unchain

D1.2 – Data Management Plan (DMP) v1

UK Research and Innovation

Funded by the European Union

CIVITAS Sustainable and smart mobility for all

Legal disclaimer

This document is issued within the frame and for the purpose of the UNCHAIN project. This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101103812. The UK participant in Horizon Europe Project UNCHAIN, is supported by UKRI grant number 10078841 Lancaster University.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

Copyright statement

Deliverable details

Deliverable No.	D1.2
Deliverable Type	DMP – Data Management Plan
Dissemination Level	PU - Public
Work Package No.	1
Work Package Title	Project Management and coordination
Task No.	1.3
Task Title	Ethics monitoring, Data Management and RRI
Author(s)	Elena García
Status (F:final; D:draft; RD:revised draft)	F
File name	D1.2 – Data Management Plan (DMP) v1
Version	0.4
Task start date and duration	01/05/2023 – 31/10/2026
Due Date	M6 (October 2023)
Delivery Date	31/10/2023

Review History of Deliverable

Version No.	Date	Revision Details
0.1	29/09/2023	Draft version
0.2	11/10/2023	Reviewed version
0.3	16/10/2023	Reviewed version
0.4	19/10/2023	Final version

Reviewers of Deliverable

Reviewers's name	Company	Date
Konstantinos Zografos	ULANC	09/10/2023 (v0.1)
Melina Zarouka	POLIS	16/10/2023 (v0.2)
Konstantinos Zografos	ULANC	18/10/2023 (v0.3)

Table of Contents

List of Tables

List of Figures

Acronyms

Acronym	Meaning
ANSI	American National Standards Institute
CERN	European Organization for Nuclear Research
DC	Dublin Core
DCMI	Dublin Core Metadata Initiative
DMP	Data Management Plan
DoA	Description of Action
DOI	Digital Object Identifier
EC	European Commission
EU	European Union
FAIR	Findable, Accessible, Interoperable, Reusable
FUA	Functional Urban Area
GDPR	General Data Protection Regulation
HE	Horizon Europe
IETF	Internet Engineering Task Force
IPR	Intellectual Property Rights
ISBN	International Standard Book Number
ISO	International Standards Organization
ISRC	International Standard Recording Code
MD5 Algorithm	Message Digest Algorithm
NISO	National Information Standards Organization
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
ORDP	Open Research Data Pilot
PMP	Project Management Plan
RFC	Request For Comments
SFTP	SSH File Transfer Protocol
SSH	Secure Shell
WP	Work Package

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 UNCHAIN 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, 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 it is presented 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.

1. Introduction

1.1. Purpose of the document

The main goal of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that should be considered by the UNCHAIN Consortium with regards to the project research data.

Since the DMP is a common deliverable in all the Horizon Europe projects, this document is produced on the basis of other deliverables already developed by ETRA for other European projects and adapted to the UNCHAIN.

The DMP is not a fixed document; on the contrary, it will evolve during the lifespan of the project. It 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 (European Commission, 2016).

This first version of the DMP aims to outline how the UNCHAIN project will try to make the research data findable, accessible, interoperable, and reusable (FAIR) and therefore contribute to knowledge discovery and innovation. Although this first version of DMP (due in M6) does not provide very detailed information on the specific data sets to be collected, generated, and processed during the project, it lays the foundations for creating an effective data management strategy covering the complete research data life cycle.

The DMP will be updated in M18, M36 and M42.

1.2. Scope of the document

This deliverable describes the initial data management plan for the project covering the complete research data life cycle. This DMP will be used by the UNCHAIN consortium as a guideline when handling the research data during and after the end of the project. This DMP, which will be continuously updated, presents the envisioned data management strategy and includes a first 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. As the project progresses, it will be possible to update the DMP by providing further information and details on 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 UNCHAIN project.

1.3. Structure of the document

This deliverable follows the DMP template designed by the European Commission (EC) (European Commission, 2021) on Data Management Plans adapted to follow UNCHAIN document procedures.

The document is organized in seven sections:

- Introduction (Section 1).
- Data Summary (Section 2).
- Fair Data (Section 3).
- Allocation of resources (Section 4).
- Data security (Section 5).
- Ethics (Section 6).
- Conclusions (Section 7).

2. Data Summary

2.1. Purpose of Data Management and relation to the project

The strategic goal of UNCHAIN is to break data silos and make the urban freight data more available and accessible. To do so, a standardised data exchange ecosystem will be implemented and 12 innovative services will be developed to contribute, from a strategic and operational perspective, to a more sustainable urban logistics and moving towards climate-neutral and smart cities. 10 European cities will participate in UNCHAIN activities: Madrid,

Berlin and Florence as living labs; Prague, Mechelen, Madeira and Riga as follower cities, and Brest metropole, Ravenna and Alba Iulia as peer-cities supporting the early adoption of the results.

UNCHAIN is fully committed to following and upholding the FAIR principles for the management of data. This applies not only to the data produced by the project as output but also to the data used during the project.

The UNCHAIN project will adopt the Open Science approach in various aspects of the project such as:

- Open access to research outputs such as publications, data, software, models, algorithms, and workflows.
- Early and open sharing of research, for example through preregistration, registered reports, pre-prints, and crowdsourcing of solutions to a specific problem.
- Use of open research infrastructures for knowledge and data sharing.
- Participation in open peer-review.
- Measures to ensure reproducibility of results.
- Open collaboration within science and with other knowledge actors, including involving citizens, civil society, and end-users, such as in citizen science.

It will comply with European recommendations regarding Data Management Plans, providing clear procedure 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 about the project research data. The DMP reflects the consortium's data management policies, systems, and procedures which will be implemented and embedded into research procedures and regularly reviewed throughout the research cycle.

The UNCHAIN consortium will strive to make data open but cannot overrule limitations that partner institutions put on data that they contribute. 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. Dataset types, formats, and standards

Throughout the duration of the project, UNCHAIN will gather (and generate) a variety of data. Broadly, the data falls into two categories:

- Organizational data – 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 the following Work Packages (WPs):

- WP2 Requirements identification, data landscaping and use cases definition. In this WP all the relevant information that will be required to set the basis and feed the services will be compiled, the public and private needs in the urban logistics ecosystem identified and the requirements for the services withing WP4 and WP5 determined. To understand the needs the urban logistic system, qualitative and quantitative analysis will be done by considering input from internal people, external and from opinions included in the social network.
- In WP3 the collaboration framework will be set: the requirements of large, medium, and small urban logistics stakeholders related to privacy, security, standards, accountability and competence will become into the "usage contracts" definition, and in the pairs of data type + conditions for sharing the data cities and logistics operators.
- WP4 and WP5 are the technical Work Packages in which the services will be developed according to the data available to achieve the requirements defined in the Description of Action (DoA).
- WP6 will demonstrate, evaluate, and analyse the impact assessment of the services in the living labs and follower cities. To do, questionnaires will have to be developed to be able to measure them.
- WP7 will define a clear go-to market strategy that includes Business Models and Plans, detailed individual exploitation strategies and policy recommendations.

Unless it is stated differently for any specific case, all the data will be anonymised before publication and will comply with the General Data Protection Regulation (GDPR) and the specific national and European laws for the protection of personal data.

A specific section about the data related issues will be included in all the deliverables, even though in some cases it will be not applicable, as summary of the main information of each dataset (type, format, standard, responsible, storage, public/confidential…). To facilitate and homogenise the documents produced, a template including this section will be provided to all the partners. This data section will address the FAIR principles, interoperability, confidentiality and the data integrity. Furthermore, when data generated/used by a WP will be used to subsequent WPs, the WP leaders of the involved deliverables should ensure data integrity and compatibility across their WPs.

2.3. Re-use of data

Throughout the duration of the project, existing research findings, publications, and other pertinent information that is accessible will be examined. The primary purpose of this analysis will be to conduct internal project assessments, and the relevant information will be included in the appropriate project deliverables with appropriate attribution to the original sources.

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

Moreover, UNCHAIN will use some best practices coming from other initiatives and projects, with special attention to CIVITAS initiatives such as MEISTER (MEISTER Project, 2018), USER-CHI (USERCHI Project, 2020), MOMENTUM (MOMENTUM Project, 2022), SUaaVE (SUaaVE Project, 2019) or PrepDSpace4Mobility (PrepDSpace4Mobility Project, 2022).

2.4. Timetable updates of the DMP

As previously stated, the DMP is a living document that will be updated along the lifespan of the project whenever significant changes arise, such as new data sets and types, modification of consortium policies, changes of consortium composition or other external factors.

Following the guidelines provided by the EC, the DMP will be updated, at least as frequently as indicated in the following timetable of Table 1.

Table 1 – Timetable updates of the DMP

DMP version	Delivery Month	Description
v1	M6	DMP First version
v2	M18	DMP including refinements
V3	M36	DMP including refinements
V4	M42	DMP Final version

3. FAIR Data

According to [11], 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.

The FAIR principles cover numerous aspects of data management, including sourcing, typology, quality control, availability, intellectual property rights considerations, compatibility and negotiability. The exact sub-topics of each of the principles can be found in the syntax tree in Figure 1 below:

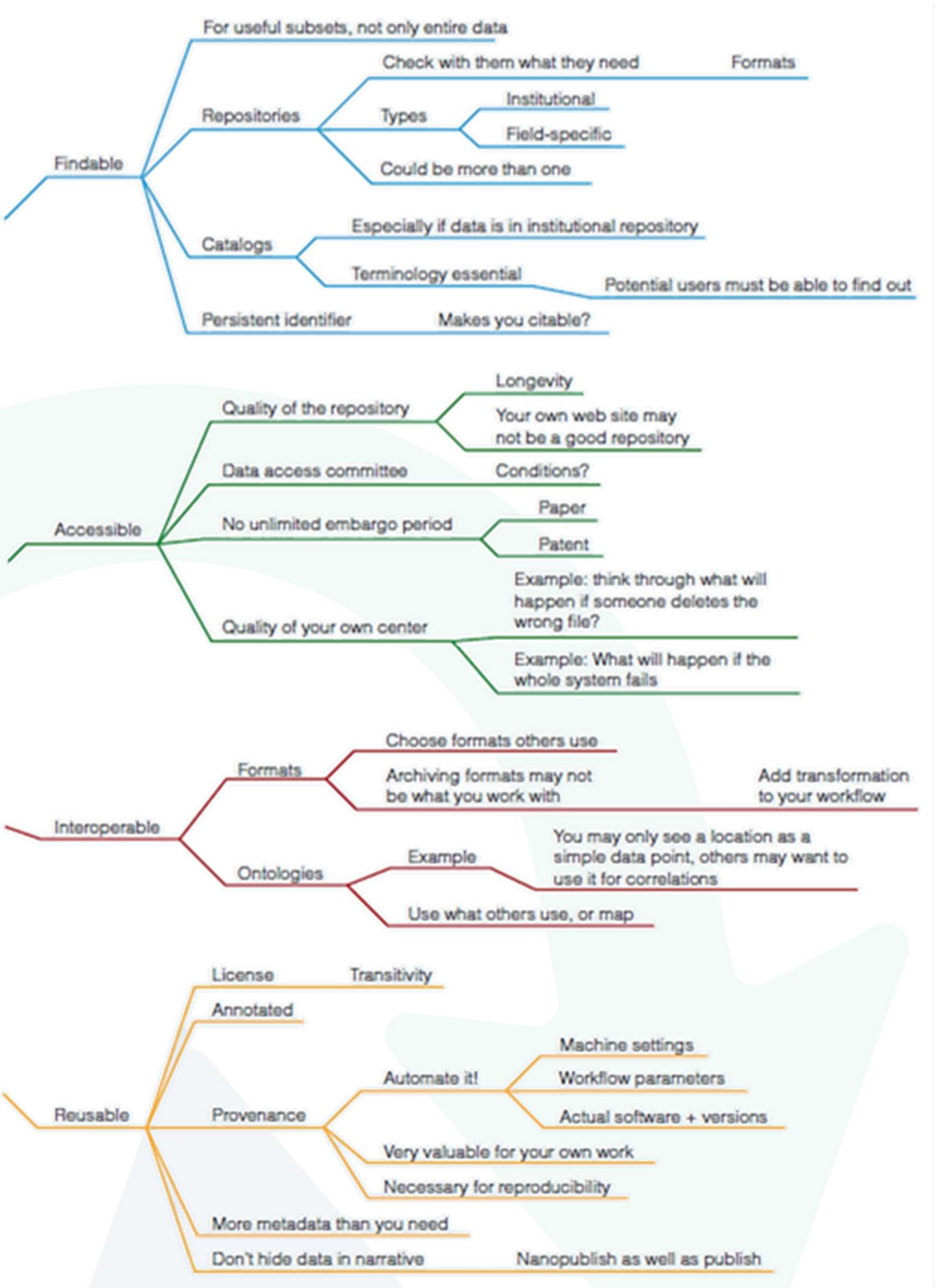

Figure 1 – Research Data Management according to the FAIR principles. Source: Elsevier.

3.1. Making data findable, including provisions for metadata

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

AS stated in section 2.2, a specific section about the data related issues will be included in all the deliverables and included in the template distributed among the partners to facilitate and homogenise the documents produced.

Following the standards promoted by the European Data Portal (European Data Portal, 2018), the UNCHAIN project research data will be inventoried and annotated with metadata, being metadata defined as follows:

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

The Dublin Core metadata standard is a straightforward and efficient element set used to describe various networked resources. The metadata landscape, as perceived by the Dublin Core community, is currently divided into 4 levels of interoperability and 15 sections for data description; these interoperability levels and sections are resumed in Table 2 and Table 3.

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

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

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

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

The fifteen elements "Dublin Core" described in this standard are part of a larger set of metadata vocabularies and technical specifications maintained by the Dublin Core Metadata

Initiative (DCMI). The fifteen element descriptions have been formally endorsed in the following standards:

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

3.2. Making data openly accessible

An analysis on which UNCHAIN research data will be made openly accessible and which data will be kept closed will be done in a later stage of the project. 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. 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 version of the DMP. At the current stage, this analysis would not make much sense since the datasets are not defined yet.

During the project lifetime, information on the following aspects will be elaborated for all datasets on a case-by-case basis, before making a consortium-wide decision on how to handle 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, UNCHAIN research data will be made available to parties with a legitimate research interest as soon as possible, without compromising privacy, ethical or commercial sustainability. Research data which is sensitive due to privacy and data protection issues will be kept confidential.

3.2.1. Open Research data repository

For assuring the open access to UNCHAIN scientific publications and research data sets, the consortium will use the ZENODO (https://zenodo.org/) repository, an online European scientific repository developed under the European OpenAIRE program and provides a commonly used Digital Object Identifier (DOI) to make them easily and uniquely citable.

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 coordinator (ULANC) and the partners that have ownership of the data.

Once 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, will be kept closed for privacy reasons.

3.2.1.1. ZENODO

ZENODO 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 (CERN, 2023) professional data centres. A detailed description of ZENODO's policies regarding the handling of the data and usage of the service can be found online (Zenodo, 2023).

As previously indicated, UNCHAIN 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 (Open Archives, 2023).

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. UNCHAIN internal repository

During the life cycle of UNCHAIN, data collected or generated by the project will be stored and systematically organised in the official project repository (ETRA, 2023).

Alfresco, the repository used in UNCHAIN, is a flexible content management web application. In the framework of the 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 and only share with the relevant partners of the different organizations.

As shown in Figure 2, folders are organized in a hierarchical and clear structure and files are uniquely identifiable and versioned by using a name convention.

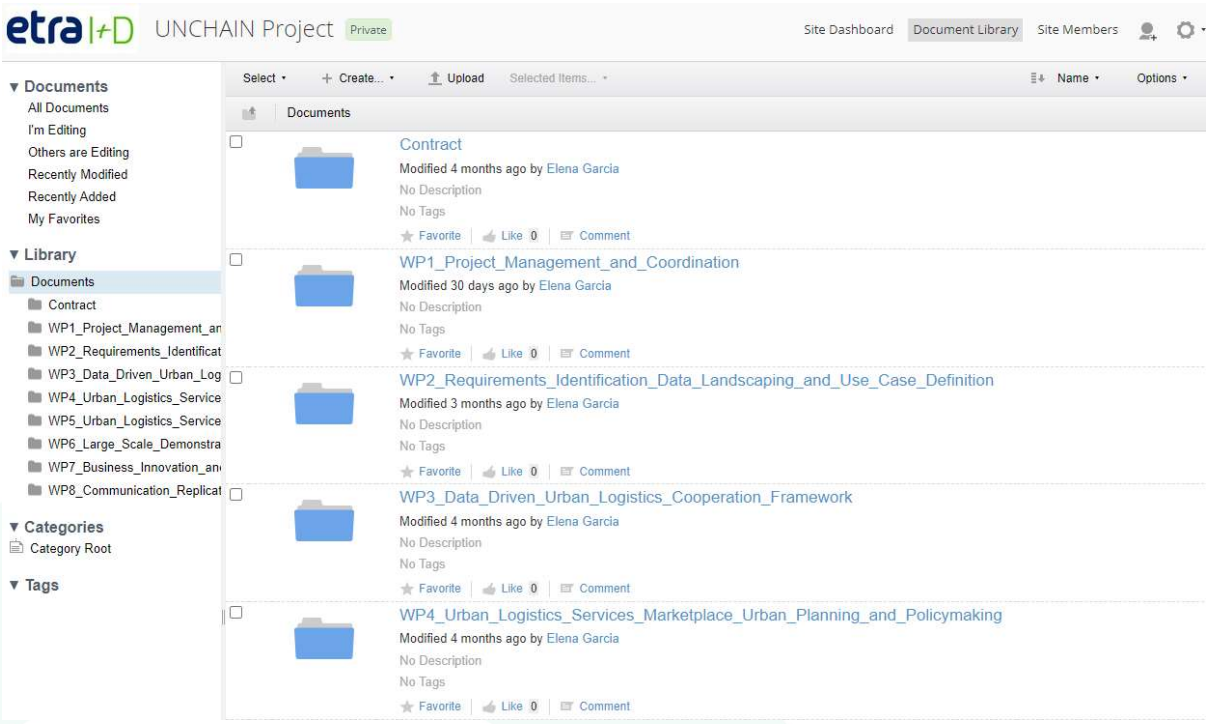

Figure 2 – UNCHAIN Repository overview.

3.2.3. Open Access publications

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

In addition, the UNCHAIN project may decide to provide other outputs and results as Open Access. As per the "Open access to publications and data in Horizon 2020 Fact sheet" (European IPR Helpdesk, 2020), there are two main routes for Open Access to scientific peer-reviewed publications.

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

Within the EU-funded UNCHAIN project, both routes of publishing (gold and green open access) are available and not mutually exclusive. Each beneficiary has the liberty to choose

the most appropriate approach for their respective publications. Presently, parallel publishing is the preferred strategy for providing open access, enabling consortium members to publish their findings in scientific journals of their choice for maximum impact while ensuring optimal dissemination through open access. It is noteworthy that most academic journals support either gold, green or hybrid open access. Hence, UNCHAIN beneficiaries can select the most appropriate publishing venue according to their preference. All publications will be assigned a DOI, making them easily discoverable and citable.

It is essential to recognize that the Open Access requirement does not obligate beneficiaries to publish their results. The decision to publish is entirely up to them. Open Access becomes an issue only if beneficiaries choose to publish their research results. The following graph illustrates the decision related to research results and possible paths for publication or other options.

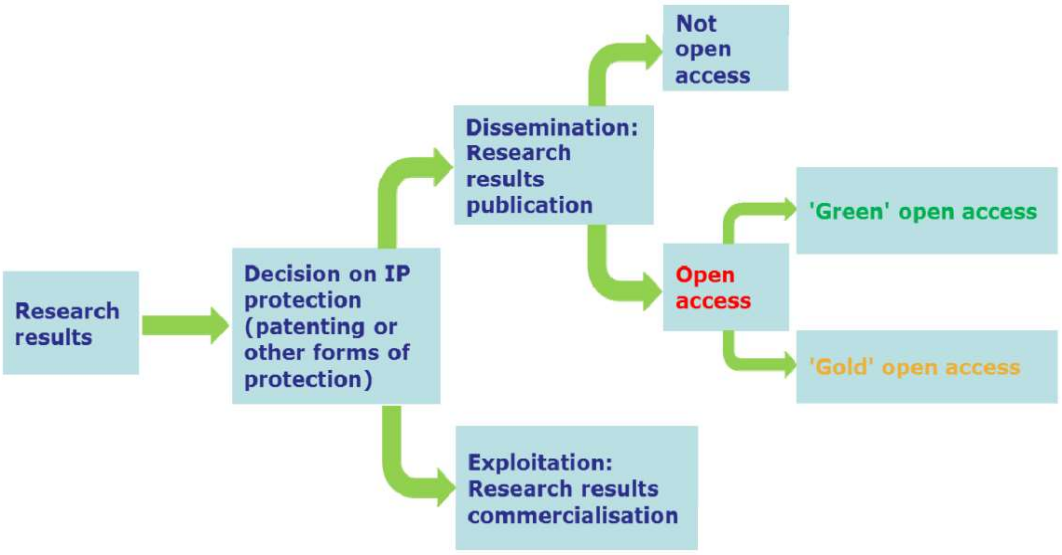

Figure 3 – Open access to scientific publication and research data in the wider context of dissemination and exploitation. Source H2020 Open Access portal.

3.3. Making data interoperable

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

3.4. Increase data re-use (through clarifying licenses)

As previously indicated, data will be treated on a case-by-case 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).

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 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., warrantied 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.

3.5. Website and public deliverables

The UNCHAIN website describes the general approach of the project, provides a short description of the project’s objective and its living labs’ description. In the following months, further development will be done and more information included.

After submission and approval from the EC, the project's public deliverables will be downloadable from the website (https://unchainproject.eu/) in a PDF format, while confidential deliverables will be kept in the repository accessible only to authorised users.

During the lifetime of UNCHAIN, data collected or generated by the project will be stored and systematically organised in the official project repository, Alfresco. The Documents Library is a project internal area only accessible to internal users to securely store and share files, making them available to the whole consortium.

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 already stated in previous sections, the DMP presented in this deliverable is just a 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. The 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 (ETRA) – prepares and leads related discussions at the relevant project meetings and to maintain the project document repository, and supports the TC and the PSC in making decisions regarding data integrity and compatibility.
- Technical Coordinator (ULANC) – identifies data collected by the project and technical project outcomes eventually suitable for publication and ensures dataset integrity and compatibility for its use during the project lifetime by different partners and WP Leaders.
- Dissemination (POLIS) – identifies publications suitable for submission in the considered repositories and maintain UNCHAIN inputs for the Open Access.
- Each individual partner – identifies own project results suitable for publication and shares the published scientific articles in advance with project coordinator, technical coordinator and dissemination manager.

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

Moreover, each UNCHAIN 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. Validation and registration of datasets and metadata is the responsibility of the partner that generates the data in a Work Package (WP). Metadata constitute an underlying definition or description of the datasets and facilitate finding and working with particular instances of data.

Additional responsibilities undertaken by the UNCHAIN project partners include:

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

5. Data Security

ZENODO and Alfresco repository will ensure secure and safe storage of both public and non-public data respectively.

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.

The project repository is hosted on a private internal server with local backup mechanism (managed by the project coordinator).

The server 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).

If required by the end users, with the aim to assure data privacy, the services developed in UNCHAIN will support advanced data anonymization and encryption mechanisms following the IDSA standard. To ensure data security, 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 (and stakeholders that have active involvement in the project activities) 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 and not allowing the data to leave their premises unencrypted;
- Store data in at least two separate locations to avoid loss of data;
- Limit the use of USB flash drives;
- Label files in a systematically structured way in order to ensure the coherence of the final dataset.

6. Ethical aspects

No data should be processed without undergoing some ethical considerations first. These considerations help to enforce privacy regulations and beyond to ensure no rights are violated and all data has been obtained in a consensual manner.

6.1. Procedures to ensure data privacy

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

- Will not be handled out to third parties outside UNCHAIN.
- Will not be exploited or commercialised.
- 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 2, 6 and 7. 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 anonymisation techniques are being applied to protect data confidentiality.

6.2. Ethical considerations

During the lifetime of the project GDPR will be in force. Furthermore, 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 UNCHAIN project and the purpose of pilot demonstrations and studies.

In case of specific user studies:

- The purpose and procedure of the research will be introduced in an understandable way.
- It will be emphasized that it is the potential participants' choice whether or not to participate in the study.
- All participants will be 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 anonymisation and data storage and security.

- Participants will be informed about duration and effort to participate in any research.
- In any survey/interview people will be informed what kinds of questions we plan to ask, and UNCHAIN will make it clear that people can choose not to answer questions.
- Participants will be 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 will be provided.
- Risks and benefits will be explained.
- If applicable, arrangements for insurance coverage for participation will be described.
- Participants will be made aware of the complaint procedure.

7. Conclusions

This document sets the guidelines and recommendations to be followed in order to make the project research data Findable, Accessible, Interoperable and Reusable (FAIR) and therefore to contribute to knowledge discovery and innovation.

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

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

8. References

CERN. (2023). Home | CERN. https://www.home.cern/

ETRA. (2023). Alfresco respository.

European Commission. (2016). H2020 Programme Guidelines on FAIR Data Management in Horizon 2020.

European Commission. (2021). Horizon Europe Data Management Plan Template.

European Data Portal. (2018). Open Data Goldbook for Data Managers and Data Holders. http://www.europeandataportal.eu

European IPR Helpdesk. (2020). Fact Sheet Open Access to scientific publications and research data in Horizon 2020: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-

MEISTER Project. (2018). MEISTER – Mobility Environmentally-friendly. https://meisterproject.eu/

MOMENTUM Project. (2022). MOMENTUM Project.

Open Archives. (2023). Open Archives Initiative Protocol for Metadata Harvesting. https://www.openarchives.org/pmh/

PrepDSpace4Mobility Project. (2022). PrepDSpace4Mobility Project.

SUaaVE Project. (2019). SUaaVE Project.

USERCHI Project. (2020). USERCHI Project.

Zenodo. (2023). Zenodo. https://about.zenodo.org/policies/