MOBIDATALAB

Labs for prototyping future mobility data sharing solutions in the cloud

D2.10 Use cases definition (V2)

26/04/2023

Author(s): Johannes LAUER (HERE), Thierry CHEVALLIER (AKKODIS), Chiara RENSO (CNR), Francesco LETTICH (CNR), Victor LEPAGE (HOVE), Hiba MECHAYAKHA (HOVE), Renée OBREGON GONZALEZ (AKKODIS)

MobiDataLab is funded by the EU under the H2020 Research and Innovation Programme (grant agreement No 101006879).

Summary sheet

Deliverable Number	D2.10
Deliverable Name	D2.10 – Use cases definition (V2)
Full Project Title	MobiDataLab, Labs for prototyping future Mobility Data sharing cloud solutions
Responsible Author(s)	Johannes Lauer (HERE)
Contributing Partner(s)	AKKODIS, CNR, HOVE, HERE
Peer Review	CNR, AETHON
Contractual Delivery Date	31-03-2023
Actual Delivery Date	29-03-2023
Status	Final
Dissemination level	Public
Version	V1.0
No. of Pages	47
WP/Task related to the deliverable	WP2/T2.6
WP/Task responsible	HERE
Document ID	MobiDataLab- D2.10 – Use cases definition (V2)
Abstract	This deliverable is a report to provide an extended overview of the Task 2.6, which consists of an extension of the version one use case descriptions. It is based on the basic use-cases described in V1, enriched with the results of discussions with the reference group, end users and pre-living lab participants, research and exchange with partner projects and the experience gained by the practical implementation. Exemplary use-cases are described in more detail with the used data sets in an end-to-end flow. This will provide triggers for additional ideas for the living and virtual labs and for further discussions within the reference group and external stakeholders.

Funded by the European Union

Legal Disclaimer

MOBIDATALAB (Grant Agreement No 101006879) is a Research and Innovation Actions project funded by the EU Framework Programme for Research and Innovation Horizon 2020. This document contains information on MOBIDATALAB core activities, findings, and outcomes. The content of this publication is the sole responsibility of the MOBIDATALAB consortium and cannot be considered to reflect the views of the European Commission.

Project partners

Organisation	Country	Abbreviation
AKKODIS	France	AKKODIS
CONSORZIO INTERUNIVERSITARIO PER L'OTTIMIZZAZIONE E LA RICERCA OPERATIVA	Italy	ICOOR
AETHON SYMVOULI MICHANIKI MONOPROSOPI IKE	Greece	AETHON
CONSIGLIO NAZIONALE DELLE RICERCHE	Italy	CNR
HOVE	France	HOVE
HERE GLOBAL B.V.	Netherlands	HERE
KATHOLIEKE UNIVERSITEIT LEUVEN	Belgium	KUL
UNIVERSITAT ROVIRA I VIRGILI	Spain	URV
POLIS - PROMOTION OF OPERATIONAL LINKS WITH INTEGRATED SERVICES	Belgium	POLIS
F6S NETWORK IRELAND LIMITED	Ireland	F6S

Document history

Version	Date	Organisation	Main area of changes	Comments
0.1	23.01.2023	HERE	All	First draft and Integration of CNR content
0.2	23.02.2023	HERE	All	Integration of HOVE content
0.3	09.03.2023	HERE	All	Added further details on Use Cases
0.4	14.03.2023	AKKODIS	All	Added use cases for the reference group
0.5	17.03.2023	HERE	All	Accepted comments
0.6	21.03.2023	HERE	All	Integrated/updated CNR content
0.7	25.03.2023	CNR + AETHON	All	Peer-Review
0.8	25.03.2023	HERE	All	Consolidation
0.9	29.03.2023	HERE + AKKODIS	All	Quality Check
1.0	29.03.2023	AKKODIS	All	Submission of Final version

Executive Summary

The deliverable D2.10 is a report providing and overview on the version 2 of the use cases. The listed use cases are an extension of the version 1 use cases. Within this document, use cases provided by the reference group are described including potential data sets and their stakeholder group. Parts of the generic use cases from the version 1 D2.9 are implemented using real world data and parts of the now existing transport cloud. Furthermore, a pre-Datathon event provided further ideas for use-cases mainly on spatial analytics and planning.

The extended set of use cases will be a further base for the living and virtual lab challenges. It will further extend the data listings and influence standard discussions and alignments. Furthermore, the described use cases will provide opportunities to connect individual stakeholders within the mobility ecosystem to combine efforts and learn from each other.

Table of contents

1. INTRODUCTION 8
2. USE CASES V1 9
3. USE CASES V2 10
3.1. PRE-DATATHON USE CASES 10
3.1.1. Educational Gap – Accessibility of universities in Germany 11
3.1.1.1. Input Data 11
3.1.1.2. Solution Architecture 11
3.1.1.3. Input Data 12
3.1.1.4. Solution Architecture 13
3.1.2. Analysis of potential school locations in Lower Saxony and Bremen 13
3.1.2.1. Input Data 13
3.1.2.2. Solution Architecture 14
3.1.3. Pre-Datathon Conclusions 14
3.2. REFERENCE GROUP USE CASES 15
3.2.1. City of Eindhoven 15
3.2.1.1. Emission Reporting and Routing 15
3.2.1.2. Urban Planning (Journey planning) 16
3.2.1.3. Input Data 17
3.2.1.4. Use case description 18
3.2.2. City of Leuven 19
3.2.2.1. Modal Shift with focus on micro-mobility (Journey Planning) 19
3.2.2.2. Bicycle lanes infrastructure (Analytics and learning) 20
3.2.2.3. Conflicts with Public Transport (Analytics & Learning) 20
3.2.2.4. Input Data 20
3.2.3. Milano AMAT (Agenzia Mobilità Ambiente e Territorio) 21
3.2.3.1. Quality assessment of published data 21
3.2.3.2. Data accessibility and interoperability 22
3.2.3.3. Emission Reporting 23
3.2.3.4. Accessibility (or Transport network adaptation) to large events 23
3.3. PROJECT GROUP USE CASES 23
3.3.1. Journey Planning and MaaS (Hove) 23
3.3.1.1. Input data 24
3.3.1.2. Solution architecture 26
3.3.1.3. Pipeline of data preparation and ingestion 27
3.3.1.4. Journey planning 28

Funded by the European Union

3.3.1.5. Journey planner statistics 29
3.3.2. Transport data sharing within the Linked Open Data vision (CNR)........ 29
3.3.2.1. Preliminary concepts 30
3.3.2.2. Definition of the use case study........ 31
3.3.2.3. Definition of the use case activities (with running example) 32
3.3.2.4. Discover datasets and services 33
3.3.2.5. Provide Datasets 34
3.3.2.6. Combine Datasets........ 35
3.3.2.7. Perform Data Analysis........ 36
3.3.2.8. Datasets and data sources of interest 37
3.3.2.9. Standards........ 38
4. CONCLUSIONS 39
5. REFERENCES........ 40
6. ANNEXES 41
6.1 Transport data sharing within the Linked Open Data vision : analyses conducted within the 'Tourism analysis' activity 41

List of figures

Figure 1: Population analysis (source: THWS student Sebastian Gut) 11
Figure 2 : Interactive result map - white spots are cities/regions, where travel time to the next university is > 60 min (source: THWS student Sebastian Gut) 12
Figure 3: Visualisation of population density, road network and existing hospitals 13
Figure 4: Thiessen polygons to calculate the catchment area for potential school locations (ref. THWS student Stefan Schönebeck)........ 14
Figure 5: Data API for NO2 measurements in Eindhoven 16
Figure 6: Use-case diagram - Urban Planning 18
Figure 7: Use-case diagram - Quality assessment of published data 21
Figure 8: Use-case diagram – Data accessibility and interoperability 22
Figure 9: Hove's use case solution generic architecture........ 26
Figure 10: Data processing pipeline before ingestion in Navitia 27
Figure 11: Navitia response to a journey request from Stradivari (Rome) to Vivaldi (Rome) 28
Figure 12: isochrone representing accessible places/ zones per duration interval from the Colosseum via Rome's public transportation........ 29
Figure 13: Use case diagram - Transport data sharing within the Linked Open Data vision 32
Figure 14: plot of the semantically enriched trajectory with ID 2615 associated with the user having ID 2115. The red dots represent the occasional stops that have been found, while the blue curve represents the moves. 41
Figure 15: SPARQL query retrieving the move occurrences associated with the semantically enriched trajectory having ID 2652........ 42
Figure 16: Moves found for the trajectory with ID 2652 (screenshot from GraphDB). 43
Figure 17: SPARQL query retrieving the occasional stop occurrences associated with the semantically enriched trajectory having ID 2652. 44

Funded by the European Union

Figure 18: SPARQL fragment retrieving the weather conditions aspect that has been associated with the semanticallyenriched trajectory having ID 2652. 45
Figure 19: SPARQL fragment augmenting the information concerning the returned POIs via the use of a linked open data source. 45

List of tables

Table 1 : Italian geographic and public transport data 25

Abbreviations and acronyms

Abbreviation	Meaning
URV	Universitat Rovira I Virgili
GIS	Geographic Information System
THWS	Technische Hochschule Würzburg Schweinfurt
GDPR	General Data Protection Regulation
X-athon	Datathon, Hackathon and Codagon (i.e. MobiDataLab Living Labs)

Funded by the European Union

1. Introduction

The use cases are a core part of the MobiDataLab strategy. They provide the base for the Transport Cloud requirements, the base for the Virtual and Living Lab challenges and they trigger activities on standardization initiatives. Further, they will be able to show a practical influence on realistic scenarios, supporting the enablement of data sharing stakeholders and increase transparency of potential incentives for data sharers.

The Use Cases Version 1 document, from the very early beginning of MobiDataLab, provided a basic set of use-cases. They were technically described by the core project team and are providing generic "place holders" for typical real-world scenarios.

Within the Use Cases Version 2 document, the use-case provider group includes the reference group and with them, typical, current real world scenarios. Further, the ideas of a Pre-Datathon have been included in the documents, which has been taken place at the THWS Würzburg. The experience of data reviews, analysis and the growing data availability since the project did start, supports a more specific use case setup.

Within this document, the use cases are described in different ways, based on their implementation phase and the providing stakeholders. Use cases from the core project team have been implemented and realized and are described in detail. Use cases from the reference group members are building on top of the version 1 use cases or bringing most recent activities as new use cases into the perspective of the MobiDataLab project. A set of practical implemented use-cases within a pre-Datathon with a student group from the Technical University of applied science in Wuerzburg is showing another potential usage of shared mobility data, provides further triggers/ideas for additional scenarios and shows possibilities on how shared mobility data can easily provide value.

Funded by the European Union

2. Use Cases V1

The Use Cases V1, described in deliverable 2.9, in the early beginning of the project, are providing a set of very generic project group focused use-cases.

The intention of the V1 use-cases is to define the initial requirements for the transport cloud, the data catalogues and the services.

Furthermore, this initial version supported communication and brainstorming with the reference group on real world use cases and contributes to challenges setup for the living- and virtual labs.

To connect the V1 use cases with the V2, a list is provided below, referring to D2.9 for further details.

1. Use cases for operations
    a. Optimisation of transport flow and estimated time of arrival (ETA)
    b. Emission reporting
    c. Analytics & Learning
    d. Re-use of transport data for journey planners / digital services
    e. Mobility as a services (MAAS)
2. Use cases for research
    a. OpenStreetMap for inclusive transport
    b. Geodata sharing applied to transport: Environmental data for sustainable transport
    c. Transport data sharing within the linked open data vision

3. Use Cases V2

Use Cases V2 extend and further specify V1 use cases. With the support of external stakeholders, version 2 of the use cases is being designed. Furthermore, a set of V1 use cases is being specified in more detail in order to have available some examples, demos and challenges for the upcoming x-athons in WP 5. Some of the use-cases are already up and running examples (e.g. MaaS by Hove), others are in a conceptual phase and enable further thinking on the requirements, important/required data and potential realisations.

3.1. Pre-Datathon Use Cases

In November 2022, HERE organized a Pre-Datathon for the students of THWS. The scope of this workshop was to retrieve initial feedback on the MobiDataLab tools as well as to stimulate ideation and co-creation activities on additional use cases with the purpose of extending v1 use cases. After an introduction on the MobiDataLab project, the available tools and services within the project context were demonstrated. Guided by the D2.9 Use-cases V1 some examples for mobility data sharing were given.

The workshop group started a two-hour project, composed of the following steps:

- Brainstorming
- Exploration of available data
- Managing the available tools
- Building a story
- Implementing and executing analysis
- Creating a presentation
- Giving a 3-minute pitch to the audience

With a first “minimum viable pitch result”, the workshop attendees presented the early outcomes. Within the semester, the student groups refined their projects and created reports with ESRI ArcGIS story maps on each topic.

The data search, exploration and investigation were done with the support of the MobiDataLab transport cloud. Especially the prepared metadata catalogues, within GeoNetwork and CKAN, have been beneficial to find and explore available data sets.

As result of the GIS analysis, ArcGIS StoryMaps were created. StoryMaps are a compilation of static and interactive geospatial content and narrative, textual descriptions. They are used to share maps as an interactive web page.

Funded by the European Union

3.1.1. Educational Gap – Accessibility of universities in Germany

This use case analyses the spatial accessibility to universities in Germany. With the "Central Place Theory" (Christaller,W., 1933), a tool for analysis of infrastructure supply exists. The used datasets have been loaded via the ArcGIS Online Portal, which is a proprietary catalogue with collected open data resources.

3.1.1.1. Input Data

- Universities in Germany
- Administrative Boundaries of Germany
- Population Density of Germany
- Cities in Germany and population
- Road network

3.1.1.2. Solution Architecture

The solution has been built using integrated algorithms within the ArcGIS toolbox. The calculation of the accessibility was done using Thiessen polygons and a more sophisticated Network analysis using the road network dataset. The overall presentation has been done via ArcGIS Story Maps, which provided an easy user interface for consumers including interactive components to review contributed maps.

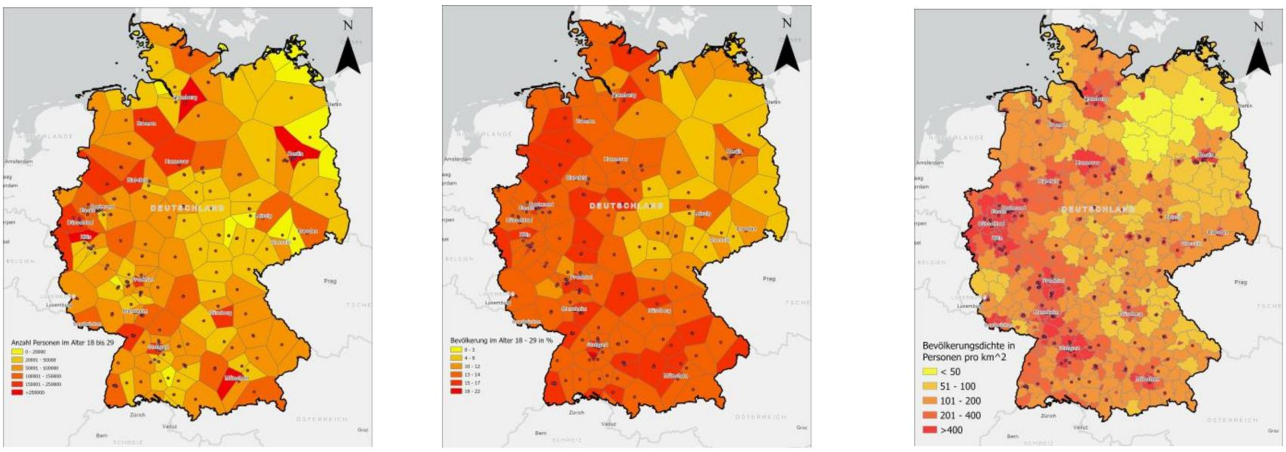

Figure 1: Population analysis (source: THWS student Sebastian Gut)

The student group did choose a classical site analysis use-case.

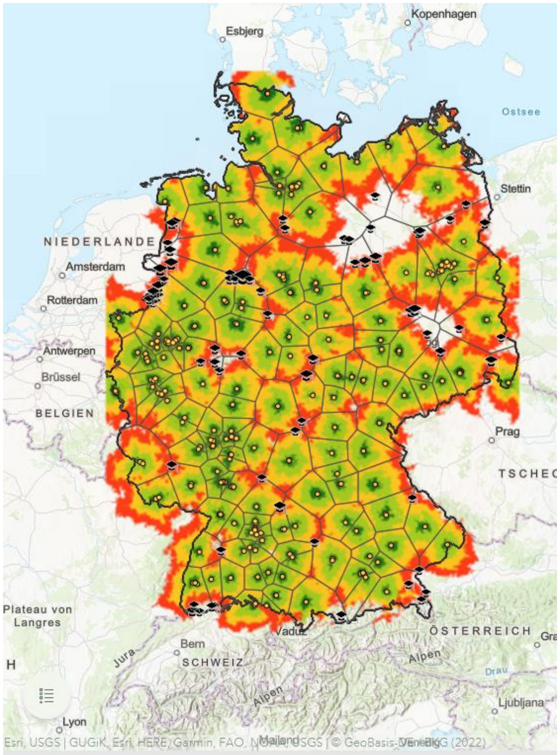

Figure 2 : Interactive result map - white spots are cities/regions, where travel time to the next university is > 60 min (source: THWS student Sebastian Gut)

Within the area of Würzburg, a location for a new hospital needs to be determined. The analysis has been made using demographic data and several map data sets. Influencing factors are population density, age distribution, existing hospitals, spatial distribution of sex and age (e.g. to provide related services, such as gynaecology, paediatrics, geriatrics, emergency medicine).

3.1.1.3. Input Data

- OpenStreetMap data of Lower Franconia: **Lower Franconia**
- Administrative boundaries Lower Saxony and Bremen
- Demographic data city of Wuerzburg (https://statistik.wuerzburg.de)

Funded by the European Union

3.1.1.4. Solution Architecture

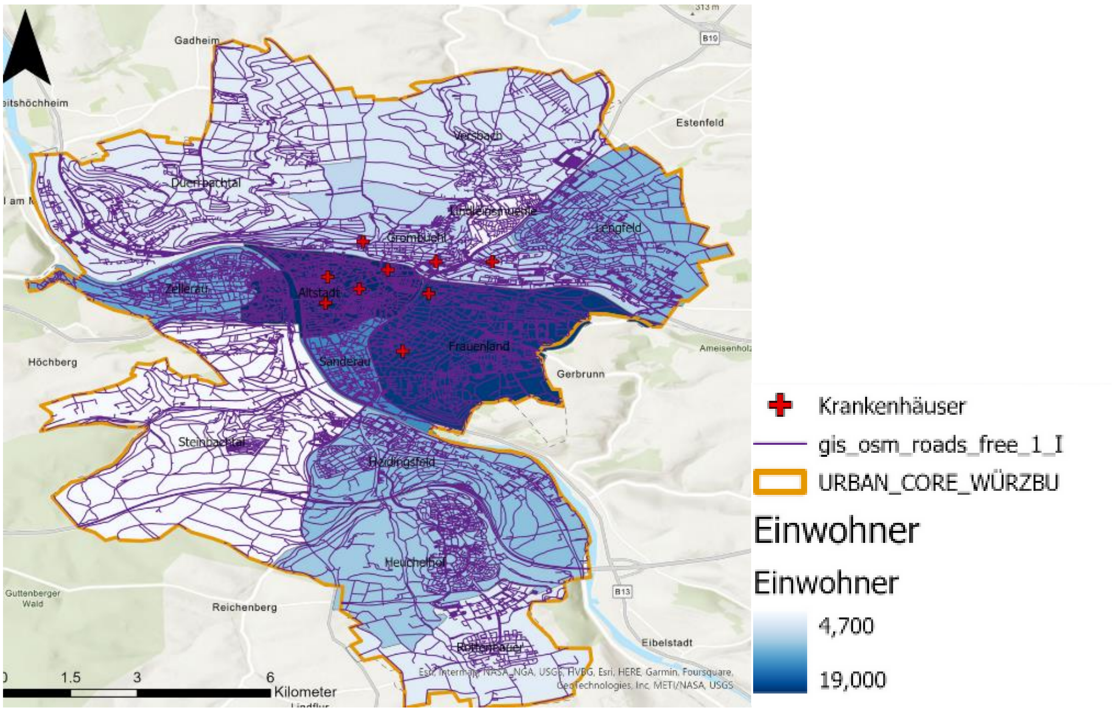

Figure 3: Visualisation of population density, road network and existing hospitals

Figure 3 shows input data sets combined in a map visualisation. The conclusion has been made by combining the datasets and using a visual reasoning approach to identify a potential place for a new hospital.

3.1.2. Analysis of potential school locations in Lower Saxony and Bremen

This use case reflects the governmental planning requirements for school locations. Population density and distribution underlies continuous changes. However, the government need to secure the availability of schools for the compulsory schooling. Shared mobility data can provide an overview on the current situation and help to predict the future need and support the infrastructural planning process accordingly.

3.1.2.1. Input Data

- OpenStreetMap data for Lower Saxony and Bremen
- Population data by ArcGIS Insights (ref: https://doc.arcgis.com/en/insights/latest/analyze/enrich-data.htm)

Funded by the European Union

3.1.2.2. Solution Architecture

The input data sets have been analysed using the ArcGIS tool stack. The calculation of the catchment areas has been made by using Thiessen polygons (cf.: Figure 4). The ArcGIS “Enrich” tool augmented the spatial data with the population statistics. With the “Optimized Hot Spot Analysis”, a proprietary ESRI tool has been used to identify the regions having high population density of school-aged children and teenagers. Finally, a network analysis determined potential locations for new schools.

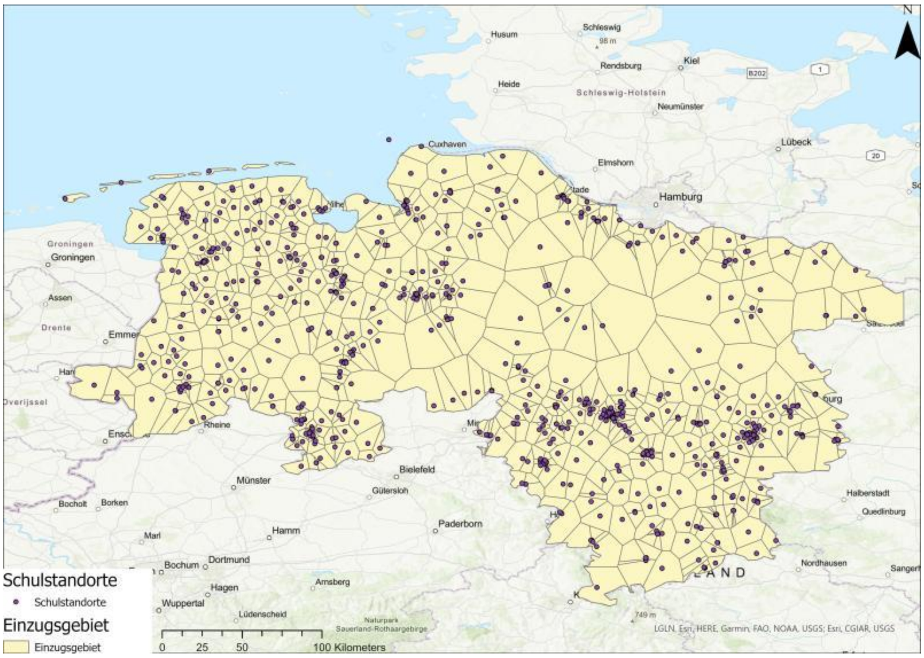

Figure 4: Thiessen polygons to calculate the catchment area for potential school locations (ref. THWS student Stefan Schönebeck)

3.1.3. Pre-Datathon Conclusions

The Pre-Datathon did provide many insights into practical usage of data and the transport cloud with the data repositories. The students did make use of the transport cloud and the data sets within the workshop and used their research skill set to get appropriate data sets to solve their challenges. This resulted in representable pitch presentations at the end of the workshop.

After the workshop, the students had 8-12 weeks time to review and enhance their projects for final submissions. Most of the students did make use of the proprietary open data catalogue, provided by ESRI via the ArcGIS open data data portal. This demonstrates on the one hand, that findability and accessibility is very important. Further, the integration within a software/analytics ecosystem is very helpful and lowers the barriers for users. On the other hand, it also demonstrates the requirement of

FAIR data access. Content is being provided via heterogeneous distribution channels, not necessarily via the ESRI proprietary ecosystem. Metadata, including license information is very restricted for some of the data sets and sources are not always transparent, which reduces the reliability and reusability of the data sets.

Overall the implemented use-cases show an additional perspective on practical usage of shared mobility data. They will help to trigger further thoughts and ideas within the living and virtual labs and demonstrates the adaption of requirements, that are also useful within more complex use-cases, e.g. where data from distributed providers need to be integrated, and/or further tools have to be used.

3.2. Reference Group Use Cases

3.2.1. City of Eindhoven

3.2.1.1. Emission Reporting and Routing

The city of Eindhoven has a Low Emission Zone (LEZ) in place and has already joined Netherland's scheme for a Zero Emission Zone (ZEZ) for logistics that will be in place from 2025. Eindhoven is keen to gather insights on the environmental impact of city's mobility and transport services.

The city of Eindhoven has a network of ~30 air quality detectors within the city. These are measuring NO2 concentration. Their data is available via an opendatasoft API (ref.: Figure 5, https://www.eindhoven.nl/stad-en-wonen/duurzaamheid/meetwaarden-luchtkwaliteit)

Funded by the
European Union

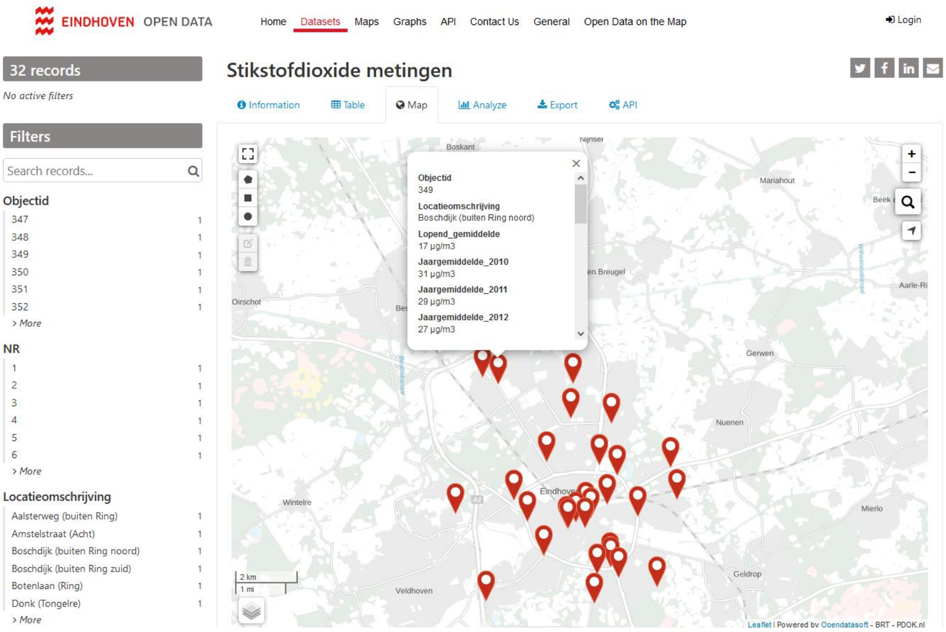

Figure 5: Data API for NO2 measurements in Eindhoven

Further data is available via the Regional Meetnet for Zuid-Oos Brabant, where PM1, PM2.5, PM10 and NO2 are being measured at ~50 locations (ref.: https://odzob.nl/meetnet). The metadata of both data sets have been integrated in the MobiDataLab Transport Cloud via the GeoNetwork instance.

Further, the city of Eindhoven did setup sensors for sound detection and cameras in the context of a research initiative at a traffic hot spot. Due to privacy and data security reasons, this data is not made publicly available.

These data sets can be a valuable source for the city of Eindhoven specific implementation of the Emission routing use case. Exported results could potentially be utilised by the city of Eindhoven during planning phases for comparing different planning choices.

The use case for data sharing is to explore where there is greatest potential for reducing city's emissions and improving the mobility concept of the city. Within this context, the compliance on the FAIR principle needs to be reviewed for the data sets. Besides the solution development, these activities will provide further insights and transparency on the data and supports the data sharing initiative.

3.2.1.2. Urban Planning (Journey planning)

A further use case has been explained within the workshop. Since the city of Eindhoven is, as many further European cities, rethinking its mobility concepts as part of the urban planning process, mobility data is playing an important role for the planning and future decisions.

Urban planning is a very important instrument to adjust the urban area based on current requirements and the requirements of the future. Within the last decades, cities, especially European cities, are shifting their mindset from a car-friendly city to a bicycle friendly city. Inhabitants see an increasing quality of life with less car traffic.

Furthermore, it will reduce emissions, such as noise, particulate matter and exhausts from combustion engines. However, the requirements of several stakeholders need to be reflected within this process, that the quality of life will not be reduced too much within the transition phase.

An urban planning process consists of manifold heterogeneous stakeholders and needs to meet many requirements of them. It is about involvement and discussions and communication. A large set of stakeholders having own, partly domain specific data sets, that are needed to contribute to a reliable and successful planning process.

3.2.1.3. Input Data

- Datasets from the city and the region
  - Cities
  - Spatial plans (IMRO)
  - Buildings (INSPIRE harmonised)
- Public transport data
  - Transport Networks: Railways - Transport Networks: Rail (INSPIRE harmonised)
  - Public Transport
  - Tracking/planning data from the construction company
  - DSI - General Construction Plans - Overprints

Funded by the European Union

3.2.1.4. Use case description

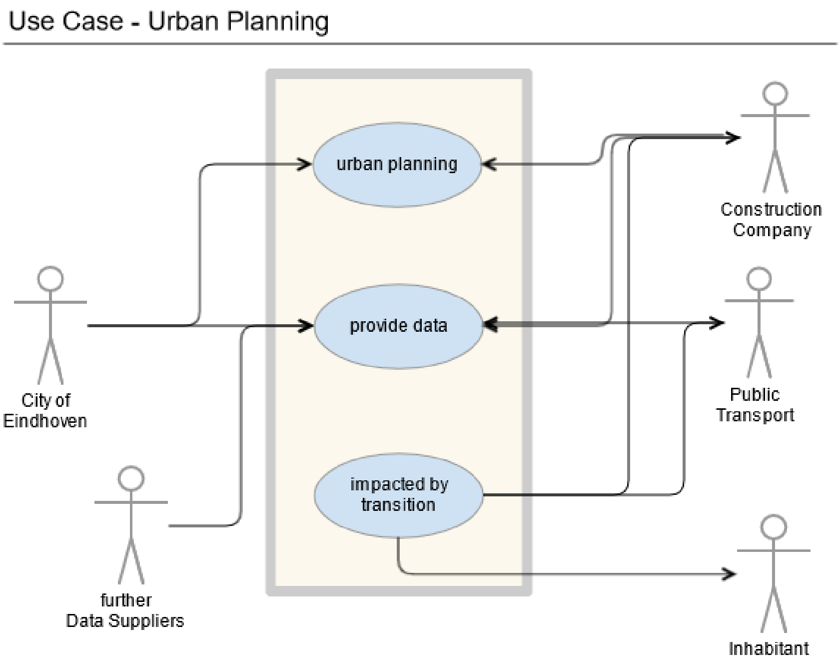

Figure 6: Use-case diagram - Urban Planning

Actors:

- The “City of Eindhoven” provides data sets and their metadata.
- “Further Data Suppliers” contributing content needed for the planning process
- The “Construction Company” is planning their logistics using the provided data and provides further data for other related stakeholders
- The “Public Transport” is impacted by the transition process and uses shared data sets to adjust routes and transportation
- Inhabitants are impacted by the transition by construction processes or traffic adjustments

Urban planning with focus on mobility and transportation features includes the two following techniques; identifying transportation patterns and analysing the impact of land use in order to imagine a completely car-free new neighbourhood.

The first step of this use case is to utilise journey planner services to assess the impact of creating new transport service lines at a neighbourhood scale. The second step of the use case is to further support urban planning decisions by proposing additional transportation-driven actions (bicycle stations, pedestrian routes etc.).

Funded by the European Union

3.2.2. City of Leuven

Several version 1 use-cases were identified to be valuable for implementation using local data from Leuven. This is based on an interview with stakeholders from the city of Leuven. The specific use cases identified for Leuven - aligned well overall with the use cases version 1 as they reflect local needs, and they can be adjusted to their specific context.

For instance, the city of Leuven wants to increase the attractivity of public transport usage by data based, reliable and user-friendly journey planning services. This need aligns with the journey planning use-case.

Additionally, data suppliers need to be identified to enable their contribution for an end-to-end system.

Adjustments and improvements to the bicycle related infrastructure are required to increase the attractivity also of shared (e-)bikes.

Another requested use case entails the data sharing between transportation stakeholders to enable more reliable and accurate ETAs. Since the request comes from the city and the needed data comes from further stakeholder, e.g. transportation agencies, there needs to be an incentive for data providers. Therefore, the benefit of sharing this data needs to be visible for the data suppliers in order to make it attractive.

The use-cases highly depend on available datasets from a heterogeneous set of stakeholders. Therefore, the open data policy of the city of Leuven is only one part of the overall strategy. By enabling further transportation chain and traffic stakeholders, more reliable and accurate ETAs can be feasible. This will also lead to a positive customer experience, save resources and results in a more sustainable traffic.

Positive examples from other regions can trigger this motivation to execute. In the context of the virtual- and living labs, stakeholders can investigate options for best practices to share data in a FAIR way. This will support solution finding for the given challenges.

3.2.2.1. Modal Shift with focus on micro-mobility (Journey Planning)

The modal shift is part of Leuven's long term policy plans and can be seen as a challenge on its own. Modal splits and monitoring of the modal split are not so easy. This translates in many projects and plans aiming to encourage more people to use public transportation, bicycles, as well as new mobility modes and/or to emphasize the combination of all of these. However, the amount of shared mobility services usage is not yet very significant.

The challenge in this case, is to study through data analysis, if there is a biggest potential for combination of public transport with shared mobility in Leuven. What is the potential for shared mobility and for which combination of modes? This potential can be studied through different perspectives.

3.2.2.2. Bicycle lanes infrastructure (Analytics and learning)

Each city has its own characteristics, for instance Leuven has a lot of students with high bike usage for local daily journeys, and most inhabitants have their own bikes. Therefore, the city of Leuven works on mobility plans regarding e.g. cycling lanes and cycling streets, but there is still room for improvement.

This use case aims to define the good criteria to decide where traffic lanes are needed (in combination with e.g. Housing, real estate, etc. using data enrichment tools). The aim is to provide a general framework helping decision making on very objective data.

3.2.2.3. Conflicts with Public Transport (Analytics & Learning)

Given data related to the areas of transport and logistics, smart parking zones for deliveries, etc. This use case aims to use mobility data in order to define what are the routes and locations where there is a potential of conflicts of pedestrians, bikes and/or light vehicles with larger vehicles.

3.2.2.4. Input Data

Input data sets can be loaded from the Belgium open data portal: https://opendata.infrabel.be/pages/home/ which is running within an open data soft instance and the corresponding API (https://opendata.infrabel.be/api/v2/console).

- List and geographical position of level crossings
- Parkings for coaches
- Accident RID (CSI)

Enabling data sharing for further stakeholders is a mandatory action that needs to happen to fulfil the requirements on the given use cases. Therefore, suppliers for public and private transport, individual bike sharing operators etc. need to be involved in the overall process.

3.2.3. Milano AMAT (Agenzia Mobilità Ambiente e Territorio)

3.2.3.1. Quality assessment of published data

The AMAT of Milano is providing several open data sets within their open data strategy. The initiative is driven and guided by the public administration. The data sets are being supplied by a heterogeneous set of data providers within the public sector. Besides following the FAIR principle, to make the data findable, accessible, interoperable and reusable, a further requirement is data quality of the individual data sets.

The reliability and the acceptance of the provided data sets depends strongly on the data quality and transparent metrics and measures.

- Input Data

The input data consists of the provided data by AMAT or a subset of the available data, that allows a classified applicable metric to describe the provided data. AMAT provides a CKAN instance which lists the available open data sets (https://dati.comune.milano.it/dataset).

An example for a contributing data set is referenced via the MobiDataLab CKAN meta data catalogue: Mobility: location of parking areas for Car Sharing (GuidaMI) . This data set provides location and number of places offered for car sharing parking areas.

- Use case description

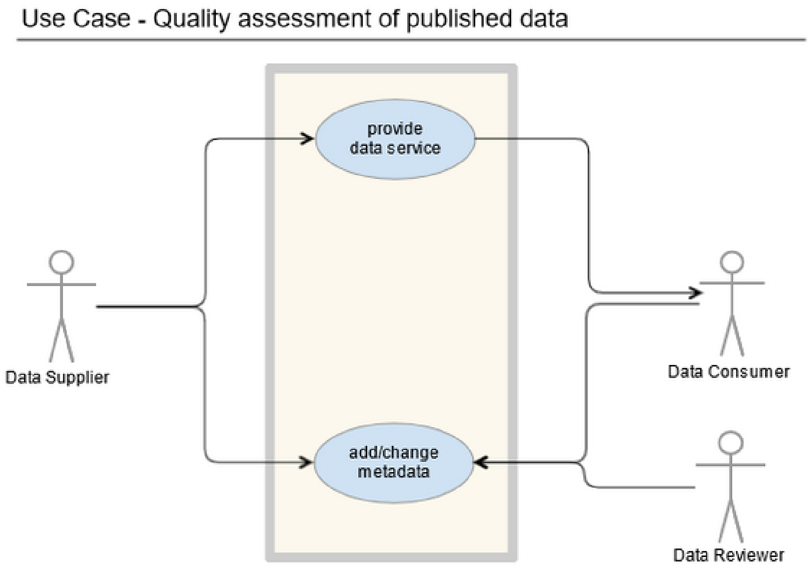

Figure 7: Use-case diagram - Quality assessment of published data

Actors:

- The "Data Supplier" provides the data sets and their metadata.
- The "Data Consumer" consumes the data and provides feedback, based on the use-case. This feedback can be integrated into the metadata.
- The "Data Reviewer" does not use the data, but analyses the data set to provide feedback, which will then be stored in the metadata.

3.2.3.2. Data accessibility and interoperability

The data sets provided via the open data portal are very heterogeneous and mostly not accessible via interoperable interfaces. This use-case describes the transition of data sets into a FAIR compliant ecosystem. The solution will enable the reusability of the data sets in different contexts.

- Input Data

Relevant input data for this use case is all available data which can be shared, and stakeholders have a benefit of using the data.

- Use Case Description

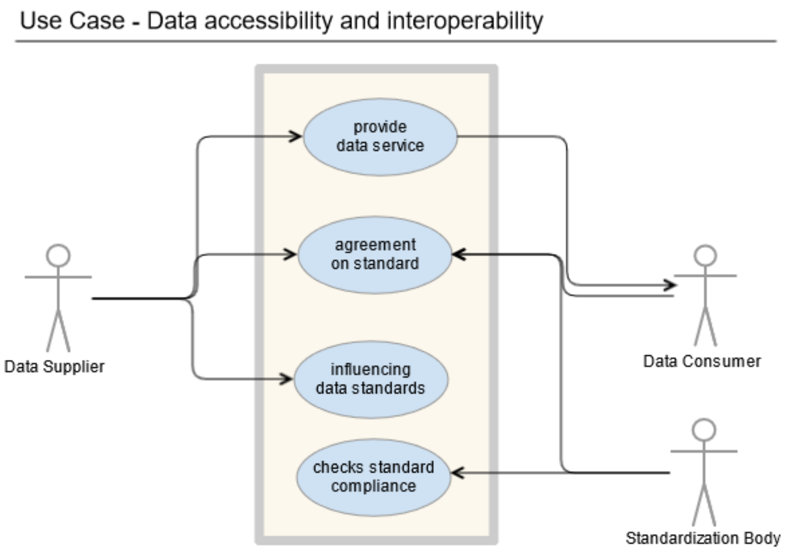

Figure 8: Use-case diagram – Data accessibility and interoperability

Actors:

- The "Data Supplier" provides data services. Standards are being influenced by the supplier and the supplier agrees with other stakeholders on the used standards
- "Data consumer" consumes the data and agrees on the used standards

- “Standardization Body” checks the standard compliance and supports the alignment between data suppliers and consumers.

3.2.3.3. Emission Reporting

Milan has a low emission zone called "area B" covering 90% of the city, and a calendar which will progressively ban certain categories of vehicles. The mobility agency of Milan (AMAT) is working on analysing the environmental impact of the LEZ thanks to e.g. air quality reports. The challenge is to further contribute to this analysis and 1) identify the zones where there is still high emission, 2) perform cross-domain analysis 3) make these reports available as a service and interoperable. Example of cross-domain analysis: comparing high emission areas with the position of locations particularly affected by high emission rates like e.g. kindergartens position.

3.2.3.4. Accessibility (or Transport network adaptation) to large events

The city of Milan regularly hosts large events (e.g. football matches, concerts, design week, fashion week, etc.) which often cause mobility-related problems like traffic congestion and have an environmental footprint (emission, poor air quality, etc.).

This use case aims to assess the most suitable adaptation of mobility choices during large events, by analysing the effects of the different possible measures, like e.g. new bus lines targeting the event or new shared mobility services around the venue, the extension of existing bus lines, etc.

- Service of Car Sharing, Bike Sharing, Scooter sharing
- Buses used for local public transport by emission class
- Pollutant monitoring stations
- Traffic volume

3.3. Project Group Use Cases

3.3.1. Journey Planning and MaaS (Hove)

The re-use of transport data is fundamental to journey planning. This involves leveraging existing transportation information to offer more effective and efficient journey planning options to users. For instance, real-time updates on public transportation schedules can be provided to users via APIs, which can greatly improve the accuracy of these solutions. Another example of a use case is the integration of traffic information in journey planning services, which can suggest alternative routes to users based on current traffic conditions. Additionally, some digital services can personalize the journey planning experience for users by offering recommendations for the quickest or most convenient modes of transportation to them specifically, which will lead to an overall better user experience.

Beyond their main purpose, journey planners could also contribute to other services:

- **Routing optimization** for public transportation systems by analysing data on ridership patterns, travel times, and other factors.
- **Improving accessibility** for public transportation systems: by analysing data on the location of stops, the availability of elevators and ramps in this case.
- **Assessing environmental impact** of public transportation using public data by analysing CO2 emissions using journey planner.

Previously in the use cases V1, we suggested creating a standard API for journey planning digital services based on Transmodal concepts and data structures. This API would allow digital service providers to access journey planning features in a standardized way, without having to handle raw datasets. We listed Navitia and Open Trip Planner (OTP) as open-source journey planners that could serve as a starting point for the standard API used eventually in the Transport cloud platform. For computing long journeys between different countries, interconnection between trip planners may be necessary and Open Journey Planning (OJP) could also be used to implement this.

In order to elaborate more on this use case, we will examine a scenario where a Transport cloud user wants to create a journey planner utilizing the data available in the platform to feed a Navitia API instance as an external calculator and explores the outcome. We will consider that all the pre-conditions detailed in the deliverable 2.9 Use cases V1, table 9 are valid and that all the needed data exists within the Transport cloud.

3.3.1.1. Input data

As a reminder, Navitia is an open-source, multi-modal public transportation routing engine and is frequently used in Mobility as a Service (MaaS) platforms to provide multi-modal transportation options to users. Its data and services can be integrated into MaaS platforms through its API, which enables developers to query Navitia's database and retrieve transit data in a standardized format. This data can then be used to power features such as journey planning, real-time transit updates, and route optimization.

This last offering multiple services dedicated to mobility such as:

- **Passenger Information API:** Offer intermodal routes to users, considering traffic information.
- **Search API:** Help users find their way by geolocating and displaying points of interests around them.

- **Autocomplete API:** Improve the user experience when entering searches with a autocomplete feature.
- **Isochrone API:** Offer the innovative search by travel time feature for the user. It calculates the accessibility of a geographical region on public transportation within a given time frame.

It can ingest a variety of data types, including:

1. **Public Transport Data**: GTFS (General Transit Feed Specification), NeTEx (Network Exchange) etc. as standardized formats for public transportation schedules and other related data.
   a. Public transport of Germany - GTFS
   b. Flixbus NeTEx
2. **Real-time data**: such as real-time vehicle location information or real-time disruption data. This information can be used to provide more accurate and up-to-date information to users.
   a. Real Time Traffic Information from VMS in Hellastron network
   b. Real-time timetables (GTFS-RT) of the Filbleu network
3. **Geospatial data**: geographical location of the transportation network, including information on the location of stops, the shape of routes, and other geographic features.

In this regard, we focused on the Italian (Milan and Rome) public transport Data from the reference group and geospatial data (extracted from OSM). These datasets were previously harvested and made available in the GeoNetwork Azure instance as shown in Table 1:

Data category	Dataset	Format	Description
Public transport data	Public transport data schedules from Rome, Milan	GTFS	- Rome's metro and bus networks' schedule - Milan's bus, tram and metro networks' schedule
Geographical data	Rome and Milan OSM data imported from GeoFabrik Italy's Center and North West partition files	OSM (.osm.pbf)	Addresses, Point of interests …
	Poly file imported from osm.boundaries or geofabric.downloads	. poly	Geographic boundaries including Rome's and Milan's

Table 1 : Italian geographic and public transport data

3.3.1.2. Solution architecture

In this scenario's solution architecture (cf.: Figure 9), the Journey Planning Service is the main component that provides the journey planning functionality. It interacts with the Navitia API to obtain data about public transit routes, schedules, and other transportation-related information. This service can be implemented using any programming language or framework that supports RESTful API integration.

The Data Ingestion pipeline is responsible for extracting data from the Transport cloud and transforming it into a format that the Journey Planning Service can use. In this use case, the data is sourced from OSM .pbf files and GTFS. This step can be implemented using an Extract, Transform, Load (ETL) pipeline that extracts the data from the OSM .pbf files, processes and transforms it, and loads it into a database or other storage system that the Journey Planning Service can access.

Finally, the architecture includes other services such as a User Interface or Backend Services. They can be developed using any technology that is appropriate for their specific requirements.

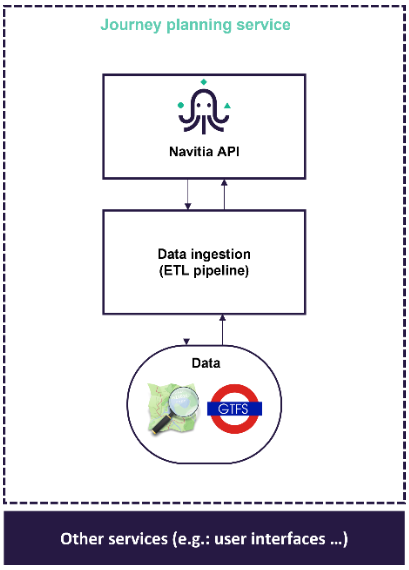

Figure 9: Hove's use case solution generic architecture

Navitia can be hosted on-premises or on a cloud provider such as Docker, AWS, and Azure. The Navitia documentation and community forums walk through the installation and deployment in more detail.

In general:

- **Data sources configuration:** Data sources need to be configured before they can be ingested. This includes specifying their type, name, URL, and other configuration details.
- **Coverage configuration:** A coverage needs to be defined before data can be ingested. A coverage represents a geographic area and includes information about the transit agencies and networks that operate within that area. It can be defined using a dedicated API, which provides endpoints to create, update, and delete coverages. The coverage area can be configured using tools in the Navitia interface or through the command line.

- **Verify data ingestion:** After data has been ingested, it's important to verify that it has been properly loaded into the Navitia instance. This can be done using the Coverage API.

After setting-up a Navitia instance, defining coverage areas, and ingesting/publishing data, it's important to test the system to ensure that it is functioning correctly.

3.3.1.3. Pipeline of data preparation and ingestion

Before publishing it into a Navitia coverage, data must be processed, indexed and converted to NTFS to make it available for routing and journey planning, by carrying the following generic data preparation pipeline:

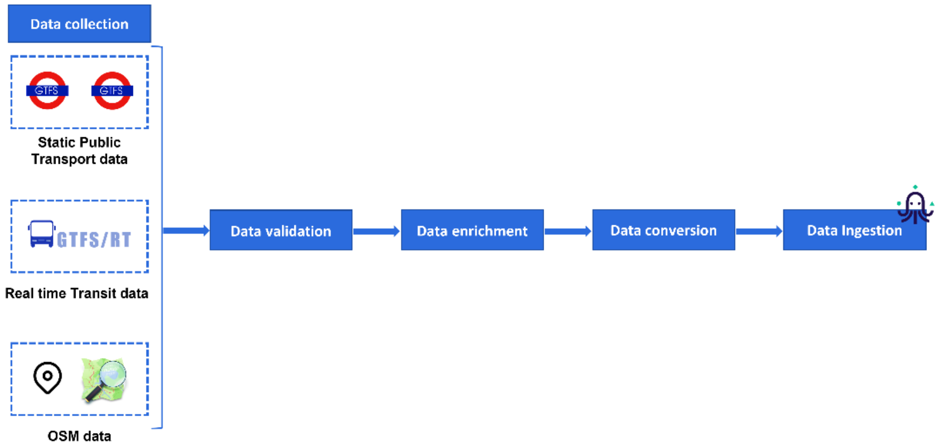

Figure 10: Data processing pipeline before ingestion in Navitia

- **Data validation:** as we cannot assume that data harvested into the transport cloud respects the standardized formats, it is necessary to confirm that it is indeed the case to ensure a successful data ingestion. For example, to validate GTFS datasets, we can use the canonical validator shared by "Mobility data" (*https://github.com/MobilityData/gtfs-validator*).
- **Data enrichments:** can vary from improving data quality such as adding missing optional information; to making custom changes to data like changing a POI category name or adding trip shapes using KML files containing the actual geographical trajectories taken by the transport lines in the public transport data in question. For instance, in our case we applied the following enrichments:
  - A spell-check of stop names using a processor ruspell that uses Aspell (https://github.com/hove-io/ruspell)
  - Generation of the file transfers.txt using https://github.com/hove-io/transfe_rs
- **Data conversion:** Navitia uses its own data model to represent transit data called NTFS (Navitia Transit Feed Standard). In this step, we used the processor GTFS2NTFS available in the transit model of Hove (https://github.com/hove-io/transit_model)

The order of data enrichment and data conversion is not fixed. It depends on the data format accepted as input of the pre-defined data enrichment processors.

3.3.1.4. Journey planning

Here are a couple of examples of a Navitia Journey requests made using the “it” coverage and transport data from the transport cloud.

1. A journey request (point to point) from: “23 Via degli Stradivari” in Rome to “Via Leonardo Umile” in Rome.

/coverage/it/journeys?from=stop_area%3AORA%3ANavitia%3A77708&to=stop_area%3AORA%3ANavitia%3AROME3603&datetime=20230215T120000&

The Navitia response is as shown in figure 2: It generates the 4 best optimal multimodal journeys, indicates the value of CO2 emissions and we can see that the fares and ticketing data are missing. To provide this information, we can add another processor in the data enrichment step before. This processor is modelled in Navitia Transit Model.


```
⊖{
    "context": ⊖{
        "car_direct_path": ⊖{
            "co2_emission": ⊖{
                "unit": "gEC"
                "value": 1970.4328570556,
            }
        }
        "current_datetime": "20230215T025103",
        "timezone": "Europe/Rome",
    },
    "disruptions": [],
    "exceptions": []
    "feed_publishers": ⊖[
        ⊖{
            "id": "it",
            "license": "ODbL",
            "name": "Italia"
            "url": "",
        },
        ⊖{
            "id": "osm",
            "license": "ODbL",
            "name": "openstreetmap"
            "url": "https://www.openstreetmap.org/copyright",
        }
    ],
    "journeys": ⊖[
        ⊕{15 items},
        ⊕{15 items},
        ⊕{15 items},
        ⊕{15 items}
    ],
    "links": ⊕[13 items],
    "notes": [],
    "terminus": ⊕[5 items],
    "tickets": [],
}
```


Figure 11: Navitia response to a journey request from Stradivari (Rome) to Vivaldi (Rome)

2. In order to find reachable places within specific time intervals, we can use isochrone computing API within Navitia to specify different time limits with the parameter boundary duration and a starting point. For instance, Figure 3 is a heat map representation of accessible zones per duration interval from the Colosseum via Rome's public transportation.

1/coverage/it/isochrones?from=stop_area%3AORA%3AITO140&boundary_duration%5B%5D=300&boundary_duration%5B%5D=900&boundary_duration%5B%5D=1800&

Funded by the European Union

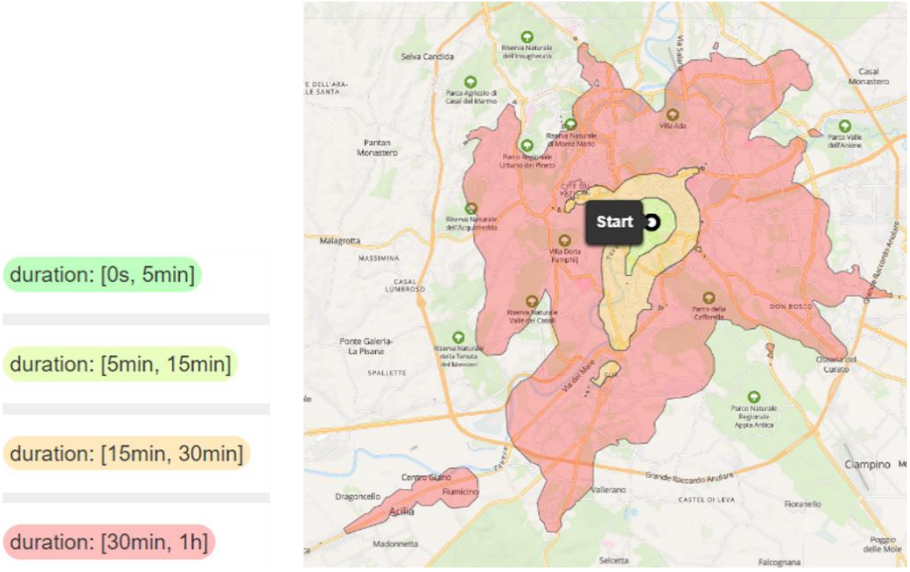

Figure 12: isochrone representing accessible places/ zones per duration interval from the Colosseum via Rome's public transportation.

3.3.1.5. Journey planner statistics

There are several types of statistics that can be deduced from journey planner-based service response logs:

- Mode Share: to track the usage of different transportation modes (e.g. bus, train, subway, etc.) over a period of time. This information can help understand which modes are most popular and patterns.
- Average Travel Time: this metric identifies which routes or modes are causing delays which can help optimize travel time.
- Capacity Utilization: of different transportation modes to understand which modes are over or underutilized.
- Origin-Destination Analysis: to identify the most popular routes and travel patterns.
- Stop Usage Analysis: to help identify which stops are most popular and which are underutilized.
- Real-Time Performance: of different transportation modes to identify delays or other issues.

3.3.2. Transport data sharing within the Linked Open Data vision (CNR)

Digital technologies in general, and data sharing in particular, have an important role in improving tourist mobility. Indeed, tourists require good travel information when planning their trips and while travelling. To this end, it is crucial to combine transport and tourism related information, especially if we consider that smartphones are nowadays massively used as multi-modal travel assistants by tourists and commuters. Unfortunately, we observe that the availability of “tourism + transport"-oriented applications is scarce – it is more common to have a tourism app on one side and a transport app on the other, than applications integrating both worlds. Local authorities that organise tourism in their area (e.g., tourist offices) would gain a lot from integrating all available mobility services (possibly in real time) with their data, especially to improve the information they provide to tourists, e.g., car parks, public transport services, tourist buses, and even data from bike sharing companies. Indeed, some transport operators are starting to adopt this vision, and quite unsurprisingly this trend has started in capital cities and major tourist cities.

In this use case study, we focus on a scenario where mobility organising authorities could improve their efficiency by integrating tourist information into their passenger information systems, as this would prove especially useful for those cities and regions that are renowned tourist destinations. This scenario is especially interesting if we consider that tourists make up a particular category of transport users that must be managed accordingly. Indeed, contrary to residents, tourists usually are not familiar with the area they’re visiting or the local transport service, they follow different travel patterns, they do not need to travel at peak times, and they often do not speak the local language. Accordingly, in the use case study we aim to exploit the services offered by the Transport Cloud to enrich mobility data with semantic information (also) provided by Linked Open Data sources (e.g., WikiData). The resulting enriched datasets should then be used to perform analyses that can be exploited by the Tourist Service provider to issue recommendations.

We assume that mobility data that need to be enriched can fall within two categories, i.e., (1) *tracks*, which are time-stamped sequences of geographic coordinates generated by the use of position-enabled devices which collect the movements of tourists visiting a city, or (2) *segments* suggested by some journey planner service, whereby a segment is a sequence of geographic coordinates satisfying some criteria, suggested by some journey planner service.

The **goal** to be achieved in this use case study is to allow a tourism service provider, via the services offered by Transport Cloud, to offer personalised services that can enrich and improve the tourists’ visiting experience by leveraging the mobility data that is being tracked, collected, and semantically enriched. We remind that the tracking, collection, and enrichment of position data of individuals are subjected to GDPR and related privacy regulations.

3.3.2.1. Preliminary concepts

The current research use case study takes inspiration from paper (Ruback et al., 2016). In (Ruback et al., 2016) the authors have shown how repositories of semantic trajectories can be built and subsequently used to answer mobility queries, in line with the Linked Data principles.

Linked Open Data (LOD) is an interesting data representation that we deem useful for semantic enrichment. Indeed, LOD principles promote the creation and publication of previously isolated databases as interlinked, reusable *knowledge graphs* by means of known Web standards (e.g., HTTP, RDF, URIs, JSON-LD, SPARQL). Even more interestingly, LOD sources can be accessed to gather contextual information that can be used to enrich mobility data. One then can see how the

Funded by the European Union

Transport Cloud vision promoted by MobiDataLab can be framed in the context of the LOD principles (c.f. (Luebke et al., 2002)).

As a side-note, we observe that representing mobility data (e.g., trajectories, trips) according to the LOD principles offers a strategy to seamlessly incorporate mobility data into a global data space, i.e., the Web of Data, in a way that it can be easily shared, (re)used, and analysed. This latter topic is, however, out of the scope of this use case study, and thus left for future work.

The final result is a *repository of enriched trajectories*, whereby the movement data and the associated semantic follow a unified formalism provided by a set of ontologies. Once a repository of semantic trajectories, in the form of an RDF knowledge graph, is in place, one can then proceed to query (and thus analyse) it – to this end, the authors of the paper employed a triple store and the SPARQL 1.0 query language.

3.3.2.2. Definition of the use case study

Similarly to the approach presented in (Ruback et al., 2016), this use case study considers a semantic enrichment process performed by the Transport Cloud, where data made available by tourism operators, transport data providers, and Linked Open Data providers, is exploited to semantically enrich trajectories.

Figure 13 illustrates the use case diagram. Figure 13 shows different actors on the right (data provisioning) side: *transport data providers* (e.g. transportation agencies), *tourism data providers* (e.g. tourism agencies) and *linked open data providers* (e.g., WikiData). These actors can interact, either actively or passively, with the transport cloud by discovering, providing, or combining data, depending on their roles and specificities. On the other hand, the actor on the left (service consumption) side, i.e., the *tourist service provider*, sends requests to the transport cloud to perform some data analyses on the enriched datasets. The tourist service provider can then build on the analyses results provided by the transport cloud to offer services to their users – for instance, recommendation, journey planning, and historical mobility data analysis.

Funded by the European Union

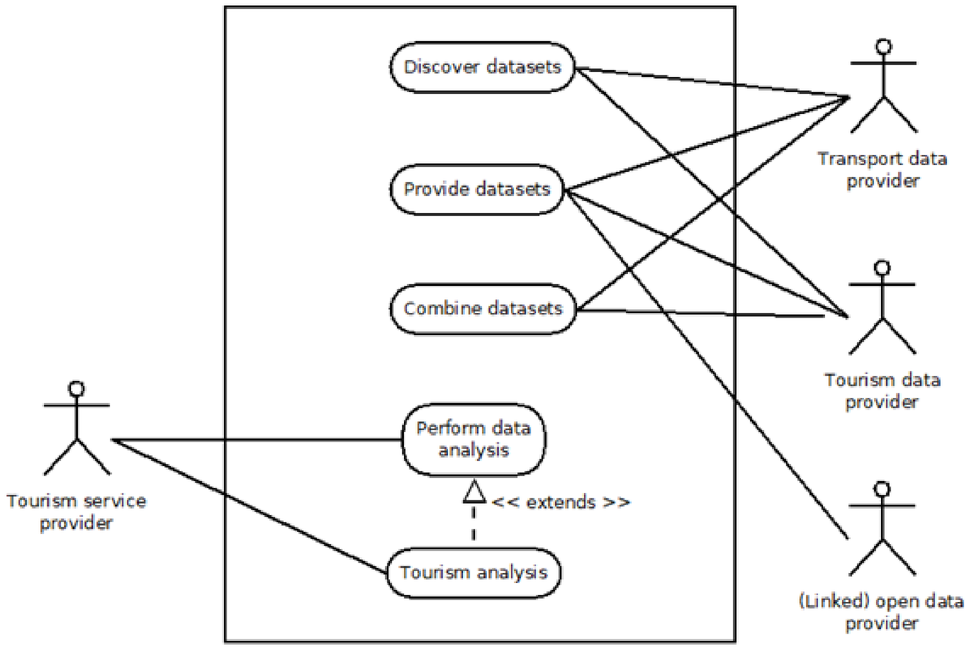

Figure 13: Use case diagram - Transport data sharing within the Linked Open Data vision

- **System**: MobiDataLab Transport Cloud

- **Actors** (who or what is going to be using the MobiDataLab Transport Cloud in this context):
  - **Primary**: The Tourism service provider
  - **Secondary**: The Transport Data provider, the Tourism Data provider, and the Linked Open Data provider.

- **Activities**:
  - Discover Datasets
  - Provide Datasets
  - Combine Datasets
  - Perform Tourism Analytics

3.3.2.3. Definition of the use case activities (with running example)

In the following we define the activities depicted in Figure 13. To better motivate the activities, we incorporate their descriptions within a running example. This enables to (1) instantiate the use case study within the context of a concrete scenario, (2) better motivate each activity, and (3) highlight the gaps that might currently exist between the goals set in this use case study and the current development status of the Transport Cloud. The order in which the activities will be presented implicitly defines the *data workflow* characterizing the use case study.

Let us suppose that the tourism service provider requires the Transport Cloud to create datasets of semantically enriched trajectories that can be analysed. More specifically, the provider intends to use such datasets to analyse the mobility behaviours of tourists within some popular touristic city. The results of these analyses might then be used by the provider to better drive the suggestions of their own recommender system – for instance, the system might use the analyses results to suggest points of interest (e.g., attractions, restaurants, hotels) that better align with the users' preferences, trips that can better satisfy the tourists' interests, and so on.

In order to be concrete, we locate our running example in the city of Rome, Italy, and aim to understand the mobility behaviours of single tourists that have visited the city. To semantically enrich the tourists' trajectories, we consider four different semantic dimensions. The first one is that of **regularity**, where during the enrichment process we capture those parts of the trajectories in which the individuals are staying at some location for some time, and try to understand if such stays exhibit some spatio-temporal regularity. For example, an individual may *regularly* stay at some location, which in turn may indicate they are staying at home, workplace, the supermarket where they usually do their groceries, and so on. We call such type of stays *regular stops*.

On the other hand, an individual may stay at some location just once, or without any apparent regularity. We call such stays *occasional stops*. The latter type of stops are particularly interesting for the purposes of this example, since they may highlight stays of that occur close to points of interest (e.g., attractions, restaurants) the tourists have visited.

The second semantic dimension is the **move**, where during the enrichment process, we capture those parts of the trajectories in which an individual has moved from some location to another, and augment such parts with the transportation means they have likely used while moving. This aspect might provide evidence on which transportation means the tourists prefer to use within a city, and which mixes of transportation means typically occur. The third dimension is the **weather**, where we enrich the trajectories with the weather conditions that the individuals have encountered during their trips. The last dimension is the **social media**, where we enrich the individuals behind the trajectories with the activities they have conducted on social media. Overall, we would like to highlight that the running example being introduced here is illustrative, and in a real-world scenario more analyses would be needed to fully understand the tourists' mobility behaviours within a city.

3.3.2.4. Discover datasets and services

The first activity required to attain the goal we set in the running example concerns the need to *find* and *select* the datasets that are required to create the dataset of semantically enriched trajectories. This would require the secondary actors to go through the activity of *providing* such datasets to the Transport Cloud. The secondary actors, however, might need to go through a preliminary activity first, whereby this activity requires the secondary actors to explore and find information or services offered by the Transport Cloud to "prepare" their datasets before providing them. Such activity is the *"Discover datasets and services".* This activity requires the Transport Cloud to give to the secondary

Funded by the European Union

actors the ability to explore and find out data or services being offered by the platform. In this case, for instance, a secondary actor may want to anonymize their data before they expose it to other actors. Accordingly, the Transport Cloud can offer anonymization services that the actors can use to satisfy this need. A secondary actor might also need to access some datasets within the Transport Cloud to gather information that is missing in their datasets, or to augment the information within their datasets.

The Transport Cloud offers technical solutions needed to implement this activity in the *Service Catalogue*, which is the focus of task 4.3, the Metadata Catalogue, which is the focus of task 4.2, and the API components, which are the focus of tasks 4.2 and 4.3. We thus refer the reader to the deliverables of these tasks for more information. The technical solutions and the data sources required to fulfil this activity clearly depend on the tourism service provider's needs. It is therefore possible that further technical solutions and datasets will be required.

3.3.2.5. Provide Datasets

This activity requires the Transport Cloud to either have the ability to access the data provided by the secondary actors, or let the secondary actors proactively provide their data to the Transport Cloud. In the context of this case study, this step is required in order to provide input to the subsequent activities. Finally, note that a secondary actor may also need to translate the datasets they want to provide in a format that is suitable for the transport cloud – to this end, the Transport Cloud can offer translation services via appropriate types of processors.

For the purposes of the running example, the datasets that need to be considered must concern the tourists' movements, as well as the four semantic dimensions introduced before. Accordingly, the first dataset required is the one that contains the tourists' trajectories moving within the city of Rome.

To this end, we identified:

- Mobility data: OpenStreetMap (https://www.openstreetmap.org/) provides publicly available trajectories voluntarily uploaded by its users. From OpenStreetMap, we therefore generated a trajectory dataset whose trajectories move within the province of Rome, Italy. The dataset contains 26395 trajectories from 3181 distinct users, spanning a time interval between March 2007 and July 2021.
- POI data: OpenStreetMap can provide the points of interests falling within any spatial region. The dataset contains 28787 POIs.
- Linked Open Data: Additional information concerning the POIs which were not available from OpenStreetMap were retrieved at query time from a Linked Open Data source, i.e., WikiData (https://wikidata.org).
- MeteoSat: For what concerns the weather conditions we identified Meteostat (https://meteostat.net/), which is an open source platform that provides free and unrestricted access to historic meteorological data.
- Twitter: Finally, for what concerns the social media semantic dimension we simulated posts published by the trajectory users on Twitter, as it wasn't possible to connect the users who generated the trajectories on OpenStreetMap to their social media profiles.

Funded by the European Union

The Transport Cloud offers technical solutions that can address this activity via the technical solutions adopted in the Data and Service Access component, the API component, and the Processors, which are covered in the tasks 4.2, 4.3, and 4.4. We therefore refer the reader to the deliverables of these tasks.

The major gap consists in ensuring that the secondary actors can provide datasets that are suitable for the creation of semantically enriched trajectories, which in turn depends on the tourism service provider's desiderata.

3.3.2.6. Combine Datasets

Once the datasets have been provided, it is possible to create datasets of semantically enriched trajectories. We call this activity "*Combine datasets*". The combine datasets activity requires the Transport Cloud to produce a dataset of semantically enriched trajectories that can be then further analysed and queried. Such dataset must satisfy three desiderata, i.e., (1) it must be produced from the datasets provided by the secondary actors, (2) it must be produced according to the semantic enrichment process required to satisfy the tourism service provider's desiderata, whereby a semantic enrichment process can be seen a pipeline (i.e., sequence) of several tasks, each dealing with a specific problem, which eventually outputs a dataset of semantically enriched trajectories, and (3) they should be represented by means of some formalism that make them queryable (and thus analysable).

In the context of the running example, we produced the semantically enriched movement data via the demonstrator of the semantic enrichment processor that has been presented in the deliverable 4.7 of Task 4.4. The semantic enrichment process used in the demonstrator is made of three steps: trajectory pre-processing, trajectory segmentation, and enrichment. We first executed the trajectory pre-processing step, then the trajectory segmentation one, and finally the enrichment step. In the pre-processing step we set the minimum number of samples to 1500, and the maximum speed threshold to 300 km/h. This yields a set of 620 pre-processed trajectories from 575 users. In the segmentation module we set the minimum duration of a stop to 10 minutes, while the maximum spatial radius a stop can have is set to 0.5 km. This yields a set of 666 stops and 1076 moves. Finally, the enrichment module enriches the segmented trajectories with all the aspects supported by the module, i.e., regularity of the stop segments, move (with transportation means estimation), weather information, and social media information. The occasional stops have been augmented with POIs found to be less than 50 metres far from the centroids of the stops. The execution of the enrichment module yields a dataset of semantically enriched trajectories.

An important feature that the enrichment module provides is the possibility to store the enriched dataset into a RDF knowledge graph. Information within the graph is structured according to the schema imposed by the STEP ontology (http://talespaiva.github.io/step/), which we lightly customized to fit the needs of the use case study needs. Such customized version has been included in the demonstrator. We therefore use this feature and store datasets of semantically enriched trajectories in RDF knowledge graphs.

In the following we report some technical solutions that might be used to implement the Combine

Funded by the European Union

Datasets activity, and a few gaps that exist between the use case study and the Transport Cloud current implementation.

To implement the Combine dataset activity, we refer to the Semantic enrichment processor being developed in Task 4.4, which has been already shown in the form of a preliminary demonstrator in the deliverable 4.7. The demonstrator implements a semantic enrichment process as a pipeline (i.e., sequence) of several tasks, each dealing with a specific problem, which eventually outputs a dataset of semantically enriched trajectories. Tasks that are commonly present in semantic enrichment processes, and that must be implemented also for this use case, are: trajectory pre-processing, trajectory segmentation, and semantic enrichment. Other tasks that address different or additional problems might be present, again, depending on the tourism service provider's needs.

The Transport Cloud has to ensure that it can support the tasks – and thus the operations – required to implement the semantic enrichment process(es) needed to satisfy the tourism service provider desiderata. Such operations are potentially many and varied, and this is where the notion of Processor (see also the deliverable 4.7 of task 4.4) within the Transport Cloud enters the scene. We recall that a Processor can be seen as a component within the Transport Cloud modelling some operation. Consequently, the notion of processor can be instantiated to implement any operation this use case requires. The implementation of a semantic enrichment process will therefore require the use of a set of cooperating processors, some of which may not be available at the present time.

Finally, the STEP ontology used to store datasets of semantically enriched trajectories might be further customized, depending on the tourism service provider's needs.

3.3.2.7. Perform Data Analysis

With the RDF knowledge graph containing the dataset of semantically enriched trajectories in place, it is now possible to conduct analyses on it. We call such activity **"***Tourism analysis***",** which is a specialisation of the "*Perform data analysis" activity* targeting this use case study. These activities require to analyse the dataset of semantically enriched trajectories according to goals that align with the tourism service provider's desiderata. This can be done by querying the dataset with an appropriate triple store of choice. The results will serve as the basis of the services that the Tourism service provider will offer to customers.

Going back to the running example, we import the RDF knowledge graph generated by the demonstrator into a triplestore, which we then use to query (and thus analyse) the graph. Our choice fell on the popular GraphDB (https://graphdb.ontotext.com/). The triplestore is then used to perform several analyses by means of the SPARQL 1.1 query language. More specifically, in the running example the activity focuses on analyses that aim to extract the mobility behaviours of a specific individual by analysing their enriched trajectories. The analyses are reported in more detail within Annex 6.1. In the analyses we report that we also made use of federated queries to retrieve further information on the POIs. Federated queries are SPARQL queries that access remote SPARQL endpoints to retrieve further information, which in turn enable to access Linked Open Data sources. In this context of this running example, we accessed data on POIs provided by WikiData (https://wikidata.org).

Funded by the European Union

The technological solution relevant to these use cases is the use of a **triple store** that can (1) import and query RDF knowledge graphs produced by the demonstrator and (2) access and query linked open data sources while querying knowledge graphs. Accordingly, our choice falls on **GraphDB**, which is a popular and widely used triple store satisfying the above desiderata. More specifically, GraphDB allows to:

- **query and analyse knowledge graphs**: RDF knowledge graphs can be imported in GraphDB[1], which can be then used to query them via the SPARQL 1.1 query language. GraphDB also supports several SPARQL extensions that may be of interest to this use case scenario, i.e., the time and date extension and the geospatial extension.
- **accessing and retrieving linked open data sources**: linked open data sources (e.g., WikiData) can be accessed and queried by means of federated SPARQL queries, i.e., SPARQL queries that access remote SPARQL endpoints. We report that federated SPARQL queries are supported since SPARQL version 1.1, which is the version supported by GraphDB.

3.3.2.8. Datasets and data sources of interest

In this section we report on the datasets that can be used for this use case study. Some of these have already been used in the running example shown in Section 1.3, while others are of potential interest since they can be used in similar contexts.

One of the most used open datasets worldwide are the Wikimedia datasets that provide quality open content at an international level from crowdsourced projects like the Wikimedia projects (Wikipedia, Wikimedia Commons, etc.). All this content is now both usable by humans and machines thanks to the Wikidata project. Wikidata is a semantic database, a structured information graph composed of "triplets", i.e., subject, predicate, and object, where everything is described according to such formalism. Data can be queried via the Mediawiki APIs and/or SPARQL endpoints (https://query.wikidata.org/).

OpenStreetMap (https://www.openstreetmap.org/) is a collaborative project to create a free editable geographic database of the world. The geodata underlying the OpenStreetMap maps is considered the primary output of the project, and consists of a wide-ranging varied types of entities, e.g., maps, countries, administrative regions, points of interests and associated information, various types of road networks, and so on. We report that OpenStreetMap also gives the possibility to access trajectories uploaded by users. OpenStreetMap let users access its data via various means – to this end we mention the JOSM Java application (https://josm.openstreetmap.de/) and the Overpass API (https://dev.overpass-api.de/)..

Meteostat (https://meteostat.net/) is an open source platform which provides free and unrestricted access to meteorological data. The project's focus is historical weather and climate data. Meteostat features multiple interfaces which provide data access for both end users and developers, i.e., a website with data visualizations, bulk data dumps of individual weather stations published under the

[1] https://graphdb.ontotext.com/

CC BY-NC 4.0 license, a Python library developed under the MIT License (https://github.com/meteostat/meteostat-python), and a JSON API (https://dev.meteostat.net/api/).

In addition to the previous datasets already used in the use case instantiation, in the following we provide a few possible datasets and data providers that may be used to implement the use case study in other contexts.

The RATP public transport operator in Paris proposes NEXT STOP PARIS (https://www.ratp.fr/en/apps/next-stop-paris), and Transport for London (https://tfl.gov.uk/) proposes content dedicated to visitors on their website. But when asking ourselves why there are yet so few hybrid tourism + transport applications, we realise that proposing quality content is difficult, as mixing tourism and transport data implies using different access channels (APIs, etc.) following different standards, and adding/creating content which transport actors are not familiar with.

Technology, data, and content are now openly available and combining transport open data, tourism open data and content is now becoming possible - at least in certain member states, notably in France. Indeed, France is making available many datasets for mobility and tourism. Besides the French National Access Points for mobility data (transport.data.gouv.fr), the French government launched in 2017 the DataTourisme platform (https://www.datatourisme.gouv.fr/) (DATAtourisme > La plateforme nationale OpenData du tourisme en France, no date). DataTourisme aims at collecting, processing and disseminating tourist information in Open Data. These Open Data are aggregated daily from local tourist information systems in the country, maintained by local authorities, regional committees, departmental agencies and tourist offices. These heterogeneous data are gathered in a semantic database following a common formalism, the DataTourisme ontology, and they can be consulted and queried via a web interface and SPARQL queries. These data are also under Open Licence (OL).

Finally, it may be interesting to explore Semantic interoperability with transport open data thanks to corresponding vocabularies (cf. the Shift2Rail project[2] and the Semantic Interoperability Framework[3]).

3.3.2.9. Standards

- Ontology Web Language (OWL)
- Resource Description Framework (RDF)
- SPARQL Protocol and RDF Query Language (SPARQL v1.1).

[2] https://transport.ec.europa.eu/transport-modes/rail/rail-research-and-shift2rail_en
[3] https://www.youtube.com/watch?v=xMV-GxWEYy4

Funded by the European Union

4. Conclusions

The use-cases are a continuously evolving, therefore two versions have been build within the course of the project. One in the early beginning to support the requirements design and the version two with the experience of the architectural design and the communication with the reference group.

Although there is a continuous improvement on the specification of the use-cases, the story is not finished at this point. However, this document provides additional connections between stakeholder groups for mobility data sharing. The use cases provided by the reference group and other stakeholders give more insights and a practical view on real world scenarios. The stakeholders are enabled with incentives (solutions as return of invest) to share their data. The real world use cases are understandable for almost all participants and reusable for other reference group members, since there are similar situations and use cases in other cities/regions.

With this deliverable, the base for the living- and virtual labs is set, and the described use-cases are providing a broad set of challenges for the Datathon, the Hackathon and the Codagon.

However, this document needs to be seen as an important milestone on the mobility data sharing journey. Within the Living- and Virtual labs, more insights, ideas and opportunities will be generated. The updates on these documents will then be transitioned into the Open Knowledge Base, where they can continuously evolve and being extended. The stakeholder group will grow and will be able to contribute their use-cases and experience to the living document to ramp down the barriers for mobility data sharing. With the transparent publication of experience within the real-world scenarios, trust and clarity within the stakeholder group(s) regarding mobility data sharing barriers will grow as the group of data sharing partners will.

Funded by the European Union

5. References

Christaller, W. (1933). Die zentralen Orte in Suddeutschland: Eine ökonomisch-geographische Untersuchung über die Gesetzmassigkeit der Verbreitung und Entwicklung der Siedlungen mit städtischen Funktionen. *Jena*.

Luebke, D. et al. (2002) ‘Level of Detail for 3D Graphics : Application and Theory.’, p. 431.

Ruback, L., Casanova, M. A., Raffaetà, A., Renso, C., & Vidal, V. (2016, July). Enriching mobility data with linked open data. In *Proceedings of the 20th International Database Engineering & Applications Symposium* (pp. 173-182).

6. Annexes

6.1 Transport data sharing within the Linked Open Data vision : analyses conducted within the 'Tourism analysis' activity

From the dataset we selected an individual (ID 2115) who produced a trajectory (ID 2652) which originates close to the Fiumicino Rome Airport in the early morning, then spends half of the day within the centre of Rome, and then goes back to the same airport in the early afternoon (Figure 14). The overall duration of the trajectory is 6 hours. All such evidence could hint that the individual is some kind of tourist passing by the city. Further analyses on the individual's mobility behaviours are, however, required to reach any conclusion.

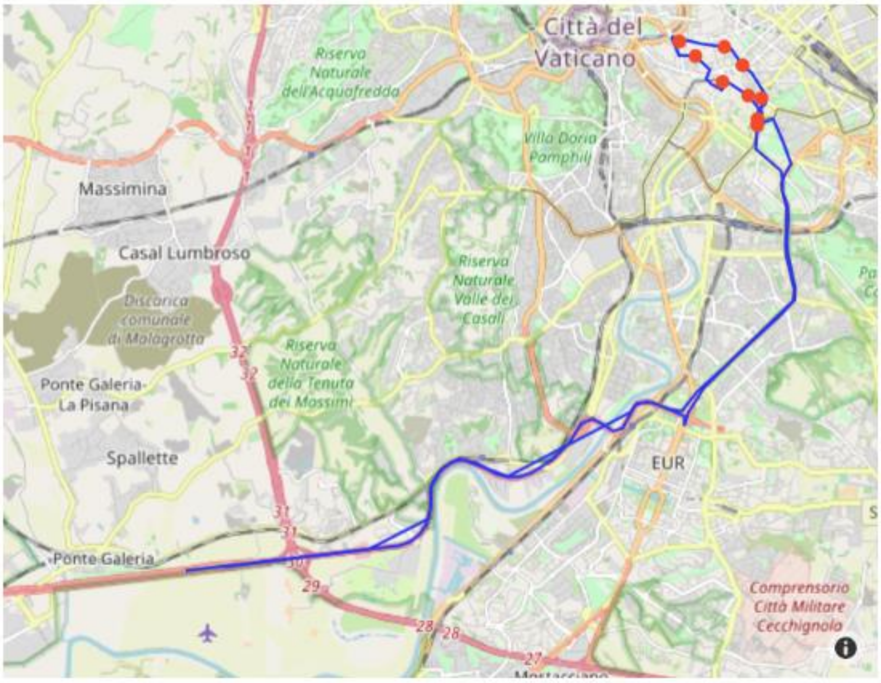

Figure 14: plot of the semantically enriched trajectory with ID 2615 associated with the user having ID 2115. The red dots represent the occasional stops that have been found, while the blue curve represents the moves.

Accordingly, we want to find out:

1. which transportation means the individual has likely used during their trip
2. the POIs the individual may have visited while staying in Rome
3. the weather conditions and social media posts related to the trip
4. finally, we access a Linked Open Data source (i.e., WikiData) to further augment the results returned by the analysis conducted at point 2. To this end we leverage the notion of federated query present in the SPARQL 1.1 query language.

Funded by the European Union

Let us first find out the transportation means the individual has possibly used. We do so by executing the SPARQL query in Figure 15, which retrieves all the occurrences of the move aspect enriching the trip, orders them temporally, and for each occurrence returns the estimated transportation means.

```
1  SELECT ?type_move ?t_start ?t_end (ofn:asMinutes(?t_end - ?t_start) AS ?duration_mins)
2  WHILE {
3     ?traj ^step:hasTrajectory / foaf:name "2115" ;
4              step:hasID "2652" ;
5              step:hasFeature ?feat .
6
7     ?feat step:hasEpisode ?ep .
8     ?ep step:hasSemanticDescription ?move ;
9           step:hasExtent ?ex .
10    ?move rdfs:subClassOf step:Move ;
11             rdf:type ?type_move .
12
13    ?ex step:hasStartingPoint / step:atTime / time:inXSDDateTime ?t_start .
14    ?ex step:hasEndingPoint / step:atTime / time:inXSDDateTime ?t_end .
15 } ORDER BY ASC(?t_start)
```

Figure 15: SPARQL query retrieving the move occurrences associated with the semantically enriched trajectory having ID 2652.

The query first finds out the trajectory of interest (lines 3-4), then retrieves all its semantic aspects (lines 5-9), and finally filters out those that are not a move (line 10). With the occurrences of the move aspect, the query first retrieves for each occurrence the estimated transportation means (line 11), and then determines its starting and ending instants (lines 13-14). The SELECT finally returns a list of tuples (Figure 16), each representing a move with the estimated transportation means, its starting and ending instants, and its duration.

	type_move	t_start	t_end	duration_mins
1	step_specialized:Train	"2014-04-10T08:30:21+00:00"^^xsd:dateTime	"2014-04-10T09:01:23+00:00"^^xsd:dateTime	"31"^^xsd:long
2	step_specialized:Walk	"2014-04-10T09:04:30+00:00"^^xsd:dateTime	"2014-04-10T09:12:09+00:00"^^xsd:dateTime	"7"^^xsd:long
3	step_specialized:Car	"2014-04-10T09:19:43+00:00"^^xsd:dateTime	"2014-04-10T09:31:13+00:00"^^xsd:dateTime	"11"^^xsd:long
4	step_specialized:Walk	"2014-04-10T09:55:34+00:00"^^xsd:dateTime	"2014-04-10T10:44:11+00:00"^^xsd:dateTime	"48"^^xsd:long
5	step_specialized:Bus	"2014-04-10T10:53:43+00:00"^^xsd:dateTime	"2014-04-10T11:14:16+00:00"^^xsd:dateTime	"20"^^xsd:long
6	step_specialized:Walk	"2014-04-10T11:47:00+00:00"^^xsd:dateTime	"2014-04-10T12:26:25+00:00"^^xsd:dateTime	"39"^^xsd:long
7	step_specialized:Walk	"2014-04-10T12:42:28+00:00"^^xsd:dateTime	"2014-04-10T12:53:40+00:00"^^xsd:dateTime	"11"^^xsd:long
8	step_specialized:Bus	"2014-04-10T13:14:48+00:00"^^xsd:dateTime	"2014-04-10T13:19:12+00:00"^^xsd:dateTime	"4"^^xsd:long
9	step_specialized:Bus	"2014-04-10T13:26:27+00:00"^^xsd:dateTime	"2014-04-10T13:38:06+00:00"^^xsd:dateTime	"11"^^xsd:long
10	step_specialized:Bus	"2014-04-10T13:57:20+00:00"^^xsd:dateTime	"2014-04-10T14:45:36+00:00"^^xsd:dateTime	"48"^^xsd:long

Figure 16: Moves found for the trajectory with ID 2652 (screenshot from GraphDB).

From the results, we report that the query finds 10 move occurrences. By looking at their characteristics, we report that the individual appears to go from the airport to the Rome's city centre by train, then mostly walked and partially used a bus while moving in the city, and finally went back to the airport by bus.

In the second analysis we want to find out the POIs the individual has possibly visited during their trip. Accordingly, the query in Figure 13 considers the *regularity* aspect – more specifically, it focuses on the individual's *occasional stops*.

Funded by the European Union

```
1  SELECT ?t_start ?t_end (ofn:asMinutes(?t_end - ?t_start) AS ?duration) ?poi_name
     ?poi_category
2  WHERE {
3    ?traj ^step:hasTrajectory / foaf:name "2115" ;
4            step:hasID "2652" ;
5            step:hasFeature ?feat .
6    ?feat step:hasEpisode ?ep .
7
8    ?ep step:hasSemanticDescription ?stop .
9    ?stop rdf:type step:OccasionalStop .
10
11   ?ep step:hasExtent / step:hasStartingPoint / step:atTime / time:inXSDDateTime ?t_start ;
12        step:hasExtent / step:hasEndingPoint / step:atTime / time:inXSDDateTime ?t_end .
13
14   ?stop step:hasPOI ?poi .
15   ?poi step:hasOSMCategory ?poi_category ;
16        step:hasOSMName ?poi_name .
17 } ORDER BY ASC(?t_start)
```

Figure 17: SPARQL query retrieving the occasional stop occurrences associated with the semantically enriched trajectory having ID 2652.

The query first finds out the trajectory of interest, and keeps only the occurrences of occasional stops (lines 3-9). For each of these occurrences the query then finds the starting and ending instants (lines 11-12). Successively, the query gathers information concerning the occurrences that have at least one POI and finally retrieves the names and categories of the POIs involved (lines 14-16). From the results we report that the query finds 12 occasional stops. By looking at the associated POIs, we report that the individual appears to have spent a good part of the morning visiting various monuments: the individual briefly stayed in the area surrounding the Palatino and then went to the Tempio della Pace. The individual then proceeded to stay for more than an hour in the area surrounding the Altare della Patria, and then stayed in the area of the Pantheon for around half an hour until lunch time. After that, the individual appears to have stayed at a restaurant for almost one hour. Finally, the individual appears to have stayed again in the vicinity of the Palatino, and then went back to the airport. All in all, the individual has been repeatedly observed nearby famous monuments, and appears to have walked and used public transportation, thus reinforcing the initial impression they were indeed a tourist.

In the third analysis we aim to find out the weather conditions and the social media posts that the individual has respectively experienced and published during the trip. Let us focus on the weather conditions, as the strategy for the other aspect is the same. To this end, we can insert in either of the two queries shown before, right before the end of the `WHILE` loop, the SPARQL fragment shown in Figure 18. The `OPTIONAL` keyword serves the purpose of not filtering out from the final results the occurrences of move (occasional stop) for which no weather information is available. The `FILTER` keyword ensures that each move (occasional stop) occurrence gets associated with an occurrence of the weather aspect only if they have a non-empty temporal overlap. Finally, the `weather_conditions` variable can be integrated in the `SELECT` to report the weather conditions. Overall, we observe that the individual experienced a sunny day during their trip.

Funded by the European Union

```
OPTIONAL {
    ?traj step:hasFeature / step:hasEpisode ?ep_w .
    ?ep_w step:hasSemanticDescription / rdf:type step:Weather ;
            step:hasWeatherCondition ?weather_conditions ;

            step:hasExtent / step:hasStartingPoint / step:atTime / time:inXSDDateTime ?tw_start ;
            step:hasExtent / step:hasEndingPoint / step:atTime / time:inXSDDateTime ?tw_end .

    FILTER((?t_start <= ?tw_end) && (?tw_start <= ?t_end)) }
```

Figure 18: SPARQL fragment retrieving the weather conditions aspect that has been associated with the semanticallyenriched trajectory having ID 2652.

Overall, the results of these analyses might represent the starting point to provide a recommender system insight on the mobility behaviours of tourists visiting a city. Clearly, more mining and analysis steps should be performed – for instance, grouping trajectories to find the most popular trips or points of interest the tourists visited, find which transportation means tourists prefer in different areas of the city, and so on.

Let us finally consider a case whereby we access a Linked Open Data source to augment the results returned by the query introduced in Figure 13, which focuses on the POIs possibly associated with the occasional stops. We turn such a query into a *federated query* by inserting the fragment shown in Figure 19, right before the end of the query WHERE clause (i.e., between the lines 16 and 17).

```
OPTIONAL {
    ?poi step:hasWDValue ?WD .
    SERVICE <https://query.wikidata.org/sparql>
    {
        OPTIONAL{?WD wdt:P18 ?img_WD .}
        OPTIONAL{?WD wdt:P1329 ?phone_WD .}
        OPTIONAL{?WD wdt:P2555 ?fee_WD .}
        OPTIONAL{?WD wdt:P571 ?year_built_WD .}
        OPTIONAL{?WD wdt:P856 ?url_WD .}
        OPTIONAL
        {
            ?WD wdt:P149 / rdfs:label ?style_WD .
            FILTER(lang(?style_WD) = "en")
        }
        OPTIONAL
        {
            ?WD wdt:P2846 / rdfs:label ?weelchair_WD .
            FILTER(lang(?weelchair_WD) = "en")
        }
    }
}
```

Figure 19: SPARQL fragment augmenting the information concerning the returned POIs via the use of a linked open data source.

Funded by the European Union

Recall that a SPARQL federated query is a SPARQL query executed over different SPARQL endpoints, which in turn allows querying Linked Open Data sources – in this case, the fragment is accessing the WikiData SPARQL endpoint. As such, the purpose behind the fragment shown in Figure 18 is to enrich with further information the POIs that the query in Figure 15 returns. First, observe that the whole fragment is wrapped within the OPTIONAL graph patterns (line 1), which prevents us to discard POIs that in OpenStreetMap do not have a WikiData identifier (line 2). For each of the POIs that have an identifier, the query attempts to retrieve from WikiData the URL of an image (line 5), the phone number (line 6), the entrance fee (line 7), the year in which the POI was built (line 8), the URL of the POI website (line 9), the architectural style (in English language, lines 10-14), and whether the POI is accessible or not (in English, lines 15-19). Finally, note that further or different information may be retrieved from WikiData, depending on the tourism service provider's needs.

Funded by the European Union

MobiDataLab consortium

The consortium of MobiDataLab consists of 10 partners with multidisciplinary and complementary competencies. This includes leading universities, networks and industry sector specialists.

@MobiDataLab
#MobiDataLab

https://www.linkedin.com/company/mobidatalab

For further information please visit www.mobidatalab.eu

MobiDataLab is co-funded by the EU under the H2020 Research and Innovation Programme (grant agreement No 101006879).