



Deliverable D 6.2

Data Management Plan

Project acronym:	ESEP4Freight
Starting date:	2023-09-01
Duration (in months):	24
Call (part) identifier:	HORIZON-ER-JU-2022-02
Grant agreement no:	101121840
Due date of deliverable:	Month 3
Actual submission date:	2023-11-30
Responsible/Author:	Celestino Sánchez (EURNEX)
Dissemination level:	PU
Status:	Issued

Reviewed: yes

Document history		
Revision	Date	Description
1	13/11/2023	First draft
2	23/11/2023	First complete version
3	24/11/2023	Final version of the document
4	28/11/2023	Document review

Report contributors		
Name	Beneficiary Short Name	Details of contribution
Celestino Sánchez	EURNEX	Structure and preparation of the document
Francisca Rosell Camps	EURECAT	Additional inputs and document review

Disclaimer

The information in this document is provided “as is”, and no guarantee or warranty is given that the information is fit for any particular purpose. The content of this document reflects only the author’s view – the Joint Undertaking is not responsible for any use that may be made of the information it contains. The users use the information at their sole risk and liability.

The content of this publication does not reflect the official opinion of the Europe’s Rail Joint Undertaking (EU-Rail JU). Responsibility for the information and views expressed in the publication lies entirely with the author(s).



Table of Contents

Executive Summary	6
Abbreviations and Acronyms	7
1 Background	8
2 Objective/aims	9
3 Data Summary	10
4 FAIR Data	13
4.1 Introduction to FAIR data principles.....	13
4.1.1 Making Data Findable (including provisions for metadata)	13
4.1.2 Making Data Openly Accessible	13
4.1.3 Making Data Interoperable.....	13
4.1.4 Increase Data Re-use	14
4.2 FAIR Data in ESEP4Freight.....	14
4.2.1 Making Data Findable (including provisions for metadata)	14
4.2.2 Making Data Accessible	15
4.2.3 Making Data Interoperable.....	17
4.2.4 Increase Data Re-use	18
5 Other Research Outputs	19
6 Allocation of Resources	21
7 Data Security	22
8 Ethics	23
9 Conclusions	24
References	25
Appendices	26
Appendix 1. First assessment of data category and FAIR Principles.....	26



List of Tables

Table 1. First identification of data needed for the Web Platform.....	10
--	----



List of Figures

Figure 1. Modules of the Web Platform	8
---	---



Executive Summary

This Deliverable is the first version of a living document that covers the entire data management life cycle even beyond the end of the project. The structure of this deliverable is based on the template provided by the European Commission for Horizon Europe projects (European Commission, 2021). In order to provide an overview to the consortium partners on the FAIR Principles, an introduction section in Chapter 4 has been included.

The purpose of the Data Management Plan (DMP) is to provide the members of the consortium and the Europe's Rail Joint Undertaking (EU JU) with an analysis of the main elements of the data management policy regarding all the datasets generated or collected by the project. It also defines the tools to be used for data management, e.g. Zenodo, Argos, and Github. Additional information about the tools can be found along the document.

It is important to highlight, however, that, as the project started recently and many questions regarding which datasets will be available and needed as well as how data will be used, many of the questions raised cannot be answered yet. Considering this, the DMP will serve as guide and handbook for the data handling within ESEP4Freight and will be the basis for the final version of the DMP to be included in the D6.3.

To offer a better overview for the consortium partners on the open questions to be addressed, it has been decided to keep the structure based on questions of the original template instead of using body text. The final version of the DMP (D6.3) is foreseen to be produced using the open-source tool Argos, that offers a structured and automated manner for the creation of DMPs for Horizon Europe.

Keywords: DMP, FAIR, railways, freight, logistics.

Abbreviations and Acronyms

Abbreviation / Acronym	Description
API	Application Programming Interface
CC0	Creative Commons
CSV	Comma-separated Values
DMP	Data Management Plan
DoA	Description of Actions
DOI	Digital Object Identifier
ER JU	Europe's Rail Joint Undertaking
EUDAT	EUropean DATa (Research Project)
FAIR	Fair, Accessible, Interoperable and Re-usable
GB	Gigabytes
GDPR	General Data Protection Regulation
GTFS	General Transit Feed Specification
JS	JavaScript
JSON	JavaScript Object Notation,
KML	Keyhole Markup Language
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
ORD	Open access to Research Data
PUB	Public
R&I	Research and Innovation
RDM	Research Data Management
RNE	RailNetworkEurope
SDK	Software Development Kit
SG	Stakeholders' Group
SNCF	National Society of the French Railways
TEN-T	Trans-European Transport Networks
TIS	Train Information System
UBA	German Environment Agency
UIC	Union International of Railways
WP	Work Package
XML	Extensible Markup Language
ZIP	Data Format

1 Background

The present document constitutes the Deliverable “D6.2 Data Management Plan” in the project European Shift Enabler Portal for Freight (ESEP4Freight). Funding body is the EU-RAIL in the call HORIZON-ER-JU-2022-02.

One of the main objectives of ESEP4Freight is to create a so-called Web Platform to provide information to relevant actors in the logistics sector to boost the modal shift to rail in the rail freight transport. The Web Platform is planned to include the modules shown in the figure below.

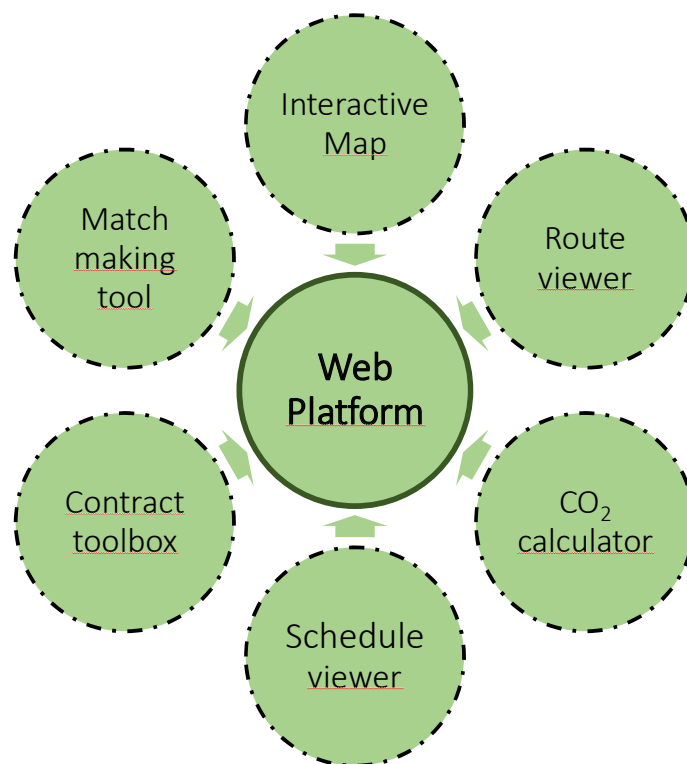


Figure 1. Modules of the Web Platform

These modules will heavily rely on an intensive use (mostly re-use) of data and during the development of the modules, new data and other research outputs will be produced. Therefore, the DMP will serve as guidelines to handle the data employed and manipulated in these modules. In addition to this, other types of data such as those linked to dissemination activities and data derived from the Stakeholder Group (SG) on their feedback on the Web Platform and other project-related inputs will be considered.



2 Objective/aims

This document provides a first assessment on FAIR Principles of the data to be generated and reused during the project as well as other topics related to the data management, such as the use of resources, the data security and possible ethical issues. This document also addresses other research outputs susceptible of being managed according to the FAIR Principles. In addition, this document will serve as guidance to produce the results of the DMP as part of the D6.3.

More precisely, the DMP addresses the following points:

- The purpose of the data collection/generation and its relation to the objectives of the project;
- Types and formats of data to be utilised and generated within ESEP4Freight. This section identifies and describes the (existing) input data that will be utilized and the output data to be generated by the project;
- The origin of the data and the expected size of the data;
- To whom might it be useful (“data utility”);
- Standards to be used, metadata and quality issues. GDPR and compliance issues are covered as appropriate;
- How data are exploited and shared/accessed for their verification and reutilisation. The exploitation of data will follow the strategies of each partner concerning their business potential, in accordance to the exploitation plan produced in WP5, and in accordance to the access to data by the partners specified in the Consortium Agreement. Specific restrictions and confidentiality aspects are clarified;
- Data storage and conservation. Where the data will be held and the arrangements and responsibilities for managing, updating and maintaining the data.
- Resources needed for data management

3 Data Summary

A first categorization of data has been already carried out. As a result of this process four high-level categories of data have been identified.

- **Category 1: Documents and dissemination materials:** deliverables, reports, demonstrations, manuscripts, productions for dissemination, communication purposes.
- **Category 2: Computer software:** including software applications (in binary form), libraries in the form of SDKs, plugins, and respective source code.
- **Category 3: Data and metadata:** materials and datasets resulting from the implementation of the developments; metadata and configuration files; bug logs and feedback logs; developer internal logs and documentation; evaluation and opinions.
- **Category 4: Data for evaluation:** consists of materials or datasets generated or collected by the project used for evaluation purposes. It may include personal data and information of participants in pilots and stakeholder engagement activities.

Strictly speaking, only the category 3, “Data and metadata”, and 4, “Data for evaluation”, can be considered data. The DMP will mainly focus on these two categories. Category 1, “Documents and dissemination materials”, and category 2, “Computer software” have been considered Other Research Outputs and will be addressed in Section 5.

- *Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.*

ESEP4Freight will heavily rely on the reuse of already available data. The final list of data to be re-used will depend on the availability of the data. However, a preliminary list of targeted data can be seen in the table below:

Table 1. First identification of data needed for the Web Platform

Category	Data Type	Data Source
Assessment of innovations	Time saved, economic savings, reliability	Previous projects such as IMPACT-1, IMPACT-2, and scientific publications
Freight market trends	Annual statistics, Publications	UIC and UIRR annual reports, International Transport Forum-Organisation for Economic Co-operation and Development
Operational, Services data	Capacities, Handling volumes & Cargo flows, Value added services, Technologies	Eurostat, International Transport Forum-Organisation for Economic Co-operation and Development, Infrastructure managers, Transport organisations, Terminal operators

Legal and contractual framework	Legal requirements for modal shift between different countries like transport documents, agreements, and operating conditions	National ministries in Europe, European Union, International Transport Forum-Organisation for Economic Co-operation and Development, UIRR
Network Infrastructure	Rail lines, main road connections, waterway network, main corridors – operator lines	Ferrmed study, OpenRailwayMap, Open Street Maps, Intermodal map, TEN-T network
Terminals data	Location, Contact, Basic Infrastructure, and Services	Intermodal Map, Rail Facilities Portal, European Transport Maps, Terminal operators
Schedules data	Timetables of intermodal services	Intermodal transport operators like HUPAC, Kombiverkehr, SNCF, Renfe, Lineas, and RNE-TIS
CO ₂ benchmarking	Reference values for different modes of transport, loading units, and infrastructure	German Environment Agency (Umweltbundesamt – UBA), EcoTransIT (as alternative/complement)

- *What types and formats of data will the project generate or re-use?*

Regarding formats the most common ones used in previous experiences are csv, json, xml, and GTFS, and also the common ones for geographical data (shapefile, KML). This list will be completed once we had identified the available data and formats.

- *What is the purpose of the data generation or re-use and its relation to the objectives of the project?*

The re-use of data is part of the project core as ESEP4Freight will gather and make accessible data from different sources in the Web Platform. In addition, it is expected that the data and resources from the Web Platform will be re-used by other users.

- *What is the expected size of the data that you intend to generate or re-use?*

A small part of the data is expected to be generated by the project. For example, the CO₂ calculator will yield estimations of CO₂ emissions based on the input data from the user. Such data directly generated by the project are expected to be in the order of magnitude of a few megabytes. However, most part of the data will be re-used from several sources and are expected to be in the order of magnitude of gigabytes. For example, the database of terminals and transport infrastructure across Europe. The major contributor to the size of the data would be the organisation profiles that could be created by the users of the platform in the Match-making tool.

An initial estimate of the server storage requirements of the different modules of the ESEP4Freight



platform are listed below. Please note that the actual values may vary based on the development and implementation phases of the project (WP 3-5).

- Interactive Map: Terminal database, Infrastructure database (0.5 GB)
- Corridor viewer: High potential freight corridors (0.5 GB)
- Schedule viewer: Combined Transport Timetables (0.5 GB)
- CO₂ Calculator: Start-destination search, Reference values, Results (0.25 GB)
- Contract toolbox: Templates and Reports (0.5 GB)
- Match-making tool: Login, Organisation profiles (25 GB)

It is therefore planned to have a server with at least 30 GB server storage, processing of the modules within the client as much as possible, hosting of the platform and a medium computing power.

- *What is the origin/provenance of the data, either generated or re-used?*

Most of the data to be used for the Web Platform are expected to be re-used from different sources. The process of sources mapping is ongoing but a preliminary list of sources has been already identified. The list of data sources can be seen in Table 1.

- *To whom might your data be useful ('data utility'), outside your project?*

Due to the nature of the project, based on providing information to relevant stakeholders through a publicly accessible on-line tools, it is expected that data will be primarily useful for stakeholders involved in the logistic sector. In addition. It is expected that data will be also of interest for researchers, policy makers, citizens, and associations linked to the freight transport.



4 FAIR Data

4.1 Introduction to FAIR data principles

According to the Guidelines on FAIR Data Management in Horizon Europe, research data should be “FAIR”: findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation-solution. Next, the main FAIR principles are listed and explained. This information has been provided by the initiative GO FAIR (2022).

4.1.1 Making Data Findable (including provisions for metadata)

The first step in (re)using data is to find them. Metadata and data should 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.

- F1 (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

4.1.2 Making Data Openly Accessible

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
 - A1.1 The protocol is open, free, and universally implementable
 - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

4.1.3 Making Data Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

4.1.4 Increase Data Re-use

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different setting.

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
 - R1.1. (Meta)data are released with a clear and accessible data usage license
 - R1.2. (Meta)data are associated with detailed provenance
 - R1.3. (Meta)data meet domain-relevant community standards

4.2 FAIR Data in ESEP4Freight

This section offers an overview of the status of the project data and the FAIR properties and the planned tools to be used during the project. For the identification, processing and assessment of each dataset during the project according to the questions included in this section, it is planned to use the webtool Argos (OpenAIRE & EUDAT, 2023). This tool supports the creation, management, and sharing of DMPs in a structured and automated manner. In addition, Argos also offers a search to Zenodo for prefilling the DMP with dataset metadata.

4.2.1 Making Data Findable (including provisions for metadata)

- *Will data be identified by a persistent identifier?*

Category 3 data from the Web Platform and Category 4 data, derived from the inputs of SG members, could be assigned a persistent identified (e.g. DOI) and make publicly accessible if considered of public interest. The data will be analysed case by case in function of the origin of the data, possible commercial interest, and GDPR protection. A repository in Zenodo¹ (Eurnex, 2023) has been created where data from Category 3 and Category 4 can be stored. Zenodo can assign a DOI number to all the documents that do not have previously received one. In addition to this, Argos will be used to collect and structure the metadata of the datasets

- *Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.*

The tools developed throughout the project will be shared on a code repository platform, such as GitHub (2023), GitLab (2023) or Bitbucket (2023). This approach will allow for the dissemination of not only the code itself but also the associated code properties (metadata) and usage instructions.

The selection of metadata standards will be decided once we have a clear overview of the data we have. As data are planned to come from different sources, data obtained must be examined first to identify common metadata standards. If no metadata standards are identified ESEP4Freight will

¹ https://zenodo.org/communities/ese4_freight

consider to follow the DataCite Metadata Schema (Vierkant, 2023).

- *Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?*

Yes, the use of keywords is foreseen. They will be defined during the project depending on the data obtained.

- *Will metadata be offered in such a way that it can be harvested and indexed?*

When possible, datasets will be interoperable according to Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Argos will index the datasets and offers the capability of harvesting metadata.

4.2.2 Making Data Accessible

4.2.2.1 Repository

- *Will the data be deposited in a trusted repository?*

It is planned to store the data for the Web Platform in a server. Depending on the origin of the data for the Web platform (e.g. downloaded or obtained via API) the data can be also deposited in a repository, such as Zenodo.

- *Have you explored appropriate arrangements with the identified repository where your data will be deposited?*

Yes, several project partners have experience working with Zenodo repositories.

- *Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?*

In the case of the Zenodo repository, it will assign an identifier (DOI) which will link the identifier to the digital object.

4.2.2.2 Data

- *Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.*

Data in the Web Platform are foreseen to be publicly accessible. If any other datasets are not be shared (or will be shared under restricted conditions) an explanation of the restrictions will be provided. This can be the case, for example, of the match-making tool as certain modules of the Web Platform may contain confidential information.

Restrictions may be applied, however, to data derived from the feedback given by SG members regarding the Web Platform and other inputs linked to the project. Possible restrictions linked GDPR, such as personal identification of members or its institutions, will be discussed with the SG members in the next months.

- *If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.*

No embargo for data from scientific publications nor from data linked to the Web Platform are foreseen.

- *Will the data be accessible through a free and standardized access protocol?*

Depending on the type of data, we can give access using an API or simply by letting users download the data in any standard format (CSV, JSON, shapefile, KML,..)

- *If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?*

No data restriction is foreseen.

- *How will the identity of the person accessing the data be ascertained?*

In general, as data in the Web Platform will be publicly accessible there will be no need of determining the identity of specific persons. However, some Web Platform services may need previous identification to avoid possible misuse. This could be the case of, for instance, the match-making tool or the contract toolbox.

- *Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?*

As the amount of personal data will be very reduce, mainly limited to the personal data from the SG, and data from the Web Platform will be openly accessible, no data access committee will be needed.

4.2.2.3 Metadata

- *Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?*

The details about the technical details on the type of license for the metadata are to be internally discussed in the consortium. This will be clarified in the D6.3 “Data Management Plan – Results” to be prepared by the end of the project.

- *How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?*

This question is linked to the future exploitation of the Web Platform once the project had been finished. Specific information about the availability and findability of the data after the end of the project will be included in the D6.3 “Data Management Plan – Results” to be prepared by the end of the project

- *Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?*

No specific software to access or read the data beyond common commercial software (e.g. Office) is foreseen.

4.2.3 Making Data Interoperable

- *What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?*

The aim of using standard formats for the different data files, like CSV, JSON, XML or other common standards for the structured data, or shapefile or KML for the geographical data and the static map layers have been set in the Grant Agreement. The questions around the data interoperability will be addressed once the available data and their formats had been identified in WP1 and WP2..

- *In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?*

In case of using uncommon or new ontologies, the provision of additional information will depend on the complexity and the tasks and the relevance of the data attached to project specific ontologies. This information will be published in the final version of the DMP in D6.3.

- *Will your data include qualified references² to other data (e.g. other data from your project, or datasets from previous research)?*

Yes, if considered relevant.

² A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: <https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata>)

4.2.4 Increase Data Re-use

- *How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?*

As developed in Section 5, a code repository platform, such as GitHub will be employed for the documentation of the code. This platform or the repository in Zenodo mentioned in section 4.2.1 will be used to publish the needed information.

- *Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?*

ESEP4Freight main objective is to provide useful information to logistics actors to promote the shift to rail in the freight transportation system through the Web Platform described in Section 3. Therefore, the project aims at making freely available as much data as possible. The licenses will be at least as open as the licences from the data received.

- *Will the data produced in the project be useable by third parties, in particular after the end of the project?*

As previously mentioned, the Web Platform aims to provide data to third parties. ESEP4Freight aims at keeping the Web Platform operative after the end of the project. The strategy to achieve this goal has been analysed in a preliminary exploitation strategy and will be further discussed during the project. The conclusions of this process will be included in the D5.4 “Communication, Dissemination and Exploitation (C&D&E) reports” at the end of the project.

- *Will the provenance of the data be thoroughly documented using the appropriate standards?*

Yes, the sources of the data will be documented.

- *Describe all relevant data quality assurance processes.*

No data quality assurance process has been set until now. This will take once the project has a clear overview of the available data, their origin as well as their formats.

- *Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.*

This topic has been addressed in Section 5.

5 Other Research Outputs

In section 3, a first categorization of project data was presented. The categories are the following.

- Category 1: Documents and dissemination materials: deliverables, reports, demonstrations, manuscripts, productions for dissemination, communication purposes.
- Category 2: Computer software: including software applications (in binary form), libraries in the form of SDKs, plugins, and respective source code.
- Category 3: Data and metadata: materials and datasets resulting from the implementation of the developments; metadata and configuration files; bug logs and feedback logs; developer internal logs and documentation; evaluation and opinions.
- Category 4: Data for evaluation: consists of materials or datasets generated or collected by the project used for evaluation purposes. It may include personal data and information of participants in pilots and stakeholder engagement activities.

From the previous categories, category 1 and 2 are (strictly speaking) not considered data and they fall (except some dissemination and communication materials) within the category Other Research Outputs. These categories will be addressed in this section.

- *In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).*

As the Web Platform is a central element of ESEP4Freight a specific plan for the Web Platform will be developed during the project. This will include a clear delimitation between those elements (e.g. software, workflows, architecture) of commercial interest and thus to be protected from those which can be publicly accessible.

As mentioned in Section 4.2.1, the tools developed throughout the project will be shared on a code repository platform, such as GitHub, GitLab or Bitbucket. This approach will allow for the dissemination of not only the code itself but also the associated code properties (metadata) and usage instructions. This will also comprise the outputs falling within the Category 2.

As mentioned in Section 4.2.1, a repository in Zenodo³ (Eurnex, 2023) has been created where elements falling within the Category 1 will be included. Zenodo will automatically assigned a DOI number to all the documents that do not have previously received one. Data from scientific papers may be subjected to temporal embargo if deemed appropriate to shield the research outputs.

- *Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.*

³ https://zenodo.org/communities/ese4_freight



A significant part of the plan for the management of Other Research Outputs will be dedicated to identify which FAIR principles could be applied to the management of the Web Platform. A first assessment of the application of FAIR principles to the categories 1 and 2, associated to Other Research Outputs was carried out during the proposal stage and can be seen in “Appendix 1. First assessment of data category and FAIR Principles”. A first draft of the plan could be produced once the architecture of the Web Platform had been designed (planned by M12). The final plan for the management of the Other Research Outputs will be included in the final version of the DMP (D6.3) at month 24.

6 Allocation of Resources

- *What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.)?*

As ESEP4Freight will heavily rely on the reuse of data, a cost assessment is not possible until the data to be reused had been received. An assessment of costs will be possible when the data size, the formats and the FAIR suitability of the data received are known. OpenAIRE (2023) guidelines will be followed for the cost estimation.

- *How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)*

The costs for making data or other research outputs FAIR will be analysed case by case.

- *Who will be responsible for data management in your project?*

The main responsible of the data management is Celestino Sánchez (Eurnex) with the support of the project coordinator and the other partners involved in the technical development of the Web Platform.

- *How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?*

This will be linked to the exploitation strategy developed at the end of the project and to be included in the D5.4 “Communication, Dissemination and Exploitation (C&D&E) reports”.

7 Data Security

- *What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?*

Zenodo: As wide used repository, has their own data protection, backup strategies and also data accessibility and authorship.

Github/Gitlab/Bitbucket: Widely used code repositories, they also have their own protection, backup strategies, and the authorship and intellectual property control by the automated definition of software licenses.

All data required for executing this project will be stored in safe environments at the designed locations of the project partners. Access rights will be restricted to the relevant partners. If data needs to be transferred among partners, the transfer needs to be done securely, via secure data channels, in an encrypted mode or by doing a physical transfer.

The data for the Web Platform will be stored and hosted on a secure server in Europe with credible data security and backup measures. The concrete hosting solution provider will be identified during the WP3 focused on defining the architecture of the platform.

In cases where sensitive data will be stored, data privacy and data protection issues will strictly follow the "user decides" principle. End-users will always have the possibility (and only the user) to decide which personal or private data to be used.

On the Web Platform, the user data will be stored in the encrypted form and also the administrators of the platform will not have access to this data, ensuring data security of the users. This also implies scenarios like the creation of user accounts within the planned Match-making tool on the platform is secure.

- *Will the data be safely stored in trusted repositories for long term preservation and curation?*

Project repositories, such as Zenodo, are planned to be maintained after the end of the project.



8 Ethics

- *Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).*

Potential ethic issues are limited to handling personal data from SG members. Personal data will be handled according to the GDPR.

- *Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?*

Yes, data for the matchmaking tools as well as data from SG members that include personal data will be handled accordingly and the person involved will be informed and their informed consent will be sought.



9 Conclusions

This document comprises the first version of the DMP. This document describes the framework through which data in ESEP4Freight will be handled. This includes storage, re-use, and interoperability and accessibility features. It has described the main tools to be used for the management of data, such as Zenodo, Argos, and GitHub (or similar).

As the project has just started and its development will heavily depend on the available data to be integrated in the Web Platform, many of the questions cannot be answered yet. This first DMP offers, however, a first overview of the data to be collected and re-used. It summarises the FAIR principles and carry out a first analysis of the suitability of the data to the FAIR principles. In addition, this document has also highlighted the relevance of other research outputs for ESEP4Freight.

D6.2, conceived as a living document, will serve as guidelines for management the data (and other research outputs) during the project and will serve as basis for the final version of the DMP to be included in the D6.3 to be produced using the open-source software Argos.



References

- Bitbucket. (2023). *Bitbucket | Git solution for teams using Jira*. <https://bitbucket.org/product>
- Eurnex. (2023). *ESEP4Freight Zenodo Community*.
https://zenodo.org/communities/esep_for_freight?q=&l=list&p=1&s=10&sort=newest
- European Commission. (2021). *Data Management Plan Template*.
https://www.openaire.eu/images/Guides/HORIZON_EUROPE_Data-Management-Plan-Template.pdf
- GitHub. (2023). *GitHub: Let's build from here*. <https://github.com/>
- GitLab. (2023). <https://about.gitlab.com/>
- GO FAIR. (2022). *FAIR Principles*. https://www.go-fair.org/wp-content/uploads/2022/01/FAIRPrinciples_overview.pdf
- OpenAIRE. (2023). *RDM costs*. OpenAIRE. <https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs>
- OpenAIRE & EUDAT. (2023). *Argos—About*. <https://argos.openaire.eu/about>
- Vierkant, P. (2023). *DataCite – Connecting Research, Advancing Knowledge*. DataCite.
<https://datacite.org/>

Appendices

Appendix 1. First assessment of data category and FAIR Principles

Data category	FAIR
(1) Documents and dissemination materials: deliverables, reports, demonstrations, manuscripts, productions for dissemination, communication purposes. <i>Public</i>	F Self-archive on website, and/or green scientific publications + Zenodo and Openaire repositories.
	A Open Access by default policy, subject to exception (confidential deliverables).
	I Common text, image, or video formats (.pdf, .docx, .jpeg, .mov, .avi, etc.).
	R Open Licences by default (attribution, non-commercial use).
(2) Computer software: including software applications (in binary form), libraries in the form of SDKs, plugins, and respective source code. <i>Public & confidential</i>	F A project repository will be used, including meta-data and mechanisms for quick search of libraries, plugins, or generated software.
	A The Project repository includes the access control mechanisms necessary during its execution, and easily accessible from the development SDKs.
	I Binary formats, and ZIP files. Common programming languages will be used for the generated code (Python, NodeJS, C, Julia, etc.) and standard communication protocols will guarantee integration by any logistics actor. Furthermore, whenever possible, open licences applications or tools will be used to build upon (such data bases, webservers, ...)
(3) Data and metadata: materials and datasets resulting from the implementation of the developments; metadata and configuration files; bug logs and feedback logs; developer internal I documentation;	F A Project repository will be used to control access to data and information that must be protected to ensure confidentiality and security in its handling.
	A Required authentication and authorization mechanisms for access to the data collected are included to guarantee the privacy and confidentiality of the data collected and generated.
	I Most extended formats in use by all the logistics actors are used in order to guarantee interoperability, replicability and scalability.

<p>evaluation and opinions. <i>Public & confidential</i></p>	R	Open Licences by default (non-commercial use).
<p>(4) Data for evaluation: consists of materials or datasets generated or collected by the project used for evaluation purposes. It may include personal data and information of participants in pilots and stakeholder engagement activities. <i>Public & confidential</i></p>	F	A repository will be used for data from the demonstration processes in the different Business Cases, so that they can be shared in different logistics projects.
	A	Required authentication and authorization mechanisms for access to the data collected are included to guarantee the privacy and confidentiality of the data collected and generated.
	I	Most extended formats in use by all the logistics actors are used in order to guarantee interoperability, replicability and scalability.