# D1.2- Data Management plan

| | |
|---|---|
| **Project number** | **101160671** |
| **Project acronym:** | **DIGITISE** |
| **Project title:** | **Digital Innovative cross-sector services for Greater citizen Integration in a just energy TransItion, and Societal Empowerment** |

| | |
|---|---|
| **Work Package** | WP1: Project Management and Consortium Coordination |
| **Responsible Partner** | UBITECH |
| **Official Submission Date** | 30/11/2024 |
| **Actual Submission Date** | 28/11/2024 |
| **Type** | DMP |
| **Dissemination Level** | Public |
| **Reviewers** | MYTIL, MP |
| **Version** | V1.0 |

Funded by the European Union

## Versioning and contribution history

| Version | Date | Author(s) | Notes |
|---|---|---|---|
| V0.1 | 09/09/2024 | UBITECH | ToC |
| V0.2 | 24/09/2024 | UBITECH | Section 1, 2, 4 |
| V0.3 | 29/10/2024 | UBITECH | Section 5, 6,7 |
| V0.5 | 15/11/2024 | UBITECH | Section 3 |
| V0.6 | 20/11/2024 | UBITECH | Conclusions, Executive Summary |
| V0.7 | 21/11/2024 | MP, MYTIL | Peer review version |
| V1.0 | 28/11/2024 | UBITECH | Final Version |



***Preface***

Funded by the European Commission under Grant Agreement number 101160671, DIGITISE is a project focused on enhancing the digital literacy and empowerment of consumers and prosumers in the energy sector. By integrating advanced technologies and fostering active engagement in digital energy activities and markets, DIGITISE aims to play a crucial role in the global energy transition.

# Executive Summary

The following document represents the Data Management Plan for DIGITISE project. Any further occurring updates in this respect will be provided during the project's duration, based on the actual development of the technical work. More specifically, in line with the F.A.I.R Principles, this delivery provides guidelines on how to make the research data collected and/or generated throughout and after the project duration Findable, Accessible, Interoperable and Re-usable. The DMP is, thus, a key document presenting the data management practices to be employed by the consortium partners, describing - among other- standards and the methodology that shall be followed for data collection and generation, and whether as well as how such data will be shared. Due to this, all partners were requested to provide input with respect to the artefacts that they will produce/provide in the DIGITISE project and the level of dissemination (Publishable / Non-Publishable).

Funded by the European Union

# Table of Contents

Executive Summary ..... 3
1. Introduction ..... 7
1.1 Scope ..... 7
1.2 Dependencies ..... 7
2. Data Management in Horizon Europe ..... 8
3. DIGITISE Data Management Overview ..... 11
3.1 Types and Formats of Artefacts Generated/Collected ..... 11
3.2 DIGITISE Artefacts and Access Rights ..... 11
3.3 Expected size of the data (if known) ..... 14
4. DIGITISE ORDP Participation ..... 15
4.1 Publishing Infrastructure for Open Access ..... 16
4.1.1 Publishing Process ..... 17
4.1.2 Publishing Platforms ..... 18
4.1.3 Access and Sharing ..... 19
5. FAIR Data ..... 21
5.1 Making data findable, including provisions for metadata ..... 21
5.1.1 Discoverability of the data ..... 22
5.1.2 Data identification mechanism ..... 22
5.1.3 Naming conventions used ..... 22
5.1.4 Clear versioning of the documents ..... 23
5.1.5 Standards for the metadata creation (if any) ..... 23
5.2 Making data openly accessible ..... 25
5.2.1 Methods or software needed to access the data ..... 26
5.2.2 Deposit of data, associated metadata, documentation and code ..... 26
5.3 Making data interoperable ..... 27
5.3.1 Interoperability of data assessment ..... 27
5.3.2 Interoperability of data assessment ..... 27
5.4 Making data re-usable ..... 28
5.4.1 Increase data re-use through clarifying licenses ..... 28

Funded by the European Union

5.4.2 Data quality assurance process ..... 28
5.4.3 Length of time for which the data will remain re-usable ..... 29
6. Allocation of resources ..... 30
6.1 Data management responsibilities ..... 30
6.2 Cost of potential value of long-term preservation ..... 30
7. Data Security ..... 31
Conclusions ..... 32
Annex – Artefact Template ..... 33

## List of Figures

Figure 1: DIGITISE Artefacts' Types ..... 12
Figure 2: Research Items Access rights ..... 15
Figure 3: Software Artefacts Access rights ..... 16
Figure 4: Dataset Artefacts Access rights ..... 16
Figure 5: Template to be used for project documentation metadata overview. ..... 25
Figure 6: Open access to scientific publication and research data in the wider context of dissemination and exploitation ..... 26
Figure 7: DIGITISE Repo Folders Structure ..... 27

## List of Tables

Table 1: Clarifications of Terms ..... 9
Table 2: Artefacts Overview ..... 11
Table 3: Partners' Research Item provision ..... 12
Table 4: Partners' Software provision ..... 13
Table 5: Partners' Dataset provision ..... 13
Table 6: Proposed Document History Table overview ..... 23
Table 7: Document History Template - Example ..... 23
Table 8: Metadata template for DIGITISE datasets. ..... 23
Table 9: Making Data Findable Template ..... 33
Table 10: Making data Accessible Template ..... 33
Table 11: Making data Interoperable Template ..... 34
Table 12: Making data Re-usable Template ..... 34

Funded by the European Union

# Abbreviations

| Abbreviation | Full Name |
|---|---|
| **DMP** | Data Management Plan |
| **EC** | European Commission |
| **FAIR** | Findable, Accessible, Interoperable, Reusable. |
| **ORDP** | Open Research Data Pilot |
| **WP** | Work Package |

# 1. Introduction

## 1.1 Scope

This deliverable presents the DIGITISE data management plan, as captured in M6 of the project. In accordance with the European Commission's Guidelines for Horizon Europe (HE) Programme, to submit a Data Management Plan (DMP) within the first six (6) months of the project, the present report forms the DMP of DIGITISE project reflecting the technical progress now of the drafting of the present document. Further updates in this respect will be provided during the project duration, based on the actual developments of the technical work. More specifically, in line with the F.A.I.R Principles, the deliverable provides for how making the research data collected and/or generated throughout and after the project duration Findable, Accessible, Interoperable and Re-usable. To this end, the deliverable outlines -among other- how the research data collected and/or generated will be handled during and after the DIGITISE project, describe which standards and methodology for data collection and generation will be followed, and whether and how data will be shared. Note that the document is largely based on the related template provided by the European Commission (EC).

This DMP outlines how data collected or generated by the DIGITISE project will be organised, stored, and shared. It specifies 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.

The present report forms a deliverable -primarily- addressed to:

- European Commission
- Partners and Advisory Group in the DIGITISE project
- EU Parliament
- Horizon Europe projects and other cyber/digital security related projects (clustering activities)
- Organizations and experts involved in DIGITISE case studies.
- Other relevant organizations, both public and private, including associations of relevant stakeholders.

## 1.2 Dependencies

The delivery of the present document falls under Work Package 1 activities and specifically Task 1.4, which extends to M36 of DIGITISE project. In the context of the related activities, it is, thus, intended that the DMP is a living document, subject to updates -to the extent necessary- based on the progress of the project activities. Also, DMP is strongly related to all WPs of DIGITISE project since it supports the data management life cycle for all research data that will be collected, processed, or generated within the project.

Funded by the European Union

# 2. Data Management in Horizon Europe

According to the EC, DMPs are a cornerstone for responsible management of research outputs, notably data and are mandatory in Horizon Europe for projects generating and/or reusing data.

The DMP is defined as:

"Data Management Plans (DMPs) are a key element of good data management. A DMP outlines the data management lifecycle for the data to be collected, processed, and/or generated during a Horizon Europe project. It specifies how data will be handled, documented, stored, and shared, ensuring alignment with the principles of open science, FAIR (Findable, Accessible, Interoperable, Reusable) data management, and compliance with applicable ethical, legal, and regulatory requirements. All Horizon Europe projects generating or collecting research data are required to submit a DMP, typically as a deliverable within the first six months of the project, with updates provided as necessary throughout its lifecycle."

The purpose of a DMP is to provide a discussion of the main elements of the data management policy that will be used by the applicants regarding all the datasets that will be generated by the project.
Overall, having considered all relevant principles regarding lawful processing of personal data, scientific research data should be easily discoverable, accessible, assessable, and intelligible, useable beyond the original purpose for which it was collected and interoperable to specific quality standards.
The DIGITISE Data Management also follows the Horizon Europe Data Management Plan Template, released by the European Commission Directorate – General for Research & Innovation. This Horizon DMP template has been designed to be applicable to any Horizon Europe project that produces, collects, or processes research data. According to these guidelines the management and organization of data should be based on four basic principles, which determine how research outputs should be processed so that they can be more easily accessed, understood, exchanged, and reused. This means that data must be findable, accessible, interoperable, and re-useable, for example by researchers interested in using the data in further research in the field.
These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation-solution. EC provides a Template with the FAIR principle. This template is not intended as a strict technical implementation of the FAIR principles, it is rather inspired by FAIR as a general concept. The template represents the set of questions that someone should answer with a level of detail appropriate to the project.

TABLE 1: CLARIFICATIONS OF TERMS

| | |
|---|---|
| **Research data** | Research data is the evidence that underpins all research conclusions (except those which are purely theoretical) and includes data that have been collected, observed, generated, created or obtained from commercial, government or other sources, for subsequent analysis and synthesis to produce original research results. These results are then used to produce research papers and submitted for publication. |
| **Open research data** | Openly accessible research data can typically be accessed, mined, exploited, reproduced and disseminated, free of charge for the user. |
| **Secondary data** | Secondary data are data that already exist, regardless of the research to be conducted. |
| **Open access** | Open access is understood as the principle that research data should be accessible to relevant users, on equal terms, and at the lowest possible cost. Access should be easy, user-friendly and, if possible, Internet-based. |
| **Metadata** | Metadata is data used to describe other data. It summarizes basic information about data, which can make finding and working with instances of data easier. |
| **Research data repositories** | Research data repositories are online archives for research data. They can be subject based/thematic, institutional or centralized. |

It is possible to develop a single DMP for any project to cover the overall approach. However, where there are specific issues for individual datasets (e.g. regarding openness), someone should clearly spell this out.
The template proposes the following issues to be addressed:

- Data Summary
- FAIR data
- Allocation of resources
- Data security
- Ethical aspects
- Other issues
- Further support in developing your DMP.

Each of the previously defined has its own set of questions that must be addressed. The proposed template states that it is not required to provide detailed answers to all the questions of the DMP that need to be submitted by month 6 of the project, subject -also- to potential future updates. Rather, the DMP is intended to be a living document -to the

extent necessary- in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur.

# 3. DIGITISE Data Management Overview

As described in the Guidelines on FAIR Data Management in Horizon Europe a Data Management Plan is a key element to ensure data is well managed. For this reason, in this section we will first identify the type of artefacts that will be generated and collected in the framework of the project. During the lifetime of the DIGITISE project, several artefacts will be produced. The artefacts that will be collected/generated are listed below in Section 3.2. As the project evolves, this list may require modifications (addition or removal of artefacts) with respect to the project developments.

## 3.1 Types and Formats of Artefacts Generated/Collected

To provide an overview of the different data sets that are currently and will be produced in the DIGITISE project, the following table shows the data type, the related WP number and the format, in which the data will be presumably stored.

TABLE 2: ARTEFACTS OVERVIEW

| # | Artefact type | Explanation | WP# | Format (indicative) |
|---|---|---|---|---|
| 1 | Research Item | Deliverables, Papers | 2-5 | .xls, .csv, .txt, .docx, .pdf |
| 2 | Software | Code, algorithms, dashboard | 3-5 | .xls, .csv, .txt, .docx, .pdf |
| 3 | Dataset | Consumer consumption Data, Metering data, Contract Related data, Demand/consumption data of users, energy generation data, Flexibility related and energy costs data | 3-6 | .xls, .csv, .txt, .docx, .pdf |

## 3.2 DIGITISE Artefacts and Access Rights

In the survey conducted during the first months of the project, the input collected from most of the partners is depicted in the following chart. The types of DIGITISE Artefacts are distributed as **53%** being datasets, **16%** research items and the remaining **31%** is of a software artefact type.

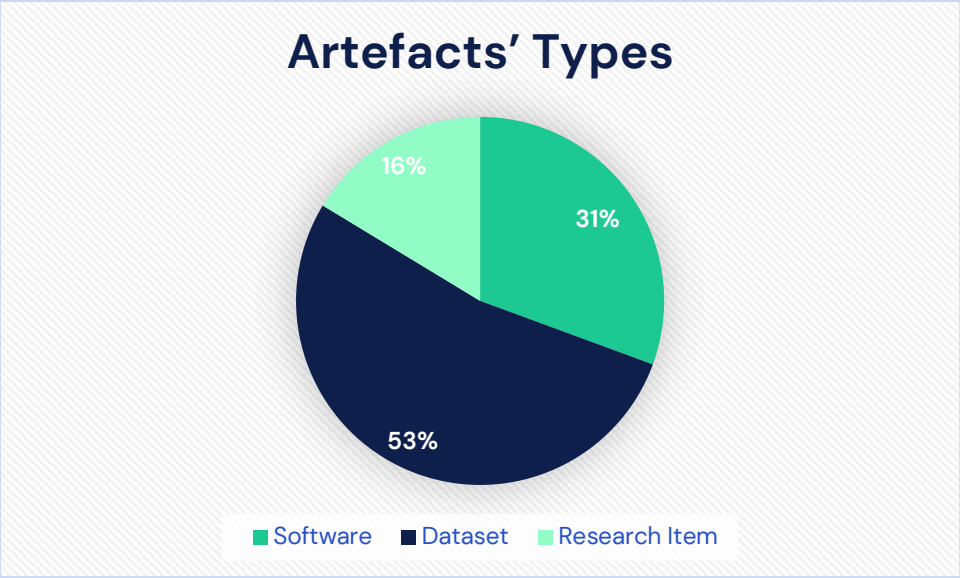



FIGURE 1: DIGITISE ARTEFACTS' TYPES

The following tables present the status and consensus within the Consortium with regards to identified artefacts and their access rights. It is provisioned that those tables are a recurring exercise, and all future updates and additions will be documented under the relevant deliverables.

TABLE 3: PARTNERS' RESEARCH ITEM PROVISION

| Partner | Artefact | Publishable/ non-publishable |
|---|---|---|
| **APPART** | Questionnaires for T2.1 (Prosumers-Consumers, Aggregators, ICT) (Deliverable D2.1) | Publishable |
| **APPART/MP** | DIGITISE End-User Requirements and Socio-Economic Analysis (Deliverable D2.1) | Publishable |
| **UBITECH** | 1st Release of DIGITISE DT Solutions and Tools for Consumers Engagement in Marketplace Environments (Deliverable 5.1) | Publishable |
| **UBITECH** | Final Release of DIGITISE DT Solutions and Tools for Consumers Engagement in Marketplace Environments (Deliverable 5.2) | Publishable |
| **UBITECH** | Energy-related household level AI analytics models | Non-publishable |
| **UBITECH** | Data Management Plan (Deliverable 1.2) | Publishable |
| **UCD** | Initial release of the DIGITISE cross-sector services and applications for end-users (Deliverable 4.1) | Publishable |

| | | |
|---|---|---|
| UCD | Final Release of DIGITISE cross-sector services and applications for end-users (Deliverable 4.2) | Publishable |

TABLE 4: PARTNERS' SOFTWARE PROVISION

| Partner | Artefact | Publishable/ non-publishable |
|---|---|---|
| CIRCE | Implicit flexibility implementation algorithms | Non-publishable |
| CIRCE | Energy Assets Optimization Strategies algorithms | Non-publishable |
| CIRCE | Implicit flexibility implementation algorithms | Non-publishable |
| CIRCE | Energy Assets Optimization Strategies algorithms | Non-publishable |
| APPART | Dashboard for the visualization of results | Publishable |
| MIW | IoT monitoring and control platform | Non-publishable |
| S5 | Data management Platform | Non-publishable |
| S5 | Flexibility management tool | Non-publishable |
| S5 | AI Analytics models | Non-publishable |
| ARDEN | IoT monitoring and control platform | Non-publishable |
| UBITECH | Household Digital Twin Application | Publishable |
| UBITECH | Flexibility Marketplace | Non-publishable |
| UBITECH | Household Digital Twin Models | Publishable |
| UCD | Health and Safety Application | Publishable |
| UCD | Behavioural Profiling and Insights Application | Publishable |

TABLE 5: PARTNERS' DATASET PROVISION

| Partner | Artefact | Publishable/ non-publishable |
|---|---|---|
| CIRCE | Historical Energy Demand/consumption data of users (billing data) | Non-publishable |
| CIRCE | Historical energy generation data (PV energy generation) | Non-publishable |
| CIRCE | Flexibility related and energy costs data | Non-publishable |
| CIRCE | Energy forecast | Non-publishable |
| CIRCE | External environmental conditions (temperature, radiation, etc.) | Non-publishable |
| CIRCE | Historical Energy Demand/consumption data of users (billing data) | Non-publishable |
| CIRCE | Historical energy generation data (PV energy generation) | Non-publishable |
| CIRCE | Flexibility related and energy costs data | Non-publishable |

| | | |
|---|---|---|
| **CIRCE** | Energy forecast | Non-publishable |
| **CIRCE** | External environmental conditions (temperature, radiation, etc.) | Non-publishable |
| **ZEZ** | Historical energy generation data (PV energy generation) (82 users / 2021-2023) cumulative and differential | Non-publishable |
| **ZEZ** | Historical consumption data of users (smart meter) (82 users / 2021-2023) cumulative and differential | Non-publishable |
| **ZEZ** | Historical data calculated for Production (kWh) Total consumption (kWh) Self-consumption (kWh) Delivered (kWh) Taken (kWh) (82 users / 2021-2023) | Non-publishable |
| **ZEZ** | PV energy generation and consumption data (ongoing) - API or CSV | Non-publishable |
| **MIW** | Consumer consumption | Non-publishable |
| **MIW** | PV Generation | Non-publishable |
| **MIW** | disaggregated loads of residential buildings | Non-publishable |
| **ARDEN** | Consumer consumption Data | Non-publishable |
| **ARDEN** | Flex Asset and RES Asset Data | Non-publishable |
| **ARDEN** | disaggregated loads of residential buildings | Non-publishable |
| **MYTIL** | Consumer consumption Data | Non-publishable |
| **MYTIL** | Demographic data | Non-publishable |
| **MYTIL** | PV Generation | Non-publishable |
| **MYTIL** | Metering data | Non-publishable |
| **MYTIL** | Contract Related data | Non-publishable |
| **MYTIL** | Disaggregated loads of residential buildings | Non-publishable |

## 3.3 Expected size of the data (if known)

It is expected that as a research outcome will generate research datasets (i.e. results of the technologies, services of the demos, etc.), publications, new services proposal, dissemination material, etc. Due to the size of the project, scope of work and complexity, the expected size cannot be estimated now.

# 4. DIGITISE ORDP Participation

In Horizon Europe, the European Commission has reinforced its commitment to open science by mandating open access to research data generated by funded projects. This initiative is built upon the Open Research Data (ORD) Pilot introduced in Horizon 2020, aiming to enhance the accessibility and reusability of research data. The Open Research Data Pilot (ORDP) of the European Commission enables open access and reuse of research data generated by Horizon Europe projects. There are two main pillars to the Pilot: a) developing a DMP and b) providing open access to research data.

A project that opts in ORDP must adhere to the following conditions:

- Develop (and keep up to date) DMP.
- Deposit the data in a research data repository.
- Ensure third parties can freely access, mine, exploit, reproduce and disseminate this data.
- Provide related information and identify (or provide) the tools needed to use the raw data to validate the research.

The ORDP applies to:

- The data (and metadata) needed to validate results in scientific publications.
- Other curated and/or raw data (and metadata) that are specified in the DMP.

From the current consensus within the consortium some of the DIGITISE Artefacts will not be publicly available as depicted in the graphics below:

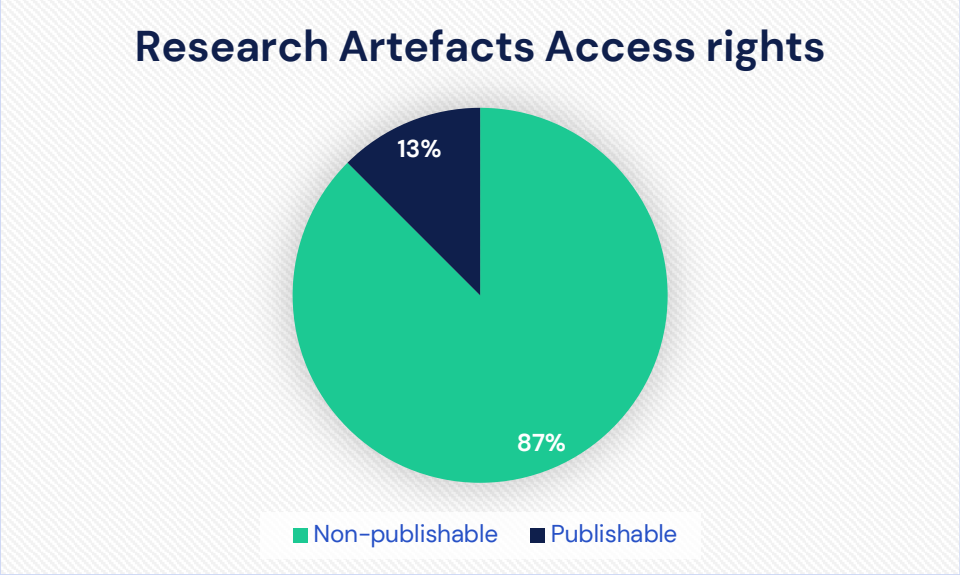



FIGURE 2: RESEARCH ITEMS ACCESS RIGHTS

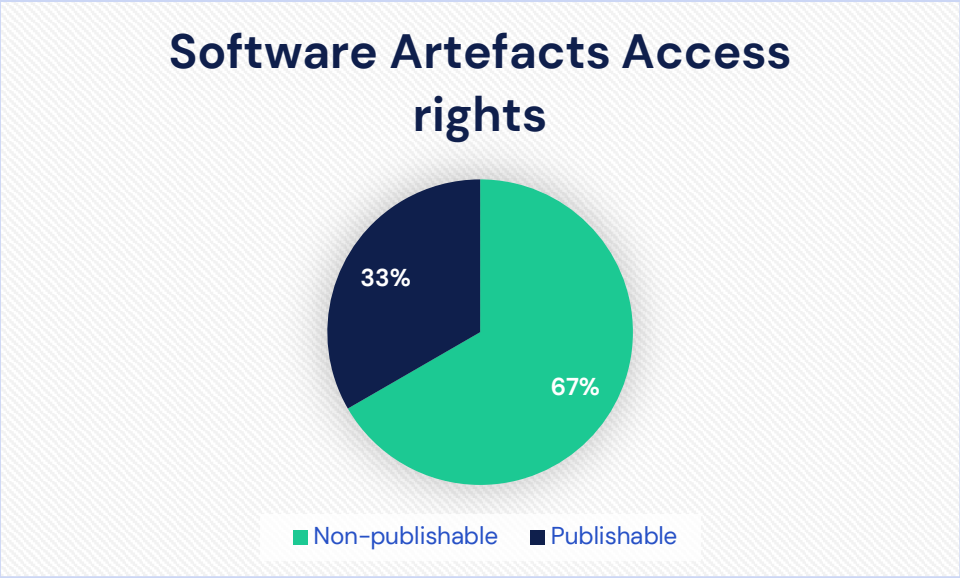



FIGURE 3: SOFTWARE ARTEFACTS ACCESS RIGHTS

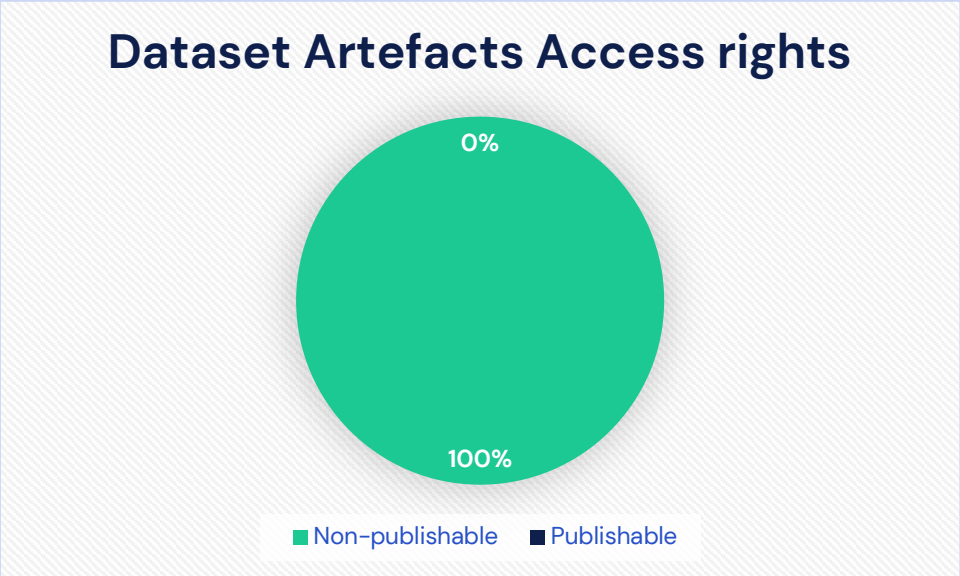



FIGURE 4: DATASET ARTEFACTS ACCESS RIGHTS

## 4.1 Publishing Infrastructure for Open Access

The DIGITISE publication infrastructure consists of a process and several web-based publication platforms that together provide long-term open access to all publishable, generated or collected results of the project. The implementation of the project will be done in accordance with the applicable regulations at the national and EU level and, especially, with the General Data Protection Regulation (GDPR) protection of personal data.

More specifically, cases where personal data information or sensitive information of internet users are collected (IP addresses, email addresses or other personal information) or further processed are not foreseen.
In the potential future case where the DIGITISE consortium will collect and/or further process personal data, this will be done in accordance with GPDR. Overall, it is aimed that DIGITISE only collects and/or further processes personal data are necessary for the attainment of the project objectives.
Both the process and the used web-based platforms are described in the following subsections.

### 4.1.1 Publishing Process

DIGITISE partners defined a simple, deterministic process that decides if a result in DIGITISE must be published or not. The term result is used for all kind of artefacts generated during DIGITISE like white papers, scientific publications, and anonymous usage data. By following this process, each result is either classified public or non-public. Public means that the result must be published under the open access policy. Non-public means that it must not be published.
For each result generated or collected during DIGITISE runtime, the following questions must be answered to classify it:

| |
|---|
| **Does a result provide significant value to others or is it necessary to understand a scientific conclusion?** |
| If this question is answered with yes, then the result is classified as public. If this question is answered with no, the result is classified as non-public. Such a result could be code that is very specific to DIGITISE platform (e.g., a database initialization) which is usually of no scientific interest to anyone, nor does it add any significant contribution. |
| **Does a result include personal information that is not the author's name?** |
| If this question is answered with yes, the result is classified as non-public. Personal information beyond the name must be removed if it should be published. This also bears witness to the repetitive nature of the publishing process, where results which are deemed in the beginning as non-publishable can become publishable once privacy-related information is removed from them. |
| **Does a result allow the identification of individuals even without the name?** |
| If this question is answered with yes, the result is classified as non-public. Sometimes data inference can be used to superimpose different user data and reveal indirectly a single user's identity. As such, to make a result publishable, the information included must be reduced to a level where single individuals cannot be identified. This can be performed by using established anonymization techniques to conceal a single user's identity, e.g., abstraction, dummy users, or non-intersecting features. |
| **Does a result include business or trade secrets of one or more partners of DIGITISE?** |

If this question is answered with yes, the result is classified as non-public, except if the opposite is explicitly stated by the involved partners. Business or trade secrets need to be removed in accordance with all partners' requirements before it can be published.

**Does a result name technology that is part of an ongoing, project-related patent application?**

If this question is answered with yes, then the result is classified as non-public. Of course, results can be published after patent has been filed.

**Can a result be abused for a purpose that is undesired by society in general or contradict with societal norms and DIGITISE's ethics?**

If this question is answered with yes, the result is classified as non-public.

### 4.1.2 Publishing Platforms

In DIGITISE, we use several platforms to publish our results openly. The following list presents the platforms used during the project and describes their concepts for publishing, storage, and backup.

#### The project Website

The partners in the project consortium decided early to setup a project-related website https://digitise-horizon.eu/ . This website describes the mission and the general approach of DIGITISE and its development status. A blog informs about news on a regular basis. A dedicated area for downloads is used to publish reports and white papers as well as scientific publications (in pre-camera ready form, or through links to the publisher's websites in case these are not open access). All documents are published using the portable document format (PDF). All downloads are enriched by using simple metadata information, such as the title and the type of document. The website is hosted by our partner AUSTRALO. All webpage-related data is backed up on a regular basis. All information on the project website can be accessed without creating an account. The website is backed up once per month.

#### Open Access Publication Platforms

**Zenodo**

Zenodo is a research data archive / online repository which helps researchers to share research results in a wide variety of formats for all fields of science. It was created through EC's OpenAIRE+ project and is now hosted at CERN using one of Europe's most reliably hardware infrastructures. Data is backed nightly and replicated to different locations. Zenodo not only supports the publication of scientific papers or white papers, but also the publication of any structured research data (e.g., using XML). Zenodo provides a connector to GitLab that supports open collaboration for source code and versioning for all kinds of data. All uploaded results are structured by using metadata, like for example the contributors' names, keywords, date, location, kind of document, license, and others. Considering the language of textual metadata items, English is preferred. All metadata is

licensed under the CC0 license (Creative Commons 'No Rights Reserved'). The property rights or ownership of a result does not change by uploading it to Zenodo.
All public results generated or collected during the project lifetime will be uploaded to Zenodo for long-term storage and open access.

**Open Research Europe**

Open Research Europe is an open access publishing platform for the publication of research that makes it easy for Horizon Europe beneficiaries to comply with the open access terms of their funding and offers researchers a publishing venue to share their results and insights rapidly and facilitate open, constructive research discussion. Uses an open research publishing model: publication within days of submission, followed by open invited peer review. Includes citations to all supporting data and materials, enabling reanalyzing, replication and reuse. Articles that pass peer review are sent to major indexing databases and repositories.

**GitLab**

GitLab is a well-established online repository which supports distributed source code development, management, and revision control. It is primarily used for source code data. It enables world-wide collaboration between developers and provides some facilities to work on documentation and to track issues. GitLab provides paid and free service plans. Free service plans can have any number of public, open-access repositories with unlimited collaborators. Private, non-public repositories require a paid service plan. Many open-source projects use GitLab 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 GitLab over provided material. For textual metadata items, English is preferred.
All source code components that are implemented during this project and decided to be public will be uploaded to an open access GitLab repository.

### 4.1.3 Access and Sharing

The accessing and sharing of data are firstly ruled by two documents: the non-disclosure agreement, which stipulates under which conditions transmitted information between the project partners is deemed confidential and must not be further disseminated; and the Description of Action (DoA) which stipulates the dissemination level of each deliverable. Moreover, the project consortium will comply with the FAIR (findable, accessible, interoperable and reusable) (European Commission, 2016) guidelines of the Horizon Europe programme.
The data necessary to successfully complete the project WPs will be shared without any restrictions amongst the WP partners either via internal repositories or direct communication. Public data will be made available on the project's website or other

repositories, as appropriate. Users will be made aware of this data primarily through research publications, patent applications, dissemination activities, invited talks, social networks and the project website. Data will be made available to the project consortium as soon as it is available; to standardization bodies when required; and to the public at the due date of the derivable, and, in case a research publication is based on that, as soon as the paper is submitted (if submission is anonymous, this will be postponed). If access to confidential data is necessary by the public, restrictive measures will be put in place.

# 5. FAIR Data

DIGITISE project supports the reuse of research data and follows FAIR principles[1]. FAIR represents a set of guiding principles to make data **Findable, Accessible, Interoperable, and Reusable**.

The international FAIR Principles have been formulated as a set of guidelines for the reuse of research data. The acronym FAIR stands for findable, accessible, interoperable, and reusable. Research data must be of a quality that makes it accessible, findable, and reusable.

**Findable:** data has a unique, persistent ID, located in a searchable resource, and documented with meaningful metadata.

**Accessible:** data is readily and freely retrievable using common methods and protocols, metadata is accessible even if the data is not.

**Interoperable:** data is presented in broadly recognized standard formats, vocabularies, and languages.

**Re-useable:** data has clear licenses, and accurate meaningful metadata conformity to relevant community standards and identifying its content and provenance.

## 5.1 Making data findable, including provisions for metadata

This document launches the data management plan to support effective collection and integration of the DIGITISE data. Storage, processing and sharing (among project participants) will occur via data exchange platforms (such as Microsoft SharePoint), whereas interaction with the wider public will be achieved through the official project website. Also, data will be stored at the coordinator's repository and will be kept for a minimum of 5 years after the end of the project. Where requested, data will be kept for 2 more years.

A naming convention will include a concise description of contents, the host institution collecting the data and the month of publication.

Version numbering will only be an issue if a participant requests withdrawal of their data in which case a version number will be added to the filename.

No specific standards or metadata have been identified for the time being for the proposed datasets.

Data will be anonymized meaning that data will not identify any individuals and therefore real names of participants will NOT be distributed.

Data will be shared only in relation to publications (deliverables and papers). As such, the publication will serve as the main piece of metadata for the shared data. When this is not seen as being adequate for the comprehension of raw data, a report will be shared along with the data explaining their meaning and methods of acquisition.

[1] Force11 (2016) The FAIR Data Principles, https://www.force11.org/group/fairgroup/fairprinciples

### 5.1.1 Discoverability of the data

To be able to use the data generated by the project it is essential to integrate data from the participants in the open calls and the activities undertaken by project partners. Considering the FAIR data principles (Wilkinson et al., 2016[2]) (meta)data should:

- Be assigned to a globally unique and persistent identifier;
- contain enough metadata to fully interpret the data, and;
- be indexed in a searchable source.

By applying these principles data becomes retrievable and includes their authentication and authorization details.

### 5.1.2 Data identification mechanism

All documents associated with the project will be identified with a project name and unique and persistent document type designator and number that will be given to the coordinator for the submission to the EC. Versioning of the document should be part of the document name and title.

As per the documents related to project activities and/or deliverables, the tasks or deliverables number will be used to identify the document followed by a brief title of the activity or deliverable.

**Example**

*DIGITISE- D1.2 - Data Management Plan -v1.0.pdf*

### 5.1.3 Naming conventions used

Each set of data produced (dataset, deliverables, etc.) will be named in a uniform way and will include a table with a version control.

The recommendations to name documents of the project are as follows[3]:

- Choose easily readable identifier names (short and meaningful);
- Do not use acronyms that are not widely accepted;
- Do not use abbreviations or contractions;
- Avoid Language-specific or non-alphanumeric characters;
- Add a two-digit numeric suffix to identify new versions of one document.
- Dates should be included back to front and include the four-digit years: YYYYMMDD.

For deliverables: **DIGITISE [Deliverable Code]-[Deliverable Title]-vA.BB** i.e.: DIGITISE D1.2- Data Management plan- v1.00 *(for submission to the Commission)*

For datasets: **WP [Work Package number] P [Pilot number; pilot activity number] - [description of the activity]** i.e.: WP6 P1.3 Results of demonstration performance.

[2] Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data3:160018 doi: 10.1038/sdata.2016.18 (2016).

[3] https://www.ukdataservice.ac.uk/manage-data/format/organising

### 5.1.4 Clear versioning of the documents

Only documents created by the consortium will be versioned, for this purpose templates include 3 descriptors to identify the versions and status of the documents:

TABLE 6: PROPOSED DOCUMENT HISTORY TABLE OVERVIEW

| Version | Date | Author | Notes |
|---|---|---|---|
| 1 | xx | xx | xx |
| 2 | | | |
| 3 | | | |

Moreover, partners, following the recommendations included in section "Naming conventions" will identify the different versions by using a two-digit number following the descriptor Draft. A document reviewed by another partner should be returned to the principal author by including **rev + acronym** of the organisation. Only the principal author will change the draft number and will add the word FINAL to documents ready to be sent to the EC or those to be used as final versions.
The document history included in the document template should be filled in as follows:

TABLE 7: DOCUMENT HISTORY TEMPLATE – EXAMPLE

| Version | Date | Author | Notes |
|---|---|---|---|
| 1 | XX/XX/2023 | ABC | Section 2.1 needs to be completed |
| 2 | XX/XX/2023 | CDE | Section 2,1 completed. Comments added to the document. |
| 3 | XX/XX/2023 | ABC | Added suggestions by TEE |
| 4 | XX/XX/2024 | XYZ | Included some topics on section 2.1 |
| 5 | XX/XX/2024 | ABC | Final version with partners contribution |

### 5.1.5 Standards for the metadata creation (if any)

Basic metadata will be used to facilitate the efficient recall and retrieval of information by project partners and external evaluators and contribute to easily finding the information requested. To this end, all documents related to the project have to include in the front-page information about author(s) & editor(s), WP, dissemination level and version.
To support the completeness of metadata, the project provides a metadata template to all stakeholders. The template will be a living document that might be expanded to fit project specific requirements.

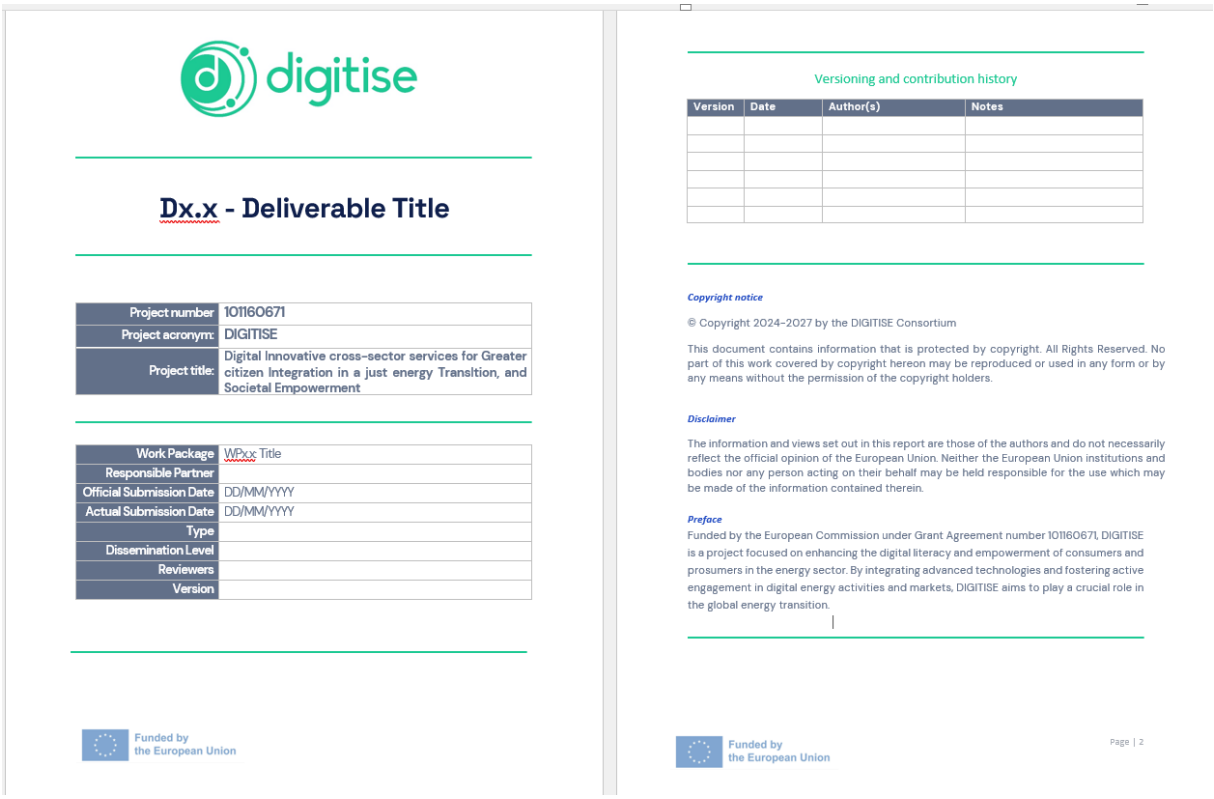
TABLE 8: METADATA TEMPLATE FOR DIGITISE DATASETS.

| # | Field | Description |
|---|---|---|
| 1 | Title | A name given to the resource. |

| | | |
|---|---|---|
| **2** | Creator | An entity primarily responsible for making the resource |
| **3** | Subject | The topic of the resource |
| **4** | Description | e.g., abstract, table of contents, graphics, … |
| **5** | Publisher | Only for published items. |
| **6** | Contributor | Entities that contributed to the making of the resource. |
| **7** | Date | The termination of the data collection period. |
| **8** | Type | [dataset, article, questionnaire, …] |
| **9** | Format | File format of the resource. |
| **10** | Identifier | e.g., ISSN if your item has been published |
| **11** | Source | Which tools were used to collect the data |
| **12** | Language | A language of the resource. |
| **13** | Relation | A related resource. |

In addition to the dataset's metadata document, dataset providers are compelled to attach additional documents such as:

1. A description of the study
2. Method of research
3. Applied questionnaires
4. Data documentation / usage manual
5. Any other information that might be of interest to a data user



digitise

Dx.x - Deliverable Title

| | |
|---|---|
| Project number | 101160671 |
| Project acronym: | DIGITISE |
| Project title: | Digital Innovative cross-sector services for Greater citizen Integration in a just energy Transition, and Societal Empowerment |

| | |
|---|---|
| Work Package | WPxx Title |
| Responsible Partner | |
| Official Submission Date | DD/MM/YYYY |
| Actual Submission Date | DD/MM/YYYY |
| Type | |
| Dissemination Level | |
| Reviewers | |
| Version | |

Funded by the European Union

Versioning and contribution history

| Version | Date | Author(s) | Notes |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

*Copyright notice*

© Copyright 2024-2027 by the DIGITISE Consortium

This document contains information that is protected by copyright. All Rights Reserved. No part of this work covered by copyright hereon may be reproduced or used in any form or by any means without the permission of the copyright holders.

*Disclaimer*

The information and views set out in this report are those of the authors and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

*Preface*

Funded by the European Commission under Grant Agreement number 101160671, DIGITISE is a project focused on enhancing the digital literacy and empowerment of consumers and prosumers in the energy sector. By integrating advanced technologies and fostering active engagement in digital energy activities and markets, DIGITISE aims to play a crucial role in the global energy transition.

Funded by the European Union

Page | 2

FIGURE 5: TEMPLATE TO BE USED FOR PROJECT DOCUMENTATION METADATA OVERVIEW.

## 5.2 Making data openly accessible

Where possible data will be made openly available subject to ethics and participants' agreement. However, the personally identifiable nature of the data collected within DIGITISE means that in most instances it would be difficult to release collected data. Where data is made available, we will do so using the Project's file repository hosted in the coordinator's premises.

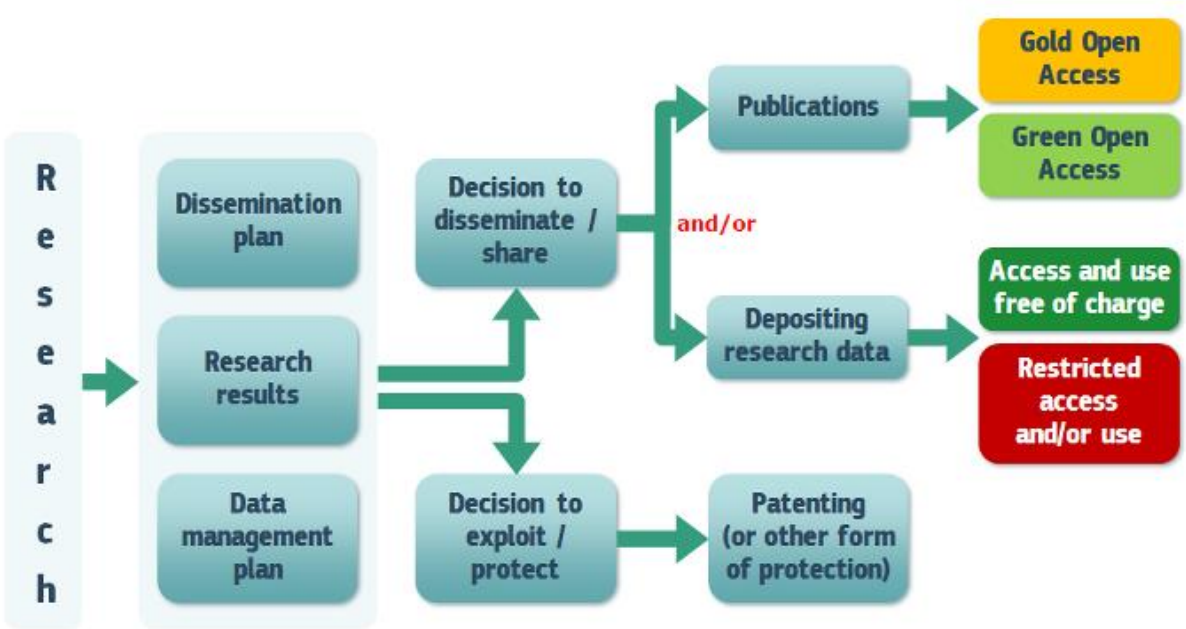


FIGURE 6: OPEN ACCESS TO SCIENTIFIC PUBLICATION AND RESEARCH DATA IN THE WIDER CONTEXT OF DISSEMINATION AND EXPLOITATION[4]

Prior to release, the requesting party will need to contact the Project Coordinator describing their intended use of a dataset. The Project Coordinator will send a terms and conditions document for them to sign and return. Upon return, the dataset will be released. Documentation will be included with the release of the data.
In alignment with the EC Guidelines on Open Access to Scientific Publications and Research Data in Horizon Europe, DIGITISE will also follow a combination of Gold and Green Open Access5 strategy to its scientific publications, which will be agreed during the first months of the project execution. Gold Access will be encouraged for high-impact journal publications while the self-archiving, Green Access will be granted for the rest of the publications. The availability through the consortium members will be considered while there will also be a relevant repository on the website of the project and in social networking sites for scientists and researchers like ResearchGate.

### 5.2.1 Methods or software needed to access the data

No specific software tools will be needed to access the data, since anonymous data sets will be saved and stored in word, pdf or excel to facilitate its exploitation and guarantee their long-term accessibility.

### 5.2.2 Deposit of data, associated metadata, documentation and code

Data will be deposited and secured on Microsoft SharePoint file repository and additional instance of all data on coordinator's account, in the following URL: https://ubitecheu.sharepoint.com/sites/DIGITISE

[4] European Commission Directorate-General for Research & Innovation (2017) Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020
[5] https://en.wikipedia.org/wiki/Open_access

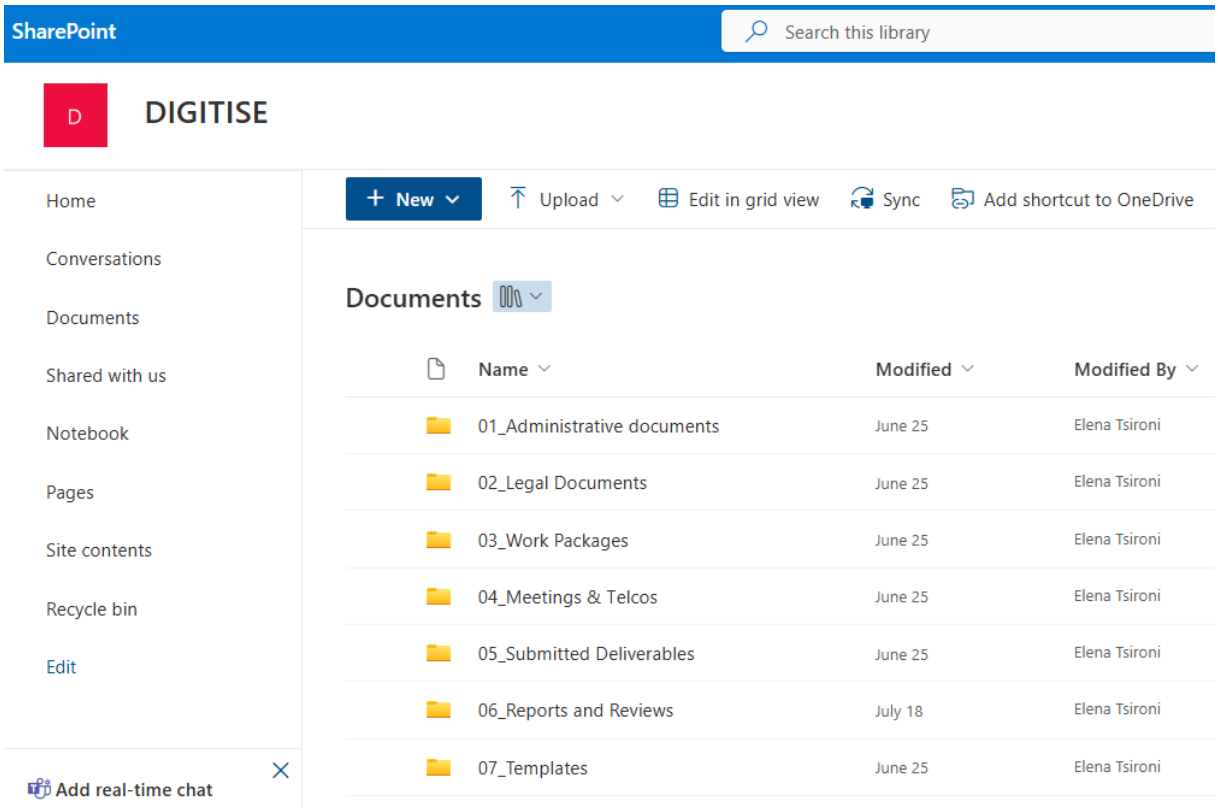



FIGURE 7: DIGITISE REPO FOLDERS STRUCTURE

## 5.3 Making data interoperable

The concept interoperable demands that both data and metadata must be machine-readable, and that consistent terminology is used.

### 5.3.1 Interoperability of data assessment

Partners will be responsible for storing the data in a comprehensive format and adapted to the real and current needs of the possible practitioners interested in using, merging, or exploiting the data generated throughout the project. The assessment of data interoperability will be updated in future reviews to guarantee the DIGITISE data fits the needs of a specific scenario (such as data infrastructures, interests or purpose of data.

### 5.3.2 Interoperability of data assessment

The vocabulary used in the project is a very standard and common language within the business creation culture and logistics. Vocabulary won't represent any barrier for data interoperability or re-use.

## 5.4 Making data re-usable

For data to be re-usable, it is -generally- considered that meta(data) have a plurality of accurate and relevant attributes and that they are released with a clear and accessible data usage license. Moreover, it is considered that (meta)data are associated with their provenance and that they meet domain-relevant community standards[6].
Note that the overall management of knowledge and the provisioning for the establishment of the related Intellectual Property Rights is dictated in detail under DIGITISE's Grant Agreement and the consortium agreement stipulating -among other- for the ownership of the background and the foreground knowledge, as well as for the commercial exploitation of the project's results.

### 5.4.1 Increase data re-use through clarifying licenses

Data will only be available on application/SharePoint platform/share portal and their use will be restricted to the research use of the licensee and colleagues on a need-to-know basis. This non-commercial licence is renewable after 2 years, data may not be copied or distributed and must be referenced if used in publications. These arrangements will be formalised in a User Access Management license which describes in detail the permitted use of the data.

### 5.4.2 Data quality assurance process

The project coordinator will be responsible for assuring the quality of the data by making sure the dataset follows the FAIR principles included in this plan, and that data is updated. Personal data processing will be done following the EU, national and international laws considering the "data quality" principles listed below[7]:
Data processing is adequate, relevant and non-excessive;

- Accurate and kept up to date;
- Processed fairly and lawfully;
- Processed in line with data subjects' rights;
- Processed in a secure manner;
- Kept for no longer that necessary and for the sole purpose of the project.

Data quality assurance process will be led in accordance with the Regulation (EU) 2016/679 (General Data Protection Regulation) on the protection of natural persons regarding the processing of personal data and on the free movement of such data.

[6] See, also, FAIR data principles (FORCE11 discussion forum) available at: https://www.force11.org/group/fairgroup/fairprinciples
[7] Wilms, G. Guide on Good Data Protection Practice in Research of the European University Institute. (March 2017). Retrieved from http://www.eui.eu/Documents/ServicesAdmin/DeanOfStudies/ResearchEthics/Guide-Data-Protection-Research.pdf

### 5.4.3 Length of time for which the data will remain re-usable

The Consortium will contribute to maintaining data re-usable as long as possible after the end of the project. The first period of 3 years has been established; however, this time can be extended under the partners' agreement. This period can vary depending on the value of the data after the end of the project.

# 6. Allocation of resources

## 6.1 Data management responsibilities

Data will be stored at the Collaboration file repository set by the Coordinator as the project's repository, and will be kept for 3 years after the end of the project. Where requested, data will be kept for 2 more years. The handling of the repository on behalf of DIGITISE, as well as all data management issues related to the project fall in the responsibility of the coordinator.

As for the publications, where the analyses of the empirical research data will be presented, the consortium will publish them in scientific journals that allow open access. The costs related to open access will be claimed as part of the Horizon Europe grant.

Regarding the data resulting from the activities of the project, each WP leader will be responsible for the storage and compliance of the data and then for uploading in the DIGITISE SharePoint web portal, or other storage systems to share the information of the project.

The DIGITISE coordinator assisted by the WP leaders will be responsible for updating this document and developing a strategy to encourage:

- the identification of the most-suitable data-sharing and preservation methods;
- the efficient use of data assuring clear rules on its accessibility;
- the quality of the data stored; and
- the storage in a secured and user-friendly interface.

## 6.2 Cost of potential value of long-term preservation

As stated in the previous section, the costs of data storage and maintenance are not going to require extra funding once the project ends. As per the value of the data, it is important to consider that the topics covered by the project respond to the current needs. Therefore, data coming out of this project will have a direct impact in the coming years but might not be of relevance as the challenges are being tackled or replaced by other priorities.

# 7. Data Security

DIGITISE data exchange platform (SharePoint) applies technological and organizational measures to secure processing of personal data against publishing to unauthorized persons, processing in violation of the law and change, loss, damage or destruction.

- **Information security**: SSL (Secure Socket Layer) certificates are applied. To ensure the appropriate level of security, the password for the account will exist on the platform only in a coded encrypted form.
- **Options for reading data**: the platform offers the possibility to make data available in a read-only or downloadable format, hindering access to information by unauthorized users.
- **Back-up policy**: complete and redundant back-ups are done every week. Moreover, every time a modification is made an older version is saved.
- **Accidental deletion or modifications**: in case of a catastrophic event that implies the partial or complete deletion of the data sets, the data from the most recent back up will be automatically restored (back-up won't be older than 60 minutes). In case of accidental deletion or modification only the most recent document will be restored, so in case of accidental changes or deletion data can be easily recovered.
- **Deletion or modification of data by users**: only administrators have the right to delete or modify the information included in the datasets.
- **Terms and conditions**: the SharePoint platform has specific terms of use and conditions that must be accepted by all users of the platform.

# Conclusions

This initial version of the data management plan provides the first information on the data to be collected and used throughout the project. The relevance of FAIR data has been described and its application on the used dataset was given. To provide for security and the handling of ethical aspects, a first outline of these topics was provided. As mentioned, the DMP is a living document and will be extended and updated throughout the whole project lifetime by all relevant partners.

Funded by the European Union

# Annex – Artefact Template

The following tables try to capture the description of the data that will be produced in the context of DIGITISE. Every use case will fill in such a template and subsequently all the templates will be collected with the beginning of WP6, the demonstration applications work package of the project.

TABLE 9: MAKING DATA FINDABLE TEMPLATE

| Making data Findable | |
|---|---|
| **Name of data set** | *Univocal identifier of the considered data [DIGITISE_Wx_Tz_01]*<br>*Please, provide one sentence description.* |
| **Data types** | *[Real time data stream, unstructured like tweets, synthetic data stream, log data of IDS, etc.]* |
| **Data generation and/or collection** | *Description of the type of input used to generate the data and the complete methodology and tools used for data collection* |
| **Purpose** | *What are the data collected/generated specifically used for?* |
| **Data origin** | *[Where applicable, information from applications to be developed by the partner.]* |
| **Version control** | *Version no, Date, Author, Description of change* |
| **Search keywords** | |
| **Responsible partner of the artefact** | |

TABLE 10: MAKING DATA ACCESSIBLE TEMPLATE

| Making data Accessible | |
|---|---|
| **Accessibility** | *Open/Confidential* |
| **Repository** | *Description/location of the available data.* |
| **Shareability restrictions / related Information** | *[Where applicable, information from applications to be developed by the partner.]* |

TABLE 11: MAKING DATA INTEROPERABLE TEMPLATE

| Making data Interoperable | |
|---|---|
| **Format** | *Data format, measuring unit, typical order of magnitude [JSON-like, CSV]* |
| **Expected size of the data** | *[To be defined, 3 TB/Day or 12 GB/day when compressed etc.]* |
| **Standards and metadata**[8] | *[The metadata attributes list. The methodologies used.]* |
| **Standard software Interfaces** | *List of the standards used to promote results replicability.* |
| **Extensions to standard interfaces** | *Extensions to the above standards as developed during the project.* |
| **Mapping to common ontologies** | |

TABLE 12: MAKING DATA RE-USABLE TEMPLATE

| Making data Reusable | |
|---|---|
| **RE-use of existing data** | *[No reuse of existing data, for the generation of synthetic datasets, it will be essential to create a recipe, reusing the existing data in logs etc.]* |
| **Data backup** | *Consistent location of the data, including previous releases* |
| **Quality Consistency** | *Constraints determining the quality/currency of the collected data.* |
| **Emulation tools** | *Description/location of possible emulation tools useful for replicating the data* |
| **Data availability** | |

8 Note that the fields pertinent to standards are, also, relevant for reusability purposes.