



Project Acronym: GREEN-LOG
Project Number: 101069892 (HORIZON-CL5-2021-D6-01 - Innovation action)
Project Full Title: Cooperative and Interconnected Green delivery solutions towards an era of optimized zero emission last-mile Logistics



DELIVERABLE

D5.2 – Data Management Plan 1st Version

Dissemination level	PU – Public
Type of Document	DMP – Data Management Plan (Report)
Contractual date of delivery	30/06/2023
Deliverable Leader	FUNDACIO EURECAT (EUT)
Status - version, date	Final, 30-06-2023
WP / Task responsible	WP5/T5.2 (EUT)
Keywords	Data Management Plan, Open Science, Data Protection

Executive Summary

The GREEN-LOG consortium shares the vision that large and unrestricted access to knowledge is a fundamental human value of scientific knowledge progress and dissemination and is essential for generating growth. Thus, open access will be the main driver for sharing scientific results under GREEN-LOG project. In this regard, GREEN-LOG ambition is to remove the industrial barriers of data-sharing to offer them the needed ecosystem to innovate, share ideas and knowledge that permit to advance sustainable transport and logistics services. Thus, GREEN-LOG embraces Open Science and Open Data as a mechanism to bring society a return of investment from our research.

The FAIR¹ Data Management guidelines are adhered to, as are the Guidelines to the Rules on the Open Access to Scientific Publications and Open Data Access to Research Data. The current data management plan describes: (i) the handling of research data during and after the end of the project; (ii) what data will be collected, processed, and generated; (iii) which methodology and standards will be applied; (iv) whether data will be publicly shared/made open access, and (v) how data will be curated and preserved (including after the end of the project). In terms of accessibility and exploitation, most datasets generated within GREEN-LOG, and as part of the demonstrations, will be available for research purposes to all partners, after being subjected to appropriate anonymisation and securitisation measures to guarantee privacy protection in accordance with General Data Protection Regulation (GDPR). Some datasets might remain private due to privacy, intellectual right protection or commercial interests. The next version of the DMP (due in M18) will provide more details on the translation of the GDPR to GREEN-LOG data collection and handling.

The description of the type of data and procedures to facilitate open science is depicted in this first version of the Data Management Plan. Updated versions of this Data Management Plan will be made available as the project progresses and more specifically at M18, M36 and M42.

The data to be collected and generated during the lifetime of the project includes network and freight models, land use data, supply/demand data, routing data, traffic data, qualitative and quantitative questionnaire and interview data, as well as assessment and evaluation outcomes.

The partners will also contribute to novel research work and transfer the generated knowledge to the general public and wider logistics community through the publication of the deliverables, webinars, scientific publications and other communication and dissemination materials.

It is envisioned to use a SharePoint repository with restricted access for members of GREEN-LOG consortium during the project lifetime to store some of the outcomes produced during the project such as project deliverables, interviews, reports, or scientific publications. This repository can store both confidential and public information under the supervision of the Data Protection Officers nominated by each partner and to be shared under NDA or bilateral agreements, if needed. However, the data sets to be used or to be generated during the project will be stored in the GREEN-LOG DataSpace to be developed within WP2, within different accessibility levels.

For the data selected to be public, after the anonymisation of any sensitive information, the datasets will be made open to the general public in Zenodo, under HE GREEN-LOG community and OpenAIRE community.

¹ F.A.I.R stands for Findable, Accessible, Interoperable and Re-usable

Open access publications will be made available in Zenodo, GREEN-LOG webpage and, also, in Open Research Europe.

Finally, for software components envision to be open source, the source code will be made available on GitHub.

Deliverable Leader (organisation)	FUNDACIO EURECAT (EUT)
Contributors (organisations)	All partners
Reviewers (organisations)	VANaPEDAL, UoW, INTRA
Approved by (organisation)	INTRA

Document History			
Version	Date	Contributor(s)	Description
0.1	06-05-2023	EUT	Table of Content
0.2	01-06-2023	EUT	Add general content
0.3	16-06-2023	EUT	Version ready for peer-review
0.4	27-06-2023	EUT, VANaPEDAL, UoW, INTRA	Comments from peer-reviewers addressed
0.5	29-06-2023	EUT, INTRA	Pre-submission review
1.0	30-06-2023	INTRA	Final QA'ed version

Table of Contents

EXECUTIVE SUMMARY	2
TABLE OF CONTENTS	5
TABLE OF FIGURES	7
LIST OF TABLES.....	8
DEFINITIONS, ACRONYMS AND ABBREVIATIONS.....	9
1 INTRODUCTION.....	10
1.1 Document scope.....	10
1.2 Document structure.....	11
2 DATA SUMMARY.....	12
2.1 Purpose and motivation	12
2.2 Dataflow.....	12
2.3 Legal framework	13
2.4 Data types, formats and size.....	15
2.4.1 Data types	15
2.4.2 Data format.....	17
2.4.3 Data sizes.....	17
2.5 Origin of data	17
2.6 Data repository	17
2.6.1 GREEN-LOG DataSpace.....	17
2.6.2 SharePoint.....	18
2.6.3 Zenodo	18
2.6.4 GitHub	18
3 FAIR DATA MANAGEMENT	19
3.1 Findability.....	19
3.2 Accessibility	20
3.3 Interoperability	21
3.4 Reusable data.....	21
4 ALLOCATION OF RESOURCES	22
5 DATA SECURITY	23
5.1 Data security as specified for restricted repository in SharePoint	23
5.2 Data security as specified for public repository in Zenodo	23
6 ETHICAL ASPECTS.....	25
6.1 Collecting personal data from the Living Labs.....	25
6.2 Information and consent form.....	26
7 CONCLUSIONS.....	27

REFERENCES	28
-------------------------	-----------

Table of Figures

FIGURE 1: DATA FLOW IN GREEN-LOG	13
--	----

List of Tables

TABLE 1: OVERVIEW OF DATA GENERATED OR USED DURING GREEN-LOG PROJECT	15
TABLE 2: OVERVIEW OF DATA TO BE COLLECTED IN GREEN-LOG PROJECT.	16

Definitions, Acronyms and Abbreviations

Acronym/Abbreviation	Title
API	Application Programming Interface
CA	Consortium Agreement
CERIF	Common European Research Information Format
DMP	Data Management Plan
DPO	Data Protection Officer
DPIA	Data Protection Impact Assessment
DOI	Digital Object Identifier
EU	European Union
EC	European Commission
FAIR	Findable, Accessible, Interoperable, Re-usable
GA	Grant Agreement
GDPR	General Data Protection Regulation
IPR	Intellectual Property Rights
JSON	Java Script Object Notation
M2M	Machine-to-Machine
NDA	Non-disclosure Agreement
REST	Representational State Transfer
RTD	Research and Technology Development

1 Introduction

1.1 Document scope

The GREEN-LOG project aims at developing Cooperative and Interconnected Green delivery solutions towards an era of optimized zero emission last-mile Logistics.

It will establish city platforms made up of inclusive stakeholder Urban Living Labs to foster social innovation, and design and implement cutting-edge delivery solutions to accelerate systemic changes to create last-mile delivery ecosystems that are economically, ecologically, and socially sustainable. The cutting-edge simulation environment for scenario building combining different solutions to enable the integration of last-mile delivery interventions is envisioned to achieve the highest possible impact on environmental sustainability and traffic reduction, also considering their financial viability.

GREEN-LOG offers Logistics-as-a-Service platforms for interconnected city logistics, automated delivery concepts using autonomous vehicles and delivery droids, cargo-bike-based innovations for sustainable micro-consolidation, and multimodal parcel deliveries integrating public transportation. The solutions are supported by networked city logistics dataspace that supply dynamic services for proactive ecosystem optimisation while respecting the interests of stakeholders including consumers, businesses, and the city.

The approach is deployed and validated in five Urban Living Labs: Athens, Barcelona, Flanders (Gent, Leuven and Mechelen), Oxfordshire and Ispra. Three follower cities – Arad, Helsingborg, and Valga – will be involved in testing the transferability of the proposed innovations.

GREEN-LOG will support knowledge sharing and research and innovation results uptake by society as part of the European Commission's plans Horizon Europe research projects. It is mandatory for the beneficiaries to manage the digital research data responsibly, in line with the FAIR principles. Data, either gathered or generated, is one of the valuable assets that will enable the scientific community to go beyond the results achieved by GREEN-LOG.

The starting point is to elaborate a Data Management Plan (DMP) and keep it updated throughout the course of the project. This is the first version of the GREEN-LOG DMP. After approval of the EC, it will be made public in GREEN-LOG website and in a public repository to be decided (most likely Zenodo). The DMP is considered to be a living document and it will be updated during the course of the project as the development of the data collection, curation and publication progress. It is envisioned to issue updated versions of this DMP in M18, M36 and M42 of the project. For the more advanced versions, GREEN-LOG will consider publishing the DMP in specialized journals like RIO².

The DMP ensures that data collected, processed, and analysed during the project lifecycle are handled efficiently, ethically, and in compliance with relevant regulations. This plan also addresses data sharing, preservation, and long-term access to facilitate reproducibility and maximize the impact of the research outcomes.

² <https://riojournal.com/>

The content of this DMP has been prepared considering the template of the “Guidelines on Data Management in Horizon Europe”³. For future versions of the DMP, it is envisioned to use an open access web tool such as ARGOS⁴ or DMPOnline⁵.

1.2 Document structure

The document consists of seven sections. Section 1 initiates with the project’s introduction. Section 2 describes the specificities of the GREEN-LOG DMP by revising the purpose and motivation for data collection and the flows data follows from the original source until the final repositories. Section 2 also provides an initial description of the data types and formats that will be handled during the project lifetime and concludes with the selection of the different data repositories. The next section specifies how the FAIR principles are applied to GREEN-LOG DMP. Section 4 reports on the resources allocated for the implementation of the DMP while section 5 revises data security measures to guarantee data protection. The activities in the Living Labs are subject to ethical principles and good practices described in section 6. Finally, the conclusions are presented in the last section (section 7).

³ https://www.openaire.eu/images/Guides/HORIZON_EUROPE_Data-Management-Plan-Template.pdf

⁴ <https://argos.openaire.eu/splash/>

⁵ <https://dmponline.dcc.ac.uk/>

2 Data Summary

2.1 Purpose and motivation

The overall motivation for data collection in GREEN-LOG is to assess the impact of the measures applied in the Living Labs and learn and provide insights that can be later used to accelerate the transition to greener and more sustainable logistics solutions.

Technology providers participating in the project require data to design and develop the technology applications included in GREEN-LOG solution as well as create the models or run simulations. Likewise, the design of business models and policies also requires data to assess their feasibility and sustainability. Finally, once the measures are applied in the Living Labs, data will be gathered to analyse the impact of GREEN-LOG solution in terms of economic, ecologic and social impact.

Beyond the scope of the project, the scientific community and technology providers can re-use the data gathered during the project as starting point for further research. Yet, the three follower cities involved in the project and any other that may be outreached during the course of the project can use the lessons learned generated by the project and applied or adapt GREEN-LOG solution or policies starting from the transferability guidelines outcomes.

Only data that is needed to perform project activities will be collected. The collection of sensitive data will be performed only if strictly necessary and with informed consent. Sensitive data will not be made public.

2.2 Dataflow

GREEN-LOG gathers data from 5 demonstration areas (Living Labs) which are used to evaluate last mile interventions and provide qualitative and quantitative information on the results of the local solutions that are implemented. Most datasets used or generated within GREEN-LOG and produced in the demonstration pilots will be available for research purposes to all partners, after being subjected to appropriate anonymization and protection measures. However, some restrictions may apply to specific data due commercialisation potential or Intellectual Property Rights (IPR) that will be handled either through Consortium Agreement (CA) or bilateral Non-disclosure Agreements (NDAs). GREEN-LOG data goes beyond purely data sets and also includes scientific publications, deliverables, training materials, technology assets, business models, policies and guidelines. GREEN-LOG considers Open Access publication both as a prime method for increasing its scientific impact and visibility, and an ethically sound treatment and dissemination of the results.

Figure 1 shows an overview of the data flow defined as part of the data process handling established in the DMP that covers the path from collection to publication. Data can be collected either manually or automatically (M2M) using different mechanisms that will be briefly described in following sections.

The data collected will be referred as Research Data and will serve to produce scientific publications, deliverables, technology assets, business models and policies as well as data sets to support future research.

According to an established procedure (analysis of data types and processes involved in the data collection and generation, etc.), it is envisioned to first store the data in a restricted data repository, only accessible by project partners according to access rights (different types of restrictions may applied). It is among the aims of the Consortium to make public as much data as possible in a public repository. However, in some cases it will not be possible. Protection

mechanisms for exploitable results will be analysed according to partners interests and market opportunities to decide which assets are meant to remain protected. Likewise, some data will be qualified as sensitive and will remain protected if it is not possible to convert it into non-sensitive after anonymisation processes to prevent the disclosure of personal or industrial protected data and references to use-cases and user profiles.

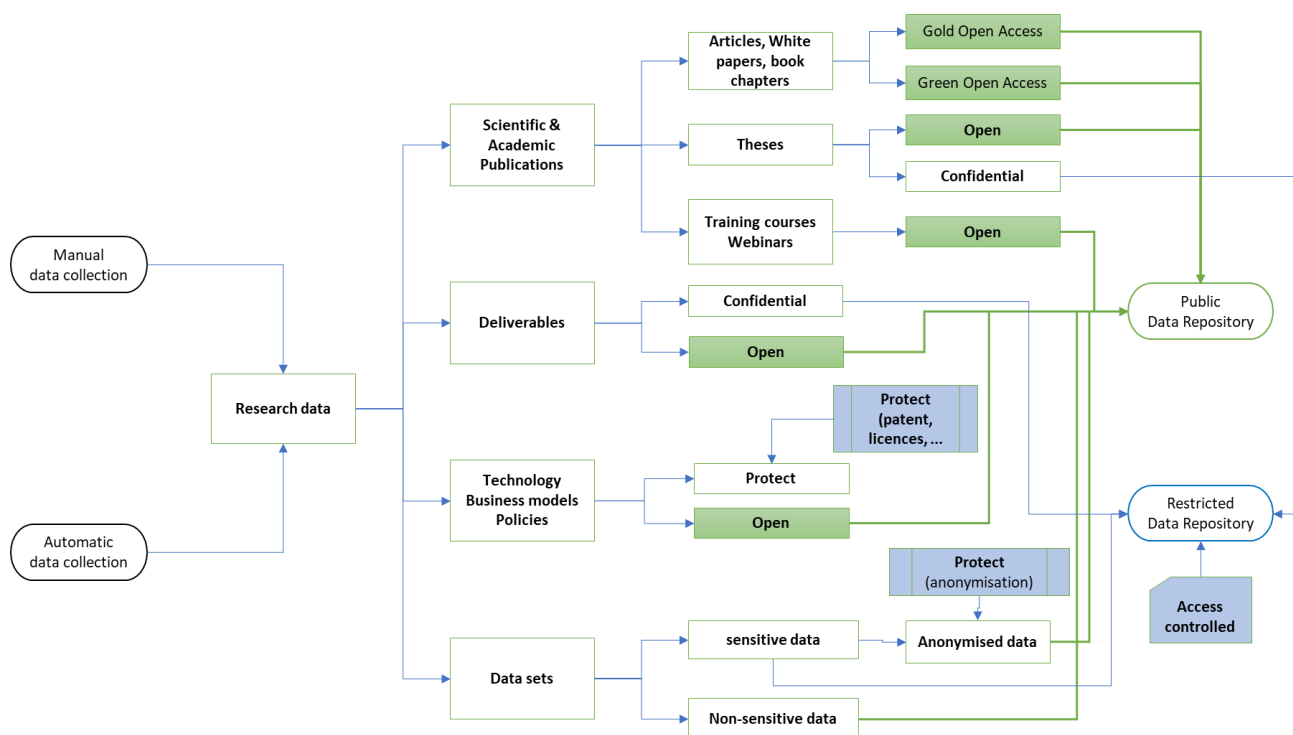


Figure 1: Data flow in GREEN-LOG

2.3 Legal framework

As of May 2018, the General Data Protection Regulation (GDPR)⁶ is applicable in all Member States in the European Union, as well as in the countries in the European Economic Area (EEA). GDPR updates and modernises existing laws on data protection to strengthen citizens' fundamental rights and guarantee their privacy in the digital age.

GDPR regulates the processing by an individual, a company or an organisation of personal data relating to individuals in the EU⁷. It does not apply to the processing of personal data of deceased persons or of legal entities. It sets down one set of data protection rules for all companies and organisations operating anywhere in the EU and European Economic Area (EEA), for two main reasons: 1) to give people more control over their personal data, 2) level the playing field for businesses and organisations operating in the EU and EEA. GDPR grant individuals a set of rights that must be protected by any actor who processes personal data. The individual rights include the right to:

- information about the processing of your personal data;

⁶ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016R0679>

⁷ https://commission.europa.eu/law/law-topic/data-protection/reform/what-does-general-data-protection-regulation-gdpr-govern_en

- obtain access to the personal data held about you;
- ask for incorrect, inaccurate or incomplete personal data to be corrected;
- request that personal data be erased when it's no longer needed or if processing it is unlawful;
- object to the processing of your personal data for marketing purposes or on grounds relating to your particular situation;
- request the restriction of the processing of your personal data in specific cases;
- receive your personal data in a machine-readable format and send it to another controller ("data portability"); and
- request that decisions based on automated processing concerning you or significantly affecting you and based on your personal data are made by natural persons, not only by computers. You can also have the right in this case to express your point of view and to contest the decision.

Concerning the UK, which is a non-EU country, it has a security of information agreement with the EU in place. The EU published two Adequacy Decisions on 28th June 2021, thereby establishing UK-EU agreements on the transfer of personal data under the EU's General Data Protection Regulations (EU) 2016/679 and the EU Law Enforcement Directive (EU) 2016/680. Both Adequacy Decisions can be found via the official EU website. In addition, the UK has published the UK/EU: Agreement concerning Security Procedures for Exchanging and Protecting Classified Information on the UK Government's official website.

As far as the Consortium is concerned, the CA and the Grant Agreement (GA) establish obligations and restrictions to partners. In particular, the members of the consortium must ensure personal data is:

- processed lawfully, fairly and in a transparent manner in relation to the data subjects;
- collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes;
- adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed;
- accurate and, where necessary, kept up to date;
- kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data is processed; and
- processed in a manner that ensures appropriate security of the data.

If materials or documents are subject to moral rights or third-party rights (including IPRs or rights of natural persons on their image and voice), the beneficiaries must ensure that they comply with their obligations under the Agreement (in particular, by obtaining the necessary licences and authorisations from the rights holders concerned).

The Data Protection Officer nominated by each partner in the Consortium will be in charge of monitoring the data handling in their organisations comply with the established obligations. During the project specific workshops will be organised to share the mechanisms in place and solve issues that may arise with the support of EUT legal department and contacting the national Data Protection Agencies, if needed. As described in section 6, personal data collection will always be done under informed consent (informed consents will be reported in later versions of DMP). More detailed procedures to ensure GDPR compliance and DPIA (in case of Public Authority participation) will be included in the second version of the DMP once a better understanding of the data to be collected and protected will be achieved.

2.4 Data types, formats and size

GREEN-LOG is a data intensive Project that will collect, process, and generate vast amounts of data of several types and formats. Table 1 summarises the information to be generated or used during the GREEN-LOG project.

Table 1: Overview of data generated or used during GREEN-LOG project

Created/ Collected/ Re-used	Type of data	Format	Dissemination level
Created	Dissemination material (Roll-ups, leaflets, posters, presentations)	PDF, PNG, JPG	Public
Created	Project website	Webdata	Public
Created	Deliverables	Docx, PDF	Confidential Public
Created	Public reports, scientific publications, articles	Docx, PDF	Public
Collected/ Re-used	Survey data	Forms, Docx, PDF, Excel	Confidential Public
Created	Workshop reports	Docx, PDF, Miro	Confidential Public
Created	Webinars	Webdata, wav, mp4	Public
Created/ Re-used	Academic outcomes: Master theses, PhD theses	Docx, PDF	Public
Re-used	Operational data from logistics operators, networks, traffic flows, ...	CSV, GFTS, GIS	Confidential Public
Collected	Operational data	Docx, CSV, DatexII	Confidential ⁸ , Public
Collected	Feedback from users	Docx, mp4, forms	Public (aggregated)

2.4.1 Data types

The types of data envisioned to be gathered during the project include the following categories:

- Operational data: Real-time logistics data collected from sensors, Internet of Things (IoT) devices, and tracking systems⁹.
- Planning data: Logistics demand data, logistics services data (i.e., fleet compositions, locations for micro-hubs, etc.), simulation models.
- Geospatial data: Maps, routes, and geographical information related to last mile logistics.
- Statistical data: Aggregate data on delivery times, costs, distances, and other relevant metrics.

Qualitative data: Interviews, surveys, and other qualitative data collected from stakeholders and end-users.

The following table (Table 2) list a preliminary set of data to be gathered, although it will be extended and detailed in the next version of the DMP.

⁸ For some operators, some restrictions apply.

⁹ Restrictions apply; only when it do not link to individuals.

Table 2: Overview of data to be collected in GREEN-LOG project.

Category	Data
Pickup point	ID Point Address (street, zip code, country) Date and time Receiver (name, id) Historical record
Delivery point	ID Point Address (street, zip code, country) Date and time Receiver (name, id) Historical record
Parcel	Volume (dimensions) Weight Packaging Goods description (incl. logistic segment and spec like fragility, perishability) Value
Time restrictions	restrictions in time for deliveries in the city, linked to the location of the local business, opening hours
Vehicle	Type Cargo volume Fuel type Length License plate
Driver/rider	ID Name Experience
Load Factor	The load factor is the ratio of the average load to total vehicle freight capacity (vans, lorries, train wagons, ships), expressed in terms of vehicle kilometres.
Delivery indicators	Number of retries Delay of delivery
Routes	Trip and Stops (geolocation)
Business rules	Type of service Pricing Cost
Hubs/Lockers	Specific case for pickup/delivery point if they are lockers or hubs: Parcels pickup Parcels delivered Timeframes
Network infrastructure	Road (geometry, traffic) Public transport (geometry, schedules, passengers flows) Logistics infrastructure locations Traffic count Traffic demand Parcel demand

2.4.2 Data format

A dataset can include different types of formats. For instance, qualitative data collected from stakeholders and end-users can consist of both written interview notes, audio files from interviews, pictures from pilot sites, and survey responses. Only parts of these datasets will be made openly available, for instance in an aggregated manner, since they include personal information.

GREEN-LOG will only use widely accepted formats for data generation, such as:

- Documents/Reports/Publications: .PDF/A, txt, doc/docx
- Spreadsheets: .xls/.xlsx
- Databases: .csv
- Audio files: .mp3, .wav, .wma, .ra
- Pictures: jpg, png
- Video: avi, flv, mov, mp4, wmv

2.4.3 Data sizes

At this stage of the project (M06) the data sizes of the different data sets are not known. This section will be completed in a later version.

2.5 Origin of data

GREEN-LOG data has different origins:

- Data provided by project partners coming from their activity;
- Data provided by stakeholders engaged in project activities;
- Data collected during the execution of the Living Labs in the corresponding 5 demonstration areas;
- Data generated by the partners after analysis of literature and processing of collected data.

The channels in which data will be collected include:

- Interviews with groups and individual participants in the Living Labs;
- Feedback from participants at stakeholder workshops;
- Survey responses;
- Market survey;
- Literature study/review and open data (re-use of existing data);
- Automated data collection by sensors and other systems in place.

2.6 Data repository

At this point of the project, the data repositories that have been chosen as candidates for restricted and public access are GREEN-LOG DataSpace, Sharepoint, Zenodo, and Github.

2.6.1 GREEN-LOG DataSpace

The GREEN-LOG DataSpace to be developed within WP2 will be the main dataset repository for the availability and management of data necessary for the functioning of the GREEN-LOG services and solutions. It will support the collection, distribution and use of mobility data and other relevant data among the last-mile ecosystem actors for driving the development of new intelligent systems. A set of data security, governance and sovereignty mechanisms will be design and implemented to handle different types of data protection including open (free to all), shared (available following agreements) and closed (data of commercial value and highly

restrictive). After the project finalisation, the open data will be transferred to Zenodo repository (see 2.6.3) to extend the impact of research.

At this stage (M06), the DataSpace is in the design phase. The next version of the DMP (M18) will include further details on the governance and data security mechanisms to be implemented.

2.6.2 SharePoint

A Microsoft SharePoint repository has been setup by INTRA to be used by Consortium partners. This is an access restricted space for collaboration where data will be stored prior to make them publicly available.

This repository is meant to store project documents (deliverables, reports, etc) communication and dissemination materials, audio and video recordings, transcriptions of interviews, surveys or alike. It is not envisioned to be used to store bulk datasets. Data with specific restrictions will be handled by the authorised parties under the directives of the DPO. The documents store in the SharePoint will follow the naming convention established in the GREEN-LOG Project Handbook [1]. Besides, the SharePoint provides a historical register of changes and editors of the document.

The non-anonymous datasets will be stored locally by the Data Protection Officer and not shared with others.

2.6.3 Zenodo

GREEN-LOG will use the open research data repository *Zenodo* to comply with the Horizon Europe mandate. All scientific publications, including public deliverables and public parts of underlying datasets will be uploaded to the *HE GREEN-LOG Community*¹⁰ in addition to the *European Commission Funded Research (OpenAIRE) Community*¹¹ in Zenodo.

Zenodo is a "catch-all" open research data repository which gathers research data across all disciplinary fields. It is for non-military purposes only, and the repository is hosted and managed by CERN. All data deposited to Zenodo is stored securely in the CERN Data Centre's cloud infrastructure¹².

2.6.4 GitHub

After an initial survey, some software components have been identified as potentially being of interest for publication as open source. A deeper analysis will be performed as development progresses bearing in mind the mandate making publicly available as many project outcomes as possible. However, it will be at the discretion of individual consortium members (and any Consortium Agreements regarding IPR on supplied data) to decide whether the source code of their developed software is openly accessible.

In the case the source code can be published, it will be done through GitHub¹³. Since the DMP is expected to mature during the project, the subsequent releases of the deliverable will specify the repositories where the data will be stored and go into more detail on how this data can be accessed by the wider research community.

¹⁰ <https://zenodo.org/communities/hegreen-log/?page=1&size=20>

¹¹ <https://zenodo.org/communities/ecfunded/?page=1&size=20>

¹² <https://zenodo.org/>

¹³ <https://github.com/>

3 FAIR Data Management

GREEN-LOG will manage data in accordance with the principles of FAIR data management as requested by the European Commission, according to OpenAire principles¹⁴ to make data:

- Findable
- Accessible
- Interoperable
- Re-usable data

The project aims to maximise access to, and re-use of research data generated by the project. However, the project will also consider the legit right of partners not to publish all the datasets and assets produced during the project for the sake of data privacy of Living Lab participants and Intellectual Property Rights.

3.1 Findability

Best practices to make data findable are to:

- assign (Meta)data a globally unique and persistent identifier;
- describe data with rich metadata;
- include in the metadata the identifier of the data they describe;
- register and index (Meta)data in a searchable resource.

The application of this principles within GREEN-LOG will be as follows:

- Use of the Digital Object Identifier (DOI) System to assign unique identifiers to the GREEN-LOG datasets.
- All datasets will be published in OpenAIRE and Zonodo repositories. Zenodo provides version control and assigns DOIs to all uploaded elements.
- Creation of HE GREEN-LOG community in Zenodo.
- Deliverables produced by GREEN-LOG will follow the naming conventions described in [1].
- Metadata will be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.

Metadata associated with each published data set in Zenodo will by default be as follows:

- Digital Object Identifiers
- Version numbers
- Bibliographic information
- Keywords
- Abstract/description
- Associated project and community
- Associated publications and reports
- Grant information
- Access and licensing info
- Language

¹⁴ <https://www.openaire.eu/openaire-guidelines-paving-the-way-to-interoperability-fairness-and-eosc>

It is envisioned to analyse the CERIF model¹⁵, specifically the part that is relevant for OpenAIRE¹⁶ to enhance standardisation of metadata.

Besides, a dataset catalogue will be build containing the following descriptors:

- Type of data set: Type of data according to categories and origin (imported or generated).
- Responsible partner: Describes who is responsible for the data generation and quality.
- Dataset name: brief description of the properties of the dataset.
- Data format description: description of the format of the data.
- (Expected) Amount: description of the amount of data.
- Confidentiality level:
 - Public: fully open, they can be shared with the general public.
 - Sensitive: data is limited under the conditions of the GA and CA.
- Additional classification: these classifiers will be revised and extended during the project and additional information will be provided as it is required during the project implementation.

3.2 Accessibility

All public datasets, scientific publications and deliverables will be uploaded to Zenodo and made openly available, free of charge. Publications and underlying data sets will be linked through use of persistent identifiers (DOI versioning). Data sets with dissemination level "confidential" (non-anonymous datasets) will not be shared due to privacy concerns. Potentially, some datasets might be restricted due to protection for commercial exploitation. If such cases arise during the project, this will be informed in the final version of the DMP. As described in 2.2, before making it publicly available in the public repository (Zenodo), data will be anonymised to avoid recognizable patterns or personal information that could identify individual. Master theses and PhD theses derived from the project will be made available at repositories provided by the corresponding academic partners participating in the GREEN-LOG consortium (i.e., UoA¹⁷, UAEGEAN¹⁸).

In the cases where private data are processed and aggregated (e.g., as part of a model, or functionality of a component) permission will be needed from the provider prior to making the altered data publicly available. Reports, deliverables, and other information for public distribution will be available on Zenodo and the GREEN-LOG website. Raw data (recordings, notes, user data, etc.) will not be available for third parties.

Metadata including licences for individual data records as well as record collections will be harvestable using the OAI-PMH¹⁹ protocol by the record identifier and the collection name. Metadata is also retrievable through the public REST API. The data will be available through Zenodo website²⁰, and hence accessible using any web browsing application.

¹⁵ <https://eurocris.org/services/main-features-cerif>

¹⁶ [CRIS information elements relevant for OpenAIRE — OpenAIRE Guidelines for CRIS Managers 1.1.1 documentation \(openaire-guidelines-for-cris-managers.readthedocs.io\)](#)

¹⁷ <https://www.uantwerpen.be/en/library/>

¹⁸ <https://www.ekt.gr/en/library/phdtheses>

¹⁹ <https://www.openarchives.org/pmh/>

²⁰ <https://zenodo.org/>

3.3 Interoperability

Data produced by GREEN-LOG will follow widely accepted standards, including DATEX-II²¹ or real-time traffic information, the GML Schema²² standards related to public transport information systems (Transmodel, NaPTAN, NPTDR, Google Transit Feed Specification), standards related to Multi-Modal Journey Planners (JourneyWeb²³), framework architectures for the transport sector²⁴.

Zenodo uses JSON schema as the internal representation of metadata and offers export to other formats such as Dublin Core, MARCXML, BibTeX, CSL, DataCite and export to Mendeley. The data record metadata will utilise the vocabularies applied by Zenodo. For certain terms, these refer to open, external vocabularies, e.g.: license (Open Definition), funders (FundRef) and grants (OpenAIRE). Reference to any external metadata is done with a resolvable URL.

3.4 Reusable data

The GREEN-LOG project will enable third parties to access, mine, exploit, reproduce and disseminate (free of charge for any user) for all *public* data sets, and regulate this by using Creative Commons Licences and corresponding data will be Open Data.

²¹ <http://www.datex2.eu/>

²² <http://www.opengeospatial.org/standards/gml>

²³ <http://www.dft.gov.uk/journeyweb/>

²⁴ <http://www.arktrans.no/>

4 Allocation of resources

EUT, with the support of FI, is responsible for ensuring the implementation of the DMP. The Data Manager is Mrs. Regina Enrich Sard. On the other hand, each partner is called to nominate a Data Protection Officer (DPO) who will handle each activity involving collection and process of data. Specific budget has been allocated for the DMP and Open Science activities (T5.2) as well as some budget to cover the cost of Green/Golden access scientific publications.

INTRA, as project coordinator has set up the internal data repository on SharePoint using its corporate Office365 license.

INTRA leading communication and dissemination activities has allocated resources for the production of communication material. POLIS is managing the project website and has allocated corresponding resources for the hosting.

Long-term preservation of the public data is ensured through Zenodo, Github and the Open Research Europe (or equivalent repositories) at no cost. Other resources needed to support reuse of data after the project ends will be solved on a case-by-case basis.

5 Data security

This section describes the security features of the research data infrastructure used to store and handle data.

5.1 Data security as specified for restricted repository in SharePoint

INTRA SharePoint is the online collaboration platform used in the GREEN-LOG project. A dedicated MS Teams environment has been established on this platform, accessible only by the partner representatives in the consortium. Data with specific restrictions will be handled by the authorised parties under the directives of the DPO setting up specific repositories under agreement of involved parties.

The GREEN-LOG SharePoint site is hosted by INTRA. INTRA applies all appropriate technical and organizational measures to ensure the safe processing of personal data and to prevent the accidental loss or destruction and unauthorized and / or illegal access to, use, modification or disclosure thereof (for more information please follow the [link](#)).

5.2 Data security as specified for public repository in Zenodo

The following list describes the security settings for Zenodo:

- Versions: Data files are versioned. Records are not versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. Records can be retracted from public view; however, the data files and records are preserved.
- Replicas: All data files are stored in the CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
- Retention period: Items will be retained for the lifetime of the repository. The host laboratory of Zenodo CERN, has defined a lifetime for the repository of the next 20 years minimum.
- Functional preservation: Zenodo makes no promises of usability and understandability of deposited objects over time.
- File preservation: Data files and metadata are backed up nightly and replicated into multiple copies in the online system.
- Fixity and authenticity: All data files are stored along with an MD5 checksum of the file content.
- Files are regularly checked against their checksums to assure that file content remains constant.
- Succession plans: In case of closure of the repository, a guarantee has been made from Zenodo to migrate all content to suitable alternative institutional and/or subject based repositories.

In the case of Github, platform for hosting and collaborating on code repositories, prioritizes security to protect users' data and ensure the integrity of their projects. Here is a summary of the security measures implemented by GitHub:

- Encryption: GitHub uses industry-standard encryption protocols (TLS/SSL) to protect data transmission between users and the platform, ensuring that sensitive information remains secure during transit.
- Access Control: GitHub provides granular access controls, allowing repository owners to manage permissions for collaborators, defining who can view, clone, modify, or merge changes into the codebase. Access can be granted at the repository, organization, or individual file level.

- Two-Factor Authentication (2FA): GitHub offers 2FA as an additional security layer for user accounts. By enabling 2FA, users are required to provide a second verification factor, such as a time-based one-time password (TOTP) generated by an authenticator app, in addition to their password.
- Vulnerability Detection: GitHub employs a security scanning feature called "Code scanning" that automatically analyses code repositories for potential vulnerabilities, including common security issues and coding errors. It helps identify and address security weaknesses early in the development process.
- Security Advisories and Alerts: GitHub provides security advisories and alerts for known vulnerabilities in dependencies used in projects hosted on the platform. Users receive notifications about vulnerabilities and recommended actions to address them.
- Security Bug Bounty Program: GitHub maintains a bug bounty program, encouraging security researchers to responsibly disclose any discovered security vulnerabilities. This initiative helps identify and address potential security flaws in the platform.
- Audit Logs: GitHub logs and tracks various activities, providing an audit trail of actions performed within repositories. This feature helps users monitor changes, identify unauthorized access attempts, and investigate any security incidents.
- Protected Branches: GitHub allows users to designate branches as "protected," enabling additional safeguards to prevent accidental or unauthorized changes. Protected branches can enforce code reviews, require status checks, and restrict direct pushes, reducing the risk of unauthorized modifications.
- Dependency Management: GitHub supports dependency management tools like Dependabot, which automatically scans repositories for outdated or insecure dependencies and creates pull requests to update them to the latest secure versions.
- Security Best Practices: GitHub provides guidance and best practices for secure coding, repository management, and collaboration workflows. This includes recommendations for secure development practices, secure password management, and safe handling of sensitive data.

GitHub uses Microsoft Azure data centres. These data centres are located in different countries and regions across the globe, spanning multiple continents. In principle, the data centre chosen is the one that is geographically closer. Microsoft Azure has data centres in EU regions.

6 Ethical aspects

The GREEN-LOG project consortium has long and extensive experience in the conduct of ethically sound research in compliance with national legislation, EU legislation, respect for international conventions and declarations, and their own institutional requirements. The overall responsibility for ethical governance of the project will rest within the Project Coordinator and concerning data management to the appointed Data Manager (EUT).

The research will be carried out in accordance with the ethics and data protection legislation of the participating countries, European legislation, and international conventions, as well as international declarations and codes of conduct relevant to research activities which include:

- Committee of Ministers Council of Europe (1990). Recommendation (No. R(90)3).
- Conventions of the Council of Europe (Convention for the Protection of Human Rights and Dignity of the Human Being)
- Universal Declaration of UNESCO (Universal Declaration on the and Human Rights, 11 November 1997).
- Charter of Fundamental Rights of the EU (signed in Nice, December 7, 2000, 2000/C364/01). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regards to the processing of personal data and on the free movement of such data and repealing Directive 95/46/EC (GDPR).

6.1 Collecting personal data from the Living Labs

Data collection activities (interviews, surveys, etc.) will be designed to maintain privacy. Personal data will not be requested unless this is absolutely necessary. Vulnerable groups like minors and individuals unable to freely provide an informed consent will be excluded. Participation is voluntary. Volunteers will be given important information, including possible risks and benefits, about their participation in the study. Participants will be given the possibility to decline and withdraw their participation at any time.

GREEN-LOG will collect pictures and video for use in communication activities (website, newsletter, social media, etc). Pictures and video can contain personal data if an individual is the focus of the image or video. Examples include: 1) pictures/video of individuals stored together with personal details (e.g. identity cards); 2) pictures/video of individuals posted on the project website along with biographical details; 3) individual images published in a newsletter. Examples of pictures and video that is unlikely to contain personal data are: 1) pictures/video where people are incidentally included in an image or are not the focus (e.g. at a big conference/workshop); 2) images of people who are no longer alive (the GDPR only applies to living people).

When collecting pictures and video GREEN-LOG will follow established guidance and best practice on collecting and processing such data to ensure that we adhere to the legal requirements (e.g. guidance established by the University of Reading²⁵, UK). Under no circumstances will pictures containing personal information be publicly shared without the subject's explicit consent.

²⁵ <https://www.reading.ac.uk/internal/imps/DataProtection/DataProtectionRequirements/imps-d-p-photographic.aspx>

Personal data will be handled in accordance with European legislation on privacy (GDPR). Under no circumstance will the deliverables or processes compromise the individual right to privacy and satisfactory handling of personal data.

All personal data will be stored on secure servers with access control managed by the Data Protection Officers. Personal data will be handled by authorised personnel, and no one will have access to the data unless this is necessary to carry out the project work.

6.2 Information and consent form

The participants will be given an information letter and a consent form (on paper or electronically). The information letter will provide information about:

- The type of data that will be collected during the study.
- How the data will be collected (interview, automatic data collection, etc.)
- What the data will be used for. The information letter will explain the purpose of the project and the expected results. It will also be explained that published information always will be anonymous, and that no personally identifiable information will be published in any way.
- How the data collected will be handled. The information letter will explain that personal data will be treated in full confidentiality and will be registered and stored in a secure manner. The data will be de-identified before it is processed (name or other characteristics serving to identify person will be replaced by a number and the list of identifiers will be kept separate from the data).
- Who will have access to the data. The information letter will state that data will be handled by a very limited number of authorised personnel and that confidentiality will be regulated by legal agreements. The data will be de-identified before it is discussed and processed within the project.
- The rights of the participants. The information letter will state that participation is voluntary and that participants have the right to see the data collected about them and that they can withdraw from the study at any time without any obligation to explain their reasons for doing so (contact information for such requests will be provided).

7 Conclusions

An Initial Data Management Plan has been defined for GREEN-LOG as a formal commitment by partners to adhere to the data management strategy and the procedures it defines. All project partners are aware of the regulations and recommendations provided by the European Commission and Funding authorities. Project partners will also comply with the principles and practices described in the GA and CA.

As Data Manager, EUT will ensure that any data management issues which may arise during the project will be handled appropriately and in a transparent and fair manner. Data Protection Officers designated by each consortium partner will be responsible for assuring data quality and protection for confidential data.

At this stage, the public repositories envisioned to share data are Zenodo and GitHub. The GREEN-LOG DataSpace will include data with different dissemination levels and will be operational during the project lifetime. SharePoint will be used as protected repository only for Consortium members to store project documents.

The DMP is a living document that will be updated during the project lifetime. Updated versions will be made available at M18 and M42 and it will be evaluated and updated as the project evolves and new information on data collection, generation and handling arise.

References

Reference	Name of document
1	D6.1 – Project Handbook, Quality Plan & Risk Management, GREEN-LOG deliverable, Netcompany-Intrasoft, 28-02-2023