projects are expected to form the basis of some of the R²D² Innovation activities, complementing and going beyond the work done in these projects, and providing new and more advanced services.

Concerning the re-use of data generated within the project activities outside the consortium, it is expected that the project R²D² will contribute to the development of a “shared knowledge basis within European area concerning threats, vulnerabilities, methods, not only for components but for entire systems and energy system technologies” as requested by the topic [5]. R²D² will provide a repository to store reports and multi-spatial and multi-temporal assessments of the events with a great impact on power systems, during T8.2 activities, following the same principles and guidelines for project’s data storage and sharing described in this document. The sharing of information is aimed at promoting a common and public discussion about the evolution of resiliency standards and regulations.

4 FAIR DATA

According to [2], the FAIR principles describe four key concepts in research data and metadata management. In accordance with the FAIR principles, data should be:

- “Findable – (Meta)data and data should be easy to find for both humans and computers. Machine-readable metadata are essential for the automatic discovery of datasets and services.
- Accessible – Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.
- Interoperable – The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.
- Reusable – The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.”

4.1 MAKING DATA FINDABLE, INCLUDING PROVISIONS OF METADATA

Means (tools) for identification and localisation will be used for processing the data during the R²D² project. The data to be generated in R²D² project will be identifiable and locatable by means of unique identification mechanisms. Files will be uniquely identifiable by using standardised name conventions and clear versioning. These conventions for the documents – and data sets – are already provided in the deliverable D1.1 Project Management Plan [4].

R²D² project research data will be inventoried and annotated with metadata following the discoverability standards promoted by the Data Catalogue (DCAT) Application Profile [12]:

“It is important to ensure that your data can be found. The term usually applied to this is the discoverability of data. Essential for discoverability is metadata. Metadata describes the dataset itself (e.g. date of creation, title, content, author, type, size). This information about the data needs to be added to the catalogues to help discover the data. Metadata needs to be both human understandable and machine readable. If it is published as Linked Data, the discoverability of the data is greatly increased. Metadata does not only serve the purposes of description and discovery, but also renders itself as essential for the scope of contextualisation (relevance, quality, restrictions (rights, costs)), as well as for matching users and software to data available on the internet.”

The use of the DCAT Application Profile for metadata [13], is highly recommended by the European Data Portal. The DCAT, which is used for datasets available on the internet, is based on the Dublin Core standards.

The Dublin Core metadata standard is a straightforward and efficient element set used to describe various networked resources. The metadata landscape, as perceived by the Dublin Core community, is currently divided into four levels of interoperability and 15 sections

for data description [14], [15]; these interoperability levels and sections are resumed in the tables below:

Table 2: Dublin Core Metadata Element Set – 4 Interoperability Levels

Dublin Core Levels of interoperability	
Level 1: Shared term definitions	Shared vocabularies defined in natural language
Level 2: Formal semantic interoperability	Shared vocabularies based on formal semantics
Level 3: Description Set syntactic interoperability	Shared formal vocabularies in exchangeable records
Level 4: Description Set Profile interoperability	Shared formal vocabularies and constraints in records

Table 3: Dublin Core Metadata Element Set – 15 Elements Overview

Contributor	An entity responsible for making contributions to the resource
Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant
Creator	An entity primarily responsible for making the resource
Date	A point or period of time associated with an event in the lifecycle of the resource
Description	An account of the resource
Format	The file format, physical medium, or dimensions of the resource
Identifier	An unambiguous reference to the resource within a given context
Language	A language of the resource
Publisher	An entity responsible for making the resource available
Relation	A related resource
Rights	Information about rights held in and over the resource
Source	A related resource from which the described resource is derive
Subject	The topic of the resource
Title	A name given to the resource
Type	The nature or genre of the resource

The fifteen elements of the "Dublin Core", described in this standard are part of a larger set of metadata vocabularies and technical specifications maintained by the Dublin Core Metadata Initiative (DCMI). The fifteen element descriptions have been formally endorsed in the following standards:

- ISO Standard 15836:2009 of February 2009 (confirmed in 2014)
- ANSI/NISO Standard Z39.85-2012 of February 2013
- IETF RFC 5013 of August 2007

The DCAT Application Profile utilizes Dublin Core standards as a foundation, but it is not a vocabulary. Rather, it is a straightforward specification used for describing metadata for EU governmental data and portals. The metadata will be made public alongside the data in a machine-readable format, utilizing standard terminology to define the metadata. The metadata will also describe the overall features of the dataset, including information about local parameters, license, origin, and quality. The European Data Portal has established best practices for using DCAT Application Profile, and the R²D² partners will adhere to them.

4.2 MAKING DATA ACCESSIBLE

An analysis of which R²D² research data will be made openly accessible and which data will be kept closed will be performed in a later stage of the project. As indicated in Section 3.2, the starting point is the definition of all types of research data to be handled and generated during and after the end of the project. This activity will be done in the framework of the ongoing task T2.2. Once this action is done, the data accessibility analysis will be carried out including the specification of software tools required to access the data. The outcome shall be implemented in the coming versions of the DMP. At the current stage, this analysis would not make much sense since the data sets are not defined yet.

During the project lifetime, information on the following aspects will be elaborated for all data sets on case-by-case base, before making consortium decision on handling of the particular data generated or collected:

- Nature and scale of the data in consideration,
- To whom it could be useful / targeted audience and its size / level of interest,
- Information on the existence of similar data and possible synergies,
- Possibility for integration and reuse of the provided data by external users / researchers, and
- Any further related issue.

In general terms, R²D² research data will be made available, when possible, without compromising privacy, ethical or commercial sustainability, to parties with a legitimate research interest. In the case of R²D², certain research data – especially considering the pilot clusters – is sensitive due to security issues and therefore will be kept confidential.

4.2.1 Open research data repository

To ensure open access to research data sets, the R²D² project will utilize Zenodo [16], an established European online scientific repository that is fully integrated with OpenAIRE [17]. Decisions on whether to upload and provide access to research data sets on Zenodo will be made on a case-by-case basis by the Project Coordinator (ETRA), the Technical Coordinator (ICCS), and the partner(s) who own the data. Datasets marked as public will be made openly available, while those containing confidential or protected information will be kept private for privacy reasons.

Zenodo [16] provides a user-friendly online service that allows researchers, scientists, EU projects, and institutions to share, preserve, and showcase research results, including data and publications, that are not already part of institutional or subject-based repositories. The service provides hosting in CERN's professional data centres, following industry best practices. Zenodo's policies on data handling and service usage are described in detail in [16].

As previously stated, R²D² intends to share datasets publicly in Zenodo, using the repository's required/provided descriptive metadata. Zenodo provides a Digital Object Identifier (DOI) for all publicly accessible uploads, making them easily and uniquely citable. It is important to note that Zenodo DOIs cannot be edited once registered. Additionally, Zenodo supports the harvesting of all content via the OAI-PMH protocol [18].

4.2.2 Open Access publications

The rules and principles of the European Commission's Horizon Europe Framework Programme clearly define that scientific results generated within projects should be made available as open-access publications, i.e. freely available online to any user. Following these rules, as enforced in the R²D² Grant Agreement – Annex 5, Article 17, open access will be ensured to all peer-reviewed scientific publications related to R²D² and its composite solutions.

In addition, the R²D² project may decide to provide other outputs and results as Open Access. As per the "Open access to publications and data in Horizon 2020 Fact sheet" [19], there are two main routes for Open Access to scientific peer-reviewed publications.

- The first route is self-archiving, also known as "Green" Open Access, where the researcher archives the published article or the final peer-reviewed manuscript in an online repository before, after or alongside its publication. Access to the article may be delayed for an embargo period of six months after publication.
- The second route is Open Access publishing, also known as "Gold" Open Access, where the scientific publisher immediately provides the article in Open Access mode, with the associated costs being charged to the researcher's affiliated research institute or funding agency.

Within the EU-funded R²D² project, both routes of publishing (gold and green open access) are available and not mutually exclusive. Each beneficiary has the liberty to choose the most appropriate approach for their respective publications. Presently, parallel publishing is the preferred strategy for providing open access, enabling consortium members to publish their findings in scientific journals of their choice for maximum impact while ensuring optimal dissemination through open access. It is noteworthy that most academic journals support either gold, green or hybrid open access. Hence, R²D² beneficiaries can select the most appropriate publishing venue according to their preference. All publications will be assigned a DOI, making them easily discoverable and citable.

It is essential to recognize that the Open Access requirement does not obligate beneficiaries to publish their results. The decision to publish is entirely up to them. Open Access becomes an issue only if beneficiaries choose to publish their research results. It will be considered also the possibility to submit scientific publications to Open Research Europe (ORE) [20], the European Commission's Open Access publishing platform for research stemming from Horizon Europe. The platform simplifies the process of complying with the open access requirements of these funding sources and enables researchers to swiftly share their findings and perspectives, as well as to promote open and constructive discussions regarding research.

4.3 MAKING DATA INTEROPERABLE

At a later stage of the project, an evaluation of data interoperability will take place to determine which data and metadata vocabularies, standards, or methodologies will be utilized to promote interoperability. The assessment will determine if standardized vocabularies will be adopted for all data types within the dataset to enable cross-disciplinary interoperability. The first step in this process will be to define all types of research data that will be generated and handled during and after the project, as well as the components or actors involved in communication within the R²D² project (as it will be outlined in the architecture deliverable D2.1).

The following actions are planned for the upcoming months:

- Identification of the main interfaces between components and actors to be developed in the scope of R²D² project.
- For each interface, assess the available standards and data models.
- For each interface, assess the available new data models based on ontologies.
- Gather the applicability of the standards and data models identified to the R²D² project.
- For each interface, assess the most appropriate standards and best suitable data models.
- Include the aforementioned process in deliverable D2.1 or D2.3.

The outcomes of the actions described above will be integrated in the coming versions of the DMP.

4.4 PROJECT'S GUIDELINES FOR DATA RE-USE

Data handling during the project will be conducted on a case-by-case basis, as previously mentioned. Once a data set is designated as public, it will be made available on Zenodo for full reuse. This may include specifying an embargo period or providing controlled access to a whitelist of individuals, following Zenodo policies [16].

As suggested by the Model Grant Agreement [3], data will be made accessible with Creative Commons Licences (CC BY or CC0) whenever possible. Depending on their characteristics, data sets may be subject to different licenses.

The Zenodo repository ensures sustainable archiving of the final research data. Items deposited in Zenodo will be retained for the lifetime of the repository, which is currently the lifetime of the host laboratory CERN and has an experimental programme defined for at least the next 20 years. All publicly available uploads on Zenodo will be stored safely for the future in the same cloud infrastructure as research data from CERN's Large Hadron Collider and using CERN's battle-tested repository software INVENIO, which is used by some of the world's largest repositories such as INSPIRE HEP and CERN Document Server.

The data will remain re-usable at least until Zenodo discontinues the dataset(s) (i.e. warranted for a minimum of 20 years).

The project envisages adopting the “data pedigree” concept, which assures that each piece of relevant information is traceable back to the original data sources. This data lineage along with metadata allows for quality audit and sensitivity analyses of the outputs.

5 OTHER RESEARCH OUTPUTS

5.1 R²D² WEBSITE AND PUBLIC DELIVERABLES

The R²D² website describes the mission and the general approach of the project and its development status, as well as provides a short description of the project's objective and its methodology, news, events and updates that are relevant to the project's activities.

After submission and approval from the EC, the project's public deliverables will be downloadable from the website, while confidential deliverables will be kept in the Alfresco [21] repository accessible only to authorised users. In certain cases, external parties may request access to confidential deliverables, and the Consortium may decide to share corresponding deliverables or specific parts with those parties. The public deliverables will be available in the commonly used PDF format, on a dedicated section of the project's website [21] (under construction when preparing this deliverable).

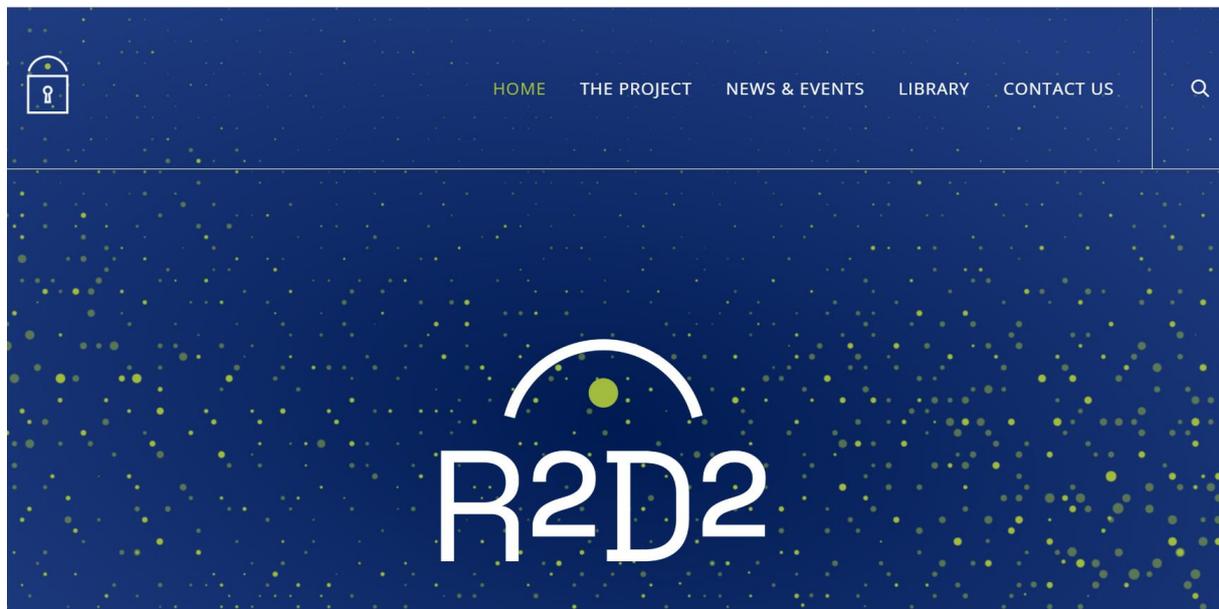


Figure 1: Project website under construction, with “library” section where to store public deliverables

During the life cycle of R²D², data collected or generated by the project will be stored and systematically organised in the official project repository on Alfresco.

Alfresco is a flexible content management web application developed using Java technology. In the framework of the R²D² project, it is used mainly as a repository to securely store and share files, making them available to the whole Consortium. The Documents Library consists of a project internal area, that is not possible to be accessed by external users.

5.2 OPEN SOURCE CODE

The open-source code for R²D² will be accessible on GitHub. Any source-code components that are developed during the project and are deemed appropriate for public access will be uploaded to an open-access GitHub repository. This repository will also be linked to the Zenodo account of the project.



GitHub [23] is an online repository which supports distributed source code development, management, and revision control. It is primarily used for source code data. It enables worldwide collaboration between developers and also provides some facilities to work on documentation and to track issues.

GitHub provides paid and free service plans. Private, non-public repositories require a paid service plan, while free service plans can have any number of public, open-access repositories with unlimited collaborators. Many open-source projects use GitHub to share their results for free. The platform uses metadata like contributors' nicknames, keywords, time, and data file types to structure the projects and their results. The terms of service state that no intellectual property rights are claimed by the GitHub Inc. over the provided material. For textual metadata items, English is preferred.

6 ALLOCATION OF RESOURCES

As this version of DMP is currently based on the use of free resources and open source software, the only costs that will be incurred are related to running the server(s) (hardware) and the working time needed to setup, maintain and evolve the different tools (efforts measured by person-months).

6.1 ROLES AND RESPONSIBILITIES

As indicated in previous sections, the DMP presented in this deliverable is just the first version. Consortium discussions will be continuously carried out to identify the relevant project outputs as well as to decide on the way and the means of their open access (if applicable). To ensure it, a dedicated time slot will be reserved at each project plenary meeting and, if needed, at selected Consortium audio conferences. The EC and the project reviewers will be informed about related work done and publications provided in the project management reports.

Individual responsibilities on data management in the project consortium are:

- Project Coordinator (ETRA) – to prepare and lead related discussions at the relevant project meetings and to maintain the project document repository Alfresco;
- Technical Coordinator (ICCS) – to identify data collected by the project and technical project outcomes eventually suitable for publication; moreover, to ensure dataset integrity and compatibility for its use during the project lifetime by different partners.
- Dissemination Manager (ETRA) – to identify publications suitable for publication in the considered repositories and maintain R²D² inputs for the Open Access;
- Each individual partner – to identify own project results suitable for publication and to share the published scientific articles in advance with project coordinator and dissemination manager.

The Project Coordinator and the Dissemination Manager have a particular responsibility to ensure that data shared through the R²D² website are easily available, but also that backups are performed and that proprietary data are secured.

Moreover, each R²D² partner has to respect the policies set out in this DMP. Data sets have to be created, managed and stored appropriately and in line with applicable legislation. Validation and registration of datasets and metadata is the responsibility of the partner that generates the data in the WP. Metadata constitute an underlying definition or description of the datasets and facilitate finding and working with particular instances of data. Additional responsibilities undertaken by the R²D² project partners include:

- Backing up data assets for sharing through Open Access repositories. It is the responsibility of the partner to possess these data assets.
- Quality control of the data assets. It is the responsibility of the demo partner to provide the data.



D1.2 – Data Management Plan

- Managing different versions in case the data assets are updated, and making sure that the latest version is available in the case of publicly available data.
- Consulting the concerned partner(s) before publishing data in the open domain that can be associated with an exploitable result. It is the responsibility of all project partners involved in this activity.

7 DATA SECURITY

The Zenodo and Alfresco repositories will ensure secure and safe storage of both public and non-public data respectively.

Zenodo provides clear security guarantees. All data files are stored in CERN Data Centres, primarily in Geneva, with replicas in Budapest. Data files and metadata are backed up on a nightly basis. Files are regularly checked against their checksums (using MD5 algorithm) to assure that file content remains constant. In case of closure of the repository, Zenodo ensures that efforts will be made to integrate all content into suitable alternatives [16].

Alfresco is hosted on a private internal server with local backup mechanism, managed by the project coordinator ETRA.

The servers hosting the research data will be accessible only by authorized system administrators. Files containing confidential data should be protected by owners using local encryption tools (i.e. password-protected archives) before being uploaded to shared repositories. Interaction through web user interfaces will use https protocol (i.e. secure). Also, a secure file transfer protocol (sftp) will be provided as the need arises.

To assure data security and privacy, the R²D² platform will support advanced data anonymization and encryption mechanisms, as well as cloud and on-premises storage on servers to which only the authorized staff have access. More specifically the servers onto which the data will be stored will have server-side encryption. This means that the server's administration personnel will be able to generate public keys for specific person who will have access to the data but will not be able to access the data themselves (since the private keys required for this access will be generated on the machine of the person with access to the data). This means that only the required personnel (and stakeholders that have active data contracts) will have access to the data and, even in the remote case of a possible data leak or server hack, the data stolen will be fully encrypted and thus fully non-accessible.

Finally, and after a retention period (to be defined), a secure deletion software will be used to destroy data, i.e. using Gutman algorithm (35-pass overwrite technique).

If deemed necessary, a full format can be used in conjunction with overwriting, to provide further assurance that data cannot be recovered, guaranteeing the destruction of the project personal data.

The following guidelines will be used in order to ensure the security of the data:

- use anonymised and aggregated data instead of individual data;
- encrypt data by the local researchers and do not allow the data to leave their premises unencrypted;
- store data in at least two separate locations to avoid loss of data;
- limit the use of USB flash drives;
- label files in a systematically structured way in order to ensure the coherence of the final dataset.

8 ETHICS

Ethical issues are covered in WP1 and also pre-assessed during the Grant Agreement Preparation phase and reported in deliverable D1.1 Project Management Plan [4]. In this deliverable, the procedure for handling R²D² ethical issues is presented. The R²D² consortium has to comply with all European and national legislation and directives relevant to the country where the data collection is taking place. The collection, processing and transmission of personal data will be analysed under principles of (a) the Universal Declaration of Human Rights and the Convention 108 for the Protection of Individuals with Regard to Automatic Processing of Personal Data, (b) the General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679), and (c) the national laws applying its provisions. Any additional regulations at national level that do not fall under the GDPR and apply to data protection or any other sensitive information, will also be taken into account.

Data managed during the project will be processed only under the following preconditions which need to be met: (a) When the data subject has given her/his consent; (b) When the processing is necessary for the performance of or the entering into a contract; (c) When processing is necessary for compliance with a legal obligation; and (d) When processing is necessary in order to protect the vital interests of the data subject.

9 CONCLUSIONS

This document sets the guidelines and recommendations to be followed in order to make the project research data FAIR and therefore contribute to knowledge discovery and innovation.

The main elements of the data management policy that is used and will be used by the R²D² project are presented and analysed in this deliverable.

As it has been stated in the document, the DMP is a living document that needs to be updated on a regular basis all along the project implementation, in order to cover all changes or progresses that might occur during the project lifetime.

10 References and Acronyms

10.1 REFERENCES

- [1] Horizon Europe, Data Management Plan Template, Version 1.0, 05 May 2021 available at: [https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/how-to-participate/reference-documents;programCode=HORIZON#:~:text=Data%20management%20plan%20\(HE\)](https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/how-to-participate/reference-documents;programCode=HORIZON#:~:text=Data%20management%20plan%20(HE))
- [2] website: <https://www.go-fair.org/fair-principles/> accessed March 2023
- [3] EU Grants AGA – Annotated Model Grant Agreement EU Funding Programmes 2021-2027 PRE-DRAFT (HE) incl. update for ALL PROGRAMMES 30 November 2021, available at: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/aga_en.pdf
- [4] R2D2 Consortium, Deliverable 1.1, Project Management Plan, December 2022
- [5] [Reliability and resilience of the grid: Measures for vulnerabilities, failures, risks and privacy | Programme | HORIZON | CORDIS | European Commission \(europa.eu\)](#)
- [6] <https://synergyh2020.eu/>
- [7] <http://trinityh2020.eu/>
- [8] <https://crossbowproject.eu/>
- [9] <https://www.wisegrid.eu/>
- [10] <https://phoenix-h2020.eu/>
- [11] <http://xflexproject.eu/>
- [12] European Data Portal, "Open Data Goldbook for Data Managers and Data Holders" public report, January 2018, available at: https://data.europa.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf
- [13] website: <https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-data-portals-europe> accessed in: March 2023
- [14] website: <http://dublincore.org/metadata-basics/> accessed: March 2023
- [15] website: <http://dublincore.org/documents/dces/> accessed March 2023
- [16] website: Zenodo - Data Repository, [Online]. Available: <https://zenodo.org/>
- [17] website: <https://www.openaire.eu/>
- [18] website: Open Archives Initiative, "Protocol for Metadata Harvesting - Tools," <http://www.openarchives.org/pmh/tools/> Accessed March 2023.
- [19] European IPR Helpdesk, Fact Sheet, Open Access to scientific publications and research data in Horizon 2020: Frequently Asked Questions (FAQs), July 2015,

available at: <https://www.horizon-europe.gouv.fr/sites/default/files/2021-07/h2020---ipr-helpdesk---open-access-3953.pdf>

[20] website: <https://open-research-europe.ec.europa.eu/> accessed: March 2023

[21] Alfresco Repository (restricted access) available at: <https://tecbox.etra-id.com/share/page/site/heu-cl5-2021-d3-02-07r2d2/dashboard>

[22] website: <http://r2d2project.eu/> (under construction so far)

[23] website: <https://github.com/> accessed March 2023

10.2 ACRONYMS

Table 4: Acronyms

Acronym	Description
DCAT	Data Catalogue Vocabulary
DCMI	Dublin Core Metadata Initiative
DMP	Data Management Plan
DoA	Description of Action
DOI	Digital Object Identifier
EC	European Commission
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
IT	Information Technology
ORE	Open Research Europe
OT	Operation Technology
WP	Work Package



**Funded by
the European Union**

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. Horizon Europe Grant agreement N° 101075714.