DES HE Data Management Plan

Deliverable ID:	**D2.3**
Project acronym:	**MultiModX**
Grant:	**101114815**
Call:	**HORIZON-SESAR-2022-DES-ER-01**
Topic:	**HORIZON-SESAR-2022-DES-ER-01-WA2-6**
Consortium coordinator:	**BHL**
Edition date:	**18 December 2025**
Edition:	**01.00**
Status:	**Official**
Classification:	**PU**

Abstract

This document describes the data management life cycle for the data to be collected, processed and generated by the MultiModX project. Through this document, the project aims to ensure that all the research data is findable, accessible, interoperable and reusable (FAIR) as well as that ethics and data security aspects are properly addressed.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Authoring & approval

Author(s) of the document

Organisation name	Date
Nommon	02/12/2025

Reviewed by

Organisation name	Date
BHL	18/12/2025

Approved for submission to the SESAR 3 JU by[1]

Organisation name	Date
BHL	18/12/2025

Rejected by[2]

Organisation name	Date

Document history

Edition	Date	Status	Company author	Justification
00.01	02/12/2025	Draft	Nommon	First Draft
01.00	18/12/2025	Final	Nommon, BHL	Submitted to the SESAR 3 JU

[1] Representatives of all the beneficiaries involved in the project

[2] Representatives of the beneficiaries involved in the project

Copyright statement

Disclaimer

MultiModX

INTEGRATED PASSENGER-CENTRIC PLANNING OF MULTIMODAL NETWORKS

MultiModX

This document is part of a project that has received funding from the SESAR 3 Joint Undertaking under grant agreement No 101114815 under European Union's Horizon Europe research and innovation programme.

Table of contents

List of figures

List of tables

List of acronyms

Acronym	Description
ADMS	Asset Description Metadata Schema
AES	Advanced Encryption Standard
ALTAI	Assessment List for Trustworthy Artificial Intelligence
API	Application Programming Interface
CSV	Comma-Separated Values
DCAT	Data Catalog vocabulary
DCAT-AP	DCAT Application Profile
DMP	Data Management Plan
DPO	Data Protection Officer
EU	European Union
FAIR	Findable, Accessible, Interoperable and Re-usable
GDPR	General Data Protection Regulation
HSM	Hardware Security Module
HTML	Hypertext Markup Language
HTTPS	Hypertext Transfer Protocol Secure

JSON	JavaScript Object Notation
KPI	Key Performance Indicator
OSS	Open-Source software
MND	Mobile Network Data
PDF	Portable Document Format
PKCS	Public Key Cryptography Standards
RAID	Redundant Array of Independent Disks
R&I	Research and Innovation
SFTP	SSH File Transfer Protocol
SHA	Secure Hash Algorithms
SQL	Structured query language
SSH	Secure Shell
SSL	Secure Sockets Layer
TLS	Transport Layer Security
WP	Work Package
W3C	World Wide Web Consortium
XML	Extensible Markup Language

Table 1: list of acronyms

1 Data summary

1.1 Purpose of the data collection/generation

The goal of the MultiModX data collection activities is to provide access to all data sources collected during the project. These data will be used to define current and future scenarios for long-distance passenger multimodal transport as well as to develop a set of innovative multimodal solutions and decision support tools for the coordinated planning and management of multimodal transport networks. The data generated by the project will consist of the data resulting from the analysis, modelling and simulation activities conducted over the data collected.

The generated data together with collected data which need to be shared between two or more Consortium Members will be stored and shared through the MultiModX Data Repository. Public results will be published in the MultiModX project website and open-access repositories such as Zenodo (https://zenodo.org/).

1.2 Relation to the objectives of the project

The relation between the project's objectives and the data used in the project is explained below:

1. **To identify and characterise current and future scenarios for long-distance passenger multimodal transport in Europe.** A scenario is defined by a set of passenger archetypes, a regional context, relevant strategic and/or tactical policies, and the expected demand and supply levels. The combination of different data sources will allow the definition of scenarios for the operation of multimodal mobility solutions.

2. **To develop a multimodal performance framework.** The objective is to develop measurement mechanisms for the proposed KPIs, by selecting the required input data and developing the data transformation and processing pipelines. The proposed KPI measurement mechanisms will take advantage of state-of-the-art techniques for the obtention of high-detail, high-resolution multimodal passenger mobility indicators through the fusion of traditional data (e.g. ticketing data, passenger surveys) and emerging big data sources (e.g. data from personal mobile devices).

3. **To develop a multimodal modelling and evaluation framework.** Each scenario is translated into high-level future traffic flows and supply patterns of both air and rail infrastructure, and thus provide the input for the assessment of changes of multimodal networks.

4. **To develop a Schedule Design Solution for the integrated planning of air and rail networks.** The data generated by the simulation experiments will be the basis for the validation and evaluation steps of the Schedule Design Solution.

5. **To develop a Disruption Management Solution based on coordinated passengers' reallocation, and tactical schedule adjustments and speed/trajectory adjustments for air and rail services.** The data generated by the simulation experiments will be the basis for the validation and evaluation steps of the Disruption Management Solution.

6. **To nurture the conditions for the transfer of the MultiModX Solutions to the subsequent stages of the R&I cycle.** This objective is to consolidate all activities performed throughout the project; hence all the collected and generated data are essential.

1.3 Features of the collected/generated data

Table 3 and Table 4 summarise all datasets identified during the project. The datasets presented here are definitive.

Table 2: DMP Updates Calendar

Deliverable	Date
D2.1 Data Management Plan	30/09/2023
D2.2 Data Management Plan update	31/07/2024
D2.3 Data Management Plan final	02/12/2025

In Appendix A, the Data Factsheets for each of the collected datasets are detailed.

Table 3: Collected MultiModX datasets

Dataset	Description	Category	Location	To be collected/ generated	Provider
Mobile network data	Anonymised mobile phone records which contain an encrypted Id of the user, the sequence of geolocated antennas to which the user is connected along the day, and the timestamp of the register. Age-gender information is sometimes available.	Transport demand	Spain	Collected	Orange
Madrid household mobility survey	Mobility survey of households in Comunidad de Madrid, from 2018. Data includes mobility patterns in cross-sectional format.	Transport demand	Spain	Collected	Madrid region survey
Spanish airport passengers' movements	Datasets that discriminate air traffic (passengers, freight, and mail) by airport and airline (either per year or per month) starting from 2020	Transport demand	Spain	Collected	AENA
UK airport passengers' movements	Data sets that discriminate, but are not limited to, aircraft movements, air transport movements, passenger movements, freight movement starting from 2015	Transport demand	UK	Collected	CAA (Civil Aviation Authority)
Passenger travel to and from the UK	Quarterly dataset (from UK International Passenger Survey) with results from a sample of passengers as they enter or leave the UK by principal air, sea and channel tunnel routes. Adult and child travellers passing through passport control are randomly selected for interview which is carried out on a voluntary and anonymous basis. Approximately 250k interviews are achieved each year, representing about 0.2% of travellers.	Transport demand	UK	Collected	ONS (Office for National Statistics)
Passenger rail usage	Statistics of passenger journeys by operator and region	Transport demand	UK	Collected	Office of Rail and Road
Eurostat database - transport demand	EU statistics on passengers. Data includes, but is not limited to, the number of rail passengers transported by type of transport, region, and country	Transport demand	EU	Collected	Eurostat
Department for Transport (DfT) statistics	UK summary statistics on passenger trips	Transport demand	UK	Collected	Department for Transport

Dataset	Description	Category	Location	To be collected/ generated	Provider
Spanish public transport schedules	Infrastructure data and frequency and schedule of public transport services. Data includes air and rail.	Transport supply	Spain	Collected	MITMA (Ministry of Transport, Mobility and Urban Agenda of Spain), RENFE, OUIGO
French public transport schedules	Infrastructure data and frequency and schedule of public transport services. Data includes rail.	Transport supply	France	Collected	SNCF
German public transport schedules	Infrastructure data and frequency and schedule of public transport services. Data includes rail.	Transport supply	Germany	Collected	Deutsche Bahn
ORR (Office of Rail and Road) data portal - UK rail regulator	UK summary statistics include, but are not limited to, rail performance, station usage, rail fare or rail emissions.	Transport supply	UK	Collected	ORR (Office of Rail and Road) data portal
Spain's national Statistical Office	Data from Spain's national Statistical Office. Datasets contain data from at the NUTS2, NUTS3 level. Datasets include, but are not limited to, • Demographic data • Socio-economic data • Labour-market data	Socioeconomic data	Spain	Collected	Spanish Statistical Office, INE
UK's national Statistical Office	Data from UK's national Statistical Office. Datasets contain data at the NUTS1 level (e.g. for London) and lower (e.g. Boroughs of London). Datasets include, but are not limited to, • Demographic data • Socio-economic data • Labour-market data • Tourism industry data	Socioeconomic data	UK	Collected	Office for national statistics

Dataset	Description	Category	Location	To be collected/ generated	Provider
Eurostat database - socioeconomic data	EU statistics at the NUTS2/NUTS3 region-level. Datasets include, but are not limited to, • Regional socio-economic data • Regional business data • Regional travel and tourism data • Regional digitalization data • Regional demographic data • Regional innovation performance data	Socioeconomic data	EU	Collected	Eurostat
EU Regional Innovation Scoreboard	The European Innovation Scoreboard provides a comparative assessment of the Research and Innovation performance of EU Member States, other European countries, and regional neighbours.	Socioeconomic data	EU	Collected	European Commission
Spanish tourism data	Datasets that include, but are not limited to, internal and external tourism, spending, hotel statistics by region starting from 2015	Socioeconomic data	Spain	Collected	Ministry of Industry, Trade and Tourism of Spain
UK tourism data	Datasets that include, but are not limited to, internal and external overnight tourism starting from 2005	Socioeconomic data	UK	Collected	VisitBritain, National tourism agency
Economic data	Datasets that include, but are not limited to, economic and financial indicators starting from 2000	Socioeconomic data	Spain	Collected	Ministry of Economic Affairs and Digital Transformation of Spain
Demand from Aviation Weekly	Aggregated demand on passengers' itineraries by air within, to and from ECAC region for September 2019.	Transport demand	EU	Collected	Aviation Weekly
SABRE air passenger demand data	Air-travel passenger demand from each NUTS2 region in Germany to each NUTS3 region in Spain, and vice-versa.	Transport demand	Germany, Spain	Collected	SABRE
Air schedules (OAG)	Global airline schedules data.	Transport supply	Worldwide	Collected	OAG

Dataset	Description	Category	Location	To be collected/ generated	Provider
Census Grid	EUROSTAT European census 2021 grid dataset, including population distribution within NUTS-2 and NUTS-3 regions.	Sociodemographic	EU	Collected	EUROSTAT
Minimum connecting times (MCT)	Standard, domestic, and international minimum connecting times at airports. Collected/modelled as in 2016 ComplexityCosts project.	Transport supply	Worldwide	Collected	University of Westminster
TUM Model data	Zonal data, simulated trip data. The TUM model is a calibrated agent-based and trip-based travel demand model developed by Pukhova et al. (2021). The model simulates travel demand within Germany and from/to its neighbouring countries.	Transport demand	Germany	Collected	TUM
Google distance API data	Service provided by Google that allows developers to calculate travel distances and times for a set of origins and destinations.	Transport supply	Germany, Spain	Collected	Google

Table 4 - Generated MultiModX datasets

Dataset	Description	Category	Location	To be collected/ generated	Provider
Passenger demand on path	Volumes of demand on specific path between an origin and a destination according to passengers' sensitivities and based on historical data.	Transport demand	Germany, Spain	Generated	MultiModX SOL1
Coordinated schedules	Coordinated schedules for air and rail, based on passenger demand flows and respecting the operational constraints.	Transport supply	Germany, Spain	Generated	MultiModX SOL2
Schedules' performance indicators	Indicators evaluating the performance of schedules in a multimodal context, in domains such as Operational Efficiency, Flexibility, Predictability, Interoperability, Cost Effectiveness, Capacity, Environment.	Quality of transport supply	Germany, Spain	Generated	MultiModX SOL1
Adjusted multimodal air-rail schedules	Passenger-centric air and rail schedules, i.e. arrival and departure times, with adjusted speeds and trajectories to optimally accommodate the demand under disruptions.	Transport supply	Germany, Spain	Generated	MultiModX SOL3

1.4 Data utility

MultiModX' results contributes to better understand, model and consequently address travellers' decisions to use a particular mode or combination of modes for their journey. In addition, the data generated contributes to the identification and assessment of new multimodal solutions, addressing changing passenger requirements and enhanced coverage of door-to-door solutions. These results enable a comprehensive characterisation of multimodal transport systems and multimodal solutions, which is useful for policy makers, the European Commission and aviation and rail stakeholders in order to make informed decisions regarding future alignment between transport modes and better use of capacity.

The SESAR 3 JU transversal performance project benefits from the metrics, indicators and data generated by the project to develop performance metrics for multimodal transport systems. Furthermore, other research projects benefit from the research and data generated by MultiModX.

2 FAIR data

2.1 Making data findable, including provisions for metadata

All the data has been documented and stored to be searchable efficiently. This philosophy increases the potential reuse of these data, both inside and beyond the project. The effective application of this principle requires the implementation of a consistent and meaningful meta-information for each dataset. The following sections explain the conventions adopted by MultiModX for metadata creation, naming, searching and control version management.

2.1.1 Metadata

In order to analyse and classify all the information provided by the different data sources, a factsheet has been designed. This factsheet includes the following metadata:

- **General information:** identification of the data source and how to access it.
    - Data source name
    - Contact information
    - Last factsheet update
- **Abstract:** brief description of the content and purpose of the data source.
- **Availability:** original source from which the information is obtained and details on the way the information is provided.
    - Access conditions (public, non-disclosure agreements, pricing plan…)
    - Data cost (per requests set, temporal unit, geographical unit…)
    - Access limitations (maximum number of queries, transferred data…)
    - Privacy/Confidentiality issues
    - Security issues
    - State (available / in process / to explore / denied)
    - Link to the data
- **Data characteristics:** temporal, geographical and size characteristics of the information provided
    - Estimated size
    - Temporal scope (From…/To…)
    - Geographical scope (worldwide, EU member states, single country…)
    - Temporal granularity (daily, monthly, yearly…)
    - Delivery frequency (time period between consecutive data updates)
    - Delivery delay (time period between data production and its availability)
    - Data format (JSON, CSV, PDF…)
- **Quality issues:** known characteristics of the data which shall be amended during the analysis process.
- **Comments:** relevant information related to data availability, data reliability and other observations.

The factsheets are based on the template included in Appendix A.

The MultiModX Data Repository shall include both data which are available for all partners and data which shall be shared between only a subset of them. The data which is only for the use of one specific partner is stored and classified in isolated and secured databases.

The public information generated by the project has been made available through the project website (see Section 2.2.2) and open repository (see Section 2.2.4) and has been organised into the following categories:

- Working papers
- Scientific papers
- Conference talks
- Stakeholder workshop material
- Other material (e.g., data visualisations)

2.1.2 Identifiers and naming conventions used

The data stored in the MultiModX Data Repository shall be organised according to a two-level classification: categories and datasets composed of files. Categories gather data related to the same area into the same section. Datasets gather files with the same type of data. Each level follows an identifier and naming convention.

2.1.2.1 Categories

The name of categories in the MultiModX Data Repository shall be as descriptive as possible in no more than a few words.

Various datasets can be gathered under the same category. The criteria to unify datasets are:

- the datasets are identical but have different format;
- the datasets contain the same information for different dates or with different aggregation levels;
- the datasets can be classified under the same label but it is convenient to have them separated;
- the datasets can be classified under the same label but they cannot be integrated in a unique format;
- a dataset is considered necessary for the interpretation of the information in another dataset and is only usable in this context.

2.1.2.2 Datasets

Datasets consist of a related set of files with meta-information that describes them. Different datasets inside a category are presented in a table format. The table has four columns: name, description, source and files. The description shall detail the content of the dataset and its granularity. Each row of the table contains a dataset comprising one or several files. The criteria to include different files in the same dataset are:

- the files contain the same information but have different format;

- the files contain the same information for different dates or with different aggregation levels.

Duplication of information shall be avoided as far as possible. In general, it is preferable to separate a dataset into different complementary datasets or files rather than copying a subset as an independent file.

2.1.2.3 Files

Files stored in the repository follow a common convention for naming. The format is ***Name_Source_Date_Version.extension***, where:

- "Name" is a word or concatenation of words within which the first letter of each word is a capital letter (camel case notation). "Name" should be descriptive of the content in the file and as brief as possible, avoiding abbreviations;
- "Source" indicates the data source from where the file was obtained or the main data source from which the information was transformed, computed or aggregated. The format should be brief (including acronyms), clear and unique for each data source. If the data is generated by the Consortium, the "Source" field can be omitted;
- "Date" stands for the string that explains the temporal scope of the dataset. This temporal scope may extend over different periods of time. Each period extension has its own nomenclature, which is unique. If the data is static or quasi-static (such as airport coordinates), "Date" can be omitted. Some examples are shown below:
    - Day: format YYYYMMDD (e.g. "20230331", 31st March 2023)
    - Month: format YYYYMM (e.g. "202303", March 2023)
    - Year: format YYYY (e.g. "2023")
- "Version" is an identifier composed by the letter "v" and a sequential number starting on 1. When a new version of the file is submitted to the MultiModX Data Repository the sequential number shall be incremented by 1.
- Finally, "extension" stands for the type of the computer file, e.g. "zip", "csv".

In order to avoid duplicity, files containing the same information using different formats shall be avoided.

2.1.3 Approach towards search keywords

The datasets registered in the repository shall be labelled with keywords. Keywords shall be descriptive, brief and unique. Unique means that if a keyword exists already with the same meaning, that keyword shall be used.

The keyword convention to be used is word-word2, where:

- "word" is a descriptive word in lowercase characters and in singular form (except if it does not make sense such as in "data" or "coordinates");
- if the keyword needs another word, the second word ("word2") can be added with a hyphen between them. Keywords shall not have more than two descriptive words.

2.1.4 Change and version control management

In the MultiModX Data Repository, change and version control management shall be carried out using the tools provided by the cloud file storage platform. Such a platform shall create backup copies of files which will be accessible by clicking on the file "Versions" section. In addition, on every new file version/restore, the collaborator shall record a message on the comment section on the file specifying the following data:

- Creation date: the date in which the file version was uploaded.
- Comments: the description of the version. This is a mandatory field for each file version.

2.2 Making data accessible

The data will be made available in different ways according to their confidentiality:

- Data which is only for the use of the Consortium is accessible through the MultiModX Data Repository or the MultiModX Information System for Consortium members, but not for general public.
- Data generated by the project and catalogued as public will be accessible through the MultiModX website and an open access repository (e.g., Zenodo).
- Data which is only for the use of specific Consortium member(s) is not accessible for any other member of the Consortium or the general public.

2.2.1 MultiModX Data Repository

The datasets used for research purposes in the framework of MultiModX shall be shared by means of the Data Repository of the project. The purpose of this repository is to provide a single set of datasets for all project members, in order to ensure that the conclusions achieved by different partners carrying out different data analysis tasks are consistent.

The repository shall consist of a shared distributed file system with one folder per dataset category (see Section 2.1.2.1). In each folder, several files shall be stored using the naming formats specified in Section 2.1.2.3.

The space shall be managed and maintained by the leader of WP2 Data Management, Nommon, which shall be responsible for granting authorisations to other members. Data quality shall be ensured by an iterative process, solving the comments and questions on the data raised by the different partners.

The set of technologies used to deploy the Data Repository for the project are further explained in Appendix B.

2.2.2 MultiModX website

The MultiModX website (accessible at https://multimodx.eu/) includes all relevant public information about the project, including the main research results and the latest news.

Apart from deliverables labelled as public, papers and other written materials, the MultiModX website will be the place where open and re-usable data will be uploaded if they exist.

2.2.3 MultiModX Information System

The personal data collected through workshops and meetings with stakeholders in the frame of MultiModX shall be shared by means of the Information System of the project.

The Project Information System consists of a private wiki managed and maintained by Nommon, the leader of T1.2 Project Information System. Security is maintained through the use of individual usernames and passwords as well as encrypted communications. Further information is also provided in D1.1 Project Management Plan.

2.2.4 Open access repository

The MultiModX beneficiaries will deposit an electronic copy of the scientific publications and public deliverables produced during the scope of the project relating to foreground in an institutional or subject-based repository at the moment of publication, e.g., via the Zenodo portal. In addition, beneficiaries will make their best efforts to ensure that this electronic copy becomes freely and electronically available to anyone through this repository (i.e., that it becomes “open access”) immediately.

The same applies for the open and re-usable data (e.g., csv files with processed data) needed for the validation of the aforementioned deliverables and publications always ensuring the commitment of ethics and GDPR requirements.

2.2.5 Data

Access and sharing rights are managed as established in the MultiModX Consortium Agreement. According to the confidentiality level of the data, access and sharing rights are provided to Consortium members and third parties in a different manner.

In the case of data which is only for the use of the Consortium, the access for all Consortium members is provided through the MultiModX Data Repository. The security is maintained through the use of individual usernames and passwords as well as encrypted communications. Nommon, as administrator of the repository, is responsible for granting authorisations to other members.

In the case of the general public, the data generated by the project and catalogued as public will be accessible through the MultiModX website (see Section 2.2.2) and the specified open access repository. Other requests will be attended on a case-by-case basis preserving the principles for dissemination established in the MultiModX Consortium Agreement.

2.2.6 Metadata

All the data is accompanied with the relevant metadata (data factsheets) as specified in Section 2.1.1. This metadata will contain the links for accessing the data generated and collected by the project and catalogued as public.

The data factsheets will specify the availability period for each dataset. Unless otherwise specified, the data which is only for the use of the Consortium will be available at least during the duration of the project and the data categorised as public will be available permanently through the open access repository.

If specific software is needed to read the data, it shall be specified in the data factsheets. The project will aim at finding open-source software to this task.

2.3 Making data interoperable

The publishable data generated by the MultiModX project is formatted in a way that can be re-used by third-party organisations by means of automated processes. In the case of visualisations and other material created for human understanding, a special effort has been applied to provide the represented data in a way that can be easily ingested and processed by computerised systems. Since these data will foreseeably have a tabular structure, the generated files are encoded using Comma-Separated Values format (CSV), which is a lightweight format to exchange this kind of data. There are many informal documents that describe "CSV" formats, but the project uses the formal description defined in the RFC 4180.

The data files are accompanied by metadata describing each dataset in the catalogue and the columns of data included on it. For the generation of these metadata, concepts already described in standard vocabularies defined by ISO 2382:2015 will be used. For general terms, concepts included in the Data Catalog vocabulary (DCAT) and the Asset Description Metadata Schema (ADMS), both published by the W3C, will be considered. In addition, the more specific DCAT Application Profile (DCAT-AP), which aims to ease the access to public sector datasets in Europe, will be employed to characterise the generated datasets.

While these specifications define XML formats for data exchange which may not be directly applicable to the expected results of MultiModX, their semantics will be employed. In further stages, the project results will be revisited to assess if any of them can be published using these formats.

2.4 Increase data re-use

2.4.1 Conditions of re-usability

2.4.1.1 Data collected

The data collected by the project may be re-used by third parties only if allowed by the data owner. Under request, MultiModX will provide the metadata to ease the identification of the datasets.

2.4.1.2 Data generated

The re-usability of the data generated by the project is subject to the general principles for dissemination and transfer of results set in the MultiModX Consortium Agreement and Grant Agreement. In particular, all the data generated will be shared for re-usability unless:

- the protection of one Consortium Member’s results or background would be adversely affected;

- legitimate academic or commercial interests of one Consortium Member in relation to the results or background would be significantly harmed;
- compliance with ethical aspects, as stated in Section 6, prevents data sharing.

2.4.1.3 Data quality assurance processes

All the data generated by MultiModX (either available for re-use or not) has gone through a quality assurance process to avoid inconsistencies and other anomalies. The process has also performed data cleansing activities to improve data quality. Checking may involve both automated and manual procedures:

- statistical analyses to detect errors and anomalous values,
- checking of out-of-range values,
- checking of data completeness,
- addition of variable and value labels where appropriate.

2.4.2 Licenses

If MultiModX provides data for re-use, open data will be licensed under the Creative Commons license CC BY-NC-SA 4.0, i.e., Attribution-NonCommercial-ShareAlike 4.0 International.

With this license, anyone will be free to:

- Share — copy and redistribute the material in any medium or format,
- Adapt — remix, transform, and build upon the material,

under the following terms:

- Attribution: appropriate credit must be given, a link to the license must be provided, and indications if changes were made must also be provided;
- Non-Commercial: the material may not be used for commercial purposes;
- Share-Alike: any contribution made by remixing, transforming, or building upon the material, must be distributed under the same license as the original.

Under these terms the data will remain usable for undetermined time. The data will be available on the project website until it is closed (see Section 4.3). After that, the project will study the provision of data on a case-by-case basis.

3 Other research outputs

Other research outcomes, such as MultiModX Performance Assessment Solution, Schedule Design Solution or Disruption Management Solution will be managed in line with the FAIR principles.

4 Allocation of resources

4.1 Budget

The tasks needed to make the data FAIR are covered with the budget included in WP2 Data Management.

4.2 Roles and responsibilities

Nommon, as the leader of WP2 Data Management, shall be responsible for WP2 deliverables (D2.1, D2.2 and D2.3).

Each Consortium Member is responsible for the observance of the conventions, standards and rules defined in the DMP to ensure that MultiModX data is FAIR.

4.3 Long term preservation

The MultiModX Data Repository (including appropriate back-up means according to Nommon Quality System) shall be maintained by Nommon during at least 6 months after the completion of the project, so that all the potentially re-usable data are available. For future reference and use the data will be transferred to long-term data storage for 4 years.

Nommon shall be responsible for maintaining the Project Information System in order to ensure that relevant data are available to Consortium Members during the lifetime of the project.

The project website shall be maintained by UIC at least during 5 years after the completion of the project, to promote the dissemination of the project results.

The costs associated to the preservation of MultiModX FAIR data are related to the costs of maintenance of the infrastructure and human resources devoted to quality assurance. The infrastructure consists of: the platform for data storage and transfer used in the MultiModX Data Repository; the Confluence collaboration platform used for the implementation of the MultiModX Information System; and the Content Management System used for developing and hosting the MultiModX Website.

5 Data security

5.1 General security

The data used in the MultiModX project shall be collected, processed and stored in reliable environments. The IT infrastructure of the Consortium partners shall provide at least three basic security elements:

- Lockable computer systems with passwords
- Firewall systems in place
- Virus/malicious software protection.

Additional security measures may be implemented to avoid specific physical or logical threats on the partners' facilities.

5.2 MultiModX Data Repository security

The datasets shared between two or more Consortium Members shall be distributed using the MultiModX Data Repository. The data will be stored on a storage server from the company OVHcloud, which is located in one of its data centres in Gravelines, France (GRA2) and is accessible via Nextcloud. This repository has been implemented using a widely used OSS client-server software for data storage and transfer. The source code of this software is freely available under the GNU AGPLv3 license, enabling fast bug removal practices (ISO/TR 14742:2010).

For the security of the Data Repository, the software implements user authentication and authorisation mechanisms allowing to verify the identity of the users accessing to the platform and manage the access policies to the different resources stored in the Data Repository. This way, it is possible to configure fine grained permissions for each Consortium partner and for each dataset.

The Data Repository executes a transparent encryption mechanism using the AES-256 symmetric key encryption algorithm over any uploaded dataset. Therefore, all the datasets stored in the system are strictly confidential, preventing data from unauthorized access without the private key. In addition, all the incoming/outgoing traffic are encrypted making use of secure application protocols, such as HTTPS and SSH, so that the Consortium partners can authenticate the system and have a guarantee of the authenticity, integrity and privacy of the transferred datasets.

Furthermore, the network in which the Data Repository has been deployed implements a firewall service which allows to discard the traffic coming from any IP address not belonging to the Consortium partners, so only the partners are able to connect to the Data Repository.

Finally, the Data Repository implements a virus scanner which analyses every newly-uploaded dataset using the open-source anti-virus engine. This engine detects all forms of malware including Trojan horses, viruses, and worms. The malware signature database is automatically updated at scheduled intervals.

5.3 MultiModX Information System security

The MultiModX Information System has been implemented using Confluence. Confluence is a web-based application developed by Atlassian that allows content sharing using wiki-formatted pages. Each member of the project team shall be provided with individual credentials (i.e., user and password), so that access to each resource within the platform can be granted individually by Nommon.

The MultiModX Information System shall store contact data, personal views and opinions of the attendees to the events organised within the project (i.e., workshops, Advisory Board meetings, etc.). These data shall be processed and stored according to the procedure described in Section 5.4.

All the actions performed on the resources hosted within the Project Information System shall be tracked in order to keep a complete history of the modifications performed over it. The history shall keep track of any resource modification: original and subsequent versions of the resource, rationale for the modification, and user responsible for it.

At the communication level, Confluence employs HTTPS, which is a version of the HTTP protocol encrypted by a SSL/TLS layer. This protocol provides bidirectional encryption between the user's browser and the system, protecting the communication contents against eavesdropping and tampering. HTTPS also provides server authentication, ensuring that the users are communicating with the actual platform and the exchanged data cannot be read of forged by any external party. The platform employs a X.509 server certificate with high-grade security issued by the Certification Authority DigiCert.

The Confluence platform in which the MultiModX Information System has been implemented explicitly addresses the most common security risks of online tools:

- **Password storage**: the user passwords are stored in the Confluence database using a salted hashing algorithm (PKCS5S2), which makes it nearly impossible to retrieve the original passwords from it. If necessary, an automatic process based on e-mail communications allows the user password to be reset.
- **Buffer overflows**: since the Confluence platform has been implemented in pure Java, the only possible buffer overruns are limited to bugs in the Java Runtime Environment itself.
- **SQL injection**: all the interactions of the Confluence platform with the database are implemented using the parameter replacement technique of the Hibernate Object-Relational mapper, which is much more resistant to SQL injection attacks than common string concatenation.
- **Script injection**: the Confluence platform is a self-contained Java application and does not launch external processes. Therefore, it is highly resistant to script injection attacks.
- **Cross-Site scripting**: as a content-management system that allows user-generated content to be posted on the web, Confluence shall implement some restrictions to avoid this type of attacks:
    - Insertion of raw HTML are disabled by default.
    - The wiki markup language in Confluence does not support potentially dangerous HTML markup.

- HTML uploaded as file attachments are served with a content-type requesting the file to be downloaded, rather than being displayed inline.
- Only system administrators can make HTML-level customisations.

Additionally, the Service Level Agreement (SLA) subscribed with Atlassian establishes that any security issue found in the platform shall be fixed within 4-8 weeks (depending on the severity of the issue) from its reporting date.

5.4 Sensitive data

Appropriate technical and organisational measures will be taken against unauthorised access to sensitive data and against accidental loss or damage. Sensitive data used in this project have been classified into three categories according to their common characteristics:

- datasets which are subject to non-disclosure agreements but do not contain personal data;
- datasets which are subject to non-disclosure agreements and contain personal data;
- contact data, personal views and opinions of any of the attendees to events organised within the project (i.e., workshops, Advisory Board meetings, etc.).

Data in the first category will be accessed only by the Consortium Member having signed the non-disclosure agreement, which will be responsible for ensuring that the required security provisions are in place. As minimum, the security procedures established in section 5.1 shall be satisfied. Additionally, National and EU legislation, as well as any specific security precautions specified by the data provider, will be observed.

An example is Nommon's data warehouse, which fulfils requirements from different data providers. This storage infrastructure is composed of arrays of magnetic disks employing RAID technology to guarantee the redundancy and avoid any information loss in case of individual disk failures. A secondary power supply system allows the disks to cleanly dismount and stop in case of power failure. The raw storage space is divided in virtual volumes which allow access to the dataset to be granted only to the personnel involved in the project using their user/password credentials.

Additionally, data in the second category, i.e., datasets containing personal information, shall be subject to irreversible pseudonymisation procedures by the data providers prior to being delivered to any Consortium Member. Specific processes for data pseudonymisation shall be defined using techniques such as hashing algorithms, key stretching algorithms and randomised keys. The combination of these techniques provides an extremely high level of trust on the unfeasibility of a reconstruction of the original values from the pseudonymised values.

When necessary, Nommon will also provide support on the design of collection processes for anonymised data from the data providers. A typical process will comprise the following steps:

1. **Generation of symmetric key**: the data provider generates a unique random key that contains at least one character of the following groups: uppercase, lowercase, numerical and special characters. For each data transfer, a new unique key is generated, therefore no key re-use is allowed.

2. **Encryption of the dataset**: the data provider uses the new symmetric key to encrypt the dataset to be transferred. A strong symmetric encryption algorithm shall be employed. An additional compression algorithm can be used in this step to reduce the amount of data to be encrypted and transferred.

3. **Encryption of the symmetric key:** once the dataset encryption is complete, the data provider encrypts the symmetric key utilised. In order to ensure that only the intended receiver of the dataset will be able to decrypt it, a strong asymmetric encryption algorithm together with the receiver's public key (usually enclosed into a certificate) shall be used.

4. **Secure transfer of the encrypted files**: the files generated in the previous steps (dataset and symmetric key, both encrypted) are transferred from the data provider to the project partner's infrastructure using a transmission service that guarantees the confidentiality and integrity of the files as well as the identity of both ends of the communication, e.g. SSH File Transfer Protocol (SFTP).

5. **Transfer to a secure storage infrastructure**: the files transferred to the partner infrastructure shall be then kept into a secure storage, usually located on an isolated network which prevents any attack from the Internet.

6. **Decryption of the symmetric key**: the partner obtains the symmetric key that the data provider used for the encryption of the dataset. To decrypt this key, the partner uses its own private key associated to the public key used by the data provider, which was securely stored.

7. **Decryption and storage of the dataset**: using the symmetric key obtained, the project partner will decrypt the dataset that will store on its secured infrastructure until its removal.

The transfer processes will employ protocols and algorithms for authentication and secure communication which are well vetted by the cryptographic community. In particular, the use of strong encryption algorithms such as AES, RSA public key cryptography, and SHA-256 or better for hashing is mandatory. The use of HSM devices which safeguard and manage private keys and support random numbers generation is not compulsory, but strongly recommended.

A scheme of the process and technical infrastructure employed is shown in Figure 1.

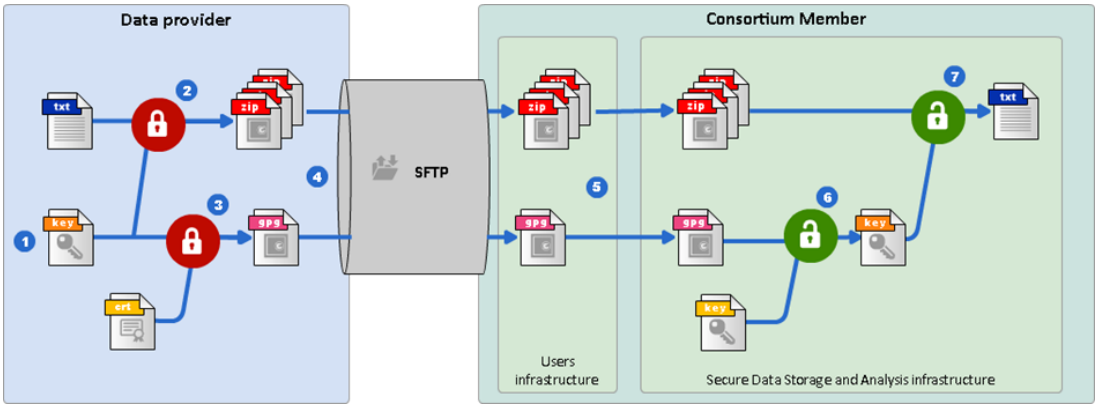


Figure 1: Flow chart of sensitive data transfer process

Finally, personal data generated by specific events will be hosted inside the Confluence platform. Access to these pages will be restricted to the members of the project participating in each particular event.

5.5 Data recovery

The storage layer of the MultiModX Data Repository has been deployed using a disk array which implements RAID 6 redundancy. RAID 6 offers dual parity, which means that the array can tolerate failures on one or two drives simultaneously without any loss of data. This kind of redundancy is commonly used in mission-critical applications, such as in the healthcare, banking and defence sectors. Thanks to the parity system implemented by RAID 6, once the damaged disks are substituted, the original data can be automatically reconstructed using the redundancy blocks stored along the array of disks. The Data Repository software monitors the health status of the disk array and notify the administrators any disk failure in order to replace the unit.

In addition to the replication and the ability to guarantee the integrity of the disk array of the Data Repository, a backup system has been set up to dump all the data contained in it to a different location. The backup system performs periodic copies which are encrypted using AES-256 and stored in a different datacentre inside the Europe region. These backups will be retained for 30 days.

Regarding the MultiModX Information System, the Confluence platform implements mechanisms to recover from unexpected data loss. The platform regularly creates back-ups of the data (including pages and files). On-site backups are performed daily and retained for seven days; tape backups are made weekly, which are then stored off-site and retained for four weeks. All backup data are encrypted.

Additionally, Nommon performs full backups of its warehouse with a periodicity of three months. These backups are stored alternatively in two different locations: on local storage means which are then stored off-site, and on a remote storage service (i.e., cloud service). Both backups are encrypted using AES-256 and are retained for six months.

6 Ethics

The goal of MultiModX is to develop a set of innovative multimodal solutions and decision support tools for the coordinated planning and management of multimodal transport networks. In this context, the objective is to nurture the conditions for the transfer of the MultiModX Solutions to the subsequent stages of the R&I cycle, by engaging relevant stakeholders throughout the full project lifecycle and disseminating the project outcomes to potential adopters; and ensuring that the components developed contain clear definition on integration and datasets requirements. For this purpose, the project draws on various data sources as well as expert involvement, as envisaged by the establishment of an Industry Board. The project has included stakeholder consultations and surveys, and as such have human participants. The respective relevant aspects and requirements relating to ethics and data protection are investigated and documented below.

In line with the ethics self-assessment as well as feedback received during the grant agreement preparation phase, the following aspects have been identified as being relevant for the MultiModX project.

6.1 Humans and personal data

As outlined in Chapter 12 of the Horizon Europe Programme Guide (version 2.0), the EC Ethics and data protection guidelines (Version 05 July 2021) and the General Model Grant Agreement (Article 14, Version 1.1), the approach pursued in this project may involve humans, or protection of personal data. The consortium understands that health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction data can be identified as "sensitive data". The consortium also understands that "Personal data" is defined as any information that could, in any way, lead to the specific identification of one unique person, such as name, surname, family, social security number, date of birth, picture, job, full address, finger print, DNA, mail address, IP or telephone number. All partners in the consortium will comply with relevant EU Directives on data protection (Directives 95/46/EC and 2002/58/EC) and the General Data Protection Regulation (GDPR, Regulation No 2016/679, which applies from 25 May 2018).

The host institution will confirm that it has appointed a Data Protection Officer (DPO) and the contact details of the DPO are made available to all data subjects involved in the research. Our activities in the project will comply with the ethics provisions set out in the above documents, highest ethical standards and applicable international, EU and national law. As such, the participation in consultations / workshops and or similar will be entirely voluntary, based on free and fully informed consent of participants. The activities and methods applied will be chosen in such a way as to maximise benefits and minimise risks / harm to participants, and will not result in discriminatory practices or unfair treatment. For our project, this entails, inter alia:

- to document and identify details on recruitment;
- the data minimisation principle;
- inclusion and exclusion criteria and informed consent procedures;
- details on unexpected / incidental findings policy;
- templates of the informed consent / assent forms and information sheets (in language and terms intelligible to the participants);

- or a description of the technical and organisational measures that will be implemented to safeguard the rights and freedoms of the data subjects/research participants.

Within the MultiModX project, there are different activities involved, which are outlined in the following section, that involve data acquisition. The potentially related ethics issues are described and it is highlighted how these are addressed appropriately.

6.2 Industry Board and expert workshops

As part of the work in MultiModX, an Industry Board is involved in discussing and evaluating results. The Industry Board members will be involved, upon consent, in different workshops that are conducted throughout the project:

1. **Workshop 1 (M6)**: Gain feedback on MultiModX' approach and Solutions - presentation of the project scope and approach, as well as a discussion on the scenarios applicable in the project (e.g. regional archetypes, passenger archetypes, policies).

2. **Workshop 2 and Industry Board round-table (M16)**: Gain feedback on the first results to help us identify a direction towards the final Solutions - presentation of initial project's results and findings, including multimodal solutions, tools, assessment, implementation.

Participants are asked in advance their permission to be recorded or their discussions documented during the workshop. According informed consent templates are provided to the participants beforehand.

6.3 Data minimisation principle

In line with the EU data minimisation principle that data processing must be lawful, fair and transparent, the MultiModX project will adhere to respective standards and regulations in its data processing activities (see EU guidelines reference). The MultiModX consortium members will ensure that the data processing procedures will only involve data that are necessary and proportionate to achieve the specific task or purpose for which they were collected (see Article 5(1) GDPR), i.e. only data will be collected that is essential to meet MultiModX research objectives. For each data acquisition activity, the consortium members will conduct a data minimisation review, i.e. reviewing whether particular data are required for a specific purpose that is relevant and limited to the project's objectives and methodology.

In case some of the data needs to be kept beyond the duration of your project (e.g. in order to finalize dissemination activities such as journal publications), the consortium will explain and justify the relevant data collection and retention arrangements respectively.

In line with the data minimisation and data protection principles, it is ensured that data is pseudonymised or anonymised wherever possible; the data are securely stored; and where appropriate, policies and procedures are established.

6.4 Measures for personal data protection

6.4.1 Storage and protection

All personal data collected and/or processed will be securely stored in the collaboration space of MultiModX, which is hosted on the Confluence platform. Under this environment, the personal data related with an event will be gathered together in a common set of pages for each particular event (e.g., a workshop). Access to these pages will be allowed only to the members of the project team involved in the organisation of that event. This information shall be exclusively used for the purpose of the MultiModX project and shall not be shared with any other third party.

Encrypted copies of the data will be generated during the backup process, as for any other data stored in the platform. All unencrypted local copies of the pages or files (if any) will be deleted after an encrypted version has been generated.

6.4.2 Retention and destruction

Personal data (i.e., contact information and opinions of the workshops' participants and the Advisory Board) shall be maintained and preserved for a period of five years after the end of the project for audit purposes. These data shall be stored following the same precautions used during the project lifetime.

Destruction of personal records shall be performed in a fashion that protects the confidentiality of the participant stakeholders. The creation of a specific record that states what data were destroyed and when is strongly recommended.

6.5 Informed consent procedures

In line with the C Ethics and data protection guidelines (Version 05 July 2021), informed consent is the cornerstone of research ethics. The MultiModX project ensures that within each activity that relates to data processing it is explained to research participants what MultiModX research is about, what their participation in the project will entail and any risks that may be involved. Only after this information has been conveyed to the participants – and they have fully understood it – the MultiModX project will seek and obtain their express permission to include them in the project (Articles 4(11) and 7 GDPR).

Whenever personal data directly from research participants is collected within the scope of the MultiModX project, the responsible consortium partner(s) will seek the participants' informed consent by means of a procedure that meets the minimum standards of the GDPR. This requires consent to be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the subject's agreement to the processing of their personal data. This may take the form of a written statement, which may be collected by electronic means, or an oral statement.

Furthermore, the MultiModX consortium will keep records documenting the informed consent procedure, including the information sheets and consent forms provided to research participants, and the acquisition of their consent to data processing. These may be requested by data subjects, funding agencies or data protection supervisory authorities.

For consent to data processing to be 'informed', the data subject must be provided with detailed information about the envisaged data processing in an intelligible and easily accessible form, using clear and plain language. As a minimum, this should include:

- the identity of the data controller and, where applicable, the contact details of the DPO;
- the specific purpose(s) of the processing for which the personal data will be used;
- the subject's rights as guaranteed by the GDPR and the EU Charter of Fundamental Rights, in particular the right to withdraw consent or access their data, the procedures to follow should they wish to do so, and the right to lodge a complaint with a supervisory authority;
- information as to whether data will be shared with or transferred to third parties and for what purposes; and
- how long the data will be retained before they are destroyed.

The data subjects must also be made aware if data are to be used for any other purposes, shared with research partners or transferred to organisations outside the EU (see article 13 GDPR).

As with any research project involving human subjects, if the data processing entails potential risks to the data subjects' rights and freedoms, they will be made aware of these risks during the informed consent procedure.

The consent process(es) and the information the MultiModX consortium partners give to the data subjects will cover all the data-processing activities related to their participation in the project's research. From a research ethics perspective, and in accordance with the principles of fair and transparent data processing, if the MultiModX consortium partners intend to use or make the data available for future research projects, the consortium will ask for the participants' additional, explicit consent to the secondary use of the data. If it is planned to use the data in multiple projects or for purposes other than MultiModX research, the consortium will give the data subjects the opportunity to opt out of the further processing operation(s).

In line with the planned activities involving data acquisition within MultiModX, informed consent procedures are implemented according to the type of data acquisition, and including information on:

- an explanation on the purpose of the data acquisition activity, and how it fits into the overall MultiModX project;
- along the data minimisation principle, it is outlined to the participants how this data contributes to meeting the objectives of the MultiModX project;
- a statement committing to the pseudonymisation of potentially personal data, and how the data will be stored and accessed.

Furthermore, the informed consent processes also include a part which participants have to agree with in case there is personal data collected, as elaborated above, including the use of all images and videos in which the participants appear requiring the approval of those participants beforehand. This approval is part of the informed consent signature process before any photo or video recording session.

6.6 Incidental findings policy

Social science and humanities research rely on methods that may unintentionally produce findings outside the scope of the original research questions. Fieldwork, observations and interviews can yield information that goes beyond the scope of the research design, thus presenting the researcher with a dilemma: whether to preserve confidentiality or to disclose the information to relevant authorities or services.

Unintended/unexpected/incidental findings may include indications of criminal activity, human trafficking, abuse, domestic violence or bullying. Researchers must inform the participants, or their guardians or other responsible people, of their intentions and reasons for disclosure, provided that doing so does not undermine the act of disclosure. A characteristic of incidental/unexpected findings is that they require the researcher to take some form of action. As a rule, followed by the MultiModX consortium, criminal activity witnessed or uncovered in the course of research will be reported to the responsible and appropriate authorities, even if this means overriding commitments to participants to maintain confidentiality and anonymity. There may be a legal obligation to report criminal activity.

Within the scope of MultiModX, the following procedures are put in place to detect, manage and communicate incidental findings:

1. Participants are informed about the possibility that incidental findings can be found;
2. Data collection and analysis;
3. Assessment by involved MultiModX consortium members whether any of the collected data relates to findings which are outside of the scope of the original research questions; if this is not the case, the incidental findings policy does not apply;
4. In case of incidental findings, these will be discussed within the consortium as well as the MultiModX Data Protection Officer(s) in order to identify the according management and communication of these;
5. Since the participants of the MultiModX data collection and processing activities are de-identified and anonymous, individual participants cannot be identified and incidental findings relating to a particular response not traced back;
6. Since incidental findings are not relevant to the MultiModX research objectives are hence not included in the results;
7. In case of criminal activity, this will be reported to appropriate authorities.

6.7 Non-EU countries

UoW, as a UK-based organisation, will perform research activities outside of the EU, and these are allowed in EU Member States (see DoA Part B for details on assigned tasks). Research activities at UoW are carried out to the highest ethical standards in line with principles of research's good practice. The activities are managed via a research governance framework of policies, processes and systems for enabling quality research and Knowledge Exchange. The policies are in line with the European Code of Conduct for Research Integrity.

6.8 Development, deployment and/or use of Artificial Intelligence systems

MultiModX will leverage the potential of geolocation data for the detailed reconstruction of the door-to-door passenger journey. Mobile Network Data (MND), which consist of the records collected by mobile network operators from the interaction between mobile devices and network antennas, provides large and well-distributed samples of passengers, and can thus be used to identify the location and duration of the activities between door-to-door trips and the stays between trip legs performed by the users included in the sample. However, contrary to passenger surveys, MND is a passively collected data source which lacks certain attributes, like key sociodemographic information about passengers (e.g., age, gender, income), relevant journey attributes (e.g., private vehicle availability, group travelling, luggage carried). Within MultiModX, **machine learning methods** will be extended with further passenger and journey features that can be gained from MND samples. The increasing computational capabilities will be leveraged to perform longitudinal analyses over long periods of time, to characterise multimodal patterns from a temporal perspective, enabling identification of alternative use of different modes by the same user for performing the same trip at different times.

The consortium understands the principles of trustworthy AI in this regard and will follow the "guidance for adopting an ethically-focused approach while designing, developing, and deploying and/or using AI based solutions" as outlined in the "Ethics by Design and Ethics of Use Approaches for Artificial Intelligence" guidelines (V1.0 of 25 November 2021), and, if applicable, the data protection guidelines previously outlined. Within the scope of this Data Management Plan (D2.3) and the related work packages, the requirements towards the applied machine learning techniques are consulted, as outlined by the "Assessment List for Trustworthy Artificial Intelligence (ALTAI)", and project specific approaches detailed. This will include explanations regarding the abilities, limitations, risks and benefits of the proposed machine learning approach, the manner in which decisions are taken and the logic behind them, as well as an outline of potential risks involved.

Appendix A Data factsheets

Only the data collected and generated during the project is included.

An example of the data factsheet to be used is shown below.

MultiModX - Data Management

1. General Information *- Identification of the data source*

Data source name:	EUROSTAT / IATA / …
Last update:	06/10/2022
Contact information:	*Jerónimo Bueno*

2. Abstract *- Brief description of the data source and its potential usefulness for the project*

EUROSTAT is the statistical office of the European Union. Amongst others, it provides air transport specific information at EU-level, which will be useful to validate the results of the project. The following information has been identified as potentially useful for the project:
- Passenger Information: number of monthly passengers per Country/Airport, disaggregated by National / IntraEU / ExtraEU passengers. Flows of passengers between the main airports.
- Flights Information: number of annual flights per Country/Airport, aircraft type, etc.

3. Availability *- Relevant information about owner and readiness to use*

Access conditions:	*Open Source, Through private agreement, Free/For sale, IPR issues…*
Data cost:	*In € per temporal / geographical unit, as applicable*
Access limitations:	*If there are any limitations about the access to the data (maximum number of queries per day, etc)*
Privacy / Confidentiality issues:	*No confidentiallity issues identified / We are not allowed to share the information with third parties / …*
Security issues:	*No security issues identified / Data is considered as confidential information. Security requirements are specified in the data access agreement / …*
State:	*Already available / Permission needs to be requested / …*
Link to the data:	

4. Data Characteristics *- Temporal, geographical and size characteristics of the information provided*

File size:	*10 GB / 20Mb / 1 Tb / …*
Temporal scope:	*February 2014 / 2016 / …*
Geographical scope:	*Geographical extent of the information provided (Worldwide, EU, Country, City…)*
Temporal granularity:	*Temporal resolution of the information provided (Daily, monthly, yearly…)*
Delivery frequency:	*How often the data are updated (it does not need to coincide with temporal granularity)*
Delivery delay:	*Gap between the date to which data corresponds and data publication*
Data format:	*csv / json / excel / …*

5. Quality Issues *- Comments about known flaws of the data source*

6. Comments *- Other relevant information*

The information comes mainly from the travel agencies, so tickets bought directly to the airlines are estimated and the estimation is not always perfect. It do not cover charter or private flights.

A.1.1 Mobile Network Data (MND)

MultiModX - Data Management

1. General Information - *Identification of the data source*

Data source name:	Mobile Network Data (MND)
Last update:	19/07/2024
Contact information:	*Jerónimo Bueno*

2. Abstract - *Brief description of the data source and its potential usefulness for the project*

The MND includes three types of data:

*- **Network events**, including: (i) Call Detail Records (CDRs), generated for billing purposes, which register the location of the user when a mobile phone connected to the network makes or receives a phone call or uses a service; (ii) network probe data, recorded for network management purposes, which register the location of the user on a periodic basis, thus increasing the temporal resolution of the CDRs. For each registered event, the data indicates the network cell where the user is located at the moment the event is registered. This means that the data does not allow us to know the exact location of the user, but it provides the antenna to which the user is connected, so we know that the user must be located within the coverage area of such antenna.*

*- **Network topology**: these data provide information on the topology of the network (location of the towers, orientation of the antennas, etc.), which allows us to estimate the coverage area corresponding to each cell. The spatial granularity of the data is therefore determined by the density of antennas deployed in each area, leading to a location accuracy of a few dozens or hundreds of metres in urban areas and several kilometres in rural areas.*

*- **Sociodemographic data**: age, gender and nationality of Orange clients, and nationality of their mobile network operators for roamers.*

This data source will be used to:
1) Identify and characterise passenger archetypes based on annual long-distance travel patterns

3. Availability - *Relevant information about owner and readiness to use*

Access conditions:	*Through private agreement*
Data cost:	*Price specified in the data access agreement on a project basis*
Access limitations:	*No limitations*
Privacy / Confidentiality issues:	*No confidentiality issues as data has been previously anonymised by Orange*
Security issues:	*Data is considered as confidential information. Security requirements are specified in the data access agreement*
State:	*Already available*
Link to the data:	*N/A*

4. Data Characteristics - *Temporal, geographical and size characteristics of the information provided*

File size:	*Several Tb, depending on the dates of study*
Temporal scope:	*From February 2016 onwards*
Geographical scope:	*Spain, potentially available for Belgium*
Temporal granularity:	*Depends on the user usage of the mobile phone. Typically one register every half an hour.*
Delivery frequency:	*Daily*
Delivery delay:	*2-3 days*
Data format:	*Csv*

5. Quality Issues - *Comments about known flaws of the data source*

Geographical granularity of mobile phone data depends on the mobile network antenna density, giving spatial uncertainty of around 200m in big cities, but up to 2km in rural areas.
User profile information may not always be accurate, as the client (who is the person that appears in the Mobile Network Operator files) may not be the user of the mobile phone (teenagers who have their phone paid by their parents, for example).
When no network coverage maps are available, Voronoi areas are used as a proxy of antenna coverage, which is not the optimal approach (they do not take into account obstacles, or antenna technology)

6. Comments - *Other relevant information*

This data source also captures the mobility of the roamers that connect to the Orange Network in the country of study and mobility of Spanish Orange clients abroad

A.1.2 Demand from Aviation Weekly

MultiModX - Data Management

1. General Information *- Identification of the data source*

Data source name:	Aviation Weekly Passengers Demand
Last update:	06/10/2022
Contact information:	*Luis Delgado*

2. Abstract *- Brief description of the data source and its potential usefulness for the project*

Aggregated demand on passengers itineraries by air within, to and from ECAC region for September 2019. Data originated by GDS and prepared and processed by Aviation Weekly. In particular:

- Per leg (o-d): total number of pax, average load factor, base fare, etc.
- Per itinerary: origin, destination and intermediate airports, total number of pax, average fare
- Per origin-destination: total number of flights, total number seats, average number seats

3. Availability *- Relevant information about owner and readiness to use*

Access conditions:	*Through private agreeement*
Data cost:	*Price specified in the data access agreement*
Access limitations:	*No limitations*
Privacy / Confidentiality issues:	*Data shall be used as part of the MultiModX project. Raw data available to UoW. Processed data, as required shared in common repository.*
Security issues:	*No security issues identified*
State:	*Already available - Processed data in shared repository*
Link to the data:	https://rd-data.nommon.es/index.php/apps/files/?dir=/MULTIMODX_CONSOTIUM_DATA/WP2%20Data%20management/Data/demand_from_aviation_weekly_sep2019

4. Data Characteristics *- Temporal, geographical and size characteristics of the information provided*

File size:	*140 Mb*
Temporal scope:	*sep-19*
Geographical scope:	*to-from-within ECAC flows*
Temporal granularity:	*Aggregated monthly passengers itineraries and demand*
Delivery frequency:	*One-Off delivery*
Delivery delay:	
Data format:	*xlsx and csv*

5. Quality Issues *- Comments about known flaws of the data source*

Data aggregated mothnly.
Some passengers itineraries from models, therefore actual number of passengers might be different.
Flows generated from tickets sales, some multicities tickets might appear as itinerary when in reality could expand on several days.
Some segments might not be air (e.g. some ferries found).

6. Comments *- Other relevant information*

A.1.3 SABRE passenger demand data

MultiModX - Data Management

***1. General Information** - Identification of the data source*

Data source name:	Sabre passenger demand data
Last update:	29/12/2023
Contact information:	*Kay Plötner*

***2. Abstract** - Brief description of the data source and its potential usefulness for the project*

Sabre data for the month of September 2019 was used to identify air-travel passenger demand from each NUTS2 region in Germany to each NUTS3 region in Spain, and vice-versa.

This data was fed as input into the strategic multimodality evaluator in work package .

***3. Availability** - Relevant information about owner and readiness to use*

Access conditions:	*Through private agreement*
Data cost:	*Price specified in the data access agreement*
Access limitations:	*No limitations*
Privacy / Confidentiality issues:	*Data shall be used for the MultiModX project*
Security issues:	*No security issues identified*
State:	*Already available*
Link to the data:	https://rd-data.nommon.es/index.php/apps/files/?dir=/MULTIMODX_CONSOTIUM_DATA/WP3%20Scenario%20definition/Strategic%20Evaluator%20Inputs/I15.%20Passenger%20demand&fileid=20638

***4. Data Characteristics** - Temporal, geographical and size characteristics of the information provided*

File size:	*292 Mb*
Temporal scope:	*September 2019*
Geographical scope:	*Germany, Spain*
Temporal granularity:	*Monthly*
Delivery frequency:	*One-Off delivery*
Delivery delay:	
Data format:	*xlsx*

***5. Quality Issues** - Comments about known flaws of the data source*

***6. Comments** - Other relevant information*

A.1.4 Spanish Rail GTFS (Renfe)

MultiModX - Data Management

1. General Information *- Identification of the data source*

Data source name:	RailGtfsRenfe_Nap_20240410
Last update:	10/04/2024
Contact information:	*Jerónimo Bueno*

2. Abstract *- Brief description of the data source and its potential usefulness for the project*

The National Access Point (NAP) is the space where, with the full backing and official character of the Spanish State, the most complete information possible on the passenger transport offer available in the national territory is concentrated. The file includes the following:
- RENFE regular medium distance, long distance and high speed train services in GTFS format.

3. Availability *- Relevant information about owner and readiness to use*

Access conditions:	*Open source*
Data cost:	*NA*
Access limitations:	*NA*
Privacy / Confidentiality issues:	*No confidentiallity issues identified*
Security issues:	*No security issues identified*
State:	*Already available*
Link to the data:	https://rd-data.nommon.es/index.php/f/20450

4. Data Characteristics *- Temporal, geographical and size characteristics of the information provided*

File size:	*~ 1Mb*
Temporal scope:	*April 2024*
Geographical scope:	*Spain*
Temporal granularity:	*Daily*
Delivery frequency:	*On demand*
Delivery delay:	*One day*
Data format:	*txt*

5. Quality Issues *- Comments about known flaws of the data source*

6. Comments *- Other relevant information*

Source: NAP - Ministry of Transport and Sustainable Mobility

A.1.5 Spanish Rail GTFS (Ouigo)

MultiModX - Data Management

***1. General Information** - Identification of the data source*

Data source name:	RailGtfsOuigo_Nap_20240410
Last update:	10/04/2024
Contact information:	*Jerónimo Bueno*

***2. Abstract** - Brief description of the data source and its potential usefulness for the project*

The National Access Point (NAP) is the space where, with the full backing and official character of the Spanish State, the most complete information possible on the passenger transport offer available in the national territory is concentrated. The file includes the following:
- Ouigo high speed train services in GTFS format.

***3. Availability** - Relevant information about owner and readiness to use*

Access conditions:	*Open source*
Data cost:	*NA*
Access limitations:	*NA*
Privacy / Confidentiality issues:	*No confidentiallity issues identified*
Security issues:	*No security issues identified*
State:	*Already available*
Link to the data:	https://rd-data.nommon.es/index.php/f/20449

***4. Data Characteristics** - Temporal, geographical and size characteristics of the information provided*

File size:	*~ 1Mb*
Temporal scope:	*April 2024*
Geographical scope:	*Spain*
Temporal granularity:	*Daily*
Delivery frequency:	*On demand*
Delivery delay:	*Weekly*
Data format:	*txt*

***5. Quality Issues** - Comments about known flaws of the data source*

***6. Comments** - Other relevant information*

Source: NAP - Ministry of Transport and Sustainable Mobility

A.1.6 Spanish Air GTFS

MultiModX - Data Management

1. General Information *- Identification of the data source*

Data source name:	AirGtfs_Nap_20240410
Last update:	10/04/2024
Contact information:	*Jerónimo Bueno*

2. Abstract *- Brief description of the data source and its potential usefulness for the project*

The National Access Point (NAP) is the space where, with the full backing and official character of the Spanish State, the most complete information possible on the passenger transport offer available in the national territory is concentrated. The file includes the following:
- Nation flight services in GTFS format.

3. Availability *- Relevant information about owner and readiness to use*

Access conditions:	*Open source*
Data cost:	*NA*
Access limitations:	*NA*
Privacy / Confidentiality issues:	*No confidentiallity issues identified*
Security issues:	*No security issues identified*
State:	*Already available*
Link to the data:	https://rd-data.nommon.es/index.php/f/20451

4. Data Characteristics *- Temporal, geographical and size characteristics of the information provided*

File size:	*~ 1Mb*
Temporal scope:	*April 2024*
Geographical scope:	*Spain*
Temporal granularity:	*Daily*
Delivery frequency:	*On demand*
Delivery delay:	*One day*
Data format:	*txt*

5. Quality Issues *- Comments about known flaws of the data source*

6. Comments *- Other relevant information*

Source: NAP - Ministry of Transport and Sustainable Mobility

A.1.7 Air schedules (OAG)

MultiModX - Data Management

1. General Information *- Identification of the data source*

Data source name:	OAG schedules
Last update:	19/07/2024
Contact information:	*Kay Plötner*

2. Abstract *- Brief description of the data source and its potential usefulness for the project*

Global airline schedules data. These schedules will be used as baseline, which will be optimise with the MultiModX SOL-2.

3. Availability *- Relevant information about owner and readiness to use*

Access conditions:	*Through private agreement*
Data cost:	*Price specified in the data access agreement*
Access limitations:	*No limitations*
Privacy / Confidentiality issues:	*Data shall be used as part of the MultiModX project*
Security issues:	*No security issues identified*
State:	*Already available*
Link to the data:	https://rd-data.nommon.es/index.php/f/20772

4. Data Characteristics *- Temporal, geographical and size characteristics of the information provided*

File size:	*200Mb*
Temporal scope:	*6 Septmeber 2019*
Geographical scope:	*Worldwide*
Temporal granularity:	*Daily*
Delivery frequency:	*One-Off delivery*
Delivery delay:	
Data format:	*csv*

5. Quality Issues *- Comments about known flaws of the data source*

6. Comments *- Other relevant information*

A.1.8 NUTS regions

MultiModX - Data Management

1. General Information *- Identification of the data source*

Data source name:	EUROSTAT NUTS regions boundaries 2021
Last update:	06/10/2022
Contact information:	*Luis Delgado*

2. Abstract *- Brief description of the data source and its potential usefulness for the project*

EUROSTAT NUTS regions boundaries 2021.

Used to determine infratructure (rail stations and airports) inside NUTS.

3. Availability *- Relevant information about owner and readiness to use*

Access conditions:	*Open Source*
Data cost:	*N/A*
Access limitations:	*No limitations*
Privacy / Confidentiality issues:	*N/A*
Security issues:	*No security issues identified*
State:	*EU countries*
Link to the data:	https://ec.europa.eu/eurostat/web/gisco/geodata/statistical-units/territorial-units-statistics

4. Data Characteristics *- Temporal, geographical and size characteristics of the information provided*

File size:	*30 Mb*
Temporal scope:	*2021 NUTS boundaries*
Geographical scope:	*EU*
Temporal granularity:	*N/A*
Delivery frequency:	*One-Off delivery*
Delivery delay:	
Data format:	*cpg, dbf, prj, shp and shx*

5. Quality Issues *- Comments about known flaws of the data source*

6. Comments *- Other relevant information*

A.1.9 Census Grid

MultiModX - Data Management

1. General Information *- Identification of the data source*

Data source name:	EUROSTAT Census GRID 2021
Last update:	06/10/2022
Contact information:	*Luis Delgado*

2. Abstract *- Brief description of the data source and its potential usefulness for the project*

EURSTAT European census 2021 grid dataset.

Used to determine population distribution within NUTS-2 and NUTS-3 regions to identify rail stations closest to the population desnity in the region.

3. Availability *- Relevant information about owner and readiness to use*

Access conditions:	*Open Source*
Data cost:	*N/A*
Access limitations:	*No limitations*
Privacy / Confidentiality issues:	*N/A*
Security issues:	*No security issues identified*
State:	*EU countries*
Link to the data:	https://ec.europa.eu/eurostat/web/gisco/geodata/grids

4. Data Characteristics *- Temporal, geographical and size characteristics of the information provided*

File size:	*1.3 Gb*
Temporal scope:	*2021 census data*
Geographical scope:	*EU*
Temporal granularity:	*N/A*
Delivery frequency:	*One-Off delivery*
Delivery delay:	
Data format:	*gpkg*

5. Quality Issues *- Comments about known flaws of the data source*

6. Comments *- Other relevant information*

Statistics on population from 2021 population and housing census broken down into a 1 km2 grid. The projection system of the 1 km2 grid is ETRS89-LAEA.

A.1.10Minimum connecting times (MCT)

MultiModX - Data Management

1. General Information *- Identification of the data source*

Data source name:	Minimum connecting times (airports)
Last update:	06/10/2022
Contact information:	*Luis Delgado*

2. Abstract *- Brief description of the data source and its potential usefulness for the project*

Standard, domestic and international minimum connecting times at airports. Collected/modelled as in 2016 ComplexityCosts project.

Used to estimate if a given connection between two flights is feasible by passengers (i.e., if tickets that require that connection can be sold)

3. Availability *- Relevant information about owner and readiness to use*

Access conditions:	*Collected by UoW*
Data cost:	*N/A*
Access limitations:	*No limitations*
Privacy / Confidentiality issues:	*Data shall be used as part of the MultiModX project.*
Security issues:	*No security issues identified*
State:	*Already available - Processed data in shared repository*
Link to the data:	https://rd-data.nommon.es/remote.php/webdav/MULTIMODX_CONSOTIUM_DATA/WP2%20Data%20management/Data/example_strategic/es/mct_air.csv

4. Data Characteristics *- Temporal, geographical and size characteristics of the information provided*

File size:	*4.1 Kb*
Temporal scope:	*N/A*
Geographical scope:	*worldwide (288 airports)*
Temporal granularity:	*N/A*
Delivery frequency:	*One-Off delivery*
Delivery delay:	
Data format:	*csv*

5. Quality Issues *- Comments about known flaws of the data source*

6. Comments *- Other relevant information*

Generic values for some airports. Only differentiation on connecting times as a function of connection type (standard, international, domestic) no consideration of airline/terminal/gates.

Values from 2016 but no significant changes expected.

A.1.11 NUTS2 Region data

MultiModX - Data Management

***1. General Information** - Identification of the data source*

Data source name:	NUTS@ Regions Data from
Last update:	-
Contact information:	*Ram Kamath*

***2. Abstract** - Brief description of the data source and its potential usefulness for the project*

Data on various NUTS2 region characteristics, such as population, population density, disposable household income, annual air traffic passengers, railway passengers originating from NUTS2 region, railway line density, share of individuals who ordered goods or services over the internet, have been obtainedfrom teh Eurostat database.

This data is used to create clusters of types of NUTS2 regions in Europ, as part of work-package 3

***3. Availability** - Relevant information about owner and readiness to use*

Access conditions:	*Open Source*
Data cost:	*N/A*
Access limitations:	*No limitations*
Privacy / Confidentiality issues:	*N/A*
Security issues:	*No security issues identified*
State:	*EU countries*
Link to the data:	https://ec.europa.eu/eurostat/web/regions/database

***4. Data Characteristics** - Temporal, geographical and size characteristics of the information provided*

File size:	*11 MB*
Temporal scope:	*2019*
Geographical scope:	*EU*
Temporal granularity:	*Annual*
Delivery frequency:	*One-Off delivery*
Delivery delay:	
Data format:	*xlsx*

***5. Quality Issues** - Comments about known flaws of the data source*

Lots of missing data

***6. Comments** - Other relevant information*

Appendix B Data repository

B.1 Requirements

In order to choose the technology that best suits the scope and necessities of the MultiModX project, a requirements collection work was carried out. A summary of the main requirements is listed below:

- Web Graphic User Interface (GUI).
- Metadata support.
- API for system integration.
- Apps for functionality extension.
- User Management: authentication via password, authorisation via role and management (add/edit/remove).
- Database and file backup.

B.2 Description

The chosen storage software technology that meets our needs is Nextcloud. This widely used platform for data storage and transfer is open source and freely available under the GNU AGPLv3 license, enabling community add-ons and support.

For the security of the Data Repository, Nextcloud implements user authentication and authorisation mechanisms that enforces the verification of the identity of the users accessing the platform. In addition, registered user administrators can control permissions to directories and files stored in the Data Repository. This way, it is possible to configure fine grained permissions for each Consortium partner and for each dataset.

The Data Repository will offer different layers of encryption for the stored data. For example, every time a file is being transfer between clients and servers. Also, all the stored data is encrypted; therefore, all the datasets stored in the system will be strictly confidential, preventing data from unauthorized access without the private key. In addition, all the incoming/outgoing traffic will be encrypted making use of secure application protocols, such as HTTPS. This way, the Consortium partners have a guarantee of the authenticity, integrity and privacy of the transferred datasets.

Furthermore, the network in which the Data Repository has been deployed implements a firewall service which allows to discard the traffic coming from any IP address not belonging to the Consortium partners, so only the partners will be able to connect to the Data Repository. Finally, Nextcloud allows administrators to create groups to assign read and write permissions to the file system.

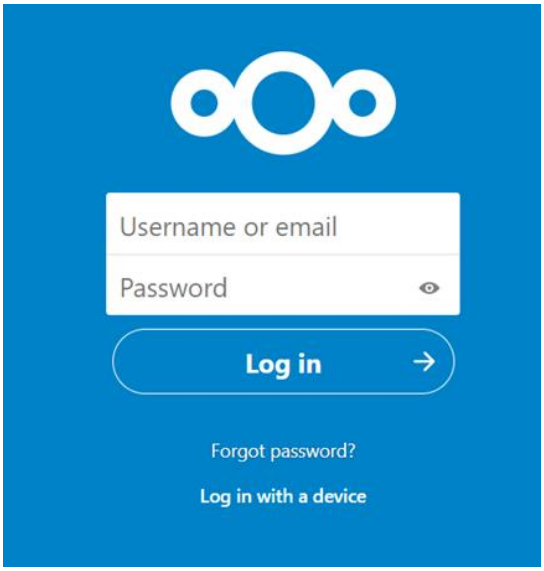
The MultiModX Data Repository (including appropriate back-up means according to Nommon Quality System) shall be maintained by Nommon during at least 6 months after the completion of the project, so that all the potentially re-usable data are available. For future reference and use the data will be transferred to long-term data storage for 4 years. **At the end of this 4-year period, all MultiModX data will be definitely deleted from the repository.**"

B.3 Main use cases

The following illustrations are the main use cases, some of which were extracted from the Nextcloud user manual: https://docs.nextcloud.com/server/latest/user_manual/en/

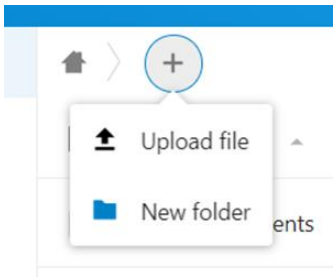
B.3.1 Logging in

The data repository admins will collect the emails, names and last names of all of the team members to register them in the data repository database. Once the new users are active, they can log in to the data repository by entering their credentials.



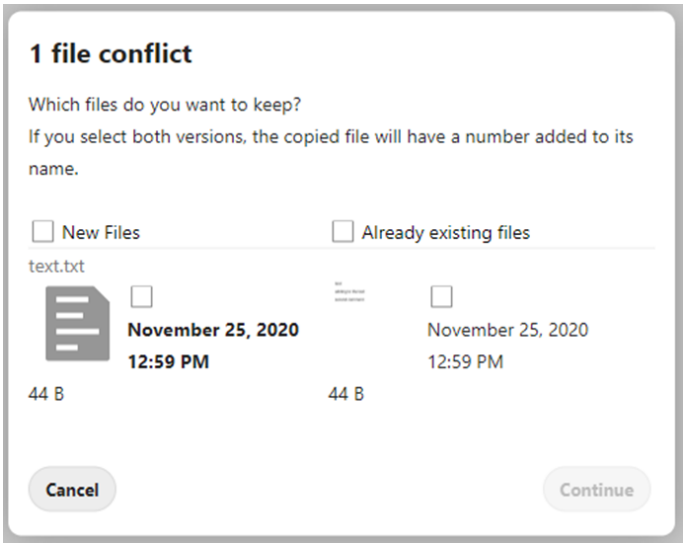
B.3.2 Uploading a file or creating a folder

Upload or create new files or folders directly in a Nextcloud folder by clicking on the “New” button in the Files app. The “New” button provides the following options:

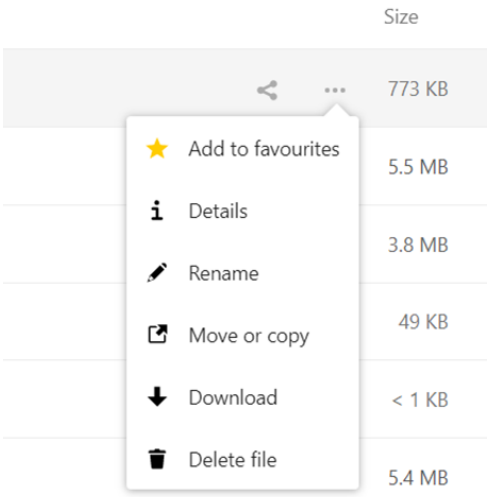


B.3.3 Updating a file

When uploading an existing file, the user needs to select the New Files checkbox in the conflict popup to update the existing file.



In addition, the user shall provide a comment on every document update with the description and date of the update. To leave a comment on the document, click on the 3 dots of the file and the Details menu, followed by the comments section.

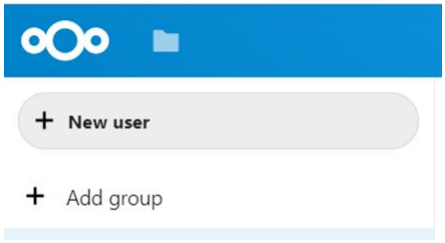


B.4 Group Administrator

Administrators can assign some accounts to group administrator on specific groups. Users with this group admin role can create and remove users. The following sections describe the main Group Administrator use cases. For more information visit the admin manual at https://docs.nextcloud.com/server/20/admin_manual/configuration_user/.

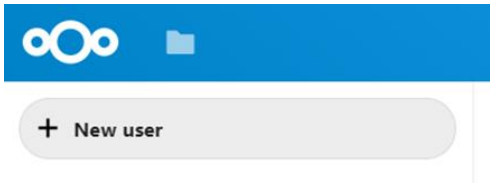
B.4.1 Creating a group

To create a new group, go to the Users settings, select Add Group button and enter the group´s name.



B.4.2 Adding a new user

Group admins can add new users to the projects group by navigating to the Users settings and click on the New User button.



User accounts have the following properties:

- **Login Name (Username)**: The unique ID of a Nextcloud user, and it cannot be changed.
- **Full Name:** The user's display name that appears on file shares, the Nextcloud Web interface, and emails. Admins and users may change the Full Name anytime. If the Full Name is not set it defaults to the login name.
- **Password:** The admin sets the new user's first password. Both the user and the admin can change the user's password at any time.
- **Groups:** You may create groups, and assign group memberships to users. By default, new users are not assigned to any groups.
- **Group Admin:** Group admins are granted administrative privileges on specific groups, and can add and remove users from their groups.
- **Quota:** The maximum disk space assigned to each user. Any user that exceeds the quota cannot upload or sync data.