MODI A LEAP TOWARDS SAE L4 AUTOMATED DRIVING FEATURES

A leap towards SAE L4 automated driving features

Report on database creation for data collection and sharing

31st July 2025

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

Co-funded by the European Union

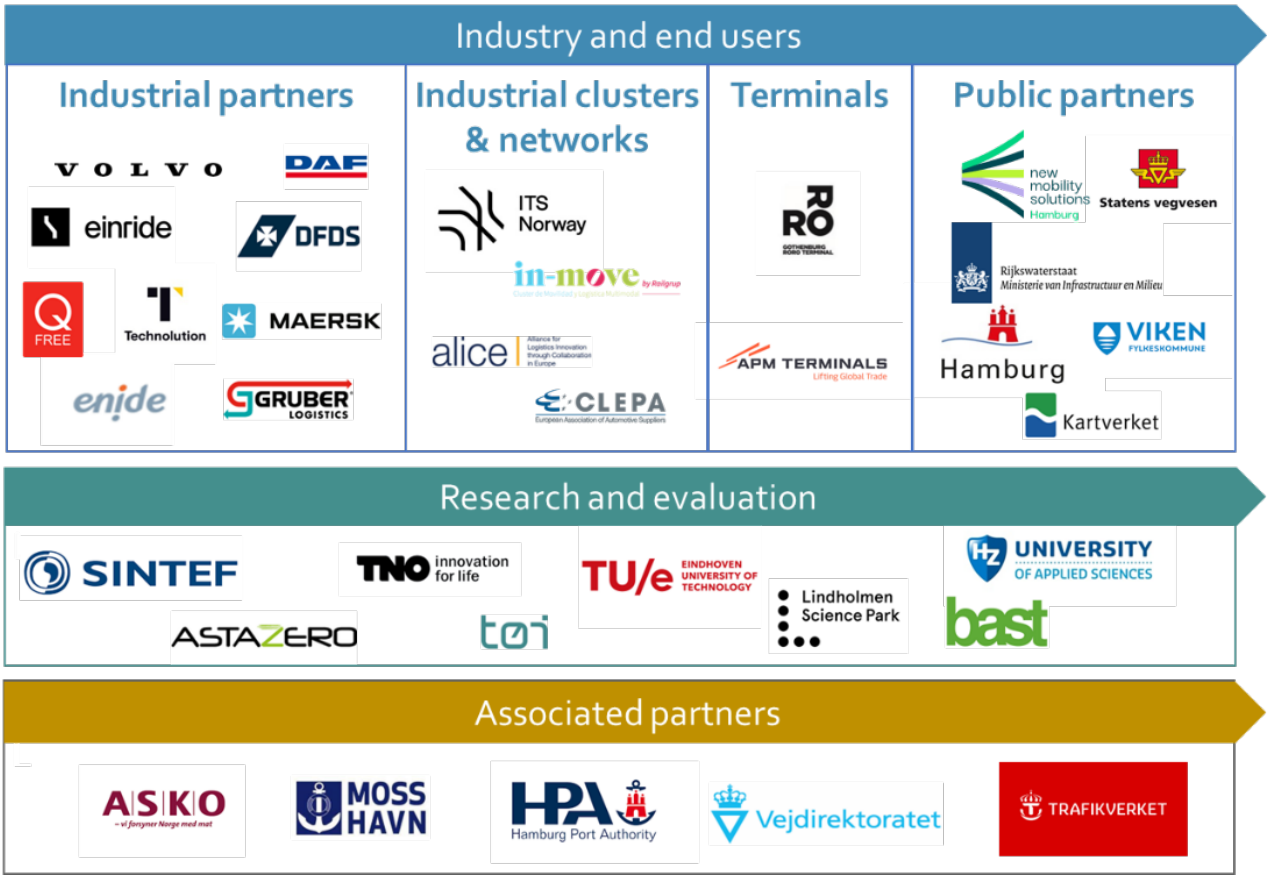
A leap towards SAE L4 automated driving features

Report on database creation for data collection and sharing

31st July 2025

Document Summary Information

Grant Agreement No	101076810	**Acronym**	MODI
Full Title	A leap towards SAE L4 automated driving features		
Start Date	01st October 2022	**Duration**	42 months
Project URL	https://www.modiproject.eu		
Deliverable	D2.3 Report on database creation for data collection and sharing		
Work Package	WP2		
Contractual due date	31st July 2025	**Actual submission date**	18st July 2025 (for version 1.0)
Nature	R – Document, report	**Dissemination Level**	PU – Public
Lead Beneficiary	TUE		
Responsible Author(s)	Jos den Ouden (TUE)		
Responsible Co-Author(s)	Pavol Jancura, Jochem Smit, Lars van der Brom, Tom Nagel (TUE)		
WP leader	Hanne Seter (SINTEF)		
Technical expert peer reviewer(s)	Simen Støa (SINTEF), Tariq van Rooijen (TNO), Sven Jansen (TNO)		
Quality peer reviewer(s)	Runar Søråsen (ITSN)		
Approved	Ragnhild Wahl (ITSN)		



Revision history (including peer reviewing & quality control)

Version	Issue Date	% Complete	Changes	Contributor(s)
V0.1	01/05/2025	5		Jos den Ouden, Jochem Smit, Tom Nagel, Lars van den Brom, Pavol Jancura (TUE)
V0.2	25/06/2025	5	Draft for Technical expert peer review	Jos den Ouden, Jochem Smit (TUE)
V0.3	08/07/2025	99	Draft for Quality review	Jos den Ouden, Jochem Smit (TUE)
V1.0	11/07/2025	100	Final version (1.0)	Jos den Ouden, Jochem Smit (TUE)

Disclaimer

The content of this document reflects only the author's view. Neither the European Commission nor CINEA is responsible for any use that may be made of the information it contains.

While the information contained in the document is believed to be accurate, the authors(s) or any other participant in the MODI consortium make no warranty of any kind regarding this material, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose.

Neither the MODI Consortium nor any of its members, their officers, employees, or agents shall be responsible or liable in negligence or otherwise, howsoever in respect of any inaccuracy or omission herein.

Without derogating from the generality of the foregoing, neither the MODI Consortium nor any of its members, their officers, employees, or agents shall be liable for any direct or indirect or consequential loss or damage caused by or arising from any information advice or inaccuracy or omission herein.

Copyright message

Table of Contents

List of Figures

List of Tables

Terms and abbreviations

Term / Abbreviation	Description
CCAM	Connected, Cooperative and Automated Mobility
DOI	Digital Object Identifier
DSF	Data Sharing Framework
FAIR	Findable, Accessible, Interoperable, Reusable
FAME	Framework for coordination of Automated Mobility in Europe
FESTA	Field opErational teSt supporT Action
FOT	Field Operational Test
GDPR	General Data Protection Regulation
GNSS	Global Navigation Satellite System
L2, L4	SAE Level 2, and SAE Level 4 of driving automation[1]
NDS	Naturalistic Driving Study
OEM	Original Equipment Manufacturer
RQ	Research Question
SAE	Society of Automotive Engineers
UC	Use Case
UC CCAM	Use Case CCAM Test Corridor
UC GE	Use Case Germany
UC NL	Use Case Netherlands
UC NO	Use Case Norway
UC SE	Use Case Sweden
WP	Work Package

It is to be noticed that there is a plethora of terms used to refer to the CCAM solutions/vehicles to be used in the project (e.g., L4 CCAM vehicles, L4 CCAM automated vehicles, CCAM vehicles, CCAM AVs). To ensure coherence, the term "CCAM vehicles" (which in this project refers to either L4 or L2 automated vehicles) is used throughout the document.

[1] https://www.sae.org/blog/sae-j3016-update

Executive Summary

The main purpose of this deliverable D2.3 "Report on database creation for data collection and sharing", is to report on the design and implementation of the MODI data sharing platform that has been created in Task T2.3 of MODI, to be used by T2.5 (Impact analyses) and T2.6 (Gap analyses) and populated with (meta)data from MODI use cases. This is part of T2.3 in WP2 of the project, in which three specific points are defined for WP2. The MODI evaluation methodology will:

a. collaborate with the FAME[2] (HORIZON-CL5-2021-D6-01-06) project.
b. tailor the evaluation methodology to the specific needs of the UCs of MODI, including identifying the input required from each UC to perform the evaluation.
c. specify common and preferably standardised data formats, in coordination with the UCs and aligned with the FAME project.

In T2.2 the methodology and data format (through collaboration with FAME) has been described (see D2.2 [1] for complete overview), which T2.3 has taken and:

- *developed of cloud-based data-sharing platform*: Section 3 explains the design and architecture of the MODI data sharing platform with using standard cloud-based technology to ensure Open Access
- *created a consolidated project database, a public database, and a data flow*: Section 4 explains how the MODI data sharing platform can be utilized by data providers and data analysts, showcasing the data flow. It also shows how the metadata is stored on public databases and how open data can be shared publicly.
- *created a data sharing service internally as well as externally*: Section 4 explains also how the open metadata (MODI data sharing platform) and data (using Zenodo) can be accessed for MODI partners and others outside of the consortium.

To support consortium members (in particular WP5 and T2.5 and T2.6), this deliverable explicitly describe how the data sharing platform can be utilized (specifically chapter 4), making this deliverable also act as a manual. So far the platform has been tested by SINTEF, TUE and NPRA and several demonstrations / information sessions (to UC leads and T2.5, T2.6 partners in June 2024, to MODI NL partners in April 2025 and in MODI monthly meeting on 30 January 2025) on how to use the platform have been organized to get utilization of the platform started.

[2] Complete project name: Framework for coordination of Automated Mobility in Europe.

1 Introduction

1.1 Project summary

MODI Ambitions: A leap towards SAE L4 automated driving features

The MODI project aims to accelerate the introduction of highly automated freight vehicles through demonstrations and by overcoming barriers to the rollout of automated transport systems and solutions in logistics. The logistics corridor from the Netherlands to Norway has been chosen for demonstration activities as the Netherlands, Germany, Denmark, Sweden, and Norway are expected to be among the first movers to implement fully automated vehicles in Europe.

MODI comprises five use cases, each describing a part of the logistics chain in confined areas and on public roads. It identifies what is already possible on an automated driving level without human interaction and what is yet to be developed. The MODI objectives are to:

- Implement new technology within the cooperative, connected and automated mobility (CCAM) spectrum.
- Define recommendations for the design of physical and digital infrastructure.
- Demonstrate viable business models for connected and automated logistics.
- Perform technical and socio-economic impact assessments.

Major challenges include regulatory aspects and standardisation, border crossings, access control, charging, coordination with automated guided vehicles, loading/unloading and handover from the public to confined areas.

MODI test sites include a CCAM test corridor from Rotterdam to Oslo with specific use cases at Rotterdam (The Netherlands), Hamburg (Germany), Gothenburg (Sweden), and Moss (Norway).

The ambition of MODI is to take automated driving in Europe to the next level by demonstrating complex real-life CCAM use cases while:

- Showing the local, national, and international context of freight transport with CCAM vehicles, both in confined areas and on public roads.
- Cooperating and co-creating with logistics companies, road operators, vehicle OEMs, providers of physical and digital infrastructure and other stakeholders to bridge the gap between R&D and market readiness.
- L4 solutions for long-distance operational design domains.
- Creating innovative business models and improved business models across the logistics chain.
- Proving that the technology can soon deliver on promised benefits at relatively high speeds and medium traffic complexity, including a coordinated CCAM system to support smart traffic management.
- Paving the way to enable highly automatic transport on important corridors, connecting main ports across Europe.
- Accelerating CCAM in Europe by setting examples of business-wise CCAM integration in logistics.

1.2 Aim of the deliverable

The purpose of this deliverable is to present an overview of the architecture and implementation of the data sharing platform that has been developed by T2.3 and provide guidelines on how to use it. This deliverable is therefore named D2.3 Report on database creation for data collection (both quantitative as well as qualitative data and metadata) and sharing. It is reflecting the final development of the data sharing platform that was

finalized by Month 34 (July 2025) and which is to be utilised by T2.5 and T2.6, and the MODI Use Cases (in WP5).

1.3 Relation to MODI output

WP2 has two goals related to T2.3:

- Perform an impact assessment of the MODI L4 CCAM solutions and systems regarding traffic, environment, safety, economy, and society in real world conditions (done in T2.5)
- Perform a Gap Analysis of the technological and the societal readiness and give recommendations about the best practices (done in T2.6)

T2.3 specifically was tasked with the development of a data sharing platform that supports those efforts, utilizing the metadata format that was co-developed in T2.2. T2.3 therefore developed this platform to be used by MODI UCs and T2.5 and T2.6 predominantly (aside from it also making available the data for external use). This document is based on several demonstrations/workshops in which we demonstrated the MODI data sharing platform towards T2.5 *Impact analysis*, T2.6 *Gap analysis of the technological and the societal readiness* and WP5 Use Case lead. The first session was given to T2.5, T2.6 and UC leads on 24 June 2024. The topics provided then was on the architecture of the data sharing platform as well as a demonstration of the platform for data providers and data analysts. Later (April 2025) we also provided a separate workshop to the Use Case Netherlands (UC NL) partners as well as a project wide introduction at the MODI monthly update meeting in January 2025. In the design of the platform, T2.3 utilised the metadata format that was provided in the FAME project and the FESTA Data Sharing Framework, which was adapted by T2.2 for use in MODI project, as defined in D2.2 [1].

This document therefore also functions as a "manual" (specifically chapter 4) for the data providers (MODI Use Cases (WP5)) as well as the data users (T2.5 and T2.6) of the MODI project, and as such it is expected that this can serve all the MODI project partners.

1.4 Structure of the report

This deliverable is structured in the following chapters:

- Chapter 2 provides the motivation for the MODI Data Sharing Platform and a description of definitions of data generated (data and metadata) and used in MODI for which the data sharing platform has been created, considering also requirements for complying with FAIR principles.
- Chapter 3 shows the design and architecture of the MODI Data Sharing Platform, considering the data format and FAIR principles defined from Chapter 2 and D2.2 [1].
- Chapter 4 describes for the two main user categories of the platform: The data providers and the data analysts. The chapter also provides a step-by-step approach on how to upload, find and download data using the FAME described metadata format.

2 Motivation for a data sharing platform in MODI

The MODI initiative participates in the Open Research Data platform, OpenAIRE, to enhance the accessibility and usability of research outputs. The establishment of a data sharing platform is therefore motivated by two core objectives:

- The data collected in MODI (by the Use Cases (UCs) in WP5) and shared on the data sharing platform is to be used in the evaluation of the project (by T2.5 and T2.6) and therefore requires input from use cases (WP5).
- Provision of Open Access to Research Data: Adhering to the FAIR (Findable, Accessible, Interoperable, Reusable) principles to maximize data utility and transparency, both internal and external of the project.

As stated in the Grant Agreement, T2.3 shall facilitate data collection/sharing for evaluation through:

- *development of cloud-based data-sharing platform*: Section 3 explains the design and architecture of the MODI data sharing platform using standard cloud-based technology to ensure open access.
- *creation of a consolidated project database, a public database, and a data flow*: Section 4 explains how the MODI data sharing platform can be utilized by data providers and data analysts, showcasing the data flow. It also shows how the metadata is stored on public databases and how open data can be shared publicly.
- *creation of a data sharing service internally as well as externally*: Section 4 explains also how the open metadata (MODI data sharing platform) and data (using Zenodo) can be accessed for MODI partners and others outside of the consortium.

During the course of the project, the data-sharing platform has been developed with the use of the data formats that were defined in T2.2 (see D2.2 [1]).

The platform has been tested and finetuned for ease of use and clarity, with help of WP2 lead SINTEF, UC lead TNO and OEM DAF (L4 CCAM vehicle).

Currently the platform is gradually being filled with data from the UCs (led by TNO, ITSH, LSP and SINTEF) in WP5 and data can be used by Tasks 2.5 and 2.6 for evaluation.

Section 3 and section 4 describes further the design of the platform and **how** data can be shared, using the MODI data sharing platform.

Deliverable D2.2 [1] already describes in full detail **what** data is shared (based on the FAME CCAM Data Sharing Framework [2] (which is largely based on FESTA methodology [3] and FOT-NET/ CARTRE DSF [4]). Therefore, we will not cover this in detail in this deliverable. We will highlight and repeat some key aspects to better explain some of these concepts (such as data vs. metadata and FAIR principles) in the context of the development of the data sharing platform.

2.1 Data and metadata

According to D2.2, we define two types of data: data and metadata.

Data refers to the recorded values resulting from the measurement of a system, environment, or actor during an experiment, field operational test, or similar activities (context). This includes raw measurements (acquired) as well as processed information (derived) or reduced data (aggregated) that reflects what occurred during the test or activity. Data can be both quantitative and qualitative.

Metadata is data that describes other data. It provides contextual information that is necessary to understand and interpret the primary data. Metadata includes information about the conditions under which

the data was collected, the format and structure of the data, the ownership and rights associated with it, and other descriptors.

Data re-use requires precise knowledge about the data. Therefore, it is vital to have extensive and high-quality metadata, providing the following information:

- the conditions in which they have been collected;
- the purpose;
- how they have been stored, cleaned up, processed and aggregated;
- and how they can be accessed.

This translates, in context of MODI, to the following data that is being collected in MODI:

- Data related to vehicles and digital infrastructure.
- Data related to the vehicle's environment.
- Data related to (test)-persons and stakeholders.

The metadata format is based on the FOT-Net Data Sharing Framework (DSF). The DSF can be consulted if anything is unclear; each section in the format is directly or indirectly based on a section in the DSF document [4]. According to FAME CCAM Data Sharing Framework (from here on named FAME CCAM DSF) [2] and D2.2, **metadata** in principle provides information about **data** and can be divided into different types, describing different traits of a dataset. In context of DSF and MODI, we define the following four subcategories of metadata:

- Descriptive metadata, describing the content of a dataset, is perhaps the most useful type for data analysis.
- Structural metadata are the prerequisite that helps the analyst understand the structure of the dataset, by describing 'data about the containers of data' (Roebuck K., 2012).
- Administrative metadata are collected for the effective operation and management of data storage.
- Execution documentation provides an overall description of how the study was performed.

See D2.2 and Annex I for a more extensive explanation of these four metadata categories.

As agreed early in the MODI project, the choice was made to use .YML file format for storing the FAME/FESTA metadata. A working metadata example is attached with this deliverable, see Annex II.

The structure and relationship between data and metadata is highlighted in Figure 1 (replicated from D2.2):

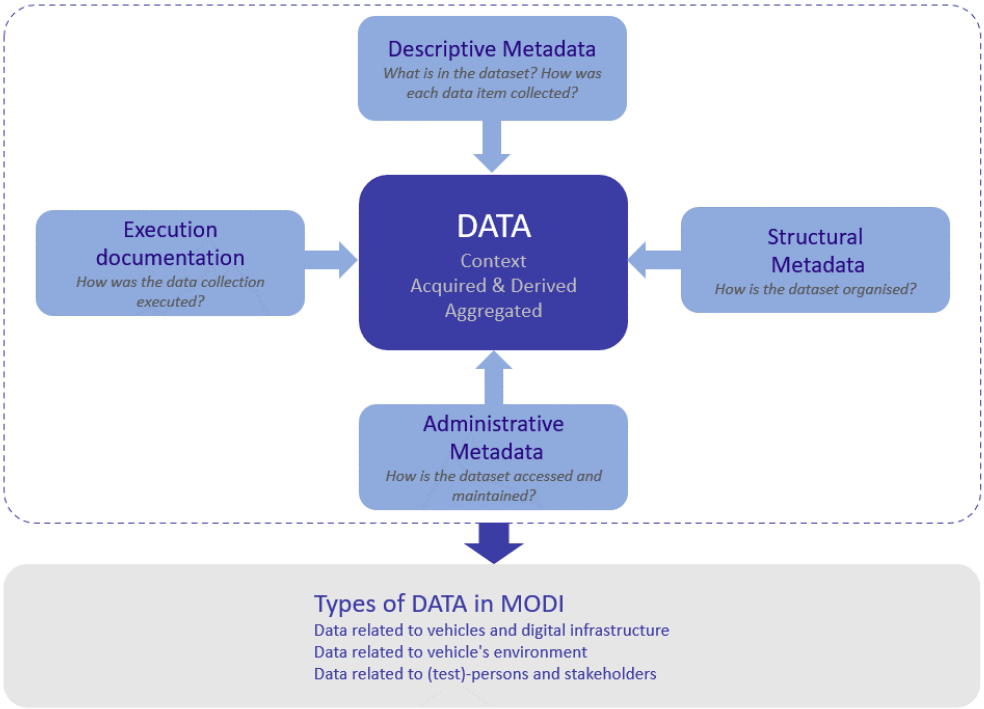


Figure 1: Relation between the data and metadata categories provided by the DSF and the types of data established for the MODI project.

2.1.1 Open vs. licensed/proprietary data

As indicated in D7.2 and D2.2, the MODI project works according to the FAIR principles [5], making data findable, accessible, interoperable, and reusable. Therefore, the project strives to store data as much as possible in an open repository to be accessible to a wider audience, outside the project consortium. Yet, it is important to remember that not all data are suitable to be openly accessible, as some data might be of a confidential commercial nature[3] or is subject to restraints due GDPR regulations. This had impact on the design of the MODI Data Sharing platform, where we chose early on to have 2 distinct data storage parts, based on data ownership (see Table 1): one server using the Zenodo platform [6] for open (public) data and the other server(s) at data providers for licensed and proprietary data (see Table 1 and D2.2 for a further explanation of these different types).

Table 1: Categories of confidential commercial data. Adapted from FOT-NET/CARTRE DSF [4].

Data category	Ownership	Access
Open	Owned by all or part of the project consortium*	Open for all or certain project partners*
Licensed	Owned by the data provider (usually the data owner who holds the IP rights)	Open for all or certain project partners, with approval by the owner**
Proprietary		This data is not open/shared

*Table notes: *Needs to be agreed upon and signed by the project partners; **Licensed data could be made open by adding signals to convert the data to non-sensitive (recommended by the DSF).*

[3] From the FAME CCAM DSF [2] and FOT-NET/CARTRE DSF [4]: "*Confidential commercial data is information which an organisation has taken steps to protect from disclosure, because disclosure might help a competitor*", p. 44 , FAME CCAM DSF [2].

3 MODI Data sharing platform - Design concept and architecture

Considering the requirements associated with the storage and management of large datasets as stated in Section 2, the MODI project had the specific task aimed to facilitate data collection and sharing (i.e., T2.3 *Development and utilisation of CCAM data sharing platform*). This task has developed a cloud-based data-sharing platform (called MODI Data Sharing Platform) that allows data access internally to the project partners for analysis (e.g., for impact and gap analyses). The data-sharing platform is also designed to be used to share data externally in the future. The management of the data available on the data-sharing platform is done in close cooperation with T7.3 *Data Management, Ethics, Security and Gender Equality Plans*, and in accordance to D7.2 *Data Management Plan* [7].

3.1 Design concept

The creation of the data-sharing platform for the project follows the initial design concept illustrated in Figure 2 below.

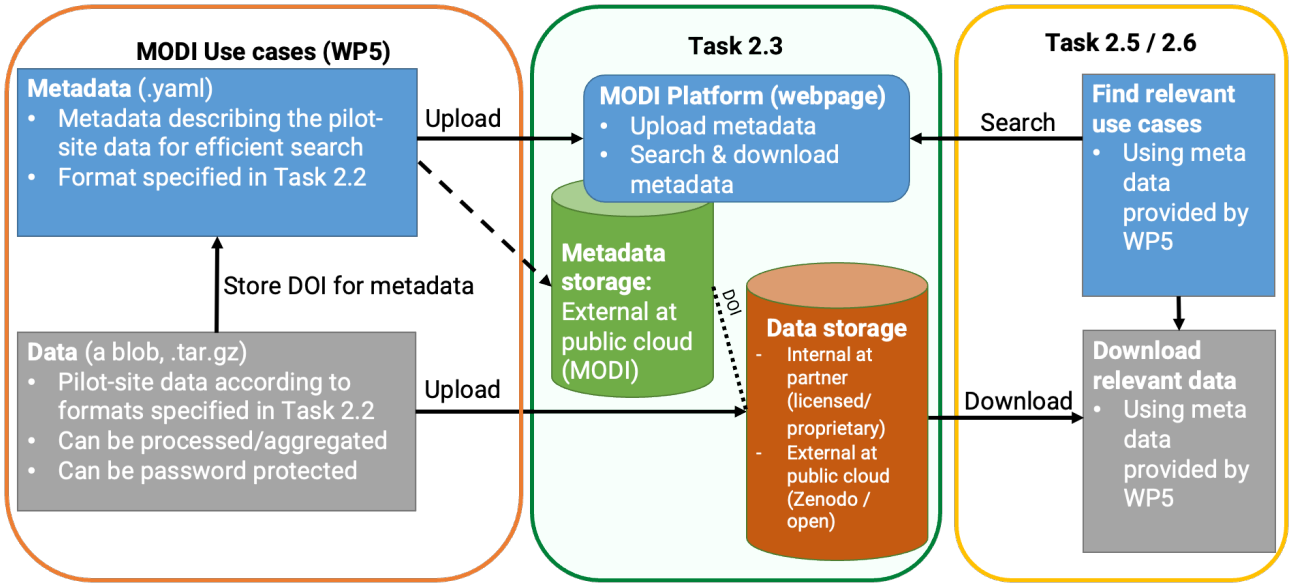


Figure 2: The design concept of the data-sharing platform.

The design concept for MODI data-sharing platform from Figure 2 can be described as follows:

- Data is the data directly collected by the project partners, particularly the by MODI UCs (part of WP5) (left grey box), as defined in [4], section 5.1 and in D2.2 [1].
  Each partner in charge of the collection of data is responsible for the organisation of their data.
  The platform assumes storage of two types of data: proprietary (stored on partner's own local server solutions) and public (stored on MODI Zenodo platform or other publicly available platforms).
- Metadata linked to the data, providing descriptive, structural and administrative metadata to the original data. Annex I provides an example of such metadata needed to describe the data. The metadata follows the categorisation indicated in Section 2.1, Figure 1 and Annex I, to sufficiently describe the data for efficient search in the MODI data-sharing platform (left blue box).
- The data can be stored both internally at each partner's own institutional database, or externally at a public cloud environment (green cylinder). For MODI, a MODI community [8] has been created on Zenodo (see Figure 6). In case of licensed/proprietary data, the data provider is responsible for its storage.

- The MODI data-sharing platform has three main interfaces) upload, search, and download (blue box with rounded corners).

It is to be noted that there is a difference between the metadata storage (green cylinder), data storage (orange cylinder) and the MODI platform (blue rounded box). While the first one entails the metadata database and repositories owned by either the project partners or public repositories services storing data (as defined in Chapter 2), the second one (MODI platform) is the front-end webpage, functioning as a search and data access platform, providing external open access to the metadata. As such, the MODI data-sharing platform will not function as (and should not be mistaken for) a database to store all the project data (which is represented here as a separate 'data storage' (orange cylinder)). The MODI data-sharing platform is linked to those different databases, either via a DOI [9], or via a URL in the metadata, to redirect the data analyst to the adequate database in which specific data can be retrieved for further processing.

3.2 Architecture

The implementation of the MODI data-sharing platform is based on an architecture using existing open-source tools such as Github as well as additional access request processes to ensure MODI partners can gain access.

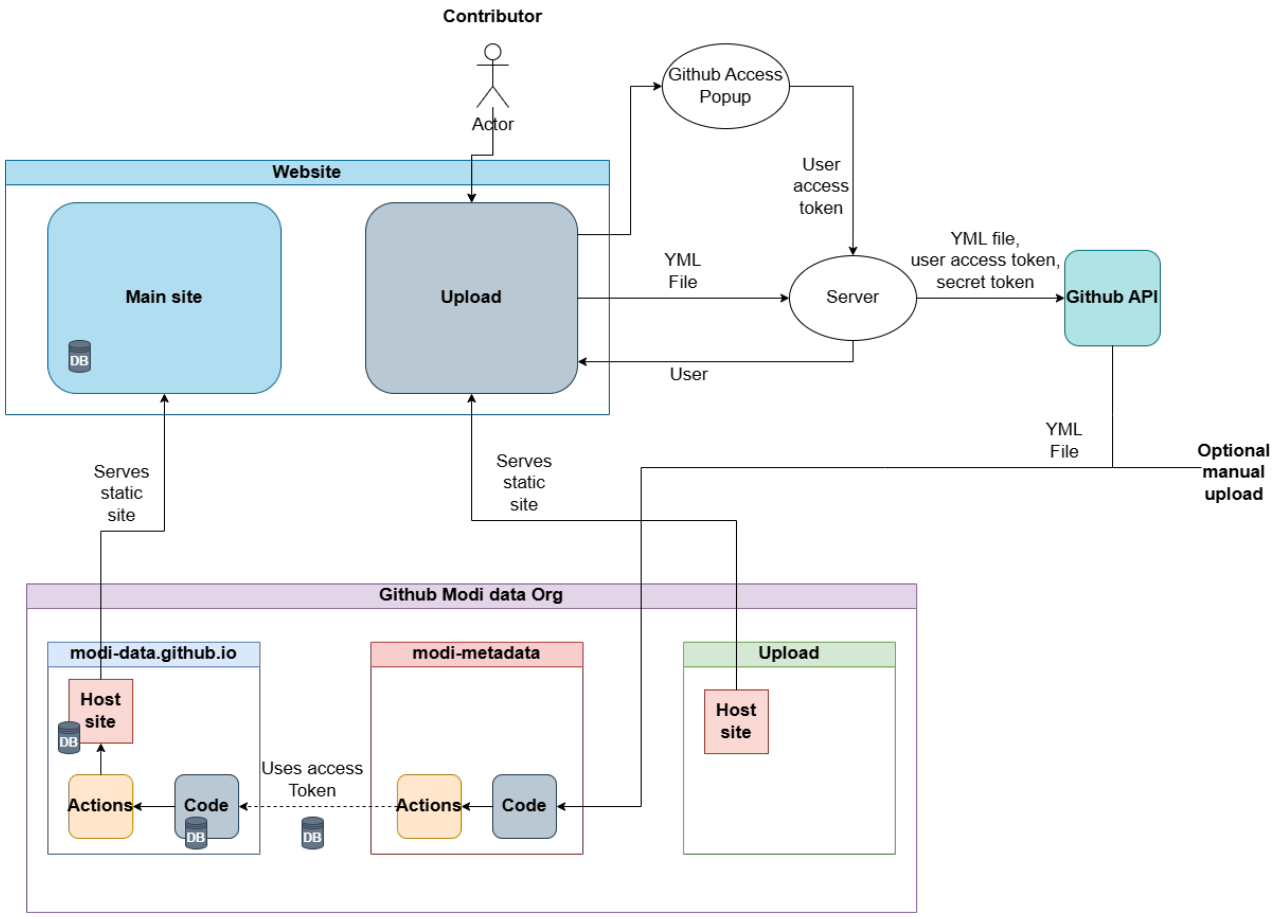


Figure 3: MODI data-sharing platform architecture

The architecture from Figure 3 describes the interaction between the different parts of the MODI data-sharing platform and can be interpreted as follows:

- A contributor provides input for the metadata via the upload section of the website which creates a .YML file with metadata. To authenticate each user and ensure they can upload the metadata properly to the MODI data sharing platform, the server requests that the contributor to log into GitHub, and then

the upload site presents a popup to log into GitHub. After the contributor has logged in, the user access token is sent to the server. Together with the secret token, the .YML file is uploaded to the MODI-metadata repository using the GitHub API. The Github API is necessary to be able to upload data to the MODI data sharing platform (e.g., as data provider), not to access and view / download the metadata on MODI data sharing platform (e.g., as data analyst). Thereby Open Access is guaranteed. Once the .YML file is uploaded to the MODI-metadata repository, it is automatically added to the database in the modi-data.github.io repository via GitHub actions and accessible to everyone (without restrictions or log in requirements (i.e., open access)).

- Alternatively, instead of using the upload site, a contributor can also manually upload a .YML file directly via the GitHub repository, making it available on modi-data.github.io.

3.2.1 Workflow for requesting access as a data provider

Initially, the data provider must create a GitHub account if they do not already have one. Subsequently, they must request access to the repository and become a contributor through a request form[4]. Within one business day, the repository owner will grant access, enabling the data provider to upload metadata via the following two methods. Getting access to the platform and getting a GitHub account are two distinct processes and Figure 4 these steps:

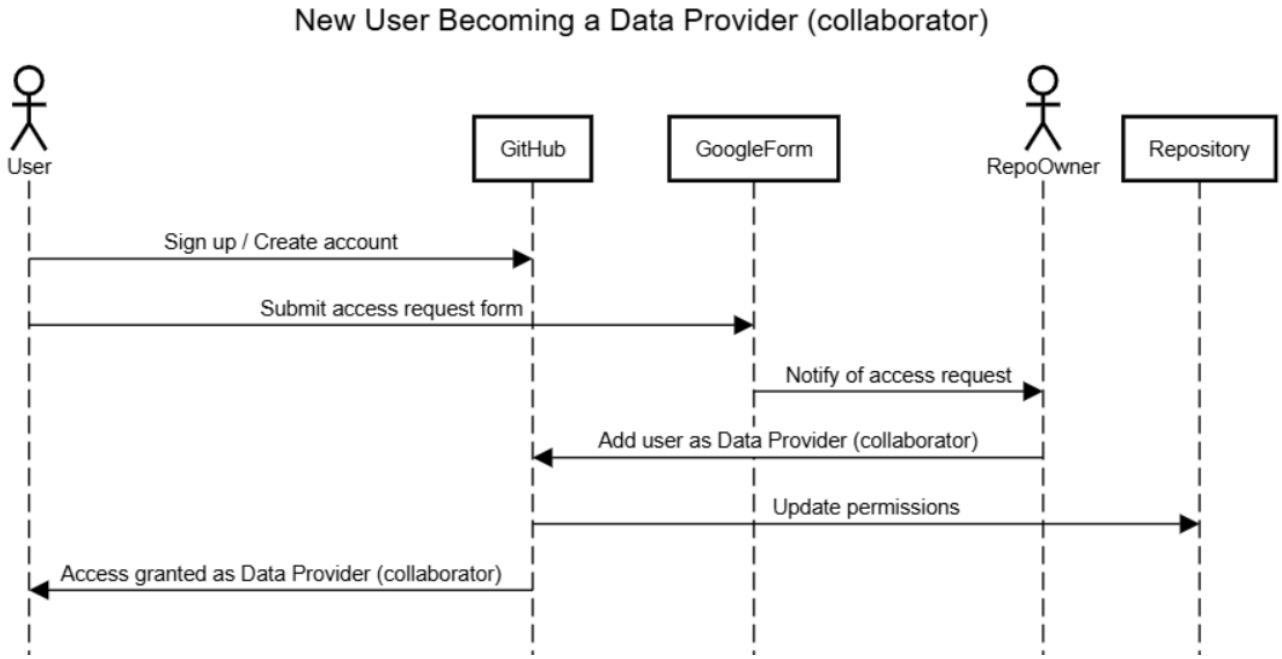


Figure 4: Process of becoming a data provider (named here also 'collaborator' according to GitHub standard terminology)

First, a new user must have a GitHub account or make a new account if needed. Second, the user needs to request access to the repositories of the MODI data-sharing platform via the a request form. Then within one business day, the user is added to the repository as a contributor by the repository owner. Finally, the new user is a contributor ready to upload metadata files on the MODI data-sharing platform.

[4] The access through the request form and feedback form has been distributed within the MODI consortium (see MODI Data Sharing Platform – launch slides only accessibly to MODI partners) and will not be made public here in this deliverable yet, to prevent unnecessary and unverifiable access.

4 Usage of the data-sharing platform

The MODI data sharing platform has been designed with two distinctive user types in mind: **data providers**, in context of MODI typically use case leaders / members, and **data analysts**, typically researchers in MODI from T2.5 and T2.6 who need the data for their evaluation tasks. Since the MODI data sharing platform is built upon open source tools, it is also accessible for external use.

This section describe how to use the platform for both user types. This description was already shared within the consortium the Data Sharing Platform kick-off event on 26 June 2024, for which all T2.5, T2.6 and Use Case leads of WP5 were invited. In total 19 members attended.

Section 4.1 describe how data providers can upload data and generate the required metadata on the MODI Data Sharing Platform to ensure FAIR principles. Since public data is uploaded to Zenodo a Digital Object Identifier (DOI) will be assigned in this process [9].

Section 4.2 describe how data analysts can find and download metadata on the MODI Data Sharing Platform and use this information to find the accompanying data, which was made available by the data providers, on either a public or proprietary cloud storage.

4.1 Data providers – how to upload data and metadata

Data providers are required to upload two types of data: data and metadata (required to make the data FAIR).

Section 4.1.1 describe the two ways to upload data. Section 4.1.2 describe how to add the accompanying metadata and link it to the data that was uploaded, to make the data comply with FAIR principles and the FAME metadata format.

4.1.1 Uploading data

For uploading data, there are two options, namely uploading public data or proprietary/licensed data. Both methods will be described below:

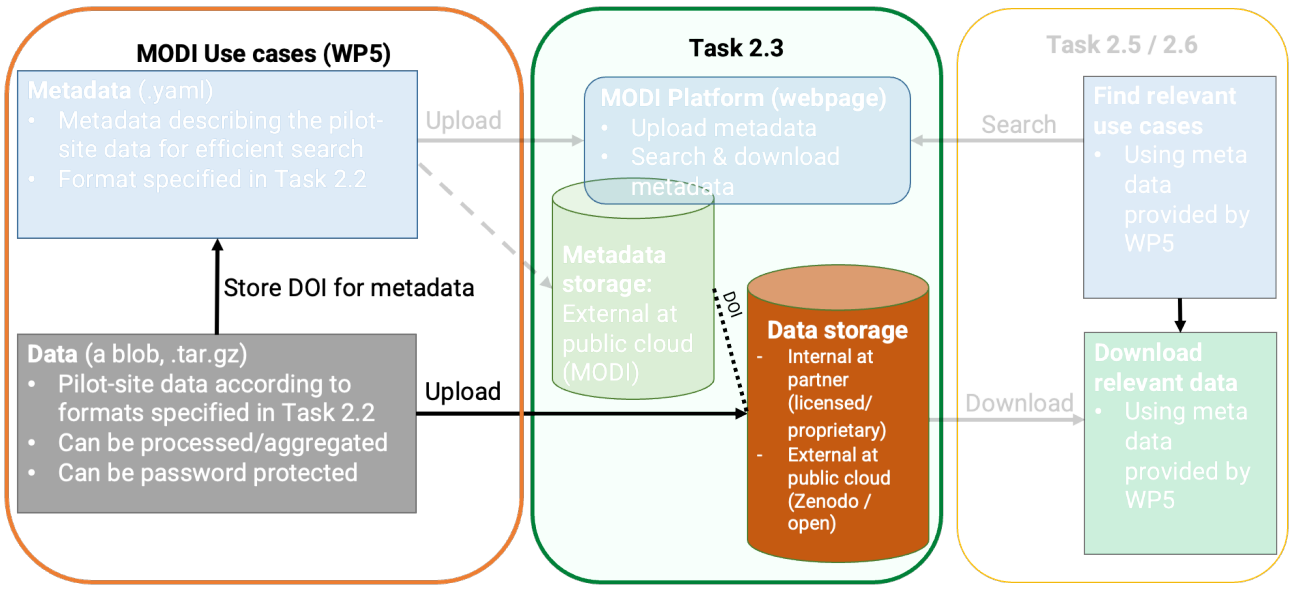


Figure 5: Uploading of data to either public or proprietary cloud databases

Uploading public data

MODI uses the Zenodo platform as a **public data** platform for storing MODI related data[5]. The benefit of using this existing platform is that it complies with all FAIR principles. On Zenodo there is a limit of 50 GB per dataset, and the availability of the data is guaranteed for at least 5 years.

In Figure 5, the MODI Zenodo page is shown. To upload data, the green 'New upload' button on the top right needs to be pressed. Then the data can be uploaded in the files section, and at least all information fields marked with a red asterisk (*) need to be filled in to be able to upload the data.

One of the required fields is the DOI (Digital Object Identifier) field. If the data provider already has one for the data, it can be added manually. If a new DOI is needed, Zenodo can generate a new one. The DOI is important because it ensures that the data can be linked to the metadata easily on the MODI Data Sharing Platform.

When all the necessary and applicable fields have been filled in, it can be selected whether the visibility of the data is public or restricted in the visibility section on the right side of the page. The standard setting is public, but it can be made proprietary. The data can be either published to the community if ready, or saved as a draft if more editing is needed.

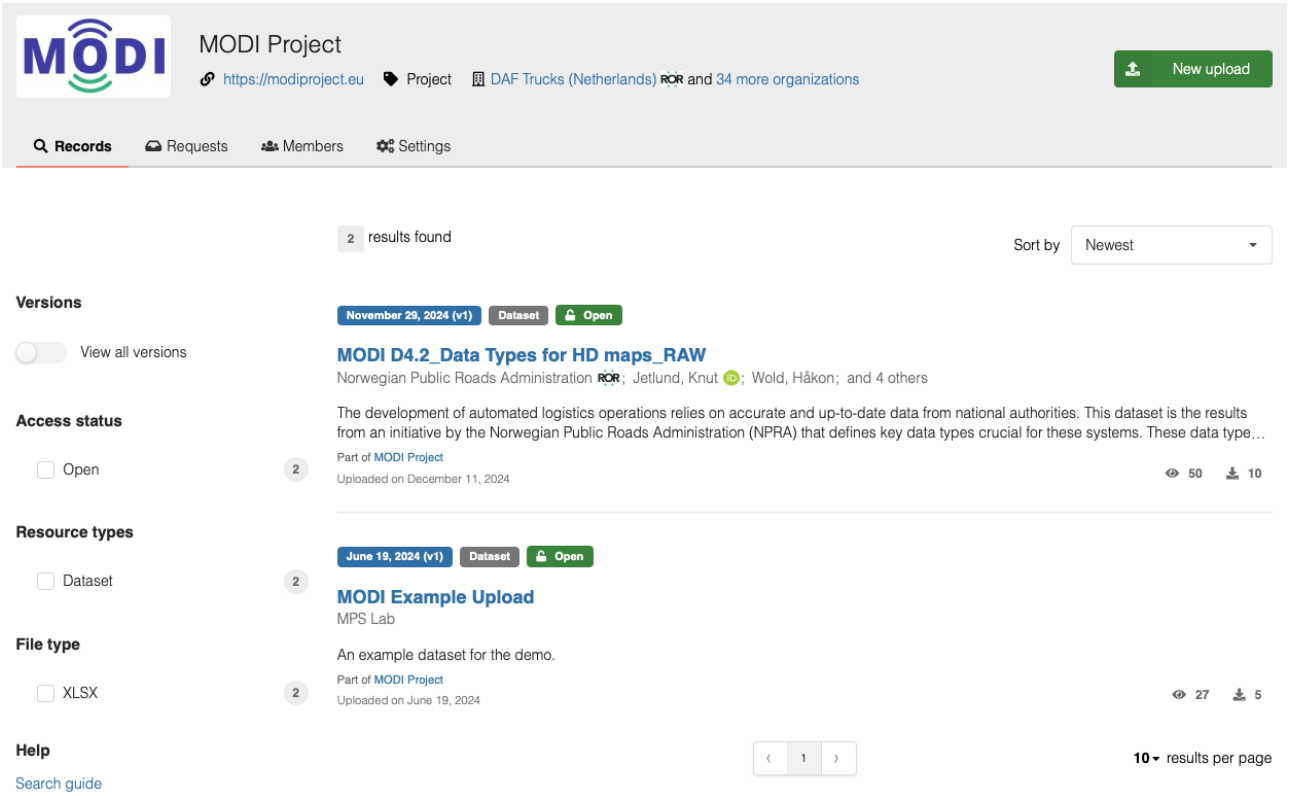


Figure 6: Overview of MODI Project community page on Zenodo for storing public data

[5] In principle, storing public data on proprietary servers would also be an option. However, it is then the owner's / data provider's responsibility to also adhere to FAIR principles. By using Zenodo instead, data providers/owners can comply with FAIR principles by definition.

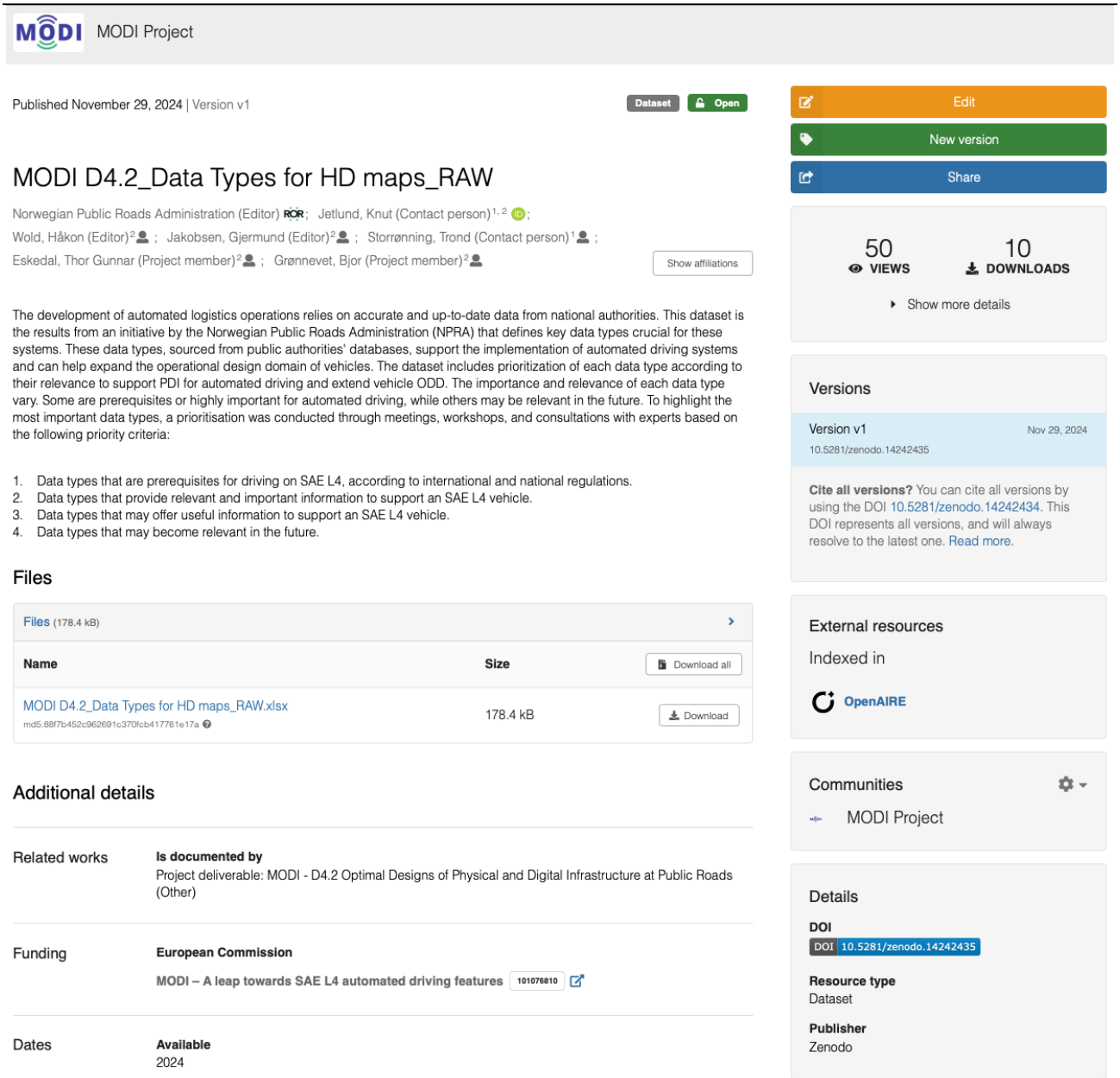


Figure 7: Example of a public data uploaded by MODI partner NPRA. In the bottom right corner, the DOI is shown, that was generated when uploading this dataset

Uploading proprietary or licensed data

Besides being able to upload data with restricted access on Zenodo as described in the section about public data, data providers can also upload data (licenses or proprietary data (see Section 2.1.1)) on their own proprietary databases, however there are some important things to note when doing this. First, the data providers themselves are responsible for the hosting of the data platform on which the proprietary data is stored. Secondly, the data provider should provide information on how to access the data. This can be done by correctly providing the metadata via the MODI data-sharing platform as described in 4.1.2.

4.1.2 Uploading metadata

For uploading data, we use the MODI platform as illustrated in Figure 8below.

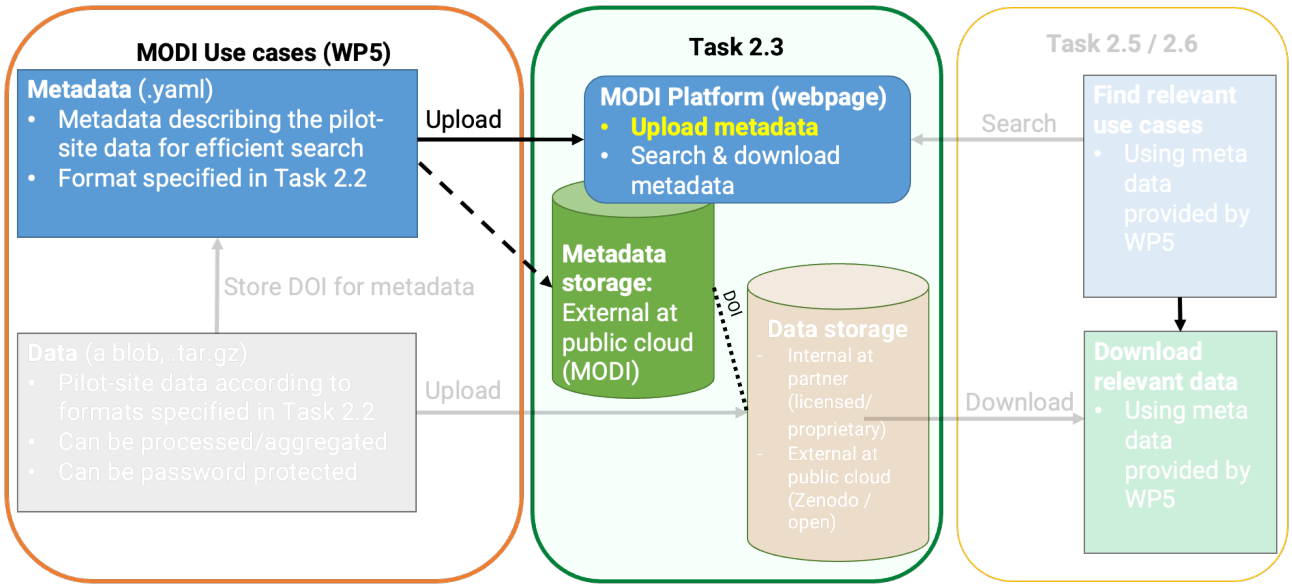


Figure 8: Uploading of metadata

The primary objective of metadata uploading is to ensure the data's accessibility, interoperability, and reusability (FAIR principles). The MODI data-sharing platform utilises a public GitHub repository to aggregate and store all metadata. Metadata can be uploaded in two ways: either through the metadata generator or manually by uploading a .YML file using the GitHub website.

However, since the MODI data-sharing platform relies on GitHub repositories, data providers must first obtain access (see Section 3.2.1 and Figure 4.).

Metadata Generator

The metadata generator is made such that data providers can easily generate and upload metadata by simply filling in a form (see Section 2.1 for reference of the metadata format and Annex II for an example of the generated metadata in .YML file format).

All data uploaded using this tool will be visible on the MODI Metadata platform. Figure 11 depicts some of the fields requested in the summary section of the metadata generator. To be able to generate a metadata file, all required fields need to be filled in (highlighted by an asterisk (*) in the field title). However, filling in the other fields will help users in finding the right data.

After filling in the fields of the summary section, the 'next button' on the bottom of the page moves to the next section. This is repeated until all sections are completed.

There are some data fields which are important to emphasize here: The 'Unique dataset id' field in the 'Administrative' section will determine the name of the metadata file, and thus also the name which data users must search for to find the metadata file. If the data provider uses Zenodo to store the data, the DOI should be provided since this links to the data. If the data provider uses a proprietary database to store the data, it is important to correctly fill in the fields shown in Figure 9 so that the data can still be FAIR.

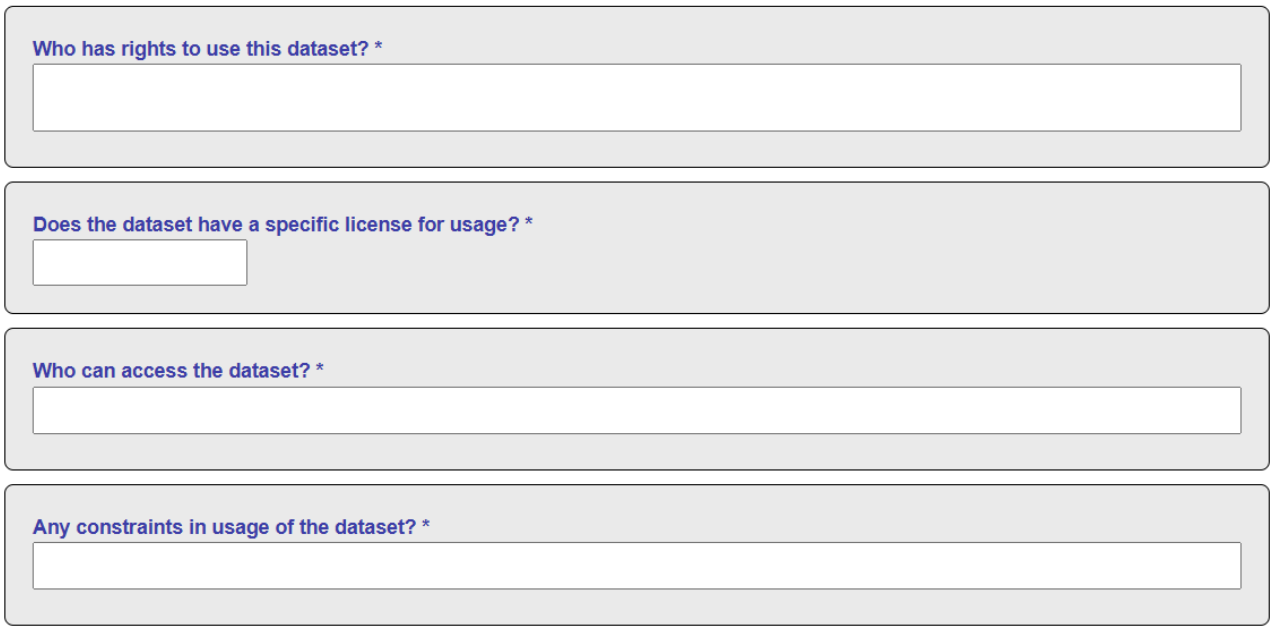


Figure 9: Fields to fill in metadata important for licensed / proprietary data

The Download and Upload buttons (see Figure 10) are used to either generate and download or generate and upload the metadata file.

Figure 10: Metadata form navigation buttons

Once the upload button is pressed, a popup message will appear asking the data provider to login to their GitHub account. When the data provider has successfully logged in, the data provider will get the popup message stating that the upload was successful.

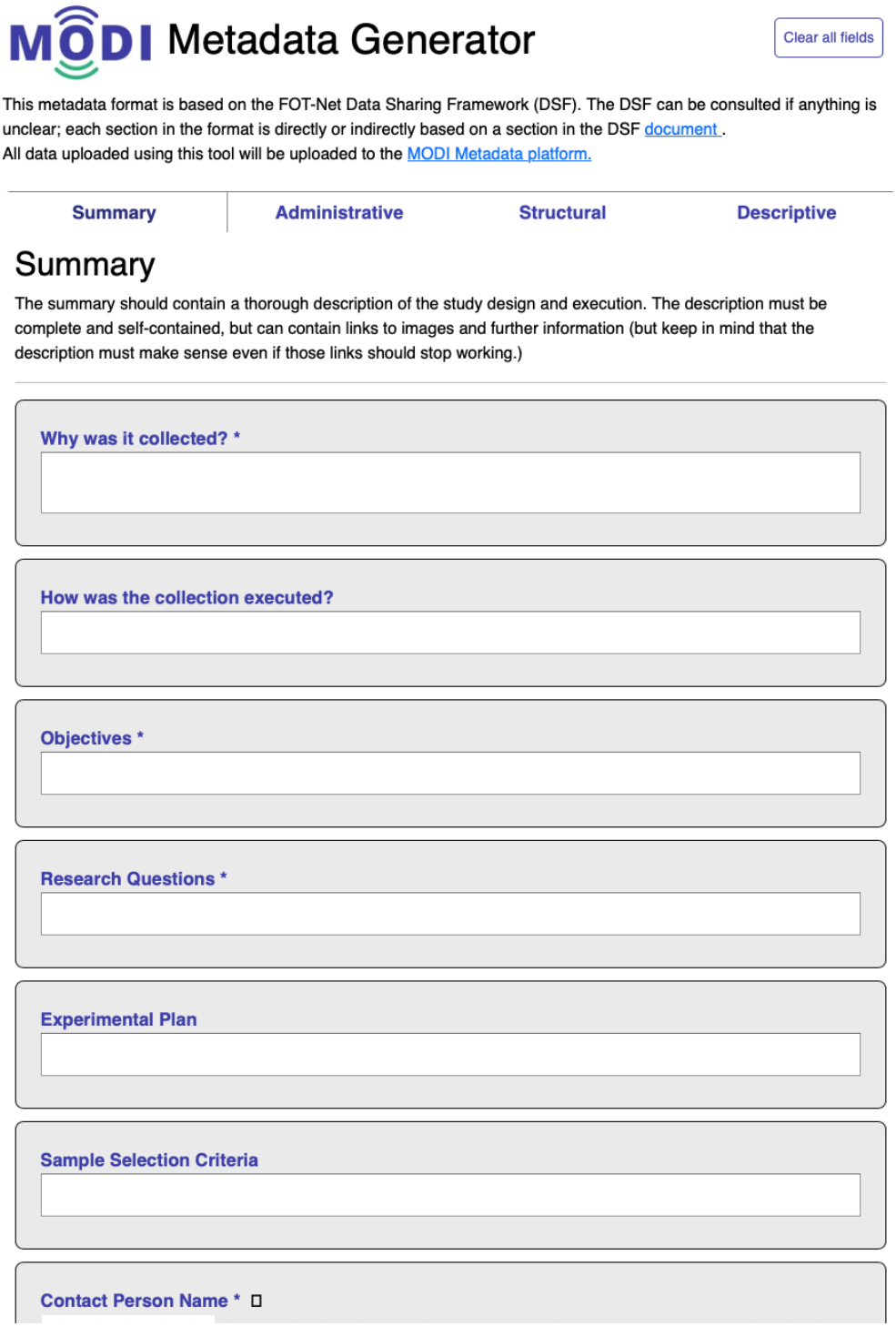


Figure 11: Overview of the metadata generator on MODI Data Sharing Platform website

4.2 Data analysts – how to download data and metadata

For data analysts, two main steps are relevant: search for metadata and downloading the relevant data for evaluation (see the right-hand side of the diagram):

Within MODI project, this is typically done by T2.5 and T2.6 as these are the evaluation tasks.

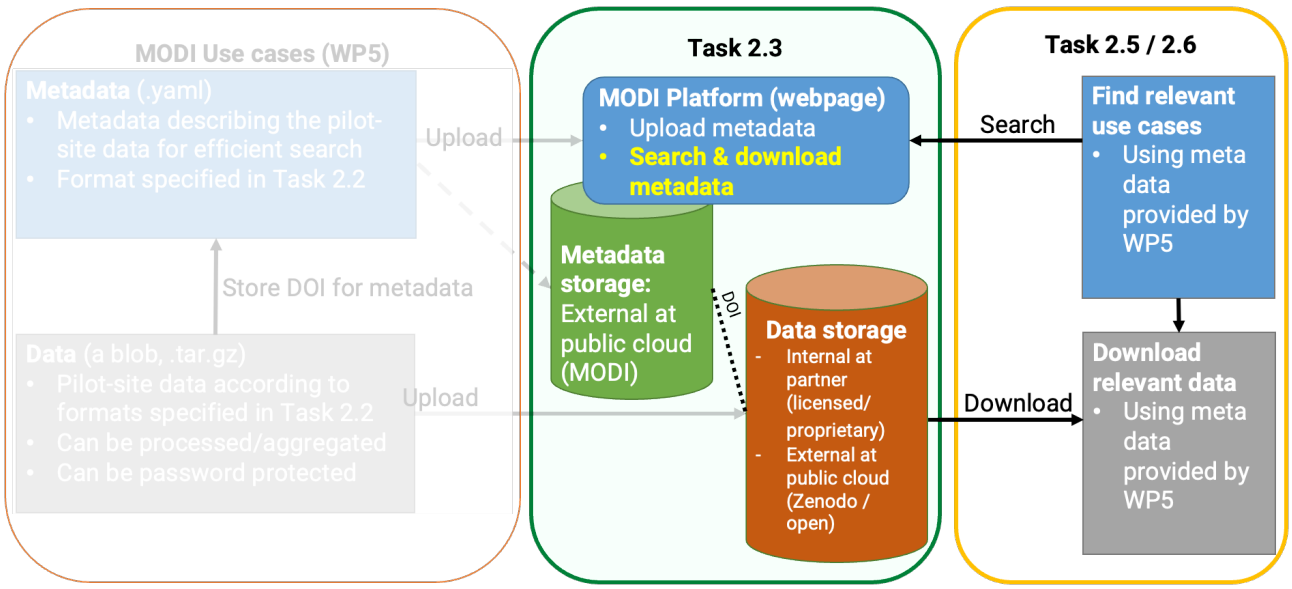


Figure 12: Finding the required data using the metadata on the MODI Platform and downloading the data from either public or proprietary cloud databases (highlighted)

4.2.1 Searching metadata

On the search page of the website of the MODI data-sharing platform (https://modi-data.github.io/search/), the search bar can be used to search for metadata using various criteria:

- Using the *Search* bar, metadata can be directly found when the title of the YAML file is known.
- The *Producer* bar searches metadata files made by the specified data provider: who produced/recorded the data. Can be company/institute/legal entity.
- The *Area* bar searches metadata files made in the specified area: this is where the data was recorded. Can be an exact coordinate, a polygon, a town name, a country name, etc. If providing coordinates/polygons, it must be encoded as Well Known Text (WKT) with an SRID of 4326 [10].
- The *Data type* bar searches for metadata files which point to data of specified data type: can be one or more of the following: survey, interviews, mobile network, itsg5, lane markings, signage, physical road infrastructure, speed/acceleration, video, pictures, GNSS, lidar, radar, weather, vehicle probe data, events or other. You may add multiple data types if relevant, separated by a comma.
- The *Modi use case* bar searches for metadata files which point to certain data gathered for a specified MODI use case: can be one of the MODI use cases: UC CCAM, UC GE, UC NL, UC NO or UC SE.

When the metadata file is found, the data user can click on the title, and then a view like Figure 13 is shown. In this view, the most important parts of the metadata are shown and a download button enables downloading the metadata.

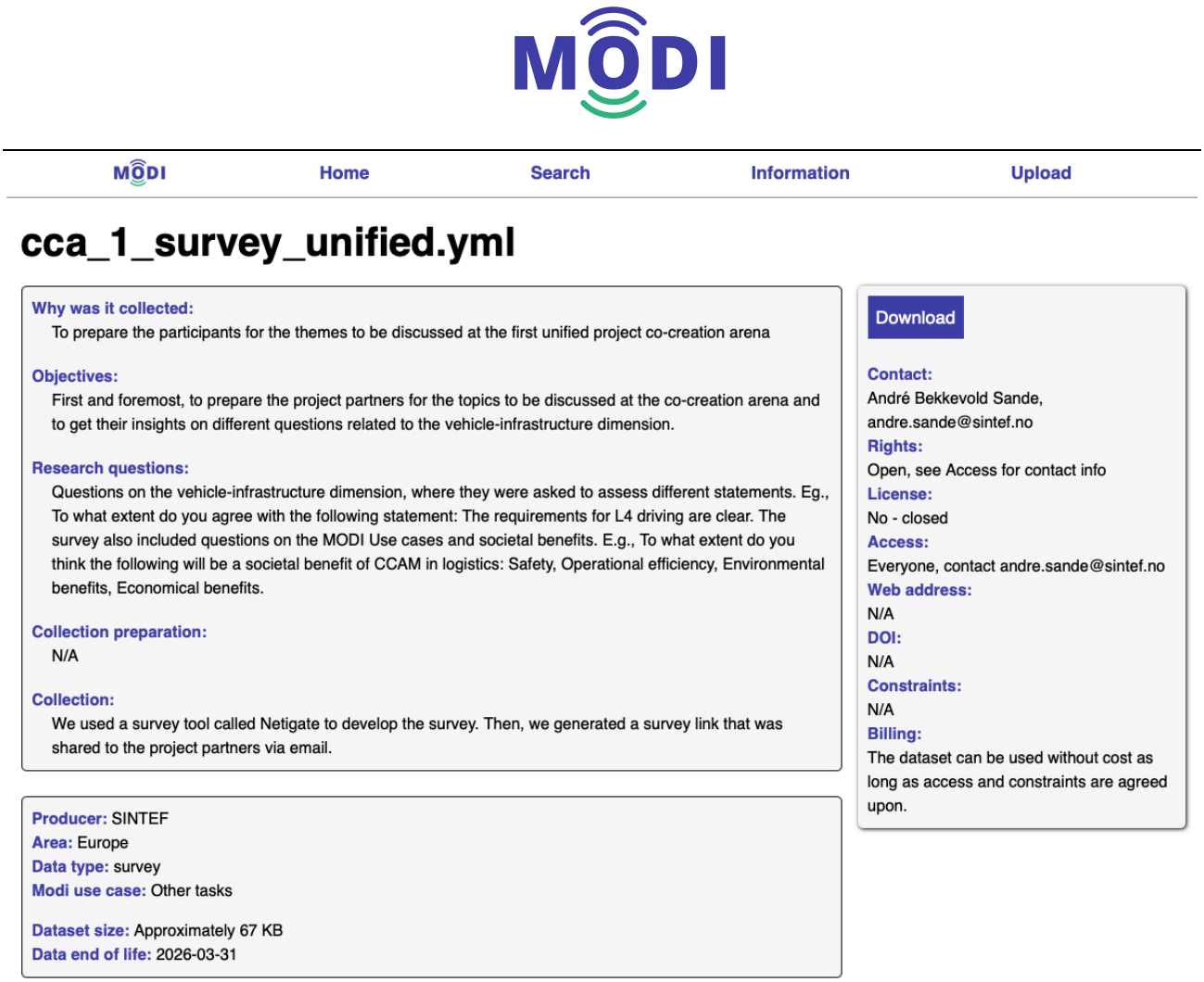


Figure 13: Example of metadata view

With the metadata acquired, the relevant DOI, URL to the actual data can be accessed. In case of proprietary data, contact information is provided (incl. email) to help the data analyst find the source of the data.

4.2.2 Issue tracking / feedback

For further improvement, we have set up a feedback form that can be utilized to provide feedback or indicate any issue. Please refer to the earlier mentioned kick-off slides[6] for this.

[6] The access through the request form and feedback form has been distributed within the MODI consortium (see MODI Data Sharing Platform – launch slides only accessibly to MODI partners) and will not be made public here in this deliverable yet, to prevent unnecessary and unverifiable access.

5 Conclusions

The scope of T2.3 included:

- *development of cloud-based data-sharing platform*: Section 3 explains the design and architecture of the MODI data sharing platform using standard cloud-based technology to ensure Open access.
- *creation of a consolidated project database, a public database, and a data flow*: Section 4 explains how the MODI data sharing platform can be utilized by data providers and data analysts, showcasing the data flow. It also shows how the metadata is stored on public databases and how open data can be shared publicly.
- *creation of a data sharing service internally as well as externally*: Section 4 explains also how the open metadata (MODI data sharing platform) and data (using Zenodo) can be accessed for MODI partners and others outside of the consortium.

This report shows why and how this data sharing platform was designed, implemented and utilized with several examples.

The data sharing platform was designed to handle both internal sharing as well as sharing externally and keeping in line with FAIR principles, by providing also links with the Zenodo platform (for public data storage). Additionally, a separate metadata platform was developed that helps describing that data, providing contextual information necessary to understand and interpret the primary data.

The platform was introduced within MODI in June 2024 for utilization by other partners and started with some initial testing of usage. Later, it was again demonstrated in the MODI monthly meeting with all WP leads, and UC leads, and again to UC NL to encourage uptake of the platform.

5.1 Best practices and lessons learned

Initially T2.3 aimed to set up early the MODI data sharing platform, to help generate early interest and encourage adoption by other MODI partners. To do so, T2.3 aligned with FAME project for data sharing insights and was hoping to adopt those to set up the MODI data sharing platform quickly. However, since also FAME project was in its early stage of development, T2.3 could not implement FAME's insights yet and instead collaborated with T2.2 towards implementing the metadata format based on previous attempts like FOT-NET/CARTRE DSF [4]. The chosen metadata format remained consistent throughout the project. This helped T2.3 being able to have a project wide introduction on 26 June 2024.

To ensure long-term platform lifetime and to facilitate testing, open-source tools such as GitHub were utilized in the developments. This decision allowed for reliance on an established authentication system and existing tooling. However, during deployment in June 2024, it was decided to restrict access only to MODI partners (during the project), to ensure utilization internally first and testing the platform, before opening the platform up for external use. A separate access request process was established (see Figure 3). Unfortunately, a institute wide cyber-attack on TU/e in January 2025 [11], caused the entire TU/e IT infrastructure to be shut down for days, disrupted this process. This eventually led to a prolonged platform outage of several months, since access to the access request machines was blocked by TU/e IT to prevent other cyber-attacks. This incident highlights the importance of hosting such data sharing platforms entirely on dedicated open platforms with redundant machines for improved reliability and reduced barriers to widespread use.

So far, only a handful of MODI partners have actively utilized the platform (mainly SINTEF, TUE and NPRA). The limited adoption of the platform so far could be attributed to several factors. Firstly, the current technical implementation requires data providers to log in separately to two platforms for uploading data and metadata, which creates an inherent barrier. Automating and centralizing this process would help alleviate this issue. Since both the metadata sharing platform as well as Zenodo already uses Github, such integration could be relatively straightforward, however this is not yet demonstrated.

Secondly, commercial sensitivity often prevents OEMs / TIER 1s etc. from sharing data. However, MODI has set up deliberately two different platforms: one for metadata and one for data. Sharing only metadata openly can already benefit CCAM related research, in which the data itself can still be kept proprietary or in licensed format. This lowers the threshold for commercial companies to share metadata and in participating in such projects. The metadata in the MODI data sharing platform, with a clear disclaimer that the actual data can still be accessed upon request from the (commercial) data provider, can be valuable to create awareness of such dataset to a broader public / external researcher.

5.2 Recommendations

To ensure long-term operability, the current metadata data sharing platform could be hosted completely in cloud, with additional redundancy measures in place. Ideally, this means that operations should be shifted to organizations such as Zenodo [8] or 4TU.ResearchData [12]. These organizations already have the proper infrastructure in place to host data according to FAIR principles. By ensuring the metadata is also stored on the same platform (linked to the data that is already stored on Zenodo for example), redundancy can be ensured, and ease of use can be improved (by having to only login, instead of two). Since the current metadata platform is already using similar access protocols, such a transfer could be relatively simple, but would require involvement of entities such as Zenodo or 4TU.ResearchData as partners in an upcoming project to ensure proper deployment. Such entities also have extensive experience with data sharing and its legal requirements (including GDPR). In MODI this was largely covered by D7.2 *Data Management Plan* [7], but still significant effort was required from both data providers and data analysts to establish the required agreements.

Data sharing is crucial for the development and deployment of Connected, Cooperative, and Automated Mobility (CCAM) solutions. While several data spaces exist for CCAM research and development, a comprehensive and interoperable CCAM data space is still needed to support the diverse needs of the industry and research community. This data space should facilitate data exchange, harmonize data formats, and establish governance mechanisms for data access and usage rights. According to [13], several CCAM related projects have already established separate data sharing platforms (including MODI), but utilization is often hampered by the lack of a centralized storage, to which all these projects can contribute. Therefore actions are proposes to set up a Federated CCAM data exchange platform [13]. This would address similar issues that T2.3 has encountered in the development of the MODI Data Sharing Platform. The above lessons learned, and recommendations could be used in the development process of such a federated platform and MODI could help transferring its data into that platform as well.

6 References

[1] MODI consortium, 'MODI - D2.2 Evaluation Methodology v.1.2'. Mar. 31, 2025.
[2] FAME Consortium, 'CCAM Data Sharing Framework v2'. Apr. 15, 2024. Accessed: Nov. 02, 2025. [Online]. Available: https://www.connectedautomateddriving.eu/wp-content/uploads/2024/02/CCAM-Data-Sharing-Framework-v2.02.pdf
[3] FOT-NET Data, 'FESTA Handbook, version 8', Updated and maintained by FOT-Net, CARTRE and ARCADE EU Coordination and Support Actions., 2021. Accessed: Feb. 22, 2023. [Online]. Available: https://www.connectedautomateddriving.eu/methodology/festa/
[4] FOT-Net Data support action and CARTRE support action, 'Data Sharing Framework v1.1'. Jan. 29, 2019. Accessed: Feb. 14, 2023. [Online]. Available: https://www.connectedautomateddriving.eu/wp-content/uploads/2021/09/Data-Sharing-Framework-v1.1-final.pdf
[5] M. D. Wilkinson, M. Dumontier, J. Aalbersberg, and et al., 'The FAIR Guiding Principles for scientific data management and stewardship.', *Scientific Data*, vol. 3, no. Article nr. 160018, 2016, doi: https://doi.org/10.1038/sdata.2016.18.
[6] 'Zenodo - Research. Shared.' Accessed: Jun. 25, 2025. [Online]. Available: https://about.zenodo.org/
[7] MODI consortium, 'MODI - D7.2 Data Management Plan'. Mar. 31, 2023.
[8] MODI consortium, 'Zenodo MODI Project Community page'. Accessed: Jun. 25, 2025. [Online]. Available: https://zenodo.org/communities/modi-project/records
[9] DOI Foundation, 'What is a DOI?' Accessed: Jun. 25, 2025. [Online]. Available: https://www.doi.org/the-identifier/what-is-a-doi/
[10] 'EPSG:4326 WGS 84 -- Spatial Reference'. Accessed: Jul. 09, 2025. [Online]. Available: https://spatialreference.org/ref/epsg/4326/
[11] 'cyberattack'. Accessed: Jan. 14, 2025. [Online]. Available: https://www.tue.nl/en/news-and-events/cyberattack
[12] '4TU.ResearchData'. Accessed: Jul. 04, 2025. [Online]. Available: https://data.4tu.nl/
[13] 'Federated CCAM data exchange platform (CCAM Partnership) | Programme | HORIZON', CORDIS | European Commission. Accessed: Jul. 07, 2025. [Online]. Available: https://cordis.europa.eu/programme/id/HORIZON_HORIZON-CL5-2026-01-D6-06

7 Annex I: Descriptions of categories for data and metadata, and data acquisition methods (from D2.2)

The following sub-sections present a brief description of the different categories of data and metadata, and different methods for data acquisition, based on the FESTA methodology and as presented also in D2.2 [1]. For a more comprehensive definition of the different data aspects, the reader is encouraged to see the FESTA handbook [3].

Data categories

- Context data: Indicate all the circumstances that surrounds a setting for a research study. This information is needed to fully understand the conditions in which the data was collected and to properly derive to conclusions. The context data can be directly collected (e.g., simultaneously with other type of data, such as weather sensors) or retrieved from external data sources (e.g., access to open weather data sources). It can also be already existent (e.g., regular and natural traffic conditions) or generated for the experimental aim (e.g., simulated and/or modified traffic conditions). Context data refer also to background information of, for example, people demographics (e.g., age, gender) and subjective attitudes towards a specific product, function, or service (e.g., people's subjective evaluations). As such, this data category can be used for either quantitative or qualitative data.
- Acquired and/or derived data: Acquired data are all the data that was purposely collected to directly address a study research question, i.e., for analyses purposes. Derived data are all data that have been under a transformation process from raw data to more usable data needed to evaluate, for example, KPIs. As an example within the MODI project, derived data would be the combination of the data for "perceived safety scale" and the data for "types of encounters with CCAM" to evaluate the indicator "Safe behaviour of users of confined area (occupational safety)" (see KPI-S02, D2.2 [1]). Acquired and/or derived data include data (not limited to):
    - Time-history data (measurements over time)
    - In-vehicle measures (such as vehicle dynamics, in-vehicle systems state, vehicle positioning and media (audio and/or video data))
    - Human behaviour measures
    - Roadside measures
    - Experimental conditions (as external factors that might impact people's behaviours)
    - Time and location segments (measurement during a delimited period)
- Aggregated data: Indicate that different kinds of reduced data (e.g., at a location segment) are used together to perform a smaller, more usable data analysis and interpretation.

Although derived data and aggregated data follow a similar principle (i.e., both combine different types of data), aggregated data is sought to be used for smaller analyses (e.g., one isolated situation), whereas derived data is sought to be used for larger analyses (e.g., to apply results to the whole project).

Metadata categories

- Descriptive metadata: Respond to the questions: *What is in the dataset?* And *How was each data item collected or calculated?* Descriptive data indicate the content of a dataset, by giving detailed information needed to understand the data. As such, it is probably the most useful type of metadata for data analysis. The descriptive metadata are used to describe context data, acquired or derived data, and aggregated data. For example, the metadata for some location segments in MODI (e.g., urban area in an UC) can indicate the measures needed to collect the data, and characteristics related to the location segment (e.g., number of exits or roundabouts, number of vehicles passing at a specific time and location segment, and/or average speed for such vehicles). The KPIs related to specific datasets should also be defined as descriptive metadata.

- Structural metadata: Respond to the question: *How is the dataset organised?* As the name indicates, structural metadata describes the structure of a data as a database schema design. For example, how data is organised in a dataset. This comprehends the definition of a database in which the system format is outlined, i.e., the data containers and formats. The database design, data container and content description are also part of the structural metadata. The storage and deposition of data for the MODI project is described in the project document D7.2 Data Management Plan.
- Administrative metadata: Respond to the question: *How is the dataset accessed and maintained?* The administrative metadata indicate the administrative information needed for the proper management of data storage, including topics such as access conditions, ownership, and preservation. The administrative metadata should be stored together with the datasets. Examples of types of administrative metadata are (but not limited to): version number, archiving date, access and rights information, contractual agreements, periodic backups, and end of data life. The creation of Digital Object Identifier (DOI) not only supports the creation for references and citations when the dataset has been used in a publication, but it also supports data management. For the MODI project, the creation of DOI will be used as a standard identification mechanism for the project's metadata (see D7.2 Data Management Plan).
- Execution documentation: Respond to the questions: *Why was the dataset collected?* This type of metadata provides an overall description of how the study was carried out, i.e., the study design, including the objectives and research questions. This is important for data analysis, as the documentation can provide information to project partners who were not a part of the data collection to properly interpret the data results, thus understanding how the data answer to specific research questions. This type of metadata also responds to the question: *How was the data collection executed?* The execution of a study should thus be documented, for example in a similar way as when it is reported in a research article. This means that the description of the study procedure should include e.g., RQs, experimental design/test plan, stimuli, monitoring equipment, study participants, description of test site/environment, sample selection criteria, and duration.

Data acquisition methods

- Background data acquisition: Includes details about drivers/users (e.g., demographic information, driving competence, specific task competences). The data can be collected using interviews or questionnaires, and due to its nature, the data usually need to be protected (due to privacy issues). As such, only parts of the data might be needed to be stored in the project's database.
- In-vehicle Data Acquisition Systems (In-vehicle DAS): Include systems (sensors) integrated in vehicles and systems attached or mounted on the vehicles, such as video cameras or data storage tools. The FESTA Handbook considers in-vehicle DAS as one acquisition method and "nomadic devices" as another acquisition method. Nomadic devices are defined as external objects purchased from the marked and installed in a vehicle. As such, these can be considered like the systems attached on the vehicles, as defined for the in-vehicle DAS. To avoid confusion between the terms, the nomadic devices are considered included in this category of in-vehicle DAS.
- Subjective data acquisition: Indicates all the data collected from study participants about for instance attitudes, personal evaluations, or opinions. Such data can thus be collected using interviews, questionnaires and/or focus groups, and usually need to be protected due to privacy issues.
- Real-time observations: Indicate all the data collected by one main observer either in real time or after an event has passed, e.g., using a recorded video. The data can be quantitative or qualitative (e.g., KPI-S06 which requires the observed behaviour of other road users). In both cases, i.e., either the data is quantitative or qualitative, the data should be saved as a digital format in the project's database.

Infrastructure data acquisition: Includes devices or sensors that are usually placed close to a road (e.g., on a road in which the study will take place). On-road devices or sensors are often used to collect data about the weather conditions that could be complementary to the data collected from in-vehicles sensors. In such cases, the FESTA Handbook recommends synchronising the data collection from

both in-vehicle DAS and infrastructure data acquisition systems (e.g., using GPS time as recommended by FESTA).

8 Annex II: Metadata example - on testing mobile connectivity when crossing borders

Below is given an example of a metadata YAML file, that is uploaded to the MODI Data Sharing Platform; it can be found on the platform under Search: https://modi-data.github.io/search/

Figure 14 shows how the metadata is (partially) represented on the MODI data sharing platform.

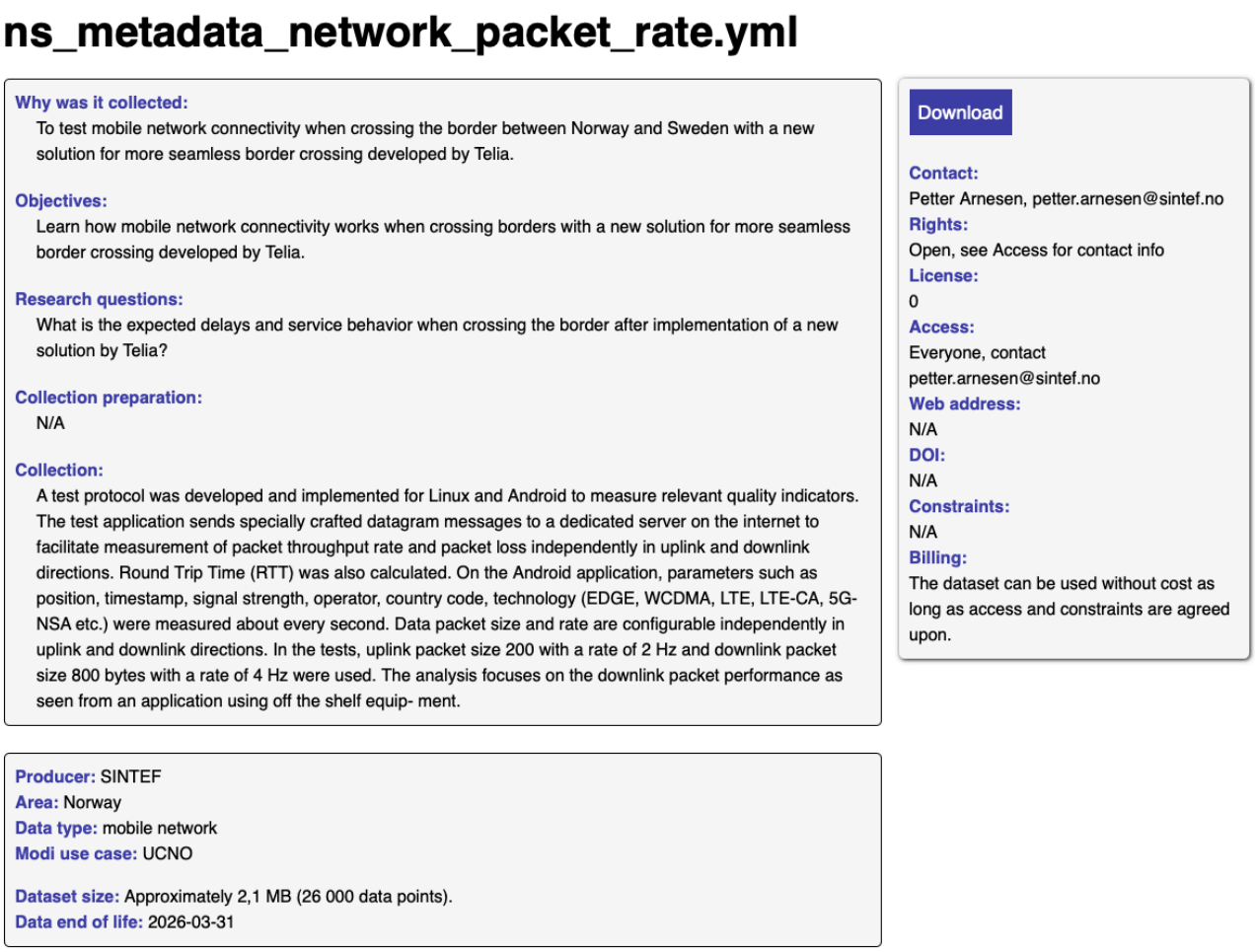


Figure 14: Partial representation of the metadata on MODI Data Sharing Platform. With the Download button in the top right-hand corner, the rest of the metadata can be accessed for further inspection

Below is the actual metadata that can be accessed when using the blue 'Download' button on the top right-hand corner (indicated in Figure 14 as well).

This metadata format is based on the FOT-Net Data Sharing Framework (DSF).

The DSF can be consulted if anything is unclear; each section in the format is directly

or indirectly based on a section in the DSF document.

https://www.connectedautomateddriving.eu/wp-content/uploads/2021/09/Data-Sharing-Framework-v1.1-final.pdf

```
# The summary should contain a thorough description of the study design and execution.
# The description must be complete and self-contained, but can contain links
# to images and further information (but keep in mind that the description
# must make sense even if those links should stop working.)
# Multiline strings can be written by starting the value with a pipe symbol (|).
summary:
  why_was_it_collected: |
    To test mobile network connectivity when crossing the border between Norway and Sweden with a
    new solution for more seamless border crossing developed by Telia.

  how_was_the_collection_executed: |
    Field test trial collecting several relevant parameters related to connectivity in
    multiple runs across the border.

  objectives: |
    Learn how mobile network connectivity works when crossing borders with a new solution for more
seamless border crossing developed by Telia.

  research_questions: |
    What are the expected delays and service behavior when crossing the border after implementation of a
new solution by Telia?

  experimental_plan: |
    Several trips with several smart phones each trip was travelled, resulting in 40 reg-
    istered trips across the Swedish Norwegian border at Svinesund in both directions. Both
    high-end Samsung S21 smart phones with 5G connectivity and low-cost Motorola E6
    (without 5G) smart phones where used, with SIM cards from all three national operators
    in Norway (Telia, Ice and Telenor) used as well. The new solution was only implemented for
    Telia, so Ice and Telenor are only used as a baseline.

  sample_selection_criteria: N/A
```

A person that can be contacted for more information about this dataset.

contact_person: Petter Arnesen, petter.arnesen@sintef.no

A web address to download the data if applicable.

web_address: N/A

Which entity produced this dataset

producer_of_data: SINTEF

Data set DOI if applicable

DOI: N/A

Which location are the data originating from. Can be an exact coordinate, a polygon, a town name, a country name, etc.

If providing coordinates/polygons, it must be encoded as WKT with an SRID of 4326.

data_area: Norway

Which type of data this is.

Must be one or more of:

survey, interviews, mobile network, itsg5, lane markings, signage, physical road infrastructure,

speed/acceleration, video, pictures, gnss, lidar, radar, weather, vehicle probe data, events, other.

You may add multiple data types if relevant, separated by a comma.

data_type: mobile network

Which of the use cases in MODI this dataset is related to.

Must be one of: UCNL, UCGE, UCSE, UCNO, UCCCAM, Other tasks

modi_use_case: UCNO

An approximate description of the size of this dataset

dataset_size: Approximately 2,1 MB (26 000 data points).

A tiny sample of the dataset to show how it looks (if possible).

```
  dataset_example: |
    2024-10-02 09:17:15.555;DeviceName: sam-s21-telia
    2024-10-02 09:17:15.557;Version: 4 debug 2023-07-02 10:52:15 UTC
    2024-10-02 09:17:15.558;RateParameters: 2 200 4 800 (NB: not always included)
    2024-10-02 09:17:15.896;11.2018029;59.1793564;1;  0.0;      0;0;  0.0;      0;0.000
    2024-10-02 09:17:16.911;11.2018029;59.1793564;3;100.0;   3090;4;100.0;  31159;0.083
    2024-10-02 09:17:17.926;11.2018029;59.1793564;5;100.0;   3130;9;100.0;  30346;0.085

# 5.3.3 Administrative metadata
# "Administrative metadata are collected for the effective operation and management of data
# storage and catalogues. This administrative information, covering various topics, is stored
# along with the datasets. From a FOT data re-use perspective, the key role of administrative
# metadata is to cover access conditions, rights, ownership and constraints."
administrative:
  # The version number of this dataset, used to keep track of different
  # versions if it is uploaded multiple times.
  # The version number can be numeric, date-based, or anything else, but
  # it must be easy to see which version is newer without further knowledge.
  version_number: 1

  # The date this dataset was uploaded. Must be in ISO 8601 format (YYYY-MM-DD).
  archiving_date: 2025-01-02

  # A unique identifier for this dataset. This will be used to identify this and
  # future updates in the metadata database, and must never be changed after the
  # dataset has been uploaded for the first time.
  unique_dataset_id: modi_uc_no_border_crossing_network_packet_rate_new_solution

  # Who has rights to use this dataset?
  rights: Open, see Access for contact info

  # Does the dataset have a specific license for usage?
```

license: No

Who can access the dataset?

access: Everyone, contact petter.arnesen@sintef.no

Any constraints in usage of the dataset?

constraints: N/A

Can the dataset be used for free, or does usage incur any costs/billing for the user?

billing: The dataset can be used without cost as long as access and constraints are agreed upon.

When the data set will stop existing (for example due to privacy regulations).

Must be in ISO 8601 format (YYYY-MM-DD).

data_end_of_life: 2026-03-31

A more detailed description of the various processes of the data collection,

for example the methods/tools used to collect the data, filtering, post-processing,

storage file structure, etc.

This should include information about relevant conditions during the data collection

(for example weather, climate, time of day, etc).

Each key under "processes" can be custom made, and there can be as many as is

necessary. The proposed keys are just examples.

processes:

collection: |

A test protocol was developed and implemented for Linux and Android to measure relevant quality indicators. The test application sends specially crafted datagram messages to a dedicated server on the internet to facilitate measurement of packet throughput rate and packet loss independently in uplink and downlink directions. Round Trip Time (RTT) was also calculated. On the Android application, parameters such as position, timestamp, signal strength, operator, country code, technology (EDGE, WCDMA, LTE, LTE-CA, 5G-NSA etc.) were measured about every second. Data packet size and rate are configurable independently in uplink and downlink directions. In the tests, uplink packet size 200 with a rate of 2 Hz

```
    and downlink packet size 800 bytes with a rate of 4 Hz were used. The analysis focuses on
    the downlink packet performance as seen from an application using off the shelf equip-
    ment.
  filtering: |
    This data set is the raw data, and has no filters.
  aggregation: |
    This data set is the raw data and has no aggregations.

# 5.3.2 Structural metadata
# "Structural metadata are used to describe how the data are structured in relation to other
# data. Data are organized into a system (e.g., a database and/or file system), a structure or
# database schema and a data content format. The aim of structural metadata is to facilitate
# the initial phase of data re-use by providing the necessary documentation about how the data
# is organized. The description should include the file system, the file structure and how to
# interpret the contents of a data container. All components of the dataset need to be
# described."
structural:
  summary: |
    Each trip was recorded as a separate CSV file. File with the corresponding id in
    'modi_uc_no_border_crossing_service_info_new_solution' was recorded on the same trip.

  # File format. Please include attributes such as delimiter for CSV, decimal separator,
  # thousand separator (if any), and any other properties that is relevant for parsing the data.
  file_format: csv with ';' as delimiter, ',' as decimal separator.

  # Please describe the storage structure of the data. If it consists of flat files, the folder
  # structure and file naming can be relevant. If it is a database, a description of the relevant
  # tables, indices, triggers and views is important.
  file_structure: |
    Flat files in a single folder, file names consists of [ID]_[date-time start of trip]-network-service-info.txt.
    Headers are not included in the files, but the fields described below describes them in the correct order.
    The first 2-3 rows are metadata for the file and does not follow the same format as the rest.
```

```
  # Which tools can be used to read the data. Especially important for non-standard formats.
  required_tools: notepad, various CSV parsing libraries

  # If the files were made using a specific tool, the tool name and version may be relevant
  # for reading the data.
  tool_versions: N/A

  # Additional keys and values can be added here if this is a non-standard format that requires
  # a more detailed explanation.

# 5.3.1 Descriptive metadata
# "Descriptive metadata shall include detailed information needed to understand each part of a
# dataset. The purpose is to describe the dataset and build trust in it—by providing not only the
# characteristics of each measure or component, but also information about how the data were
# generated and collected."
descriptive:
  # The descriptions can vary by data type.
  # For each data field in your dataset, please fill out all relevant parameters from section
  # 5.3.1 of the DSF, and enter the attributes using the "Data description item" in lower-case
  # as the key. The following list lists the most relevant attributes from the tables, see
  # tables 3 to 9 in the DSF for full explanations.

  # - description
  # - data_precision
  # - unit
  # - sample_rate
  # - filter
  # - origin
  # - bias
  # - type
  # - definition
```

```
# - range
# - error_codes
# - quality
# - offset
# - enumeration_specification
# - availability
# - srid (for coordinates)
# - time_zone (for time stamps)
# - time_format (for time stamps)

# Please note the following additions to the DSF:
# - When dealing with coordinates, the SRID must be specified with the srid key.
# - When dealing with times, the time zone must be specified with the timezone key.
# - When dealing with dates or times, the format must be specified with the format key.

fields:
  timestamp:
    description: Timestamp in format yyyy-mm-dd hh:mm:ss.fff UTC.

    timezone: UTC
    data_precision: 10
    unit: s
    sample_rate: 1hz
    origin: The data collection app
    type: datetime
    range: 2023-06-29 - 2023-07-27
    quality: High

  latitude:
    description: The GNSS latitude from the data collection app.

    srid: 4326
```

data_precision: 8

unit: degrees

sample_rate: 1hz

origin: The data collection app.

type: float

range: 0-180

error_codes: Repeated value from the previous row means that there was no new position for this row.

quality: Usually within 3 meters.

longitude:

description: The GNSS latitude from the data collection app.

srid: 4326

data_precision: 8

unit: degrees

sample_rate: 1hz

origin: The data collection app.

type: float

range: 0-180

error_codes: Repeated value from the previous row means that there was no new position for this row.

quality: Usually within 3 meters.

number_of_packets_sent:

description: The number of packets sent since start of trip.

sample_rate: 1hz

origin: The data collection app.

type: float

quality: High

percent_of_packets_received_by_server:

description: The percent of packets received by the server.

unit: percent

sample_rate: 1hz

origin: The data collection app.

type: float

quality: High

effective_bitrate_measured_by_the_server:

description: |

The effective bitrate measured by the server.

sample_rate: 1hz

origin: The data collection app.

type: float

quality: High

number_of_packets_received_from_server_by_app:

description: The number of packets received from server by app since start of trip.

sample_rate: 1hz

origin: The data collection app.

type: float

percent_of_packets_received_by_app_from_server:

description: |

The percent of packets received by app from server.

unit: percent

sample_rate: 1hz

origin: The data collection app.

type: float

quality: High

effective_bitrate_measured_by_the_app:

description: The effective bitrate measured by the app.

sample_rate: 1hz

origin: The data collection app.

type: float

quality: High

round_trip_time:

description: Round trip time, time to send a packet from app, process it on the server, and then send it back to the app.

sample_rate: 1hz

origin: The data collection app.

type: float

quality: High