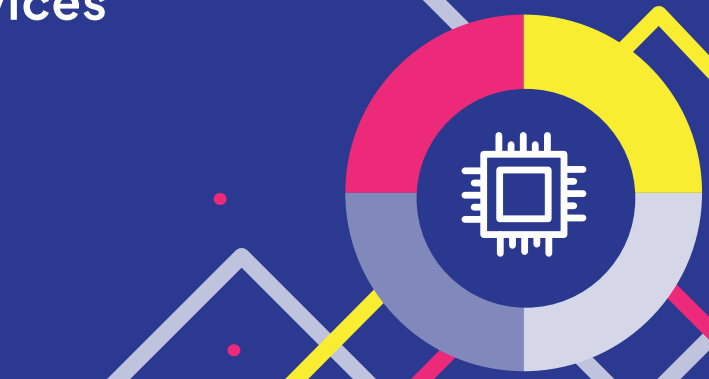
Advanced traffic management solutions for synchronized and resilient multimodal transport services

SYNCHROMODE



D1.5: Data Management Plan - updated version 2

Responsible author: E. Mitsakis
Contributing authors: D. Tzanis
Date: 31/10/24

This project has received funding from the European Union's Horizon Europe research and innovation program under grant agreement No 101104171

Table of Contents

Executive Summary

This deliverable presents a comprehensive analysis of the data management policy designed for implementation in the SYNCHROMODE project. It outlines the anticipated data categories and types, while providing a detailed list of associated metadata elements for the project's data. Additionally, this document describes the framework for data sharing, with a strong focus on adhering to FAIR (Findable, Accessible, Interoperable, and Reusable) principles.

In this updated version of deliverable "D1.4 Data Management Plan - updated version 1", improvements have been made in terms of data categorization and the processes for ensuring high-quality data management across the project. The deliverable expands on the procedures for data sharing and reuse, reflecting the project's progress and the integration of updated protocols for safeguarding data security. It also includes more detailed guidance on ethical considerations, with specific reference to privacy and data protection regulations such as the GDPR.

Furthermore, updated resource allocation strategies and enhanced security measures have been incorporated, ensuring the SYNCHROMODE project complies with Horizon Europe's data management standards throughout its lifecycle.

Summary sheet

Deliverable No.	1.5
Project Acronym	SYCHROMODE
Full Title	D1.5 – Data Management Plan (Updated Version 2)
Grant Agreement No.	101104171
Responsible Author(s)	Evangelos Mitsakis (CERTH)
Contributing Author(s)	Dimitris Tzanis (CERTH)
Peer Review	Leire Serrano (DEUSTO)
Date	31/10/2024
Status	Final
Dissemination level	Confidential
Version	1.0
Work Package No.	WP1
Work Package Title	Project management
Work Package Leader	ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS (CERTH)
Project Coordinator	ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS (CERTH)
Website	https://synchromode.eu

Project partners

Organisation	Country	Abbreviation
ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS	Greece	CERTH
UNIVERSIDAD DE LA IGLESIA DE DEUSTO ENTIDAD RELIGIOSA	Spain	DEUSTO
NOMMON SOLUTIONS AND TECHNOLOGIES SL	Spain	NOMMON
YUNEX GMBH	Germany	YUNEX
MAP TRAFFIC MANAGEMENT BV	Netherlands	MAPTM
AIMSUN SLU	Spain	AIMSUN
BE-MOBILE	Belgium	BE-MOBILE
VMZ BERLIN N BETREIBERGESELLSCHAFT MBH	Germany	VMZ
ARRIVA PERSONENVERVOER NEDERLAND BV	Netherlands	ARRIVA NL
RUPPRECHT CONSULT-FORSCHUNG & BERATUNG GMBH	Germany	Rupprecht
POLIS NETWORK	Belgium	POLIS
PNO INNOVATION SL	Spain	PNO
REGION OF CENTRAL MACEDONIA	Greece	RCM
CITYLOGIN IBERICA SL	Spain	CITYLOGIN
PROVINCIE ZUID-HOLLAND	Netherlands	PZH
UNIVERSITY COLLEGE LONDON	United Kingdom	UCL
EMPRESA DE BLAS Y COMPAÑÍA, S.A.	Spain	ARRIVA ES

Document history

Version	Date	Organisation	Main area of changes	Comments
Version 0.1	01/10/2024	CERTH-HIT	Initial Draft	Internal
Version 0.2	24/10/2024	CERTH-HIT	Initial Draft	To be reviewed
Version 0.3	28/10/2024	DEUSTO	Reviewed Draft	
Version 1.0	31/10/2024	CERTH-HIT	Final	

List of Figures

List of Tables

List of acronyms

API	Application Programming Interface
CAN	Controller Area Network
CERN	Conseil Européen pour la Recherche Nucléaire
C-ITS	Cooperative Intelligent Transport Systems and Services
CKAN	Comprehensive Knowledge Archive Network
CSV	Comma-separated values
DATEX	Digital Analog Telecommunications Experimenter
DMP	Data Management Plan
DMS	Data Management System
DPIA	Data Protection Impact Assessment
EC	European Commission
ETA	Estimated Time of Arrival
ETSI	European Telecommunications Standards Institute
FAIR	Findable, Accessible, Interoperable, and Reusable
GDPR	General Data Protection Regulation
GPS	Global Positioning System
ICT	Information and Communication Technology
JSON	JavaScript Object Notation
MD	Message-Digest
OBU	On-Board Unit
RDF	Resource Description Framework
RSU	Road-Side Unit

1. Introduction

1.1. Project Overview

The SYNCHROMODE project aims to develop a data driven ICT toolbox for improving the management of transport operations from a multimodal perspective, to manage the transport network. SYNCHROMODE will provide transport managers with new predictive and network optimization capabilities for balancing the transport supply and demand, enabling efficient reaction to different types of events. SYNCHROMODE will demonstrate via well-chosen Case Studies the effectiveness of integrated multimodal and multi-actor traffic and transport management solutions and the SYNCHROMODE Toolbox, able to balance the demand load of both people and goods and, at the same time, reduce individual journey times.

1.2. Purpose of the document

This deliverable presents an in-depth analysis of the critical components within the data management policy intended for implementation in the SYNCHROMODE project. The document serves as a comprehensive repository pertaining to the overarching objective of SYNCHROMODE of guidelines and recommendations, facilitating adherence to the principles of making research data Findable, Accessible, Interoperable, and Reusable (FAIR). By doing so, the project aims to contribute significantly to the advancement of knowledge discovery and innovation. The adoption of the Horizon 2020 FAIR Data Management plan template, as outlined in Annex 1 of the EC guidelines [1], has been instrumental in laying the groundwork for the development of an effective data management strategy that encompasses the entire research data lifecycle and the various cross-cutting activities in data management (Figure 1).

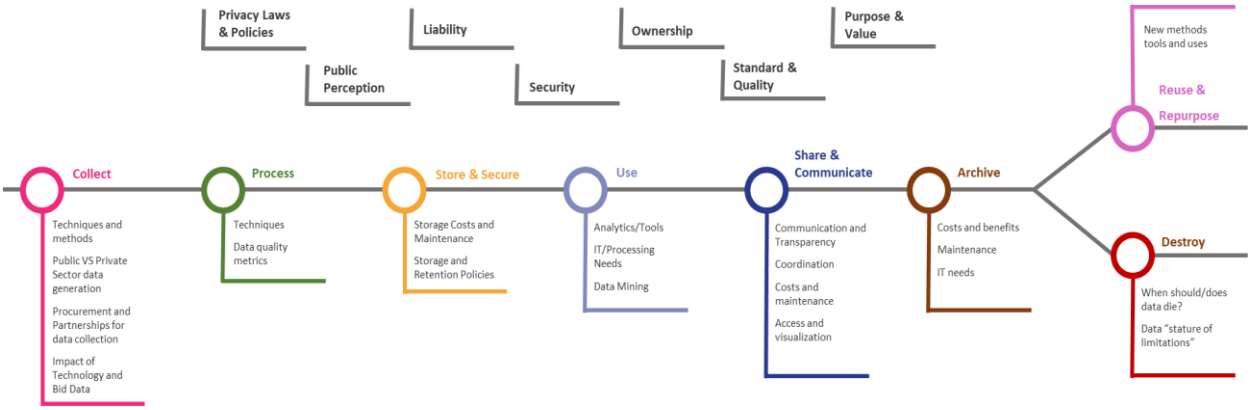


Figure 1: Example of the Data Management Life Cycle and the various Cross-Cutting Issues in Data Management (Based on [2]).

1.3. Intended audience

This deliverable is disseminated internally within the project's consortium. The intended readership comprises of the members of the SYNCHROMODE consortium and the EC's project officer.

1.4. Document structure

The structure of this document is as follows:

- Project description, intended audience, and Deliverable Interrelations
- Overview of data within SYNCHROMODE
- FAIR data principles
- Allocation of Resources
- Data Security
- Ethical Aspects
- Conclusions

It is important to note that this deliverable, entitled Data Management Plan (DMP) Version 2 and scheduled for completion by Month 18, remains a dynamic and evolving document. As with the first version, it is not static but will continue to adapt throughout the project's lifecycle. This approach ensures that any updates to datasets or changes in Consortium policies are incorporated, allowing the deliverable to accurately reflect the project's ongoing developments. A final version of the DMP will be produced by Month 36.

1.5. Interrelations

The DMP is part of WP1, meaning Project Management. As such it is a cross-cutting document and closely linked to other WPs and Tasks that include data collection and processing. An initial identification of such tasks is provided in Deliverable D1.3 “Ethics requirement and monitoring plan”.

2. Data Summary

2.1. Objectives of data summary section

The objectives outlined in this section are as follows:

- Commence the initial categorization of the data.
- Compile a comprehensive list of the data types that will be generated.
- Establish a catalogue of metadata elements to describe the generated data, simplifying their future reuse.
- Provide guidance and recommendations pertaining to data collection and sharing procedures throughout the project's duration and beyond.

2.2. Initial data categorization

Throughout the course of the SYNCHROMODE project, a diverse range of data sets will be gathered and generated (Figure 2). It is imperative that these data sets are stored in a manner that ensures easy accessibility for both human users and software applications.

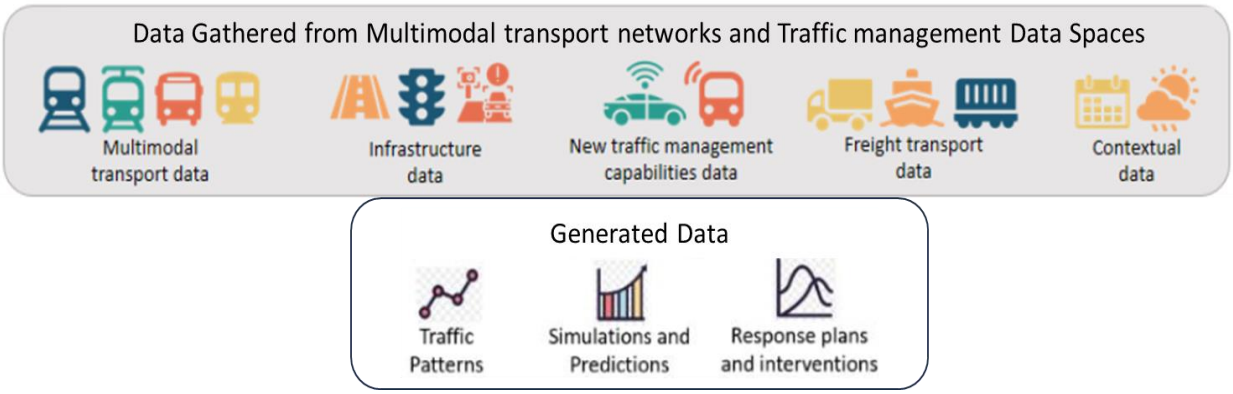


Figure 2: Examples of Gathered and Generated Data within SYNCHROMODE.

In a broad sense, these datasets can be categorized into two distinct groups:

- Organisational Data: This pertains to data concerning the characteristics of organizations involved, their internal structures, and processes that are relevant to the implementation of the Research and Innovation Action.
- Technical and Scientific Data: This category encompasses raw and processed experimental data, scientific analyses and publications, as well as software code and algorithms.

The detailed specifications regarding the various types, formats, and standards for the data to be collected, processed, and/or generated during the project will be developed within the framework of the different work packages.

Over the project's lifecycle and within its work packages, the consortium will collaborate on defining and refining all types of research data to be managed during and after the project's conclusion. This

includes determining the methodologies and standards to be applied, as well as strategies for sharing, curating, and preserving the data.

2.3. List of Data Types

The SYNCHROMODE project will collect a large amount of raw data including, among others, real-time GPS-data, vehicle sensor data, traffic plans, routes, occupancy levels, Public Transport ETAs, road infrastructure data etc. Table 1 depicts an updated list of Technical and Scientific data to be gathered and/or generated through SYNCHROMODE. In this updated version of the deliverable there are new additions under category "Other data".

Table 1 List of technical and scientific data within SYNCHROMODE.

Transport Data	Demand Data
Traffic lights status (real-time)	Public transport ticketing data
Traffic lights programs (library)	Public transport smart card data
Variable Message Signs (real-time)	Origin-Destination matrices unimodal
Variable Message Signs (library)	Shared mobility usage/demand data
C-ITS message broadcasted (real-time)	On-demand mobility usage/demand data
Traffic management plans, incl. triggers, targets & actions (real-time)	Survey data
Traffic management plans, incl. triggers, targets & actions (library)	Official statistics data
Traffic counts (real-time, from any traffic measuring system)	Delivery demand data (freight)
Traffic counts (historical, from any traffic measuring system)	Parking data (real-time)
Travel times (real-time, from any traffic measuring system)	**Supply Data**
Travel times (historical, from any traffic measuring system)	Public transport schedules
Speed (real-time, from any traffic measuring system)	Public transport lines
Speed (historical, from any traffic measuring system)	Public transport stops
Floating Car Data (raw data, real time)	Transport network supply data
Floating Car Data (raw data, historical)	On-demand mobility supply data
C-ITS OBU Data (raw data, real time)	Shared mobility supply data

C-ITS OBU Data (raw data, historical)	Parking supply data
C-ITS RSU Data (raw data, real time)	Freight fleet data
C-ITS RSU Data (raw data, historical)	
Shared mobility data (real-time)	
Bridge-opening (historical, real-time)	
Events (planned and unplanned, historical, real-time)	

Other data	
Data from mobile-phone records	Weather related data
Smartphone data	Mobile network data
Mobile app data	Ticket sales Keukenhof data
CAN data	Holidays (national, school) of Netherlands
Land use data	Census data
Household travel survey data	Logistics consolidation centers

Personal data
Name
Email
Organization/Role

2.4. Metadata Elements

Each data set collected or generated within SYNCHROMODE will be accompanied by properly formatted metadata that will contain the information summarized in Table 2.

Table 2 Metadata of data gathered and/or generated within SYNCHROMODE.

Metadata	Description
Identifier	Unique identifier index for each dataset and Case Study.
Category	The general category to which the dataset belongs (Traffic, Demand, Supply, Other).
Sub-category	A distinct and specialized category within the general category.

Metadata	Description
Source Name	The name of the dataset.
Source Description	A brief description of the dataset and its characteristics.
Modes	The modes of transportation associated with the dataset (Road, Rail, Maritime, Air).
Type	The type of transportation data represented in the dataset, specifying whether it pertains to passenger transportation or freight transportation.
Date created	The date when the dataset was initially created or compiled.
Date last updated	The most recent date when the dataset was modified or updated.
Temporal Coverage	The temporal coverage of the dataset indicates if data is static (fixed data), historical, real-time, or prediction.
Geographic Coverage	The geographic area or region to which the dataset pertains (Urban, Inter-urban, Rural, National).
Granularity	The level of detail/resolution at which the data is aggregated (Yearly, Quarterly, Monthly, Weekly, Daily, Hourly, Minute-level, Second-level).
Standard Used	The standard or protocol followed for data collection, formatting, ensuring consistency and compatibility with established practices.
Language	The language in which the dataset is documented/ recorded.
Format	The file format in which the dataset is stored, such as Excel, CSV, JSON, or other applicable formats.
Sample Size	Percentage of the total population/trips/etc. captured by the data source.
Access Type	The type of access granted to the dataset, indicating whether it is open, restricted, or requires specific authorization or permissions.
API Link	If applicable, a link to an API that provides access to the dataset programmatically.
Provider	The organization, entity, or individual that provides the dataset.
Owner	The entity or individual who owns or has ownership rights over the dataset.
License Type	The type of license or legal agreement governing the usage and distribution of the dataset, ensuring compliance with applicable terms and conditions.
Responsible Partner	Project's partner responsible for managing the dataset.

2.5. Data Sharing

The SYNCHROMODE project will strive to offer open access to data but cannot overrule limitations that partner institutions put on data that they contribute. Moreover, an ethical approach will be adopted and maintained throughout the project's lifetime. The responsible partners will assure that the EU standards regarding ethics and data management are respected and applied. Figure 3 depicts the process though which data are to be uploaded.

Data controller refers to the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of data. In SYNCHROMODE, indicative data controllers include the Project Coordinator as well as Work Package and Task Leaders, as they have been defined under deliverable "D1.1 Project & Innovation Management plan".

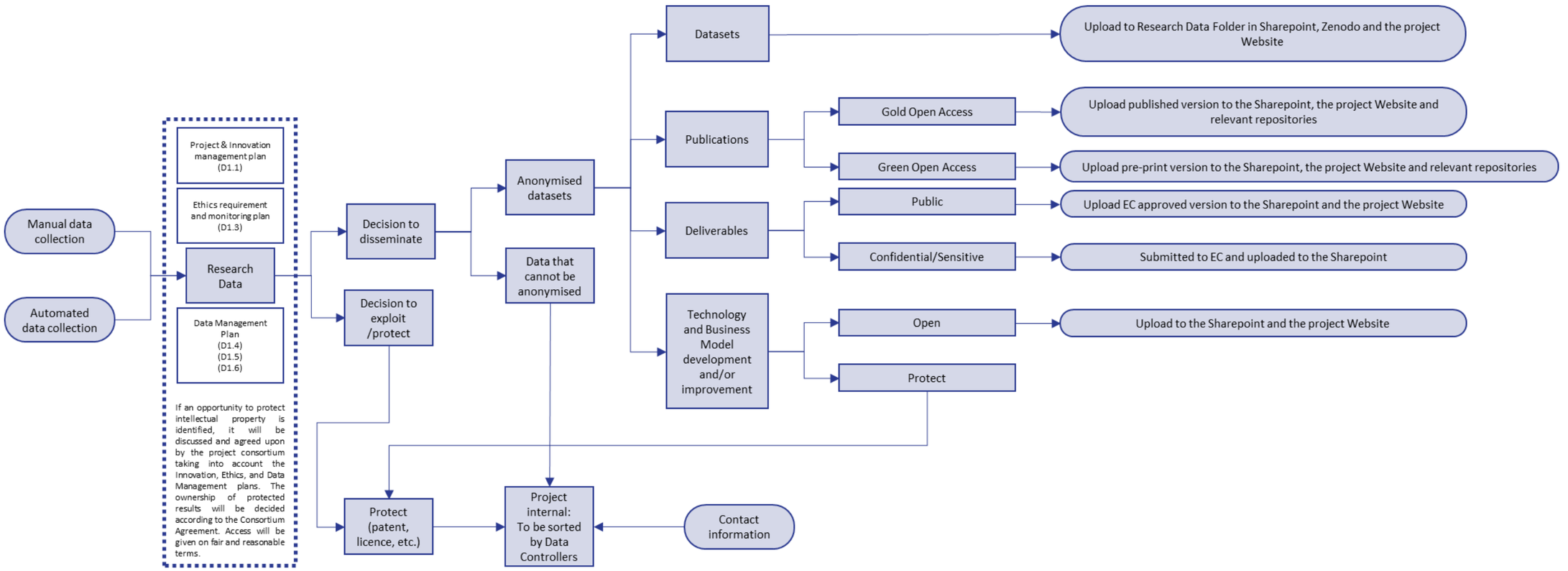


Figure 3: Process for uploading data sets (Based on: (3)]

3. FAIR Data

3.1. The concept of FAIR data

In the context of SYNCHROMODE, it is essential that the generated data adheres to the FAIR principles, which stand for Findable, Accessible, Interoperable, and Reusable. These principles (Figure 4) should not be seen as prescriptive standards or mandates that dictate specific technologies or implementations. Instead, they serve as a flexible framework to guide us when designing a Data Management Plan. A proposed workflow for achieving data "FAIRification", as recommended in the relevant literature, is summarized in Figure 5. The following subsections will elaborate on how these principles will be applied to ensure that the critical aspects of data lifecycle management are comprehensively addressed.

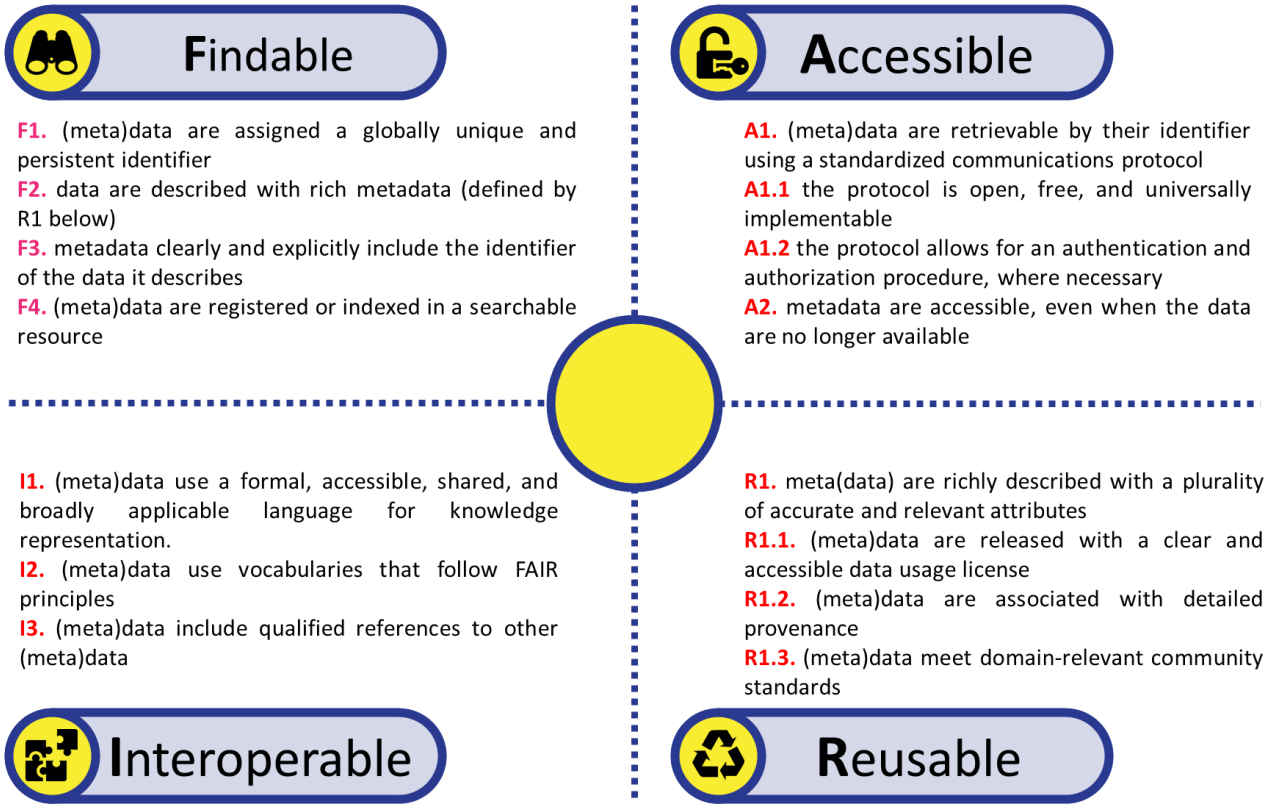


Figure 4: Principle of FAIR data (Based on [4]).

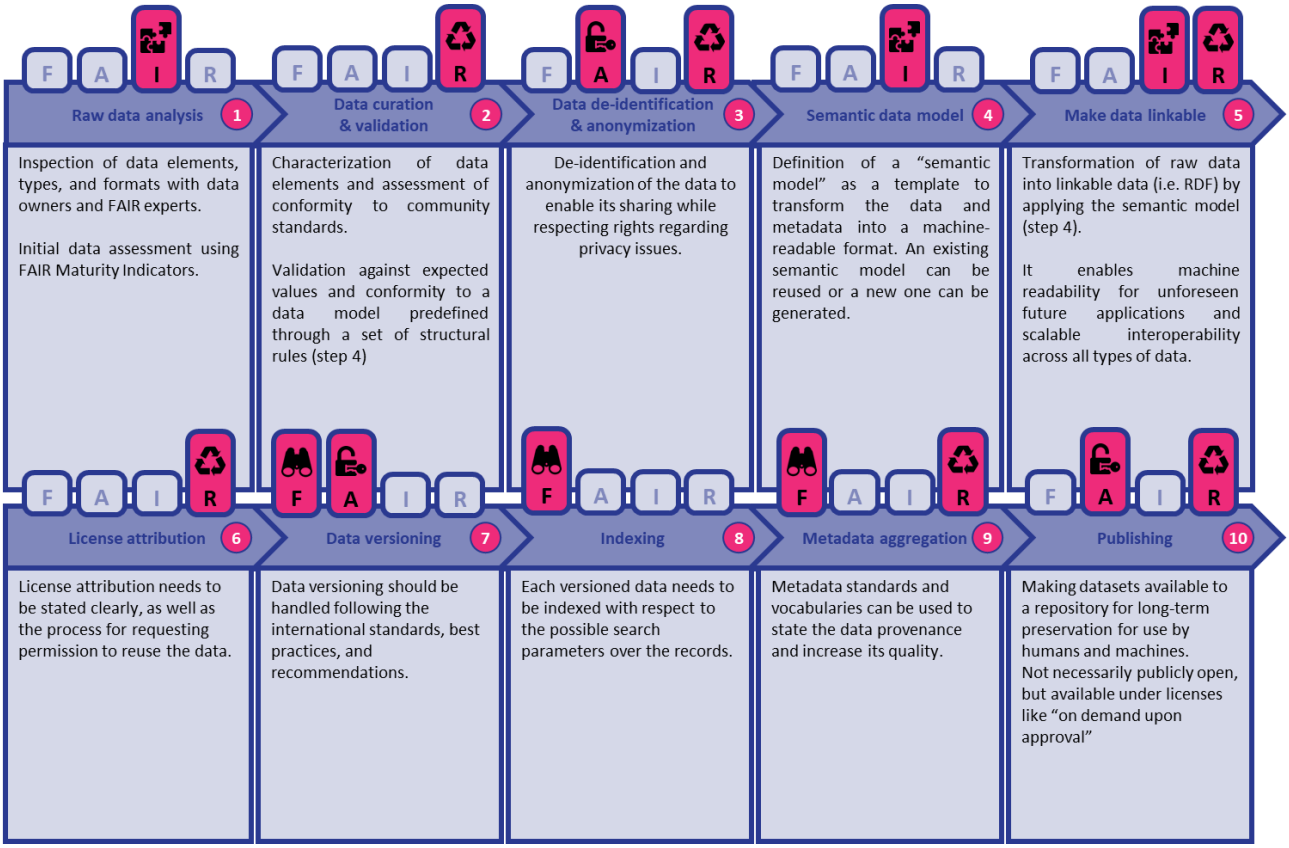


Figure 5: Indicative workflow for the "FAIRification" of data (Based on: [4],[5],[6]).

3.2. Making Data Findable

The data gathered and generated along the SYNCHROMODE project will be identifiable and locatable by means of unique identification mechanisms. Files will be uniquely identifiable by using standardised name conventions and clear versioning.

Both internal and external data will be systematically collected and documented, with the inclusion of relevant metadata, as previously shown in Table 2, and in accordance with the standards set by promoted by the European Data Portal, which defines metadata as follows [7]:

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

3.3. Making data Accessible

As already mentioned, an effort will be made in SYNCHROMODE for data to be made available with open access. This will not apply to the datasets that will be deemed by the Consortium as IPR protected such as proprietary data owned by consortium beneficiaries or third legal parties.

During the lifespan of SYNCHROMODE Project, any data that collected or generated will be stored and systematically organised in the official project repository. This may be applied specifically for static data, whose nature implies that the information will change over time. Also, this is the case for data collected during any dissemination activity of the project or under any other conditions. All This repository is hosted in a dedicated SYNCHROMODE space on SharePoint. SharePoint is a web-based collaborative platform developed by Microsoft. It is primarily a document management and storage system, but the product is highly configurable and contains team collaboration groupware capabilities, so it can be used as collaborative software. The SYNCHROMODE space will be used both for knowledge exchange and collaboration purposes by the project members and as a repository to securely store and share files, making data available to the whole consortium. The repository offers access only to project partners and to authorised system administrators and cannot be accessed by external users.

As indicated in Figure 3, data/information that have been deemed as not to be publicly shared will be stored solely on SharePoint.

As shown in Figure 6, folders withing SharePoint are organised in a hierarchical and clear structure. Files are uniquely identifiable and versioned by using the name convention adopted by the project.

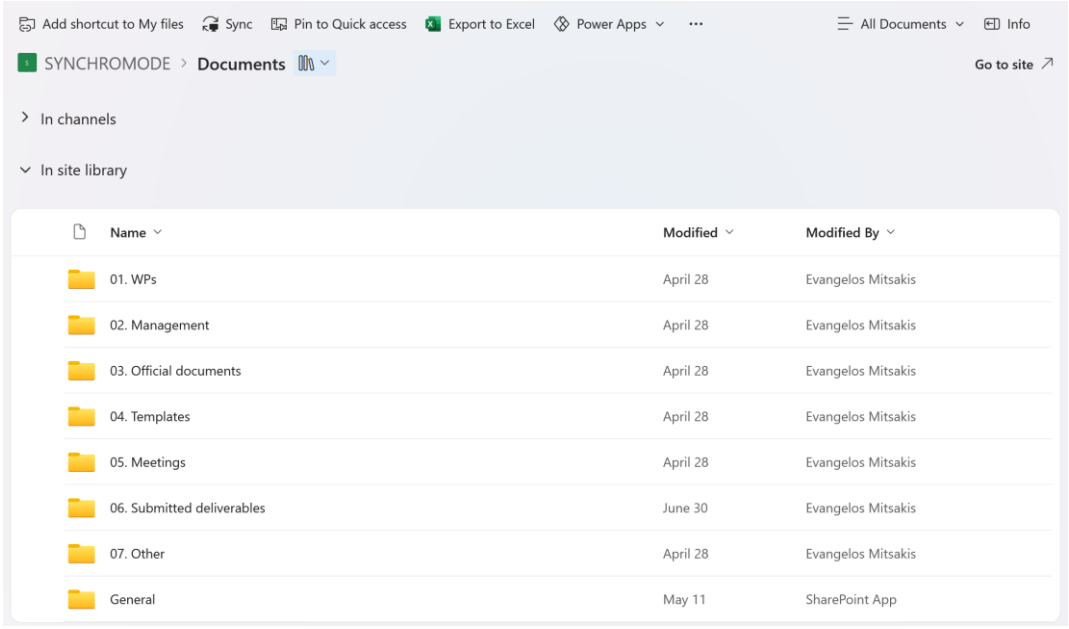


Figure 6: SYNCHROMODE SharePoint

Since the SharePoint has as a main functionality the sharing of documents, a CKAN based repository has been developed for the exchange of data within the consortium. CKAN is an open-source Data

Management System (DMS) that offers an intuitive and secure means of accessing data and its metadata, along with a variety of plugins that expand its functionality.

The developed repository, named the "SYNCHROMODE Data Exchange Repository", is accessible through the following link: https://data.synchromode.eu/. This platform serves a dual purpose. First, it functions as an exchange platform, enabling organizations and users to exchange datasets via a User Programmatic Interface (UPI), commonly known as software APIs. Through this platform, data shared among the project consortium can be accessed and downloaded, facilitating the development and testing of module results without the need for the fully developed SYNCHROMODE Toolbox databases. Additionally, the platform acts as a data catalogue, providing access to all available datasets generated during the development, testing, and validation phases of the various modules.

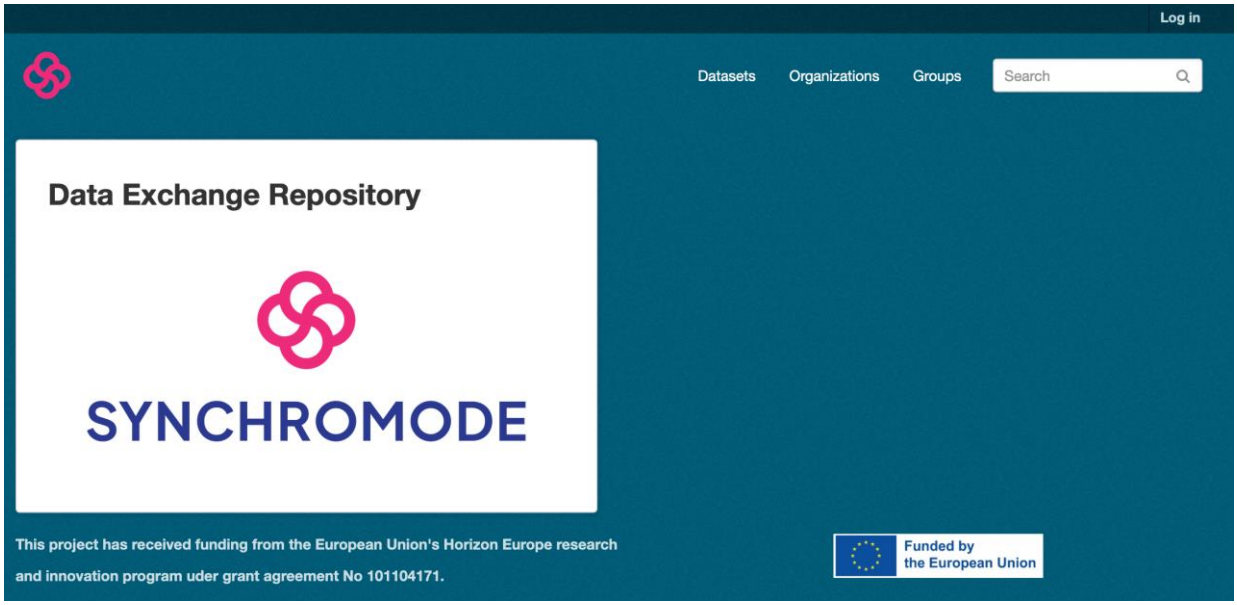


Figure 7: Landing page of the SYNCHROMODE Data Exchange Repository

Users must be registered and provided with a username and password to access this platform. External users, outside the project consortium, cannot access the platform as the sign-in functionality is hidden. Detailed information about accessing the SYNCHROMODE Data Exchange Repository, as well as instructions for working with the platform, can be found in deliverable "D3.2: APIs for Data Harmonisation and Fusion – Version 1". The manual for using the platform is located in Annex I of deliverable D3.2.

The rules and principles of the European Commission's Horizon Europe Framework Programme clearly establish that scientific results generated within Horizon 2020 projects will be made available as open access publications, i.e., free of charge online access for any user. Following these rules, open access will be ensured to all peer-reviewed scientific publications related to SYNCHROMODE results. Furthermore, other project outputs and results may be provided in the Open Access section upon respective decision of the project consortium.

According to the "Open access to publications and data in Horizon 2020 Fact sheet" [8], two main routes exist for open access to scientific peer-reviewed publications:

- Self-archiving (also called Green open access) means that the published article or the final peer reviewed manuscript is archived by the researcher – or a representative – in an online repository before, after, or alongside its publication. Access to the article is often – but not necessarily –delayed ('embargo period') as some scientific publishers may wish to recoup their investment by selling subscriptions and charging pay-per-download view fees during an exclusivity period.
- Open access publishing (also called Gold open access) means that an article is immediately provided in open access mode by the scientific publisher. The associated costs are shifted away from readers, and instead to (for example) the university or research institute to which the researcher is affiliated, or to the funding agency supporting the research.

Since these two routes described previously are not mutually exclusive in an EU-funded action, within the SYNCHROMODE project each beneficiary will be able to choose the most suitable approach for each publication concerned following the guidelines mentioned in Figure 3.

Moreover, all project datasets deemed as shareable will be securely stored in a private cloud-based repository, such as ZENODO (https://zenodo.org/), a free service developed by CERN under the EU FP7 project OpenAIREplus (grant agreement no. 283595). This repository will contain comprehensive information about the software, tools, and instruments used by the dataset creators, ensuring that secondary data users can access and validate the results. The SYNCHROMODE data collection will be accessible in the ZENODO repository via a similar address structure, like the following link: https://zenodo.org/collection/<<synchromode>>.

To facilitate access, datasets in the cloud repository will be linked to the management and exploitation portals of the project and assigned DOIs (Digital Object Identifiers) to allow third parties to retrieve them.

Deliverables, publications, and, when appropriate, datasets will also be uploaded to the project website (https://synchromode.eu/).

To guarantee the security of these datasets, we will employ the latest security features as provided by the repository platform, including but not limited to firewall protection, password security, and encryption. If any intellectual property rights (IPR) issues arise during data sharing, they will be handled according to established protocols. Additionally, all necessary supplementary materials, such as software for parsing the datasets and standards documents, will be made available by the consortium through the data management portal.

3.4. Making Data Interoperable

Data assets collected in and stored in SharePoint and CKAN-based repository, are accompanied by their appropriate metadata, as mentioned in previous sections, and the reasons for any extensions that may be made to this will be appropriately documented.

Data interoperability is foreseen in the project through conformance to standards. Specifically, SYNCHROMODE data will aspire to conform to a number of standards, including the following:

- ETSI EN 302 637-2
- ETSI EN 302 637-3
- ETSI TS 102 894-2
- OCIT-C
- DATEX II
- GTFS
- NETEX
- DCAT-AP (D3WC initiative)
- Coordinated Metadata Catalogue
- CAM ETSI EN 302 637-2
- DENM ETSI EN 302 637-3
- CDD (Common Data Dictionary for all messages)
- SPAT-MAP ETSI TS 103 301
- IVIM ISO/TS 19321
- Road Sign codes ISO/DIS 14823

3.5. Making Data Reusable

As previously indicated, data will be treated on a case-by-case basis during the project. Once a data set is marked as public, and, therefore, made publicly available on ZENODO, it will be fully reusable (with the possibility of specifying embargo period or with controlled access to whitelist of persons; see ZENODO policies).

The ZENODO repository ensures sustainable archiving of the final research data. Items deposited in ZENODO will be retained for the lifetime of the repository, which is currently the lifetime of the host laboratory CERN and has an experimental programme defined for the next 20 years.

All publicly available uploads on ZENODO will be stored safely for the future in the same cloud infrastructure as research data from CERN's Large Hadron Collider and using CERN's battle-tested repository software INVENIO [9], which is used by some of the world's largest repositories such as INSPIRE HEP [10] and CERN Document Server. The data will remain re-usable at least until ZENODO discontinues the dataset(s) (i.e., warrantied for a minimum of 20 years). The project envisages adopting the 'data pedigree' concept, which assure that each piece of relevant information is traceable back to the original data sources. This data lineage along with metadata allows for quality audit and sensitivity analyses of the outputs.

The public project deliverables will be available for download on the website www.synchromode.eu after their submission to the EC and/or through the ZENODO repository.

Data that are indicated as public may also, when possible, be made publicly available through the associated National Access Points (NAPs). Figure 7 depicts a snapshot of the Greek NAP, in which various transport related data are made available.

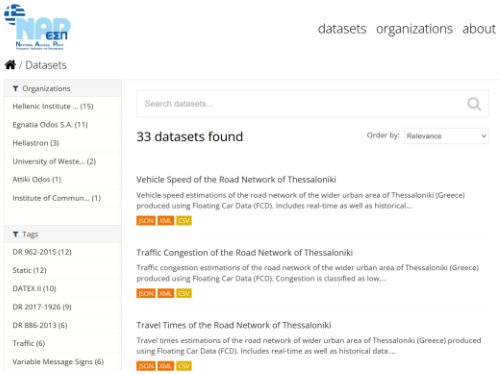


Figure 8: The Greek NAP.

Confidential deliverables are stored on the project's internal repository (SharePoint), to which access is restricted. Similarly confidential datasets will be stored on the project's CKAN based repository. Confidential deliverables and datasets might be requested by external parties, in which case the consortium might make decision to disseminate corresponding deliverables or specific parts of the deliverables or data to external parties. The project deliverables on the website will be provided in widely adopted PDF format.

3.6. FAIR Maturity Indicators

In the relevant literature FAIR maturity models have been suggested. One such model [11] is based on a list of FAIR Maturity Indicators, as those are summarised in Table 3. Each project partner will strive to evaluate their project related data against these indicators.

Table 3 Data FAIR maturity assessment (based on [14])

ID	FAIR	Indicator		Priority
1	F1	Metadata is identified by a persistent identifier	□□□	Essential
2	F1	Data is identified by a persistent identifier	□□□	Essential
3	F1	Metadata is identified by a globally unique identifier	□□□	Essential
4	F1	Data is identified by a globally unique identifier	□□□	Essential
5	F2	Rich metadata is provided to allow discovery	□□□	Essential
6	F3	Metadata includes the identifier for the data	□□□	Essential
7	F4	Metadata is offered in such a way that it can be harvested and indexed	□□□	Essential
8	A1	Metadata contains information to enable the user to get access to the data	□□	Important

ID	FAIR	Indicator		Priority
9	A1	Metadata can be accessed manually (i.e. with human intervention)	□□□	Essential
10	A1	Data can be accessed manually (i.e. with human intervention)	□□□	Essential
11	A1	Metadata identifier resolves to a metadata record	□□□	Essential
12	A1	Data identifier resolves to a digital object	□□□	Essential
13	A1	Metadata is accessed through standardised protocol	□□□	Essential
14	A1	Data is accessible through standardised protocol	□□□	Essential
15	A1	Data can be accessed automatically (i.e. by a computer program)	□□	Important
16	A1.1	Metadata is accessible through a free access protocol	□□□	Essential
17	A1.1	Data is accessible through a free access protocol	□□	Important
18	A1.2	Data is accessible through an access protocol that supports authentication and authorisation	□	Useful
19	A2	Metadata is guaranteed to remain available after data is no longer available	□□□	Essential
20	I1	Metadata uses knowledge representation expressed in standardised format	□□	Important
21	I1	Data uses knowledge representation expressed in standardised format	□□	Important
22	I1	Metadata uses machine-understandable knowledge representation	□□	Important
23	I1	Data uses machine-understandable knowledge representation	□□	Important
24	I2	Metadata uses FAIR-compliant vocabularies	□□	Important
25	I2	Data uses FAIR-compliant vocabularies	□	Useful
26	I3	Metadata includes references to other metadata	□□	Important
27	I3	Data includes references to other data	□	Useful
28	I3	Metadata includes references to other data	□	Useful
29	I3	Data includes qualified references to other data	□	Useful
30	I3	Metadata includes qualified references to other metadata	□□	Important
31	I3	Metadata include qualified references to other data	□	Useful
32	R1	Plurality of accurate and relevant attributes are provided to allow reuse	□□□	Essential
33	R1.1	Metadata includes information about the licence under which the data can be reused	□□□	Essential

ID	FAIR	Indicator		Priority
34	R1.1	Metadata refers to a standard reuse licence	□□	Important
35	R1.1	Metadata refers to a machine-understandable reuse licence	□□	Important
36	R1.2	Metadata includes provenance information according to community- specific standards	□□	Important
37	R1.2	Metadata includes provenance information according to a cross- community language	□	Useful
38	R1.3	Metadata complies with a community standard	□□	Important
39	R1.3	Data complies with a community standard	□□□	Essential
40	R1.3	Metadata is expressed in compliance with a machine-understandable community standard	□□□	Essential
41	R1.3	Data is expressed in compliance with a machine-understandable community standard	□□	Important

Essential	Such an indicator addresses an aspect that is of the utmost importance to achieve FAIRness under most circumstances, or, conversely, FAIRness would be practically impossible to achieve if the indicator were not satisfied.
Important	Such an indicator addresses an aspect that might not be of the utmost importance under specific circumstances, but its satisfaction, if at all possible, would substantially increase FAIRness.
Useful	Such an indicator addresses an aspect that is nice-to-have but is not necessarily indispensable.

4. Allocation of resources

SYNCHROMODE partners have to observe the policies set out in this DMP and datasets have to be created, managed and stored appropriately, to respond to the data management challenges in an efficient manner.

When deemed necessary a preliminary (DPIA) will be conducted at case study sites in collaboration with the data collectors and processors.

As this DMP is currently based, mostly, on the use of free resources and open-source software, the main costs that will be incurred are related to the server(s) (hardware) required to run them and the working time needed to setup, maintain, and evolve the different tools (efforts measured by person-months).

5. Data Security

ZENODO and the project repositories will ensure secure and safe storage of both public and non-public data. ZENODO provides clear security guarantees. All data files are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest.

Data files and metadata are backed up on a nightly basis. Files are regularly checked against their checksums (using the MD5 algorithm [12]) to assure that file content remains constant. In case of closure of the repository, ZENODO ensures that efforts will be made to integrate all content into suitable alternatives.

The project repository is hosted on a SharePoint site, which is only accessible by authorised users (the project coordinator maintains strict control as to who is able to access the system). Files containing confidential data should be protected by owners using local encryption tools (i.e., password-protected archives) before being uploaded to shared repositories. Interaction through web user interfaces will use https protocol (i.e., secure). To assure data privacy, all sensitive data will be anonymised, encrypted, and stored on servers to which only the relevant staff have access. More specifically the servers onto which the data will be stored will have server-side encryption. This means that the server's administration personnel will be able to generate public keys for specific personnel who will have access to the data but will not be able to access the data themselves (since the private keys required for this access will be generated on the machine of the person with access to the data). This means that only the required personnel will have access to the data and, even in the unlikely case of a possible data leak or server hack, the data stolen will be fully encrypted and thus fully non-accessible. After a retention period (to be defined), a secure deletion software will be used to destroy data. If deemed necessary, a full format can be used in conjunction with overwriting, to provide further assurance that data cannot be recovered, guaranteeing the destruction of the project personal data.

Similarly, CKAN offers built-in permissions and access controls that can be used to ensure that only authorized users have access to data. It also has several security features that can be used to secure the data and the communication between the CKAN server and the clients. Additionally, the SYNCHROMODE CKAN instance will be backed up on a monthly basis.

The following guidelines will be used in order to ensure the security of the data:

- Keep anonymised data and personal data of pilot participants separate,
- Encrypt data if it is deemed necessary by local researchers,
- Store data in at least two separate locations to avoid loss of data (this is already handled via the SharePoint's internal cloud backup system),
- Limit the use of USB flash drives,
- Label files in a systematically structured way in order to ensure the coherence of the final dataset.

6. Ethical Aspects

No data should be processed without undergoing some ethical considerations first. These considerations will be in accordance to the guidelines set in D1.3 (Ethics requirement and monitoring plan) and help to enforce privacy regulations and, moreover, to ensure no rights are violated and all data has been obtained in a consensual manner.

The project consortium is aware of the importance to keep privacy and protect the personal data, so the following standards are stablished to ensure that no personal data:

- will be handed out to third parties outside the SYNCHROMODE project,
- will be exploited or commercialized without explicit consent,
- will be kept for longer than absolutely necessary,
- will be accessible for use or diffusion outside the project framework,
- will not be subject to retrieval in case it is requested,
- will not be destroyed as the relevant scientific purpose is fulfilled.

Ethical aspects, specifically those related to protection of personal data, are addressed in detail in D1.3 “Ethics requirement and monitoring plan”.

Moreover, during the first half of the project’s lifetime (M1 – M18), and based on the initial estimation of tasks requiring ethics clearance, only Task 2.1 “User Needs & System Requirements Definition”, was required to obtain ethics approval. For the other tasks of the project that have already started or finished, and which were marked in deliverable “D1.3 Ethics Requirement and Monitoring Plan” as potentially requiring an ethics clearance request, no submission was necessary, as the data collected, and the activities performed do not require ethics clearance. The ethics clearance request for Task 2.1 can be found in Annex I.

7. Conclusions

This deliverable presents a comprehensive analysis of the data management policy that will guide the SYNCHROMODE project. It consolidates key guidelines and information related to SYNCHROMODE data, resource allocation, data security, and ethical considerations (in conjunction with D1.3). The primary objective is to establish a framework to ensure that the project's research data is Findable, Accessible, Interoperable, and Reusable (FAIR), thereby enhancing knowledge discovery and promoting innovation.

In this updated version, additional emphasis is placed on ensuring the integration of evolving data standards, improving data quality management, and further refining the sharing mechanisms within the project consortium and beyond. Furthermore, advancements have been made in the data sharing process, security protocols, and ethical compliance, all of which align with the Horizon Europe guidelines.

The Horizon Europe (formerly Horizon 2020) FAIR Data Management Plan template, provided as Annex 1 of the EC guidelines, continues to serve as the foundation for crafting an efficient data management strategy. This updated deliverable reflects the most recent changes in the project's data management approach, covering the complete research data lifecycle and accounting for the dynamic nature of data collection, processing, and dissemination in the context of SYNCHROMODE.

8. References

[1] EC, H2020 templates: Data management plan v2.0, 2018, Available at: https://ec.europa.eu/research/participants/data/ref/h2020/other/gm/reporting/h2020-tpl-oa-data-mgt-plan-annotated_en.pdf

[2] K. Miller; M. Miller; M. Moran; B. Dai, Data Management Life Cycle, 2018, Texas A&M Transportation Institute

[3] C. W. Hanssen et al. GREENCHARGE-D1.1 Data Management Plan, 2021

[4] M. Wilkinson et al., The FAIR Guiding Principles for scientific data management and stewardship, 2016, Scientific Data

[5] A.A.Sinaci et al., From Raw Data to FAIR Data: The FAIRification Workflow for Health Research, 2020, Methods Inf Med, doi: 10.1055/s-0040-1713684

[6] N. Tachos et al. ProCancer-I: D3.5 Initial Data Management Plan Initial Data Management Plan, 2021

[7] Dublin Core Metadata Element. http://dublincore.org/

[8] https://ec.europa.eu/programmes/horizon2020/sites/default/files/FactSheet_Open_Access.pdf

[9] https://inveniosoftware.org/

[10] https://inspirehep.net/

[11] C. Bahim et al. The FAIR Data Maturity Model: An Approach to Harmonise FAIR, 2020, Data Science Journal

[12] https://www.sciencedirect.com/topics/computer-science/message-digest-algorithm-5

Annex I – Ethics Clearance Submissions

The following form has been filled in by partner Rupprecht regarding Task 2.1 “User Needs & System Requirements Definition”.

Task Activity Details
Briefly describe the activity within the task for which clearance is requested.
T2.1 under WP2 involves an online survey designed to evaluate user perspectives about each case study in the SYNCHROMODE project. This survey aims to understand perspectives from different user groups including their needs, expectations, and concerns regarding multimodal traffic management.

Data Collection Procedures
What is the primary objective of your data collection?
To evaluate user perspectives about each case study within the SYNCHROMODE project in Greece, Netherlands and Spain.
What type of data collection method will you be using (e.g., survey, interview, observation, measurements, etc.)?
The data collection will be carried forward via online survey, using the Qualtrics platform, which will be anonymous.
How will the data actually collected (in-person, online, via phone, via equipment (e.g. detectors) etc.)?
Data will be collected online via Qualtrics platform. The dissemination will take place via social media platforms, QR codes, and emails (participant’s email address will not be recorded).
What steps will be taken to ensure the accuracy and reliability of the data collected?

To ensure the accuracy and reliability of the collected data, we took the following steps while designing the survey:

1. We conducted a literature study of user needs towards multimodal transport network through projects in the EU, Eurobarometer surveys, and guidelines focusing on survey methods and data management.
2. We revised the initial draft survey multiple times based on feedback from internal experts, case study leaders, and consortium partners in the project.
3. We will make the survey available in the local language of each case study location, including Greek, Dutch, Spanish, and English, to ensure that the survey respondents can easily understand the questions.
4. The survey follows 'display logic' which narrows down questions related to the case study areas, based on the user's response in each section.

Do you expect any challenges during data collection?

Since the survey takes between 10-15 minutes to complete, it may pose certain challenges to the overall response rate in each case study location. The level of dissemination, along with unexpected situational factors, will influence the number of respondents for the survey.

Are there any limitations in your data collection method that should be noted?

The digital format of the survey may pose a challenge for individuals who have limited access to digital tools such as smartphones, tablets, or laptops.

Anonymisation Procedures

Which types of data do you anonymize (e.g., personal identifiers, sensitive data)?

The survey does not collect direct personal data to identify a person. No names, address, email addresses, or telephone numbers are requested within the survey. 'User characteristics' within the survey have questions, where the variables such as age or salary are anonymized. The only data requested within the survey is that of a postal code (for the case study in Netherlands) which is also made optional for the user to respond to. This assists in masking the direct linkability to the user.

How do you determine which data elements require anonymization?

The indirect variables within the user survey which relate to the user directly or indirectly are the 'user characteristics'. The 'user characteristics' section of the user survey includes questions with

multiple options in the format of 'range', which creates a masking effect on the overall outcome. For instance, for the user's age, options are available in the format '25-34 years old', '35-44 years old', and so on. This follows the principle of 'generalization' as one of the anonymization techniques mentioned within the Article 29 Data Protection Working Party. In addition, it also allows the participant to select the option 'prefer not to say', enabling voluntary participation.

Are your data anonymization procedures compliant with relevant data protection laws and regulations (e.g., GDPR, national regulations)?

Yes.

Are there any instances where data utility is significantly impacted due to anonymization?

The level of significance is subjective considering the user characteristics, as it has already gone through generalization process. For instance, the user groups falling within the category of 'persons with disability' or 'persons with health issues' have been grouped together to mask any links with the physical characteristics of a person.

At what stage in the data processing cycle is anonymization applied?

The data is anonymized prior to any data processing cycle through generalization of the survey questionnaire.

Are there any internal or external audits conducted to ensure the integrity of the anonymization process?

The survey has been subject to multiple revisions based on the feedback from the project's case study leaders (including other internal partners) and the questions have been revised to address different case study solutions and involve the least user characteristics where possible, which indirectly assists in reducing any linkage with survey respondents.

Data Storage Procedures

What types of data storage solutions are utilized (e.g., on-premises, cloud-based, hybrid)?

Cloud-based data storage options are available via Qualtrics and SharePoint platforms.

What security controls are in place to protect stored data (e.g., encryption, access controls)?

The servers onto which the data will be stored will have server-side encryption. This means that the server's administration personnel will be able to generate public keys for specific personnel who will have access to the data but will not be able to access the data themselves (since the private keys required for this access will be generated on the machine of the person with access to the data).

How do you classify the sensitivity of the data being stored?

Considering personal data, and the description from the European Commission, considers following data as sensitive:

- personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs;
- trade-union membership;
- genetic data, biometric data processed solely to identify a human being;
- health-related data;
- data concerning a person's sex life or sexual orientation.

The survey does not address any of these topics specifically but only categorizes people with disabilities or health issues in one group, to focus on vulnerable user needs towards multimodal traffic management. The question does not ask the user for any medical conditions or details related to it in the past or present.

Data Access

Who will have access to the stored data, and how is this access managed and tracked?

Rupprecht will have the initial access to the stored data from Qualtrics, which will be made available via SharePoint platform to other partners upon request (following agreement from the project coordinator).

What procedures are in place for granting, reviewing, and revoking access to data storage systems?

The project repository is hosted on a SharePoint site, which is only accessible by authorised users (the project coordinator maintains strict control as to who can access the system).

What is the name and formal qualifications of the responsible researcher?

Dr.-Ing. Lakshya Pandit (from Rupprecht) will be the responsible researcher for the task. He has a PhD in the domain of measuring multimodal accessibility and is a two-time DAAD research scholar.

Data Destruction Procedures

Please describe your organization's data destruction policy.

Following D1.4, after a retention period (to be defined), a secure deletion software will be used to destroy data.

What specific methods will you use for data destruction (e.g., shredding, degaussing, overwriting)?

If deemed necessary, a full format can be used in conjunction with overwriting, to provide further assurance that data cannot be recovered, guaranteeing the destruction of the project personal data.

How will you handle the destruction of physical media (e.g., paper records, hard drives)?

No physical media will be utilized for the survey data records.

How do you handle situations where data is not properly destroyed or is accidentally recovered?

Only the required personnel will have access to the data and, even in the unlikely case of a possible data leak or server hack, the data stolen will be fully encrypted and thus fully non-accessible.

Informed Consent Procedures

Please describe the process by which informed consent will be obtained from non-consortium members.

For the survey participants, the initial section of the survey includes introduction about the project and consent which follows a survey logic. This section allows the participant to either consent to a voluntary participation or not. The latter option will terminate the survey immediately.
Will there be any specific forms or templates used for this documentation?
The survey includes the consent section inspired from the D1.3 Annex I example.
How will informed consent be documented?
This will be included as a separate section within the survey data repository.
Will the informed consent documentation be provided in the native language of the non-consortium member?
Yes, this will relate to the local language of the case study areas.
Who is responsible for obtaining informed consent from non-consortium members?
The local lead partners from the case study areas are responsible for the dissemination and result collection, which includes obtaining informed consent (part of the survey).
What is the procedure if a non-consortium member wishes to withdraw their consent?
Survey participants can opt out of the survey voluntarily at any point before submission. After submission, the data is anonymized, and it becomes difficult to retract the data corresponding to the exact individual since no name, email address, or address is obtained from them. Depending on the duration between the survey submission and withdrawal request, the data can be erased more accurately.