



<b>Title:</b>	<b>Document Version:</b>
D1.2 Data Management Plan	1.0

<b>Project Number:</b>	<b>Project Acronym:</b>	<b>Project Title:</b>
H2020- 863927	X-Flex	LC-SC3-ES-1-2019 Flexibility and retail market options for the distribution grid

<b>Contractual Delivery Date:</b>	<b>Actual Delivery Date:</b>	<b>Deliverable Type*-Security*:</b>
M6 (March 2020)	M6 (March 2020)	R-ORDP

\*Type: P: Prototype; R: Report; D: Demonstrator; O: Other.

\*\*Security Class: PU: Public; PP: Restricted to other programme participants (including the Commission); RE: Restricted to a group defined by the consortium (including the Commission); CO: Confidential, only for members of the consortium (including the Commission).

<b>Responsible:</b>	<b>Organisation:</b>	<b>Contributing WP:</b>
Manuel Serrano	ETRA	WP1

<b>Authors (organisation):</b>
Manuel Serrano (ETRA), Esteban Pastor (ETRA)

<b>Abstract:</b>
The Data Management Plan (DMP) clarifies the handling of research data during and after the project. It includes data that will be collected, processed or generated during the project, methodology and standards will be applied, whether data will be shared and how data will be curated and preserved, taking into account all data-related aspects of the project.

<b>Keywords:</b>
Data Management Plan, DMP, FAIR, interoperable, open .



## Revision History

Revision	Date	Description	Author (Organisation)
V0.1	30.01.2020	New Document	Manuel Serrano (ETRA)
V0.2	13.01.2020	Document for review	Esteban Pastor (ETRA)
V1.0	25.03.2020	Final revision and approval	Manuel Serrano (ETRA)



This project has received funding from the  
European Union's Horizon 2020 Research and Innovation Programme  
under Grant Agreement № 863927.

More information available at <http://xflexproject.eu/>



## Copyright Statement

The work described in this document has been conducted within the X-Flex project. This document reflects only the X-Flex 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 X-Flex 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 X-Flex Consortium or the Partners detriment and are not to be disclosed externally without prior written consent from the X-Flex Partners.

Each X-Flex Partner may use this document in conformity with the X-Flex Consortium Grant Agreement provisions.



## 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 X-FLEX Consortium regarding the project research data.

This document will evolve during the development of the project, when the project progresses and when 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 is the first iteration in which we present the envisioned data management strategy and make 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.



## Index

<b>1 INTRODUCTION .....</b>	<b>7</b>
1.1 Purpose of the Document .....	7
1.2 Scope of the Document .....	7
1.3 Structure of the Document .....	7
<b>2 Data Summary.....</b>	<b>8</b>
2.1 Purpose of data management and relation to the project.....	8
2.2 Data set types, formats and standards .....	8
2.3 Re-use of data .....	9
2.4 Timetable for updates of the DMP.....	9
<b>3 FAIR data.....</b>	<b>9</b>
3.1 Making data findable, including provisions for metadata.....	10
3.2 Making data openly accessible .....	12
3.2.1 Open research data repository .....	13
3.2.2 X-FLEX internal repository.....	13
3.2.3 Open Access publications .....	14
3.3 Making data interoperable .....	16
3.4 Increase data re-use (through clarifying licences).....	16
3.5 Public deliverables .....	17
<b>4 Allocation of resources.....</b>	<b>17</b>
4.1 Responsibilities and decision making.....	17
<b>5 Data security .....</b>	<b>18</b>
<b>6 Ethical aspects.....</b>	<b>19</b>
6.1 Procedures to ensure data privacy.....	19
6.2 Ethical considerations .....	19
<b>7 REFERENCES AND ACRONYMS.....</b>	<b>20</b>
7.1 References .....	20
7.2 Acronyms .....	21
<b>8 ANNEX A. DATA MANAGEMENT PLAN SCOPE – SUMMARY TABLE .....</b>	<b>22</b>



## LIST OF FIGURES

Figure 1 – Research Data Management according to the FAIR principles .....	10
Figure 2 – Summary of Metadata Best Practices.....	12
Figure 3 – X-FLEX repository overview .....	14
Figure 4 – Example of folders and files in X-FLEX repository .....	14
Figure 5 – Open access to scientific publication and research data in the wider context of dissemination and exploitation. <i>Source H2020 Open Access portal</i> .....	15

## LIST OF TABLES

Table 1 – Interoperability Levels .....	11
Table 2 – Dublin Core Metadata Element Set – 15 Elements Overview .....	11
Table 3 – Data Management Plan Scope Summary .....	23



## 1 INTRODUCTION

### 1.1 PURPOSE OF THE DOCUMENT

The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used by the X-FLEX Consortium with regard to the project research data.

The DMP is not a fixed document; on the contrary, it will evolve during the lifespan of the project. This DMP will be a living document in which information will be available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur [1].

This first version of the DMP aims to outline how the X-FLEX project will try to make the research data findable, accessible, interoperable and reusable (FAIR) and therefore contribute to knowledge discovery and innovation. Although the first version submitted by month 6 of the project did not provide very detailed information on the specific data sets to be collected, generated and processed during the project, the objective was to lay the foundations for creating an effective data management strategy covering the complete research data life cycle.

### 1.2 SCOPE OF THE DOCUMENT

This Data Management Plan (DMP) has been prepared by taking into account the “Guidelines on FAIR Data Management in Horizon 2020 - Version 3.0. 26 July 2016 [1].

This report describes the data management plan for the project covering the complete research data life cycle. This data management plan, to be continuously updated, will be used by the X-FLEX consortium as a guideline when handling the research data during and after the end of the project. The previous version was the first iteration in which we present the envisioned data management strategy and make 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. This updated version of the DMP gets into more detail and describes 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 by the X-FLEX project.

### 1.3 STRUCTURE OF THE DOCUMENT

Following the Horizon 2020 FAIR Data Management plan template provided as Annex 1 of the EC guidelines, this document is organised in five main sections:

- Data Summary (Section 2)
- FAIR Data (Section 3)
- Allocation of resources (Section 4)
- Data security (Section 5)
- Ethical aspects (section 6)

Final sections include references, acronyms, and annexes containing results of the X-FLEX data management and ethical related documents.



## 2 Data Summary

### 2.1 PURPOSE OF DATA MANAGEMENT AND RELATION TO THE PROJECT

The X-FLEX project has as its main goal offering a set of integrated solutions that facilitate the optimum combination of decentralised flexibility assets, of both on the generation (DER) and on the demand side (V2G, power-to-heat/cold, batteries, demand response) enabling all parties, including final prosumers, to offer their flexibility in the local and wholesale market creating benefits to all the actors in the smart grid value chain.

X-FLEX does participate into the Open Research Data Pilot (ORD Pilot) and therefore Open access to research data is applicable (X-FLEX Grant Agreement 863927 [2] – Article 29.3). It will comply with European recommendations regarding Data Management Plans [1], providing clear procedure for findable, accessible, interoperable and re-usable (FAIR) data and updating the current document along the development of the project.

The purpose of the DMP is to provide an analysis of the main elements of the data management policy that will be used by the Consortium with regard to the project research data. The DMP reflects consortium data management policies, systems and procedures - which will be implemented and embedded into research procedures and regularly reviewed throughout the research cycle.

We will strive to make data open but cannot overrule limitations that partner institutions put on data that they contribute (as specified in their Background included in the X-FLEX Consortium Agreement- Attachment 1). Moreover, an ethical approach will be adopted and maintained throughout the fieldwork process. The responsible partners will assure that the EU standards regarding ethics and Data Management are fulfilled.

### 2.2 DATA SET TYPES, FORMATS AND STANDARDS

Throughout the duration of the project, X-FLEX will generate a variety of data. All of it has to 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 6, more specifically in task T6.1 “Standards and data models analysis”.

As indicated in the X-FLEX Description of Action (DoA), the deliverable D6.1 “Data standards and models” to be submitted in Month 18, will provide the description of the open, secure and flexible customer-centric architecture including details of the types and formats of data to be handled.

According to the project schedule, task 6.1 starts in Month 7 and lasts until Month 18. During this period, 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 would be shared, curated and preserved.

The starting point of these tasks is to list the types of data to be handled in the framework of X-FLEX 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.

### 2.3 RE-USE OF DATA

As explained in the DoA (Part B section 1.3.3), GINETTE will use and exploit some of the preliminary solutions and technologies developed in five H2020 projects: WISEGRID, COMPILE and INVADE projects. These three projects will be the basis of some of the X-FLEX Innovation activities, complementing and going beyond the work done in these projects, providing new and more advanced services.

Moreover, the project of course analyses available results from other research activities, publications, and further relevant information available. This information will be mainly used for internal project analysis and will be provided in respective project deliverables with appropriate references to origins of the gathered information. However, as the analysed information has not been created by X-FLEX, the project is not considering provision of these data as public data sets, because the X-FLEX project does not own the relevant information and results.

### 2.4 TIMETABLE FOR UPDATES OF THE DMP

As previously indicated, the Data Management Plan is a living document that will be updated over the course of the project whenever significant changes arise, such as new data, modification in consortium policies, changes in consortium composition, external factors, etc.

## 3 FAIR data

According to [3], the FAIR principles describe four key concepts in research data management. Data should be:

- **Findable** – Easy to find by both humans and computer systems and based on mandatory description of the metadata that allows the discovery of interesting datasets;
- **Accessible** – Long term storage so data can be easily accessed and/or downloaded with well-defined license and access conditions, whether at the level of metadata, or at the level of the actual data content;
- **Interoperable** – Ready to be combined with other datasets by humans, as well as computer systems;
- **Reusable** – Ready to be used for future research and to be processed further using computational methods.

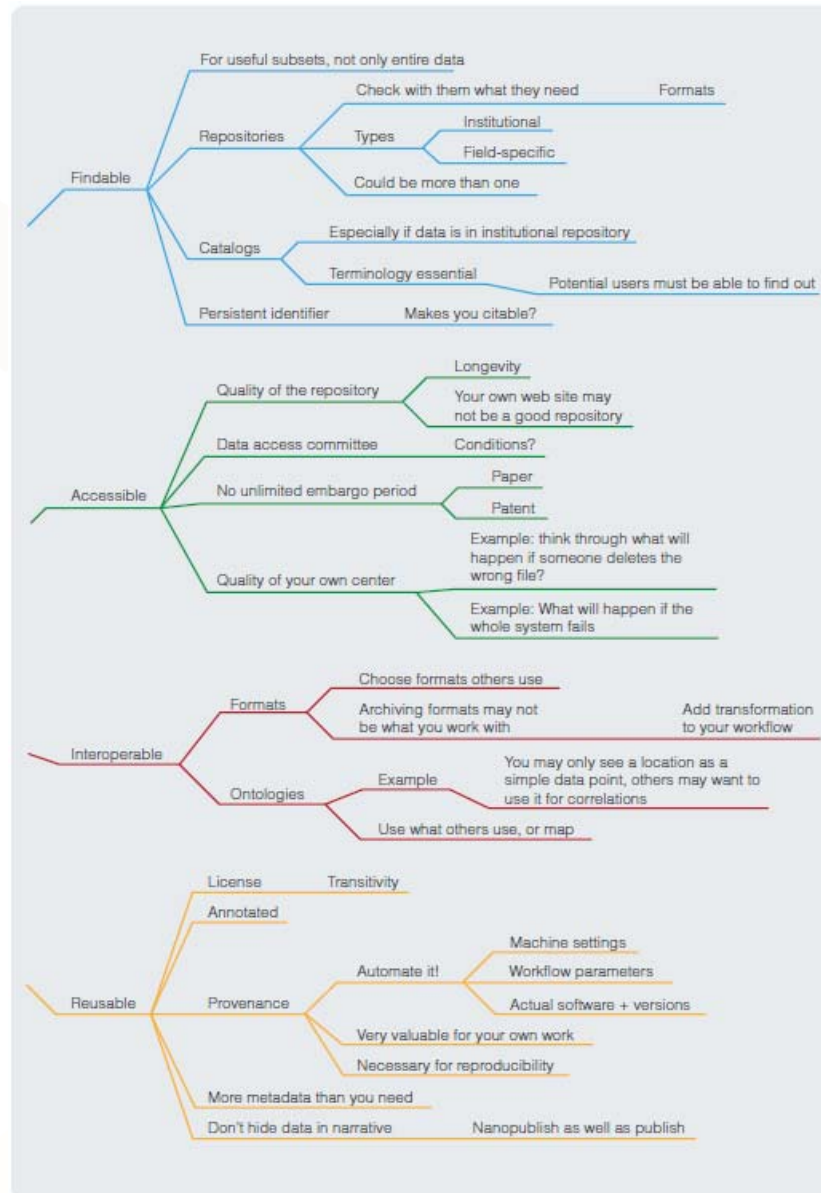


Figure 1 – Research Data Management according to the FAIR principles  
Source: Elsevier

### 3.1 MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA

Identification and localisation means will be used for the data to be processed during X-FLEX project.

The data to be generated in X-FLEX 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 - were already provided in D1.1 [4].

X-FLEX project research data will be inventoried and annotated with metadata following the standards promoted by the European Data Portal [5].

*“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."*

European Data Portal

The European Data Portal strongly recommends the use of the DCAT Application Profile for metadata [5, p. 47]. The Data Category Vocabulary (DCAT) – used for datasets on the internet, is based on the Dublin Core standards.

The Dublin Core metadata standard is a simple yet effective element set for describing a wide range of networked resources. From the perspective of the Dublin Core community, the metadata landscape is currently characterized in terms of four "levels" of interoperability and 15 sections for data description [6] [7]:

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 1 – Interoperability Levels**  
Source: DCMI

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

**Table 2 – Dublin Core Metadata Element Set – 15 Elements Overview**

The fifteen element "Dublin Core" described in this standard is 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:

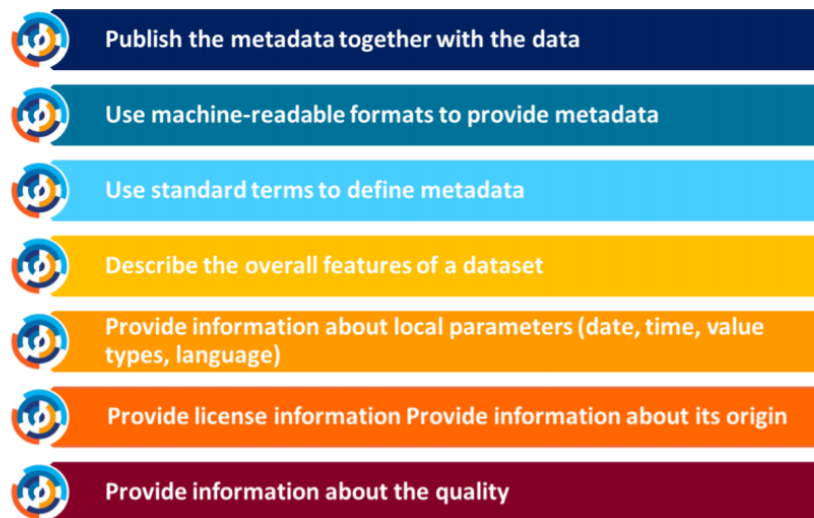
- ISO Standard 15836:2009 of February 2009 (confirmed in 2014)

- ANSI/NISO Standard Z39.85-2012 of February 2013
- IETF RFC 5013 of August 2007

The DCAT Application Profile is based on Dublin Core standards but is not a vocabulary. It is a simple specification for metadata descriptions of EU governmental data and portals.

The metadata will be published with the data using a machine-readable format and standard terms to define the metadata; the overall features of the dataset will be described with information about local parameters, licence, origin and quality.

The figure below summaries the best practices outlined by the European Data Portal which will be followed the X-FLEX partners when using the DCAT-AP.



**Figure 2 – Summary of Metadata Best Practices**  
*Source DCAT-AP*

### 3.2 MAKING DATA OPENLY ACCESSIBLE

An analysis of which X-FLEX research data will be made openly accessible and which data will be kept closed will be done in a later stage of the project. As indicated in section 2.2, 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 and this activity will be done in the framework of the ongoing tasks in WP6. 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 updates 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, X-FLEX 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 X-FLEX,

certain research data – especially considering the pilot clusters – is sensitive due to security issues and therefore will be kept confidential.

### 3.2.1 Open research data repository

For implementing open access to the research data sets, X-FLEX will use Zenodo, an established online European scientific repository that is fully integrated with OpenAIRE. The decision on whether a research data set will be uploaded to and opened for access in Zenodo will be made on a case-by-case basis between the coordinator (ETRA), the technical manager (ICCS) and the partner(s) that have ownership of the data.

When a data package study has been marked as public, it will be made openly available. Data gathered by partners outside of the project work plan and protected by IPR, or inside the work plan but containing confidential information (e.g. related to personal data), will be kept closed for privacy reasons.

#### Zenodo

Zenodo [8] (<https://zenodo.org/>) offers a simple online service that enables researchers, scientists, EU projects and institutions to share, preserve and showcase multidisciplinary research results (data and publications), that are not part of the existing institutional or subject-based repositories of the research communities. It provides service hosting according to industry best practices in CERN's professional data centres. A detailed description of Zenodo's policies regarding the handling of the data and usage of the service is found at <https://zenodo.org/policies>



As previously indicated, X-FLEX intends to share datasets in the publicly accessible disciplinary repository Zenodo using descriptive metadata as required/provided by that repository. Zenodo assigns all publicly available uploads a Digital Object Identifier (DOI) to make the upload easily and uniquely citeable. (It is NOT possible to edit a Zenodo DOI once it has been registered). Zenodo further supports harvesting of all content via the OAI-PMH protocol<sup>1</sup>.

A digital object identifier (DOI) is a character string (a "digital identifier") used to uniquely identify an object such as an electronic document. Metadata about the object is stored in association with the DOI name and this metadata may include a location, such as a URL, where the object can be found. The DOI for a document remains fixed over the lifetime of the document, whereas its location and other metadata may change. Referring to an online document by its DOI provides more stable linking than simply referring to it by its URL, because if its URL changes, the publisher need only update the metadata for the DOI to link to the new URL. A DOI name differs from standard identifier registries such as the ISBN (ISO 2108:2005) and ISRC (ISO 3901:2001). The purpose of an identifier registry is to manage a given collection of identifiers, whereas the primary purpose of the DOI system is to make a collection of identifiers actionable and interoperable.

### 3.2.2 X-FLEX internal repository

During the life cycle of X-FLEX, data collected or generated by the project will be stored and systematically organised in the official project repository on ALFRESCO.

Alfresco [9] is a flexible content management web application developed using Java technology. In the framework of the X-FLEX project, it is used mainly as a repository to securely store and share files, making data available to the whole Consortium. The repository (documents section) consists of a project internal area, not possible to be accessed by external users.

---

<sup>1</sup> Open Archive Initiative - Protocol for Metadata Harvesting (OAI-PMH)

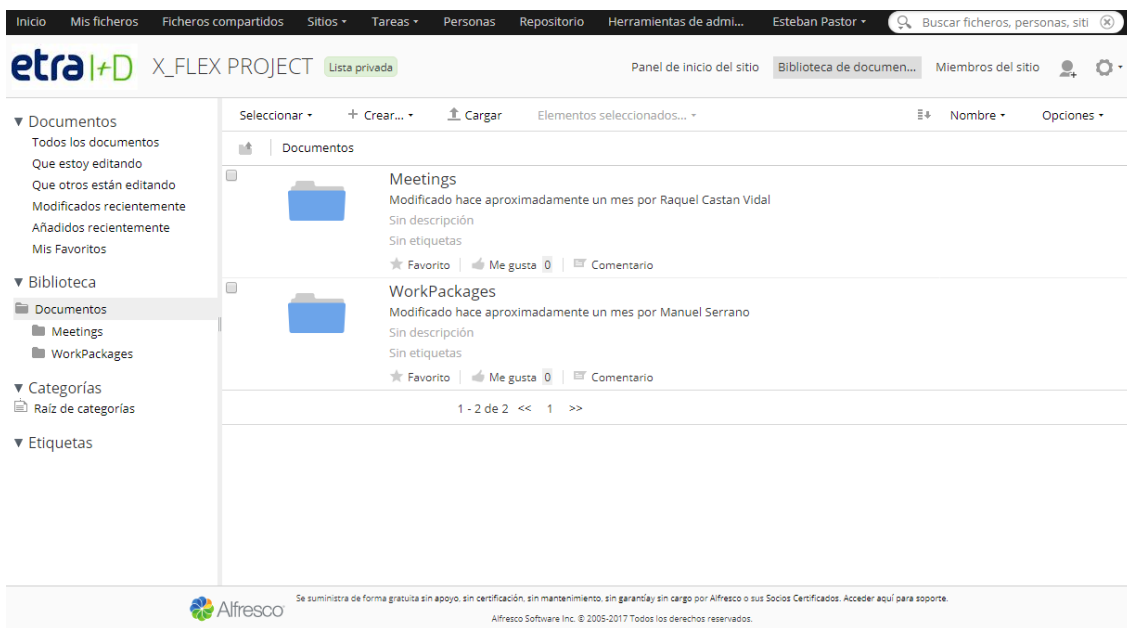


Figure 3 – X-FLEX repository overview

As can be seen in the snapshots below, folders are organized in a hierarchical and clear structure and files are uniquely identifiable and versioned by using a name convention.

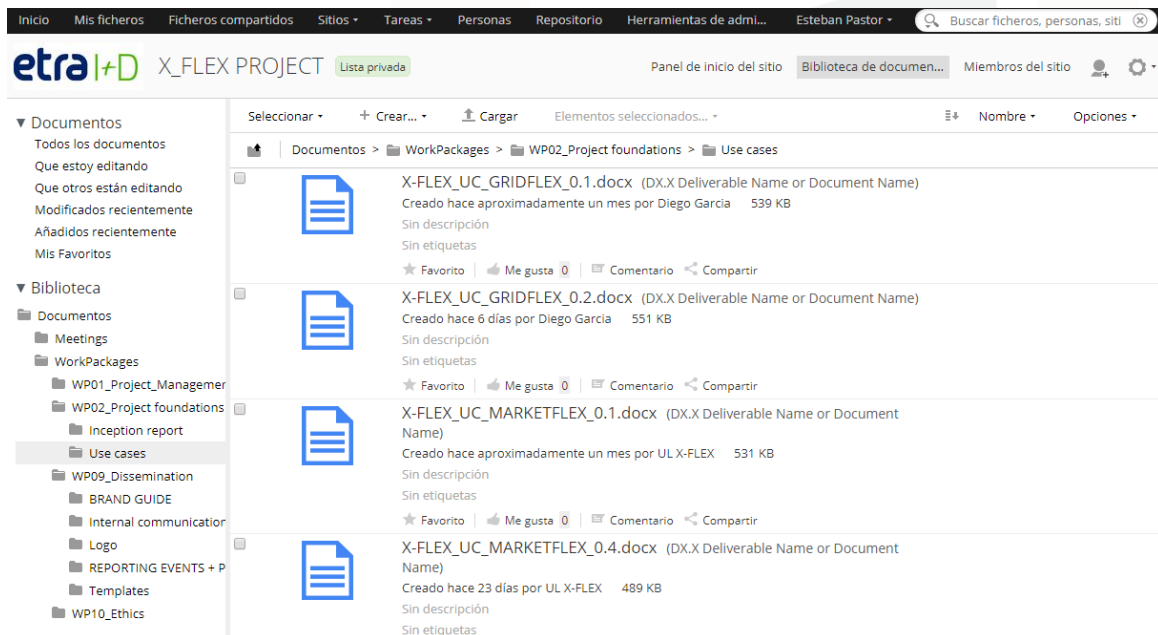


Figure 4 – Example of folders and files in X-FLEX repository

### 3.2.3 Open Access publications

The rules and principles of the European Commission’s Horizon 2020 Framework Programme clearly establish that scientific results generated within H2020 projects will be made available as open access publications, i.e. freely available online to any user. Following these rules, enforced in the X-FLEX Grant Agreement – Article 29.2, open access will be ensured to all peer-reviewed scientific publications related to X-FLEX and its



composite solutions. Furthermore, other project outputs and results may be provided in the Open Access upon respective decision of the project consortium.

According to the “Open access to publications and data in Horizon 2020 Fact sheet” [10], two main routes exist for open access to scientific peer-reviewed publications:

- Self-archiving (also called ‘Green’ open access) means that the published article or the final peer-reviewed manuscript is archived by the researcher in an online repository before, after or alongside its publication. Access to the article is often – but not necessarily - delayed (‘embargo period’ of six months of publication) as some scientific publishers may wish to recoup their investment by selling subscriptions and charging pay-per-download view fees during an exclusivity period
- Open access publishing (also called ‘Gold’ open access) means that an article is immediately provided in open access mode by the scientific publisher. The associated costs are shifted away from readers, and instead charged to the research institute to which the researcher is affiliated, or to the funding agency supporting the research

Since these two routes described above are not mutually exclusive in an EU-funded action, within the X-FLEX project each beneficiary will be able to choose the most suitable approach for each publication concerned. In any case, this specific aspect will be further discussed and described in an updated version of the Data Management Plan

For the moment, parallel publishing has been chosen as the primary strategy for providing open access. This allows consortium members to publish their results in the scientific fora and journals of their choice for maximum impact and still ensure optimal dissemination of the results by open access. It is worth to mention that the overwhelming majority of academic journals support either the gold, the green or a hybrid open access route. It means that X-FLEX beneficiaries have freedom to publish where they feel it is the most appropriate. All publications will have a DOI making them easily findable and citable.

It is important to note that the open access requirement does not mean an obligation to publish results. As stated by the guidelines: “The decision to publish is entirely up to the grant beneficiaries. Open access becomes an issue only if publication is chosen as a means of dissemination”. The graph below illustrates the decision related to research results and possible path for publication or other options.

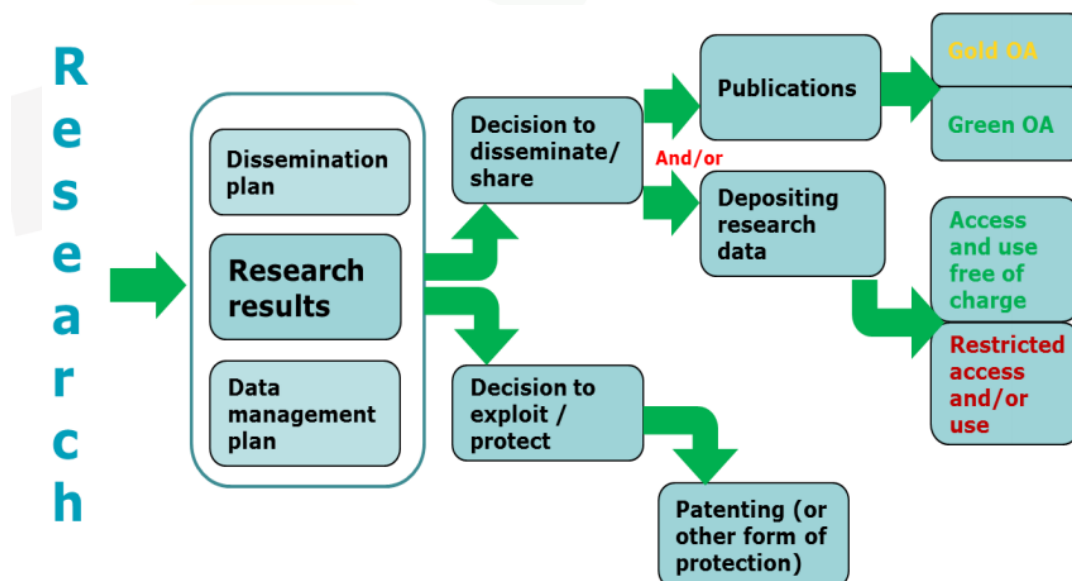


Figure 5 – Open access to scientific publication and research data in the wider context of dissemination and exploitation. Source H2020 Open Access portal



### 3.3 MAKING DATA INTEROPERABLE

An assessment of the data interoperability, specifying what data and metadata vocabularies, standards or methodologies will be followed in order to facilitate interoperability will be carried out in a later stage of the project. The assessment will address whether standard vocabulary will be used for all data types present in the data set in order to allow inter-disciplinary interoperability. Again, for the development of this task, 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 and the definition of the components or actors-components that will be communicating in the scope of the X-FLEX project (this is one of the outputs of the architecture).

Concerning standards and interoperable data models within the project (WP3); 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 X-FLEX 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 X-FLEX project;
- For each interface, assess the most appropriate standards and best suitable data models;
- Include the aforementioned process in WP3 deliverables.

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

### 3.4 INCREASE DATA RE-USE (THROUGH CLARIFYING LICENCES)

As previously indicated, data will be treated on a case study basis during the project. Once a data set is marked as public, and, therefore, made publicly available on Zenodo, it will be fully reusable (with the possibility of specifying embargo period or with controlled access to whitelist of persons; see Zenodo policies in [11]).

When possible, and as recommended by European Commission's guidelines [12], data will be made available with Creative Commons Licences (CC BY or CC0). The data sets may be given different licenses according to their specificities<sup>2</sup>.

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 assure 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

---

<sup>2</sup> The EUDAT B2SHARE tool includes a built-in license wizard that facilitates the selection of an adequate license for research data (<https://eudat.eu/services/userdoc/license-selector>)





### 3.5 PUBLIC DELIVERABLES

The public project deliverables will be available for download on the website after their submission and approval from the EC. The X-FLEX website will have a private section locked for users that do not have login details. Confidential deliverables will be stored in this private section only accessible for restricted users. Confidential deliverables might be requested by external parties, in which case the Consortium might make decision to disseminate corresponding deliverables or specific parts of the deliverables to particular external parties. The project deliverables on the website will be provided in widely adopted PDF format.

## 4 Allocation of resources

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

### 4.1 RESPONSIBILITIES AND DECISION MAKING

As indicated in previous sections, the Data Management Plan presented in this deliverable is just the first version, and the related Consortium discussions will be continuously carried out, to identify the relevant project outputs as well as to decide on way and means of their open access (if applicable). To ensure it, a dedicated time slot will be reserved at each project plenary meetings and, if needed, at selected Consortium audio conferences. EC and 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 – to prepare and lead related discussions at the relevant project meetings and to maintain the project document repository ALFRESCO;
- Scientific and Technical Project Manager – to identify data collected by the project and technical project outcomes eventually suitable for publication;
- Dissemination (WP18) Leader – to identify publications suitable for publication in the considered repositories and maintain X-FLEX 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.

Moreover, each X-FLEX partner has to respect the policies set out in this DMP. Datasets have to be created, managed and stored appropriately and in line with applicable legislation.

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

UNEW, Scientific and Technical Project Manager, will ensure dataset integrity and compatibility for its use during the project lifetime by different partners.

Validation and registration of datasets and metadata is the responsibility of the partner that generates the data in the WP. Metadata constitutes an underlying definition or description of the datasets, and facilitate finding and working with particular instances of data.

Backing up data for sharing through open access repositories is the responsibility of the partner possessing the data.



Quality control of these data is the responsibility of the relevant WP Leader, supported by the Project Coordinator.

If datasets are updated, the partner that possesses the data has the responsibility to manage the different versions and to make sure that the latest version is available in the case of publicly available data.

Last but not least, all partners must consult the concerned partner(s) before publishing data in the open domain that can be associated to an exploitable result.

## 5 Data security

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

Zenodo provides clear security guaranties. All data files are stored in CERN Data Centres, primarily 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 [11].

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 privacy, all data will be anonymised, encrypted and stored on servers to which only the relevant 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 personnel 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 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 virtually 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;
- 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.



## 6 Ethical aspects

### 6.1 PROCEDURES TO ENSURE DATA PRIVACY

The project is aware of the importance to keep privacy and protect the personal data, so the following standards are established in relation to personal data:

- will not be handled out to third parties outside X-FLEX;
- will not be exploited or commercialized;
- will be kept for no longer than necessary;
- will not be accessible for use or diffusion outside the project framework;
- will be subject to retrieval in case it is requested;
- will be destroyed as the relevant scientific purpose is fulfilled.

Engagement with end users and stakeholders is mostly taking place in Work Packages 1, 2 and 7; being the last one in which demonstration activities are carried out. Information managed by the project consortium during such activities may be of a private or confidential nature.

Starting from these considerations, some procedures must be adopted to ensure that the privacy of the involved end-users is safeguarded. Access to sensitive information is being carefully controlled with restriction policies (where appropriate), and anonymization techniques are being applied to protect data confidentiality.

### 6.2 ETHICAL CONSIDERATIONS

- Ethical standards and guidelines will be rigorously applied, regardless of the country in which the research is carried out.
- Participants in the pilot activities must receive introductory descriptions about the X-FLEX project and the purpose of pilot demonstrations and studies.

In case of specific user studies:

- The purpose and procedure of the research are being introduced in an understandable way.
- It is being emphasized that it is the potential participants' choice whether or not to participate in the study.
- All participants are being informed of their right to privacy and the extent to which participation in this research may impact on their lives – and the mechanisms the researchers have put in place to protect participant privacy through processes of anonymization and data storage and security.
- Participants are being informed about duration and effort to participate in any research.
- In any survey/interview people are being informed what kinds of questions we plan to ask, and we will make it clear that people can choose not to answer questions.
- Participants are being made aware of their 'withdrawal rights': that they can withdraw from the research at any time and that, if they wish, any personal data, recordings or images can be destroyed.
- Contact information to the project's stakeholders is being provided.
- Risks and benefits are being explained.
- If applicable, arrangements for insurance coverage for participation are being described.
- Participants are being made aware of the complaints procedure.



## 7 REFERENCES AND ACRONYMS

### 7.1 REFERENCES

- [1] “Guidelines on FAIR Data Management in Horizon 2020,” [Online]. Available: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf).
- [2] X-FLEX, *Grant Agreement number: 863927 — X-FLEX*, 2019.
- [3] «Research Data Management plans. Elsevier.,» [En línea]. Available: [https://www.publishingcampus.elsevier.com/websites/elsevier\\_publishingcampus/files/Guides/Research%20Data%20Management%20Plan.pdf](https://www.publishingcampus.elsevier.com/websites/elsevier_publishingcampus/files/Guides/Research%20Data%20Management%20Plan.pdf).
- [4] X-FLEX, «D1.1 Project Management Plan,» March 2020.
- [5] European Data Portal, «Open Data Goldbook for Data Managers and Data Holders,» [En línea]. Available: [https://www.europeandataportal.eu/sites/default/files/european\\_data\\_portal\\_-\\_open\\_data\\_goldbook.pdf/.](https://www.europeandataportal.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf/)
- [6] “Dublin Core Metadata Initiative,” [Online]. Available: <http://dublincore.org/metadata-basics/>.
- [7] D. C. M. E. Set. [Online]. Available: <http://dublincore.org/documents/dces/>.
- [8] Z. -. D. Repository. [Online]. Available: <https://zenodo.org/>.
- [9] ALFRESCO, [En línea]. Available: <https://www.alfresco.com/>.
- [10] “Fact sheet open access to publications and data in Horizon 2020 - Fact sheet,” [Online]. Available: [https://www.iprhelphdesk.eu/sites/default/files/newsdocuments/Open\\_Access\\_in\\_H2020.pdf](https://www.iprhelphdesk.eu/sites/default/files/newsdocuments/Open_Access_in_H2020.pdf).
- [11] “Zenodo policies,” [Online]. Available: <https://www.zenodo.org/policies>.
- [12] EC, «H2020 Programme Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020,» [En línea]. Available: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2).



## 7.2 ACRONYMS

Acronyms List	
CERN	European Organization for Nuclear Research
D	Deliverable
DCMI	Dublin Core Metadata Initiative
DMP	Data Management Plan
DoA	Description of Action
DOI	Digital Object Identifier
DPIA	Data Protection Impact Assessment
EC	European Commission
FAIR	Findable, Accessible, Interoperable and Reusable (FAIR).
H2020	Horizon 2020. The EU Framework Programme for Research and Innovation
IPR	Intellectual Property Rights
ISBN	International Standard Book Number
ISRC	International Standard Recording Code
OAI-PMH	Open Archive Initiative-Protocol for Metadata Harvesting
ORD	Open Research Data
SP	Sub-Project
TSO	Transmission System Operator
URL	Uniform Resource Locator
WP	Work Package
OAI-PMH	Open Archive Initiative-Protocol for Metadata Harvesting
ORD	Open Research Data

## 8 ANNEX A. DATA MANAGEMENT PLAN SCOPE – SUMMARY TABLE

This table provides a summary of the Data Management Plan (DMP) issues to be addressed, as outlined in the Horizon 2020 Data Management Plan Template<sup>3</sup>.

DMP component	Issues to be addressed
<b>1. Data summary</b>	<ul style="list-style-type: none"> <li>• State the purpose of the data collection/generation</li> <li>• Explain the relation to the objectives of the project</li> <li>• Specify the types and formats of data generated/collected</li> <li>• Specify if existing data is being re-used (if any)</li> <li>• Specify the origin of the data</li> <li>• State the expected size of the data (if known)</li> <li>• Outline the data utility: to whom will it be useful</li> </ul>
<b>2. FAIR Data</b>	<ul style="list-style-type: none"> <li>• Outline the discoverability of data (metadata provision)</li> </ul>
2.1. Making data findable, including provisions for metadata	<ul style="list-style-type: none"> <li>• Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?</li> <li>• Outline naming conventions used</li> <li>• Outline the approach towards search keyword</li> <li>• Outline the approach for clear versioning</li> <li>• Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how</li> </ul>
2.2 Making data openly accessible	<ul style="list-style-type: none"> <li>• Specify which data will be made openly available? If some data is kept closed provide rationale for doing so</li> <li>• Specify how the data will be made available</li> <li>• Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?</li> <li>• Specify where the data and associated metadata, documentation and code are deposited</li> <li>• Specify how access will be provided in case there are any restrictions</li> </ul>
2.3. Making data interoperable	<ul style="list-style-type: none"> <li>• Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.</li> </ul>

<sup>3</sup> [http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan\\_en.docx](http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx)



---

	<ul style="list-style-type: none"><li>Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?</li></ul>
2.4. Increase data re-use (through clarifying licences)	<ul style="list-style-type: none"><li>Specify how the data will be licenced to permit the widest reuse possible</li><li>Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed</li><li>Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why</li><li>Describe data quality assurance processes</li><li>Specify the length of time for which the data will remain re-usable</li></ul>
3. Allocation of resources	<ul style="list-style-type: none"><li>Estimate the costs for making your data FAIR. Describe how you intend to cover these costs</li><li>Clearly identify responsibilities for data management in your project</li><li>Describe costs and potential value of long term preservation</li></ul>
4. Data security	<ul style="list-style-type: none"><li>Address data recovery as well as secure storage and transfer of sensitive data</li></ul>
5. Ethical aspects	<ul style="list-style-type: none"><li>To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former</li></ul>
6. Other	<ul style="list-style-type: none"><li>Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)</li></ul>

---

**Table 3 – Data Management Plan Scope Summary**