30 November 2024

SUM4Re

Creating materials banks
from digital urban mining

D16.2 Data management plan (version 1)

VERSION 1.0

PUBLIC

Associated with document Ref. 101129961

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

Funded by the European Union

SUM4Re

Creating materials banks from digital urban mining

Deliverable ID	D16.2
Deliverable name	D16.2 Data management plan (Version 1)
Lead partner	UVIGO
Contributors	-
Beneficiary	-
Distribution List	Consortium

PUBLIC

Funded by the European Union

DOCUMENT INFORMATION

Document name	D16.2 Data management plan (Version 1)
Version	1.0
Due date	30/11/2024
Report date	30/11/2024
Number of pages	31
Lead Author	Ana Sánchez (UVIGO)
Other Authors	-
Dissemination level	Public

REVISION HISTORY

Rev.	Date	Description of revision
0.1	07/10/2024	Initial version
0.2	21/11/2024	Version after technical revision
1.0	30/11/2024	Final version (approved)

DOCUMENT APPROVAL

Written	Ana Sánchez (UVIGO)	28/10/2024
Revised	Petr Hradil (VTT) Ignacio Piñero (TECN) Inès Kahin (R2M)	12/11/2024 15/11/2024 20/11/2024
Approved	Pedro Arias Sánchez (UVIGO, Project Coordinator)	30/11/2024

EXECUTIVE SUMMARY

The SUM4Re project, funded under Horizon Europe (Ref. 101129961), aims to advance circular construction practices, reduce construction and demolition waste, and promote sustainable resource use within the built environment. To achieve these objectives, the Data Management Plan (DMP) outlines procedures for the effective collection, storage, and management of research data, ensuring that all data generated aligns with the FAIR principles. This plan details the methods for making data findable, accessible, interoperable, and reusable, supporting transparency and open access in research.

Data security and participant privacy are central components, with procedures established to protect personal and sensitive information, as required by GDPR and Horizon Europe standards. Roles and responsibilities for data management are clearly defined across the SUM4Re consortium, and adequate resources are allocated to support ongoing data curation and preservation throughout the project's lifespan.

The DMP also addresses ethical requirements, including informed consent and compliance with legal regulations, ensuring that all data-related activities adhere to best practices and legal obligations. This document will be regularly updated to accommodate changes in project needs and to maintain alignment with data management standards.

GLOSSARY

Terms, Abbreviations, and Acronyms

AI	Artificial Intelligence
ALTAI	Assessment List for Trustworthy Artificial Intelligence
BIM	Building Information Modelling
CC-BY	Creative Commons Attribution
CC-BY-SA	Creative Commons Attribution-ShareAlike
CDW	Construction and Demolition Waste
DMP	Data Management Plan
DPO	Data Protection Officer
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
IPR	Intellectual Property Rights
ISO	International Organisation for Standardisation
R2M	R2m Solution
TL	Task Leader
UVIGO	University of Vigo
WP	Work Package
WPL	Work Package Leader

TABLE OF CONTENTS

LIST OF TABLES

1. Data summary

The SUM4Re project, funded by Horizon Europe, will collect and manage diverse datasets to support the development of circular construction practices and the reduction of construction and demolition waste (CDW).

This section identifies SUM4Re project datasets in terms of format, origin, size, purpose, and utility. But first, a short explanation is provided for the understanding of these characteristics of each dataset, in general.

1.1. Types and formats of data generated or re-used

The SUM4Re project will generate and manage a diverse group of research data, all of which are crucial for achieving the project's objectives and supporting its impact. In the context of SUM4Re, research data refers to various forms of information—such as statistics, experimental results, measurements, observations from fieldwork, survey responses, and interview recordings—collected and analysed to underpin scientific reasoning, discussions, and decision-making processes. This research data will be systematically organised into datasets. A dataset[1], in this context, is a structured collection of related data elements associated with a unique body of work, designed to facilitate data analysis, sharing, and reuse.

The sections that follow provide an overview of the primary data types and formats expected to be used in SUM4Re, highlighting how they will contribute to meeting project goals and ensuring compliance with data management standards.

1.1.1. Data types expected to be generated

Throughout the project, the following main types of data will be generated, each serving specific roles in fulfilling project goals:

- **Personal Data:** During project interactions, basic personal information (e.g., names, contact details, professional backgrounds) of project consortium partners and stakeholders may be collected. This data collection is strictly limited to essential information required for project activities and will be handled in compliance with the General Data Protection Regulation (GDPR) and relevant national privacy regulations, ensuring data protection and ethical handling.
- **Research Outputs and Literature Reviews:** SUM4Re will produce various research outputs, including peer-reviewed publications, abstracts, posters, and presentations. All peer-reviewed publications will be made available through open access channels in alignment with Horizon Europe guidelines. Where open-access publication is not feasible, these outputs will be deposited in institutional or general-purpose repositories to maintain accessibility.
- **Technical Datasets**: This category encompasses diverse types of data generated or used within the project:
  - **Experimental data**: Data obtained directly from experimental activities, field measurements, and observations at project demonstration sites.
  - **Derived data**: Data generated from analyses, simulations, and interpretation of previous datasets, including processed results and conclusions drawn from raw or experimental data.
  - **Native data from tools and platforms**: This includes data inherently produced or logged by tools, platforms, and software utilized in the project.
  - **Training Data for AI Development**: SUM4Re will generate, collect and/or use datasets specifically for training AI models. These datasets will be curated to ensure

[1] H2020 Programme Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 Version 3.2 21 March 2017.

data quality, representativeness, and compliance with ethical standards, supporting reliable and unbiased AI model development.

1.1.2. Data formats and standards

To ensure data accessibility, compatibility, and compliance with FAIR (Findable, Accessible, Interoperable, and Reusable) principles, SUM4Re will adopt widely recognized data formats and adhere to established standards where applicable.

- **Data Formats:** Data formats refer to the structure in which data is encoded, stored, and exchanged. These formats ensure that data can be read, understood, and used by software tools and systems across different platforms. SUM4Re will employ formats that facilitate interoperability and support the specific requirements of the datasets generated or reused during the project. While raw test data may be generated in proprietary or specific formats, the focus will be on ensuring interoperability for results and processed data exchanged between project partners.
- **Standards:** Standards are agreed-upon guidelines or rules that define the format, structure, and methodologies for managing data. In SUM4Re, standards will ensure data quality, interoperability, and compatibility with industry practices, including compliance with Horizon Europe requirements. This includes, where relevant, ISO standards for BIM (e.g., ISO 19650) and standards for geospatial and environmental data. Metadata will be documented thoroughly to enhance data discoverability and reusability, adhering to FAIR principles and Horizon Europe's open data requirements.

A summary table for formats and standards (Table 1) will be maintained as part of this Data Management Plan and updated throughout the project as specific data formats and standards are defined and used. This table will provide a comprehensive overview of the data formats, their purpose, and their associated standards, with particular attention to ensuring interoperability for shared results.

As this Data Management Plan is designed as a living document, this table will be completed and updated in subsequent versions throughout the project's duration, reflecting ongoing developments in data collection and generation activities.

Table 1. Overview of Data Formats and Standards in SUM4Re

Id	Format/Standard	Description	Purpose/Use	Associated tools or systems	FAIR compliance
1		*What is it?*	*Why is it needed?*	*What software or hardware uses it?*	*How does it support FAIR principles?*
2					

1.2. Purpose of data generation or re-use and project objectives

The subsections below explain the strategic link between the collected data and the project objectives, highlighting how this data will contribute to advancing sustainable construction practices and ensuring alignment with European resource efficiency initiatives for a lasting impact.

1.2.1. Relation of data to project objectives

The diverse data types collected within SUM4Re—including technical data on material properties, structural integrity, environmental impacts, personal data from stakeholder interactions, and research outputs—will directly support the project's objectives. Technical datasets will provide the basis for actionable recommendations on reducing CDW, enhancing material reuse, and fostering secondary materials markets. Personal data will enable effective stakeholder engagement and relationship management, critical to achieving project goals and KPIs. Additionally, research outputs will validate the project's methodologies and disseminate

findings to a broader audience, reinforcing SUM4Re's impact on circular construction practices.

1.2.2. Contribution of data to circular construction practices

The data generated on reusable materials, stakeholder engagement, and efficient CDW management will underpin the development and refinement of circular construction models. These models will promote sustainable practices across various construction stages, from material selection and design to demolition and reuse. Through these efforts, SUM4Re aims to make meaningful contributions towards the European Union's resource efficiency goals and the broader adoption of circular construction principles.

1.3. Expected data volume

At this stage of the SUM4Re project, it is not yet possible to determine precise data volume projections, as data collection is still in its initial phases. The expected data volume will depend on the scope and intensity of project activities, including stakeholder engagement, technical data collection, and research outputs. Estimates will be updated periodically as data generation progresses and more accurate measurements can be made.

1.3.1. Projected data size for generated data

Given the current phase of SUM4Re, projected data sizes are preliminary and will be refined as data is collected (Table 6). Data is expected to range from smaller datasets (e.g., personal and survey data) to potentially larger files from technical datasets, such as 3D models, sensor outputs, and BIM data. These initial estimates will be reviewed and adjusted throughout the project.

1.3.2. Storage considerations and impact of data volume

While specific storage requirements cannot yet be fully anticipated, the project will ensure that storage solutions are scalable to accommodate data growth as the project progresses. Consideration will be given to selecting repositories and storage platforms that can efficiently manage various data formats and volumes. Data storage will be designed to support accessibility and long-term preservation in line with FAIR principles, and adjustments to storage plans will be made as needed based on actual data generation.

1.4. Data origin and provenance

Data for the SUM4Re project will originate from multiple sources, including primary data collection, stakeholder contributions, and internal project analyses.

1.4.1. Sources of data collected within the project

The sources of data in SUM4Re include stakeholder and personal data, technical datasets, and research outputs. For detailed descriptions and categorisations of these data types, please refer to Section 1.1.1.

1.4.2. Data provenance and quality control measures

SUM4Re's approach to data management includes distinct practices for both provenance and quality control, ensuring that each dataset is traceable, reliable, and aligned with project standards. Table 2 and Table 3 summarise the main measures for each aspect.

- **Provenance Measures:** Data provenance practices will ensure that the history of each dataset is accurately recorded, from its origin through to any transformations it undergoes.
- **Quality Control Measures:** Quality control practices ensure that data collected is consistent and meets SUM4Re's standards (see deliverable D16.3 Quality Assurance Plan). The measures in Table 3 outline the project's approach to maintaining data quality.

These quality control measures aim to uphold the integrity of SUM4Re's datasets, ensuring that all data collected and processed is reliable, accurate, and aligned with project goals.

Table 2. Data provenance measures for SUM4Re

Provenance measure	Purpose
Metadata Documentation	Every dataset will include metadata that captures its origin, collection methods, and processing history.
Standardised Data Collection Protocols	Consistent protocols will be applied across data activities to ensure traceable and reproducible methods.
Personal Data Provenance	Personal data records will document consent, intended use, and security measures in line with GDPR.

Table 3. Quality control measures for SUM4Re's data

Quality control measure	Purpose
Cross-Verification of Technical Data	Data accuracy will be assessed through comparisons with established benchmarks and secondary sources, where possible.
GDPR and Ethical Compliance Checks	Compliance checks will ensure that personal data is handled securely and only used within project scope.

1.5. Data utility beyond the project

The datasets collected and organized in SUM4Re will provide significant long-term value to stakeholders beyond the immediate project consortium. These key datasets will be structured to support future use by researchers, policymakers, and industry professionals in areas such as circular construction, CDW management, and sustainable building practices. By establishing a repository of data focused on circular economy principles in construction, SUM4Re aims to enhance knowledge-sharing and practical applications across Europe's construction sector, fostering innovations and best practices.

In considering data utility in the long term, the project acknowledges the potential for integration and reuse to catalyse new insights and advancements. Data utility will be a driving factor in determining the final data management plan, influencing decisions related to:

- Long-term preservation and curation
- The types of data to be shared (e.g., raw data, processed, or analysed datasets)
- The usability and accessibility of data for third parties (e.g., data licensing and usage restrictions).

1.5.1. Potential external stakeholders and beneficiaries

Potential users of SUM4Re data include:

- **Academic Researchers**: Scholars and research institutions focused on circular economy, material science, and construction sustainability will benefit from technical datasets and insights on CDW management and material reuse.
- **Government Bodies and Policymakers**: Local and regional authorities involved in construction, waste management, and environmental regulation can leverage SUM4Re data to inform policies and regulatory frameworks supporting resource efficiency.
- **Construction Industry Stakeholders**: Industry leaders, contractors, and construction firms interested in adopting sustainable practices and secondary materials in projects can use SUM4Re data for practical insights and benchmarks.
- **Standards and Certification Organisations**: Organisations focused on developing or updating standards related to sustainable building practices may find SUM4Re's data valuable in setting or refining criteria for material reuse and resource conservation.

1.5.2. Anticipated applications for external use

The data produced by SUM4Re is expected to support future projects and practices in circular construction, offering benchmarks and guidelines for sustainable materials usage. Anticipated applications for this data include:

- **Benchmarking and Best Practices**: Data will serve as a basis for establishing industry benchmarks and best practices for CDW reduction, material reuse, and lifecycle assessment in construction.
- **Policymaking and Regulations**: Data insights can inform EU-wide directives, policies, and resource efficiency standards by providing evidence of the benefits and feasibility of circular construction models.
- **Standards Development and Updates**: The project's datasets can support standards organisations in defining or enhancing certification criteria related to the circular economy, material recovery, and reuse.
- **Educational and Training Resources**: Academic institutions and training programmes may use SUM4Re data as educational resources, contributing to skills development and knowledge dissemination in sustainable construction.

1.5.3. Background Data Considerations

While SUM4Re does not anticipate re-using external data, the project may utilise background data provided by consortium partners as appropriate. Background data refers to any data, information, know-how, or intellectual property that:

1. Was owned or controlled by a partner prior to the start of the project; or
2. Is developed or acquired independently by a partner outside the scope of the project but introduced into SUM4Re by the owning partner.

All provisions established in the Data Management Plan will apply to any background data introduced to SUM4Re, ensuring alignment with FAIR principles and project objectives. Additionally, background data will be subject to the terms outlined in the Consortium Agreement, covering access rights, dissemination, and protection measures as required for responsible and secure data use.

1.6. Data management plan guiding principles

The Data Management Plan (DMP) for the SUM4Re project, funded under Horizon Europe, adheres to the principles of Open Access and FAIR data (Findable, Accessible, Interoperable, and Reusable), which support responsible data management throughout the project lifecycle. This plan outlines how data produced will be accessible for verification, re-use, and preservation. The main principles for the SUM4Re project DMP are as follows:

- **Continuous Updates and Compliance with Horizon Europe Standards**

  The DMP is a dynamic document that will be updated throughout the project. An initial version will serve as the baseline, with adjustments made as data collection progresses and at significant project milestones. Updates will ensure alignment with the evolving needs of SUM4Re and compliance with Horizon Europe requirements, with data made "as open as possible, as closed as necessary".

- **Commitment to Open Science**

  SUM4Re will adopt open science[2] practices to offer access to its scientific results and engage the civil society and end users in the co-creation, in particular: (1) open access (OA) to research data via a trusted repository under the principle 'as open as possible, as closed as necessary'; (2) early and open sharing of research (for example, through

[2] For more information, please visit: https://rea.ec.europa.eu/open-science_en

preregistration, registered reports, pre-prints, or crowd-sourcing); (3) research output management; (4) measures to ensure reproducibility of research outputs; providing OA to research outputs beyond publications and research data (5) participation in open peer-review; (6) citizen, civil society and end-user engagement (i.e., Advisory Board, targeted workshops & MOOCs for co-creation in the workforce upskilling & social awareness, engagement of the inhabitants & stakeholders in SUM4Re).

Commitment to Open Access

SUM4Re commits to making data openly accessible in a compliant, public repository whenever possible. Open Access routes include:

- **Green Open Access** (Self-archiving): Authors deposit peer-reviewed manuscripts in trusted repositories without embargoes.
- **Gold Open Access:** Publications are immediately accessible with eligible APCs (Article Processing Charges) covered by project funds, limited to full Open Access journals.

- **FAIR Principles**

  Data management aligns with the FAIR principles, requiring metadata-rich, interoperable formats. Datasets will be stored in trusted repositories, ensuring long-term accessibility for validation and further use.

- **Personal Data Protection and GDPR Compliance**

  The consortium is dedicated to protecting personal data in line with the General Data Protection Regulation (GDPR), ensuring that all personal data collected is securely processed and managed exclusively for project purposes. Any sensitive data is treated under strict guidelines to uphold participant confidentiality and data protection standards.

- **Data Ownership, Confidentiality, and Intellectual Property**

  Procedures for data collection, storage, access, and sharing policies, as well as protection, retention, and destruction, are aligned with EU standards as outlined in the Grant Agreement.

This Data Management Plan has been prepared following the template provided in the "Horizon Europe – Data Management Plan Template". This is an official project deliverable (D16.2), due in month 6 (November 2024), but it will undergo updates throughout the project duration. This initial version will be adapted as significant changes arise, and periodic updates will occur at major reporting stages.

1.7. Data management strategy

This strategy outlines the methods and processes for managing, preserving, and sharing data in a way that maximizes accessibility and usability throughout and beyond the project lifecycle. Each dataset will be managed according to the following elements:

- **Dataset Reference and Name:**

An internal identifier will be assigned to each dataset in the project, using the following format: '*ProjectName_#Task_PartnerShortName_DataSubset_DatasetName_Version_Date*' (Table 4). This naming convention allows for clear identification and tracking of datasets, facilitating data management across project teams.

- **Dataset Description:**

Each dataset will include a detailed description that outlines its origin, nature, and potential usefulness to the research community or industry stakeholders. This will specify whether the data supports any scientific publications and note the existence of similar data, identifying opportunities for integration and reuse.

- **Standards and Metadata:**

SUM4Re will adhere to existing standards whenever possible to ensure consistency and compatibility. If no applicable standards exist, the project will establish internal guidelines to outline the metadata to be included, ensuring that each dataset is thoroughly documented to facilitate reuse.

- **Data Sharing:**

The strategy for sharing data includes clear protocols on access procedures, any applicable embargo periods, and technical mechanisms for data dissemination. Necessary software or tools required for accessing and reusing the data will be provided to the intended users, and datasets will either be openly accessible or restricted to specific project partners and stakeholders, depending on data sensitivity.

- **Archiving and Preservation:**

SUM4Re will implement long-term archiving solutions, specifying the duration of data preservation, anticipated storage volumes, and associated costs. Plans will be put in place to secure adequate funding for ongoing data storage and backup to ensure data remains accessible beyond the project's completion.

Table 4. Dataset reference specifications

Field	Description
ProjectName	Refers to the project name (SUM4Re).
#Task	Refers to the specific task number associated with the data generation (i.e., T16.1).
PartnerShortName	Identifies the data custodian (e.g., WP Lead or Task Leader).
DataSubset	Refers to the database or subset within the repository where data is stored.
DatasetName	The name of the specific dataset.
Version	Indicates the release or version of the dataset.
Date	The storage or modification date.

1.8. Roles and responsibilities

Data management in the SUM4Re project is an integral part of the entire project lifecycle, and the activities to this regard need to be coordinated and monitored at project level, but also at work area and work package level. The coordinator will hold high-level responsibilities, specifically in T16.2, T17.2, and T18.2, ensuring overall governance. Main researchers or dataset owners will be accountable for low-level responsibilities, including designing the dataset lifecycle, making key decisions on long-term provisions, storage, recovery security, and handling sensitive data. Oversight of potential impacts in dissemination and exploitation will be regularly monitored to maximise project outcomes. Datasets will include standardized metadata (DataCite Metadata Schema as default) and will be deposited (Ref. Grant Agreement).

This approach distributes responsibilities among the Project Data Managers, Work Package Data Managers, and the Dissemination Manager, as outlined in Table 5, with each role covering distinct areas of accountability.

Table 5. Data management roles

Role	Contact person	Partner
Project Data Managers	**Project Coordinator**, Pedro Arias Sánchez (parias@uvigo.gal)	UVIGO

	Project Manager, Ana Sánchez Rodríguez (sum4re@uvigo.gal)	
Work Package Data Managers	WPs leaders, ref. to D16.1	WPs leaders, ref. to D16.1
Dissemination Data Manager	Inès Kahin, (ines.kahin@r2msolution.com)	R2M

1.8.1. Project Data Managers

The Project Data Managers are responsible for overseeing the technical implementation of the SUM4Re DMP and ensuring consistency across WPs. Their primary tasks include:

- Coordinating the development and updates of the DMP in collaboration with project management and technical partners.
- Managing and maintaining the technical infrastructure for the SUM4Re metadata catalogue.
- Monitoring data collection, publication activities, and deadlines, and providing timely reminders to WP Data Managers.
- Offering support and guidance to WP Data Managers to ensure compliance with data management standards.

1.8.2. Work Package Data Managers (WP Leaders)

Each WP Leader serves as the Data Manager for their respective work package, tasked with implementing the DMP policy and ensuring compliance within their WP. Their main responsibilities include:

- Gathering required data information and clarifications from partners within the WP.
- Ensuring that metadata for all data used and produced at the WP level is accurately documented and aligned with the SUM4Re DMP policy.
- Monitoring open-access requirements, ensuring that open results are deposited in the designated repository or an OpenAIRE[3]-compliant repository as needed.
- Providing support within their WP for data publication and assisting partners in meeting data-sharing requirements.
- Contributing to the DMP deliverables by summarizing WP-specific datasets that need to be listed in the DMP.
- Sending reminders to WP partners to ensure timely data submission and compliance.

1.8.3. Dissemination Manager

The Dissemination Manager plays a critical role in managing the publication and open-access policies for project outputs. Key responsibilities include:

- Ensuring that all publications comply with Horizon Europe open-access policies.
- Monitoring that self-archiving publications (Green Open Access) are correctly deposited in repositories and that metadata is accessible.
- Verifying that research data associated with publications is made available in repositories and properly linked to the respective publication.
- Providing assistance to project partners in selecting the most suitable publication path (Green or Gold Open Access).
- Sending periodic reminders to partners to ensure compliance with open-access requirements and proper linking in OpenAIRE where applicable.

[3] https://guidelines.openaire.eu/en/latest/

1.9. Overview of project datasets

This section presents a summary of the key datasets to be collected, generated, and managed throughout the SUM4Re project. Table 6 outlines each dataset in terms of its identifier, description, responsible partner, associated work package or task, data format, origin, and estimated size. This overview serves as a reference for dataset characteristics and contributes to a structured data management approach that aligns with the objectives and data utility needs of SUM4Re.

As this Data Management Plan is designed as a living document, this table will be completed and updated in subsequent versions throughout the project's duration, reflecting ongoing developments in data collection and generation activities.

Table 6. Overview of SUM4Re Project Datasets

Id	Dataset name	Description	Partner	WP/Task	Format	Origin	Size
3							
4							
5							

2. FAIR data

The FAIR principles—Findable, Accessible, Interoperable, and Reusable—are designed to improve the usability and longevity of research data. Adhering to these principles ensures that data generated within the SUM4Re project is not only available to project partners but can also be leveraged by external stakeholders, contributing to broader impacts within the field of circular construction.

In SUM4Re, applying FAIR principles involves implementing structured metadata, open-access repositories, and standardised formats to make data easier to find and access. Data will be organised to support interoperability with existing datasets and tools, while specific guidelines will be followed to facilitate data reuse across sectors. Through these efforts, SUM4Re aims to maintain high data quality standards that allow the project's datasets to serve as a valuable, long-term resource for research, policy, and industry innovation in sustainable construction practices.

This section addresses the assessment of FAIR principles within the SUM4Re project and provides additional guidelines and context for their application.

2.1. Making data findable, including provisions for metadata

In accordance with the FAIR principles (Findable, Accessible, Interoperable, and Reusable), SUM4Re will ensure that data generated throughout the project can be easily located, accessed, and reused by both internal and external stakeholders. Key actions to enhance data findability include the assignment of persistent identifiers, structured metadata documentation, consistent naming conventions, and indexing for searchability in trusted repositories.

2.1.1. Dataset Identification, Persistent Identifiers, and Repository Selection

To ensure compliance with FAIR principles and Horizon Europe requirements, each dataset generated within the SUM4Re project will be assigned a persistent identifier, such as a DOI, upon deposit in a suitable repository. This enables long-term accessibility, citation potential, and traceability, making each dataset findable and reusable.

Datasets will include standardised metadata, following the DataCite Metadata Schema as the default, to enhance discoverability and interoperability. Deposits will be made in trusted repositories, such as institutional resources, with Zenodo as the default option and the European Open Science Cloud (EOSC) as an additional alternative. These platforms support machine-readable metadata and comply with OpenAIRE standards, ensuring findability and alignment with Horizon Europe's guidelines for open access.

For the data that should be open, as soon as possible and within the deadlines set out in this DMP, ensure open accessibility —via the repository — to the deposited data, under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a licence with equivalent rights (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles (in particular machine-actionable).

In cases where sensitive or restricted-access data is involved, the principle of "as open as possible, as closed as necessary" will apply, with access levels clearly defined in the Data Management Plan (DMP). This repository selection approach ensures that datasets are preserved and accessible in the long term, supporting SUM4Re's open science objectives.

2.1.2. Metadata Standards and Content

The project will utilise the DataCite Metadata Schema as the default standard, following Horizon Europe guidelines (Ref. Grant Agreement). Metadata associated with each dataset will include (Ref. Grant Agreement, Article 17):

- Datasets (description, date of deposit, author(s), venue and embargo)

- Horizon Europe or Euratom funding
- Grant project name, acronym and number
- Licensing terms
- Persistent identifiers for the dataset, the authors involved in the action, and, if possible, for their organisations and the grant. Where applicable, the metadata must include persistent identifiers for related publications and other research outputs.

This structured approach to metadata ensures alignment with FAIR principles, allowing datasets to be indexed by search engines and harvested by aggregators. Where discipline-specific standards are required for technical datasets (e.g., BIM standards or ISO standards for environmental data), these will be specified to facilitate data integration with industry tools.

2.1.3. Naming Conventions and Version Control

As specified in section 1.7, SUM4Re will apply a standardised naming convention for all datasets and files to ensure consistency and traceability across project outputs. This naming structure provides clear identification and facilitates version tracking, ensuring transparency and accessibility for project partners and external users. The naming convention will follow this structure (Table 4):

'ProjectName_#Task_PartnerShortName_DataSubset_DatasetName_Version_Date'

2.1.4. Search Keywords and Discoverability

To optimise discoverability, datasets will be annotated with relevant search keywords in the metadata, such as "circular construction," "secondary materials," and "construction and demolition waste", among others. These keywords will align with SUM4Re's thematic focus, enhancing dataset visibility across repositories and enabling users to locate data through standard or advanced search functionalities.

2.1.5. SUM4Re datasets overview – Data findable

Table 7 identifies SUM4Re datasets in terms of metadata provisions as well as search keywords.

Table 7. Metadata of SUM4Re datasets

Id	Dataset name	Metadata (Information available)	Metadata standard	Identifier	Search keywords
1					
2					
3					

As this Data Management Plan is designed as a living document, this table will be completed and updated in subsequent versions throughout the project's duration, reflecting ongoing developments in data collection and generation activities.

2.2. Making data accessible

To maximise the impact of SUM4Re research data, the project adopts the principle "as open as possible, as closed as necessary," in accordance with Horizon Europe guidelines. Data will be shared within the consortium and beyond, with measures in place to protect sensitive or proprietary information.

2.2.1. Repository and Data Access

Data will be deposited in trusted repositories such as Zenodo (https://zenodo.org) and institutional resources that comply with OpenAIRE standards. These repositories assign

persistent identifiers (e.g., DOIs) to ensure stable referencing and long-term availability. Arrangements with these repositories guarantee that the data and its identifiers are accessible and resolvable to digital objects.

2.2.2. Data Accessibility and Restrictions

SUM4Re datasets will generally be made openly available. However, some datasets may be subject to restricted access due to legal, ethical, or contractual reasons, such as protecting intellectual property or ensuring participant confidentiality. In these cases, access will be evaluated and approved through appropriate measures, potentially involving a data access committee to review requests.

If an embargo is necessary (e.g., to publish findings or secure intellectual property protection), the reason and duration will be clearly stated, aiming for the shortest possible embargo period. Data will be accessible through standardised protocols, and methods or software needed for access, as well as relevant documentation, will be provided.

2.2.3. Methods and Software for Data Access

For datasets that require specific software or methods for access, information and documentation will be included. When feasible, open-source software or reference material will be provided to facilitate data use and verification. If required, licensing terms will be detailed, and access restrictions will be managed through agreements such as non-disclosure agreements (NDAs).

2.2.4. Metadata Accessibility

Metadata will be openly available and licensed under a public domain dedication (CC0) as per the Grant Agreement. Metadata will include information to enable users to access the data, even if the data itself is restricted. The metadata will remain available even after the data is no longer accessible, ensuring continued discoverability and reference.

The project will regularly review data access arrangements to ensure they meet legal, ethical, and contractual requirements while maximising data utility.

2.3. Making data interoperable

SUM4Re will adhere to FAIR guiding principles to make data interoperable[4], enabling seamless data exchange and reuse both within and across disciplines. To achieve this, standard vocabularies, formats, and methodologies will be followed as outlined in Horizon Europe and OpenAIRE guidelines[5].

2.3.1. Data and Metadata Standards

To ensure interoperability, SUM4Re will use community-endorsed standards for data and metadata, such as the DataCite Metadata Schema. The project will employ formats such as CSV for tabular data, LAS and OBJ for 3D modelling, JPEG and PNG for images, and JSON for structured sensor data. Standards like CEN standards, including CEN TC 442 for semantic data modelling in construction, and the ISO 19650 standard for BIM will also be applied where relevant, allowing for interdisciplinary compatibility.

SUM4Re will observe OpenAIRE guidelines for online interoperability, including:

- OpenAIRE Guidelines for Literature Repositories
- OpenAIRE Guidelines for Data Archives
- OpenAIRE for other research products

[4] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[5] https://guidelines.openaire.eu/en/latest/

These guidelines ensure that data and metadata comply with widely accepted practices, supporting integration with the European Open Science Cloud (EOSC) and facilitating data discovery.

2.3.2. Use of Ontologies and Vocabulary Mapping

In cases where SUM4Re creates project-specific vocabularies or ontologies, mappings to more commonly used and accepted ontologies will be provided to ensure broader interoperability. If uncommon or novel vocabularies are used, the project will openly publish these to allow for reuse, refinement, and extension by the research community. Ontologies related to material properties and environmental performance will adhere to recognised CEN and ISO standards to ensure consistency.

2.3.3. Qualified References

SUM4Re will include qualified references in the metadata, linking datasets to other relevant research outputs both within the project and from prior research where applicable. These references will facilitate data contextualisation and enhance reusability by clearly indicating relationships between different datasets. References will also connect to related AI models and analysis techniques described in the Grant Agreement to promote comprehensive understanding and reusability.

Further details on interoperability practices will be provided in updated versions of the Data Management Plan as the project progresses and new datasets are defined.

2.4. Increase data re-use

SUM4Re is committed to maximising the reusability of its datasets, in line with the FAIR guiding principles and Horizon Europe's open data requirements. To this end, the project will implement clear licensing provisions, provide comprehensive documentation, and ensure data quality through internal review processes.

2.4.1. Licensing Provisions and Data Usability

SUM4Re will use standard Creative Commons licenses to protect the ownership of datasets while enabling broad re-use. Where appropriate, licenses such as CC-BY (Attribution) and CC-BY-SA (Attribution-ShareAlike) will be applied, allowing others to freely use, share, and adapt the data while ensuring proper attribution and adherence to similar licensing terms. In cases where data sensitivity or intellectual property concerns arise, more restrictive licenses, such as CC-BY-NC (NonCommercial), will be considered.

Some references for selecting an adequate license are:

- Public License Selector: A tool to assist in choosing the appropriate license
- B2SHARE License Wizard: This tool provides guidance and facilitates the selection of a suitable license for research data

It is recommended to use Creative Commons Attributions, such as **CC-BY-SA-NC-ND** (ShareAlike, NonCommercial, NoDerivs), as applicable. Provisions to specify required data citations may also be considered in the metadata, ensuring proper credit is given to original data creators.

Access to underlying data will be provided immediately upon publication or within a maximum embargo period of six months if necessary to protect intellectual property or align with publication policies, as recommended by the European Commission. The European Commission's template letter for requesting publisher amendments[6] will be used if needed to negotiate open access terms.

[6] https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-oa-guide-model-for-publishing-a_en.pdf

2.4.2. Documentation for Data Validation and Re-use

Comprehensive documentation will be provided to facilitate data validation and re-use. This includes:

- **Readme files** outlining data collection methodologies, data cleaning processes, and analysis procedures
- **Codebooks** detailing variable definitions, units of measurement, and data structures
- **Software documentation** and references to necessary tools for data access, with open-source software provided where feasible

2.4.3. Quality Assurance and Internal Review

Quality assurance measures will be in place throughout the project to maintain high data standards. An internal peer review process will be conducted for all deliverables. This process will involve assigning internal reviewers from project partners to assess deliverable quality, ensuring compliance with FAIR principles.

2.4.4. Embargo Periods and Restrictions

While SUM4Re aims to make data as openly accessible as possible, certain datasets may be subject to restrictions due to:

- Embargo periods imposed by journal publication policies (e.g., for Green open access)
- Intellectual property protection or data sensitivity considerations
- Project-specific conditions agreed upon by consortium members

2.4.5. Duration of Data Re-usability

Data will be made re-usable for a minimum of **one year** following the project's completion, with extensions considered based on data demand and relevance. Metadata will remain available even after data access ends to ensure long-term discoverability.

2.5. DMP Review process and Data inventory

This Data Management Plan will be treated as a living document, subject to periodic review and updates throughout the project's duration. This ongoing review process ensures that the DMP remains aligned with the project's evolving data needs and adheres to FAIR principles.

- **Internal Quality Evaluation and Reporting**

An internal quality evaluation process will be implemented to assess both project data and processes continuously. This process will involve regular data quality checks and reporting mechanisms managed by work package leaders (WPLs) and task leaders (TLs). Updates to the DMP will be coordinated with the overall project management strategy, ensuring that any new data requirements or changes in data handling practices are promptly incorporated.

- **Data Inventory and Reporting**

All research data generated, as well as any scientific publications or outputs, will be documented and tracked using a **Data Inventory Table**. WPLs, TLs, and authors of publications will be required to update this inventory every six months. The table will capture essential details, such as data types, formats, metadata standards, licensing provisions, and links to any associated publications. These updates will ensure comprehensive documentation and enable effective data sharing and reuse.

The Dissemination and Communication Plan (refer to Deliverable D13.1) will also guide partners in reporting updates on articles, conference papers, and other dissemination activities. This coordinated approach ensures that all data-related outputs are appropriately documented and integrated into the project's open-access strategy.

- **Updating the Online Data Repository**

Periodic updates will also involve revising the content and structure of the online research data repository (e.g., Zenodo, institutional resources, or the European Open Science Cloud). This ensures that data collected and shared are up-to-date, accurately documented, and remain accessible for re-use. Each update will align with project milestones and significant data collection or publication events.

3. Other research outputs

Beyond data, SUM4Re will generate additional research outputs that are critical for achieving project goals. These outputs may include digital resources, such as software, workflows, protocols, models, and visualisations. In line with the principles and strategies outlined in earlier sections, SUM4Re will ensure these outputs are managed effectively to maximise their impact and reusability.

3.1. Management Strategy

The management of other research outputs will adhere to the same FAIR principles described in Sections 2.1–2.4. This includes ensuring findability through proper metadata, accessibility through open licensing where possible, interoperability using standardised formats, and reusability supported by comprehensive documentation.

3.2. Licensing and Repositories

Licensing and repository selection for these outputs will align with the practices outlined in Section 2.4. Where applicable, outputs such as software will be deposited in platforms like GitHub, with archiving in Zenodo to ensure long-term availability and proper citation.

3.3. Specific Considerations

While the general management practices apply, certain research outputs may require specific handling:

- **Software and Code:** Published under open-source licenses to facilitate reuse and integration.
- **Models and Protocols:** Documented with sufficient detail to enable replication and adaptation by other researchers.

Any physical outputs, if generated, will be handled in accordance with institutional policies, with an emphasis on reusability and proper documentation.

4. Allocation of resources

Ensuring that data and other research outputs in SUM4Re are FAIR will involve both direct and indirect costs. These expenses are eligible for reimbursement under the Horizon Europe grant, provided they align with the conditions set in the Grant Agreement.

The costs for making data and outputs FAIR include:

- **Metadata Creation and Curation:** Time and resources allocated to creating detailed metadata for datasets to ensure discoverability and reusability.
- **Repository Use and Data Archiving:** The primary repository for data storage will be Zenodo, which offers services free of charge. However, other indirect costs may include ensuring secure storage, implementing anonymisation procedures for sensitive data, and long-term archiving.
- **Open-access Publication Fees:** Fees associated with publishing in “Gold” Open Access journals will be covered by the project budget. Cost-sharing arrangements will be made among authors and their respective institutions for multi-authored papers.
- **Security Measures:** Implementing data security protocols, especially for sensitive or personal data, will be managed using institutional resources, minimising additional expenses.

Table 8 identifies allocation of resources for data management focusing of the costs. To cover these costs, SUM4Re will utilise project funds allocated for data management and leverage institutional support where necessary.

Table 8. Allocation of resources for SUM4Re datasets

Id	Dataset name	Estimated costs (total)	Estimated costs (long term provisions)	Mechanism to cover costs	Responsible
4					
5					
6					

As this Data Management Plan is designed as a living document, this table will be completed and updated in subsequent versions throughout the project’s duration, reflecting ongoing developments in data collection and generation activities.

4.1. Responsibility for Data Management

Responsibility for data management in SUM4Re is distributed among project partners as described in Section 1.8. The Project Coordinator will oversee high-level data management activities, while Work Package Leaders will manage data within their work packages. Task Leaders will handle day-to-day data-related activities, ensuring compliance with the DMP.

4.2. Long-term Preservation

Long-term preservation of datasets will be crucial for sustaining the project's impact. Key datasets identified for their long-term value, uniqueness, or potential reuse will be stored in trusted repositories like Zenodo and the European Open Science Cloud (EOSC). The decision on which data to preserve and for how long will be made collaboratively by the Project Coordinator and WPLs, considering the dataset's relevance and potential future use. Preservation plans will prioritise open formats and adhere to standards that ensure data accessibility over time.

Resources required for long-term preservation, such as storage and maintenance, will primarily be covered through institutional support, with contingency plans to address any unforeseen expenses. The value of preserved data will be continually assessed to ensure that only data with significant reuse potential is maintained.

5. Data security

SUM4Re will implement robust data security measures to ensure the protection, integrity, and safe transfer of all project data[7]. The following provisions will be in place:

- **Secure Storage and Archiving:** Project data will be stored securely in at least two separate locations: the SUM4Re SharePoint system, which will serve as the primary storage solution with restricted access to consortium members, and backup storage on institutional servers of participating partners. Trusted repositories, such as Zenodo and institutional resources, will be used for long-term data preservation and curation.
- **Data Recovery and Backup:** Regular backups will be scheduled to prevent data loss, and recovery plans will be in place to address any potential data breaches or losses swiftly. These backups will be verified periodically to ensure data can be restored as needed.
- **Secure Transfer of Sensitive Data:** Sensitive or personal data will be transferred using secure, encrypted methods, minimising the risk of unauthorised access. The use of unencrypted USB drives or insecure methods will be avoided.
- **Data Encryption and Access Control:** Where necessary, data will be encrypted, and access will be limited to authorised personnel only. Permissions will be reviewed periodically to maintain the highest level of security and to ensure that only relevant team members have access to sensitive information.

Table 9 outlines the data security provisions for SUM4Re, detailing the strategies for storage, recovery, handling of sensitive data, and long-term preservation.

Table 9. Overview of Data Security Provisions for SUM4Re

Id	Dataset name	Storage	Recovery	Sensitive data handling	Long-term provisions
1		*(SUM4Re's SharePoint, partner servers)*	*(Regular backups, recovery plans)*	*(Encryption, secure transfer methods)*	*(Zenodo, institutional repositories)*
2					
3					

As this Data Management Plan is designed as a living document, this table will be completed and updated in subsequent versions throughout the project's duration, reflecting ongoing developments in data collection and generation activities.

[7] Refer to https://www.openaire.eu/guides for additional information

6. Ethics

The SUM4Re project adheres to the highest ethical standards and complies with the General Data Protection Regulation (GDPR) and Horizon Europe guidelines, as well as relevant national and international laws.

An Ethics Mentor and a Data Protection Officer (DPO) have been appointed within the coordinator organisation (T16.1) to oversee and guide all ethical considerations throughout the project. The Ethics Mentor provides ongoing support, reviews all deliverables for ethical compliance, and ensures that project activities align with the Grant Agreement's ethics provisions. Regular meetings with the Ethics Mentor occur at least every six months, and any ethical issues are promptly addressed. In addition, ethics reports are prepared and updated throughout the duration of the project.

- **Informed Consent:** Informed consent for data sharing and long-term preservation has been (and will be) systematically included in all questionnaires conducted within the project, covering interactions with both internal and external stakeholders. These consent forms, along with detailed privacy policies and GDPR information, ensure that participants are aware of how their data will be used, stored, and shared. The Ethics Mentor also reviews these processes to maintain compliance with ethical guidelines.
- **Risk Assessment and Mitigation:** SUM4Re has conducted a preliminary assessment for Trustworthy AI (ALTAI) and will continue to monitor ethical risks throughout the project. A risk assessment and mitigation plan are in place, addressing concerns such as potential accuracy issues and the effect of environmental changes on AI systems' reliability. Regular updates to the risk assessment will be documented, and guidance will be provided by the Ethics Mentor.

Ethical considerations are embedded in the development and deployment of AI systems, ensuring transparency, fairness, and respect for human autonomy. SUM4Re will avoid bias, respect fundamental rights, and ensure that the AI systems do not negatively impact human well-being or the environment. Mechanisms are in place to report vulnerabilities or ethical concerns, and end-users will be informed about their interactions with AI through clear documentation and warnings.

6.1. Archiving and preservation

Research data and other outputs generated by SUM4Re will be archived and preserved according to ethical and legal standards. Data will be securely stored in trusted repositories, such as Zenodo and institutional resources, with long-term preservation strategies aligned with GDPR and FAIR principles. The Ethics Mentor will oversee these strategies, ensuring that data governance and archiving practices are compliant with all ethical guidelines.

Additionally, anonymisation measures will be applied to image datasets if personal data is inadvertently captured. Continuous monitoring of data management practices will be performed, with updates to the Data Management Plan (DMP) as needed to reflect new requirements or ethical considerations. The associated costs for data archiving and preservation will be supported by the project budget and institutional resources.

OpenAIRE infrastructure provides for a tool of pseudo-anonymization for text data that can be reached here: https://amnesia.openaire.eu/.

7. Other issues

At this stage, SUM4Re has not identified any additional national, funder, or sector-specific procedures for data management beyond those already outlined in the Data Management Plan. However, the project will remain vigilant and responsive to any relevant guidelines or requirements that may emerge.

If additional procedures become necessary, such as compliance with new regulations or sector-specific standards for managing construction and environmental data, they will be documented and integrated into the DMP accordingly.

8. Conclusion

The SUM4Re Data Management Plan developed in this document meets the requirements set out in the Grant Agreement by establishing a comprehensive and effective framework for managing research data in line with the FAIR principles and an open-access approach. This DMP not only facilitates adherence to Horizon Europe guidelines but also aims to maximise the impact and accessibility of research outputs generated throughout the project.

This document serves a dual purpose: Firstly, it provides a detailed analysis of SUM4Re's research datasets, emphasising long-term utility, reusability, and compliance with ethical and legal standards. Secondly, it offers clear guidelines and recommendations to ensure that data handling practices are robust, efficient, and aligned with best practices in data management and preservation. These guidelines are designed to support the project's objectives and facilitate seamless data exchange and reuse, both within the consortium and with external stakeholders.

As a living document, the DMP will be regularly reviewed and updated under the responsibilities outlined within WP16, WP17, and WP18. Updates will reflect new developments in data management practices, technological advancements, and evolving project needs, ensuring that the DMP remains relevant and effective. The Ethics Mentor and the project's Data Protection Officer will play crucial roles in ensuring that data governance, ethical considerations, and compliance with GDPR are maintained. In addition, Intellectual Property Rights (IPR) and any potential ethical issues will be monitored and managed in line with SUM4Re's policies and legal requirements.

In conclusion, this document adopts a lifecycle approach to research data management, detailing standardised curation, preservation, and access measures to ensure that the impact of SUM4Re extends beyond the consortium and the duration of the project. As the project progresses, further efforts will be made to refine and contextualise the management of datasets, ensuring continuous improvement and alignment with emerging standards and requirements.

SUM4Re

Creating materials banks
from digital urban mining

www.sum4re.eu

sum4re@uvigo.gal

SUM4Re

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

Funded by
the European Union