



Deliverable

D8.5 Data Management and Modelling Plan (release 2)

Project Acronym:	DUET	
Project title:	Digital Urban European Twins	
Grant Agreement No.	870697	
Website:	www.digitalurbantwins.eu	
Version:	1.0	
Date:	27 January 2022	
Responsible Partner:	AIV	
Contributing Partners:	IMEC, KUL, ATC, DAEM, VCS, TNO, P4ALL, PLZ, ISP, GFOSS	
Reviewers:	Andrew Stott, Yannis Charalabidis, Michiel Van Peteghem, Pieter Morlion	
Dissemination Level:	Public	X
	Confidential – only consortium members and European Commission	

Revision History

Revision	Date	Author	Organisation	Description
0.1	13.01.2022	Lieven Raes	AIV	Initial structure
0.9	25.01.2022	Internal and external expert review		Review
1.0	27.01.2022	Lieven Raes	AIV	Final version

Table of Contents

1. Data summary	6
2. Fair data	7
2.1. Data summary	7
2.2. Data collection	8
2.3. Data processing	9
2.4. Data storage	9
2.5. Data sharing	10
3. Allocation of resources	11
4. Data security	12
5. Ethical aspects	13
6. References	14
7. Annex 1: Datasets & models overview	15

List of tables and figures

Table 1: DUET Data usage scenarios - data summary overview	7
Table 2: DUET data usage scenarios - data collection overview	8
Table 3: DUET data usage scenarios - data processing overview	9
Table 4: DUET data usage scenarios - data storage overview	9
Table 5: DUET data usage scenarios - data sharing overview	11
Figure 1: Taxonomy of DUET security measures and respective TCs	12

Executive Summary

DUET aims to establish the use of big data and data visualisation as an integral part of policy making, particularly, but not limited to, the local government level and the mobility and transport policy domain. The project's relation with data is therefore essential and connatural to its experimental research objectives and activities.

Additionally, the consortium has adhered to the H2020 ORDP (Open Research Data Pilot) convention with the EC, which explicitly caters for the delivery of a DMP (Data Management Plan).

According to the DUET DoA (2019), data management planning, monitoring and reporting is part of WP8 - the Project risk and quality management work package.

This second edition describes the situation at the time of the third open beta release and builds further on the principles described in the first release, and focuses on changes compared to the first report in month 6. In the DMP we basically envisage three main data usage scenarios, which jointly compose DUET data management lifecycle:

- Original data produced by the DUET consortium and/or individual members of it (e.g. during a dissemination action or a pilot activity)
- Existing data already in possession of the DUET consortium and/or individual members of it prior to the project's initiation and open data
- Existing data sourced/procured by the DUET consortium or obtained from external partners during the project's timeline

For completeness of information, the reader interested in getting to know how the DUET consortium plans to deal with data may also refer, in addition to this DMP, to the following deliverables: D1.1 Legal Landscape and Requirements Plan [February 2020], as well as D1.2 Cities Guide to Legal Compliance for Data-Driven Decision [September 2020] and the D4.2 Duet data integration (November 2020) deliverables.

1. Data summary

The DUET project builds a reusable and interoperable Digital Urban Twin solution based on existing datasets, simulation models and creates evidence-based results. DUET is, as such, not a project producing new datasets or data layers. Instead, DUET mainly combines and re-uses what already exists innovatively.

The end-product of DUET is an open Urban Digital Twin platform using, processing and visualising existing data and simulation model outputs. These outputs are available as static data layers visualised as part of the Urban Digital Twin. A good example is a processed result layer containing the calculation of an air quality map as a result of the cooperation between the output of a traffic model delivering the input for the air quality model calculation.

Other data views are dashboard views combining the results of different datasets and delivering insights in spatial and time-related data (geo-time series).

In this DMP, we differentiate between input data and output data (e.g. simulation model outputs). Input data can be derived from external sources (e.g. external APIs or web services) or can be uploaded on the consortium's data infrastructure. The DUET Digital Twins uses a combination of non-geospatial and geospatial datasets. Non-geospatial formats are open formats like XML¹, JSON(LD)², CSV³ as non-geospatial data formats. The used geospatial formats combine data file formats and services like Shapefiles⁴, WMS⁵, WFS⁶, GeoTIFF⁷, CityGML⁸ and GeoJSON⁹. APIs are also used and are especially relevant for live data streams. Simulation models use a combination of model-specific formats, e.g. Cube Voyager¹⁰ and more generic scripting formats like GDX¹¹ scripting.

The DUET output data can't be downloaded in this 3rd open beta release. However, the consortium is discussing the possibility to make model output results available as pictures or open GIS formats like GeoJSON, WMS or WFS.

Finally, there is the need for new standardisation initiatives. An essential element related to improving Digital Twins is the further standardisation on the data input side. Data interoperability and semantics could drastically reduce the need for data integration, e.g., using standardised sensor data formats like Sensor Observation Services (SOS)¹² and SensorThings API¹³.

¹ <https://www.w3.org/XML/>

² <https://www.w3.org/TR/json-ld11/#introduction>

³ <https://www.w3.org/TR/tabular-data-primer/>

⁴ <https://desktop.arcgis.com/en/arcmap/10.3/manage-data/shapefiles/what-is-a-shapefile.htm>

⁵ <https://www.ogc.org/standards/wms/introduction>

⁶ <https://www.ogc.org/standards/wfs>

⁷ <https://www.ogc.org/standards/geotiff>

⁸ <https://www.ogc.org/standards/citygml>

⁹ <https://www.ogc.org/standards/eo-geojson>

¹⁰ <https://www.bentley.com/en/products/product-line/mobility-simulation-and-analytics/cube-voyager>

¹¹ https://www.gams.com/35/docs/UG_GDX.html

¹² <https://www.ogc.org/standards/sos>

¹³ <https://www.ogc.org/standards/sensorthings>

2. Fair data

Very pragmatically, the European Commission (2016) considers the FAIR principles fulfilled if a DMP includes the following information:

- A. *“The handling of research data during and after the end of the project”*
- B. *“What data will be collected, processed and/or generated”*
- C. *“Which methodology and standards will be applied”*
- D. *“Whether data will be shared/made open access”, and*
- E. *“How data will be curated and preserved (including after the end of the project)”*.

In the case of DUET, the above information is provided in the DMP R1 (DUET Deliverable D8.3) and Section 2 of this document, which consists of five paragraphs, respectively:

- 2.1. Data summary (*typologies and contents of data collected and produced*)
- 2.2. Data collection (*which procedures for collecting which data*)
- 2.3. Data processing (*which procedures for processing which data*)
- 2.4. Data storage (*data preservation and archiving during and after the project*)
- 2.5. Data sharing (*including provisions for open access*)

The DUET consortium makes a difference between dataset and simulation models. The input and output of the simulation models are also data and are covered under the topics below.

Specific background for what concerns (A, E), the handling of research data during and after the end of the project - the principles of the DMP R1, section 3.9 are still valid and will be further described in the DMP R3. The applied data standards (C) are listed in section one and further described in deliverable D3.2 IoT stack and API specification chapter 3, deliverable D3.4 Smart City domains, models and interaction frameworks chapter 4.1 and 4.3.

2.1. Data summary

The following table summarises the typologies and contents of data collected and produced. For each distinct category, you can find the status by month 27.

Nature of datasets	Confidential	Anonymised and Public	Non anonymised
Data usage scenarios			
Original data produced by the DUET consortium	Login data into the DUET application	N/A	N/A
Existing data already in possession of the DUET consortium and/or partners and open data	Flanders: Flanders traffic model data, TNO noise model Pilsen: n/a, TNO air quality model, Athens: /	Flanders: Flanders traffic model data, Antwerp city flows data Pilsen: Pilsen traffic model, traffic density (magnetic loops), air quality certified sensor data, Pilsen noise model	N/A

		Athens: /	
Existing data sourced/procured by the DUET consortium or obtained from external suppliers	Flanders: Air quality model (VITO, TNO)), Noise model (TNO) Pilsen: / Athens: /	Free and open datasets: Flanders: KUL Dyntapy traffic model, 2D/3D LOD 1 model Pilsen: floating car data, 3D model of the buildings, Digital surface model, Digital terrain model Athens: /	N/A

Table 1: DUET Data usage scenarios - data summary overview

2.2. Data collection

The following table summarises the procedures for collecting project related data. For each distinct case, some concrete examples are provided. Below you find an overview of the outputs created by DUET.

Nature of datasets	Confidential	Anonymised and Public	Non anonymised
Data usage scenarios			
Original data produced by the DUET consortium	Internal meeting minutes Confidential deliverables Login data Subscriptions data	DUET visualisation results like simulation model results, citizen science sensor data	DUET Publications list DUET Newsletters
Existing data already in possession of the DUET consortium and/or partners and open data	Flanders: Flanders traffic model data, TNO noise model Pilsen: Pilsen traffic model, TNO air quality model, Pilsen noise model Athens:	Flanders: Flanders traffic model data, Antwerp city flows data Pilsen: Pilsen traffic model Athens:	N/A
Existing data sourced/procured by the DUET consortium or obtained from external suppliers	Licensed access and use during project execution e.g. the Flanders air quality model by VITO, TNO and the Flanders Traffic Model by the Flanders road administration.	Free and open access and use during project execution - via obtaining data from open data portals and via application forms (see dataset list).	N/A

Table 2: DUET data usage scenarios - data collection overview

An implication of the above table that may not have been evident in the previous one is that **every partner is responsible for the behaviour of all team members**, which may also include subcontracted organisations (e.g. specialised press agencies) or even volunteers. For example, the Flanders Traffic Model contract has been signed by all the partners involved in the implementation process and makes them responsible for fulfilling the contract conditions.

2.3. Data processing

The following table summarises the procedures for processing DUET related data as applied during the first 27 months. Some of the listed measures were accomplished during the project. Other measures were already applied before by the data supplier (e.g. anonymisation of census data as part of the traffic model). Every dataset has at least been checked by the pilots, involved partners or the ethics manager. The data processing measures in the table provide an overview of actions that were effectively needed to realise a GDPR compliant and ethical use of the data in each situation (represented in the table cells).

Nature of datasets Data usage scenarios	Confidential	Anonymised and Public	Non anonymised (temporary data)
Original data produced by the DUET consortium	Analytics	Anonymisation Statistical evaluation Metadata generation Visualisation Analytics	Selection/destruction
Existing data already in possession of the DUET consortium and/or partners and open data	Anonymisation Statistical evaluation Metadata generation Visualisation Analytics	Anonymisation Statistical evaluation Metadata generation Visualisation Analytics Publication as map services	N/A
Existing data sourced/procured by the DUET consortium or obtained from external suppliers	Anonymisation Statistical evaluation Metadata generation Visualisation Analytics	Anonymisation Statistical evaluation Metadata generation Visualisation Analytics Publication as map services	N/A

Table 3: DUET data usage scenarios - data processing overview

2.4. Data storage

The following table summarises the procedures for storing project-related data during the first 27 months of the DUET project and provides an overview of the most frequently used repositories. The DUET data broker component processes the data itself as part of the DUET core technical solution. The data broker component itself doesn't store data content permanently (only for a very short period when processing the data).

Nature of datasets Data usage scenarios	Confidential	Anonymised and Public	Non anonymised (temporary status)
Original data produced by the DUET consortium	Common project repository	Common project repository, or repositories of the data provider Common	Individual partner repositories Common project repository

		repositories are used in the DUET project landing pages, the DUET Digital Twin data Open access repository	
Existing data already in possession of the DUET consortium and/or partners and open data	Common project repository, or repositories of the data provider. Visualisation results are used in the DUET Digital Twin.	Common project repository, or repositories of the data provider Common repositories e.g. used in the DUET project landing pages, the DUET Digital Twin data Open access repository	N/A
Existing data sourced/procured by the DUET consortium or obtained from external suppliers	Individual partner repositories Third party repositories of the data provider. Visualisation results are used in the DUET Digital Twin.	Common project repository, or repositories of the data provider Common repositories e.g. used in the DUET project landing pages, the DUET Digital Twin data Open access repository	N/A

Table 4: DUET data usage scenarios - data storage overview

Data storage and management of the use DUET datasources

The main principle is that DUET will use data sources stored on external data platforms as much as possible. These can be hosted in the cloud or on-site by the data providers. In case there is a strict need to host the data ourselves, DUET (pilots) will host the data itself.

Several technical components will process and align the data for use in the Digital Twin. In the current architecture, the access & authorisation services, the knowledge graph API, the data & model catalogue API and the context graph API will take up that role. In addition, other data related messages will come from the IoT event services, the IoT time-series service and the geo-data service.

All those services will deliver information to a data broker combining message broker, message mapping & validation and message streaming functionalities.

The data itself will follow standardisation principles based on ISA² and W3C. The OGC/Inspire standards will be used for managing geospatial data. W3C Linked open data principles will be used for linking data semantically.

Data management of the DUET consortium project data

Google Drive™ is the selected tool as DUET's data and information repository. This includes the project deliverables (including relevant references utilised for their production or generated from them as project publications, e.g. journal articles, conference papers, e-books, manuals, guidelines, policy briefs etc.) and any other related information, including relevant datasets. It also implies that the privacy and security measures of Google Drive™ must be GDPR compliant. The verification of such circumstances is the responsibility of the coordinator. The procedures are described in the D8.2 Project Management Handbook.

2.5. Data sharing

Last but not least, the following table summarises the procedures for sharing DUET related data in a useful and legitimate manner. When sharing, it is of utmost importance to keep in mind not only the prescriptions and recommendations of extant rules and norms (including this DMP), as far as confidentiality and personal

data protection are concerned, but also the risk of voluntary or involuntary transfer of data from the inside to the outside of the European Economic Area (EEA). The principles and approach are part of deliverable D1.2, and D1.3 Cities Guide to Legal Compliance for Data-Driven Decision Making iteration 1 and 2. The specific impact of cloud storage and computing are described in D1.5 Ethical Principles for using Data-Driven Decision in the Cloud.

Nature of datasets	Confidential	Anonymised and Public	Non anonymised (temporary status)
Data usage scenarios			
Original data produced by the DUET consortium	Personal email communication Shared repositories Confidential deliverables Log-in and subscription data	Project website Open access repository DUET landing pages visualisation DUET Digital Twin visualisation Data catalogue	N/A
Existing data already in possession of the DUET consortium and/or partners and open data	Personal email communication Shared access to software repositories	Open access repository DUET landing pages visualisation DUET Digital Twin visualisation Data catalogue	N/A
Existing data sourced/procured by the DUET consortium and/or partners	Personal email communication Shared access to software repositories	Shared repository between the involved partners and 3rd party supplier. Results visible on the DUET Digital Twin visualisation and results are available in the Data catalogue.	N/A

Table 5: DUET data usage scenarios - data sharing overview

As for the previous tables, Annex 1 (“Running list of data sources”) provides some relevant case descriptions.

3. Allocation of resources

The costs for implementing the fair principles in the project are fully integrated into the project budget. They are shared amongst the pilot partners responsible for the data and involved in data modelling and simulation models (KUL, TNO, P4All, AIV, VCS, IMEC). The coordinator has a budget to finance open access to research data and research results like articles published in scientific journals like IEEE (Raes e.a.) and the upcoming Springer Nature book “: Decide Better: Open and Interoperable Local Digital Twins”. Data management is the responsibility of every consortium partner. Still, there are specific tasks for each pilot lead and the consortium coordinator responsible for the data management plan and ethics. The final long-term preservation decision is on the project direction level (consortium coordinator, senior user representative, and senior supplier representative).

4. Data security

In the DUET architecture, data security is a vertical architectural layer affecting the other layers of the DUET architecture. The D3.10 and D3.11 reports “Multi-layered security model specifications” are available. The following security measurement categories are in the implementation phase:

- Run-time authentication
- Run-time authorisation
- Secure and trusted communications
- Run-time monitoring and auditing
- Data protection and compliance
- Software development lifecycle security

The security approach is built around Technical Controls (TCs), and TCs 18 to 34 are specifically addressed to data security and data protection. These measures assure safe data storage for long-term preservation and curation.

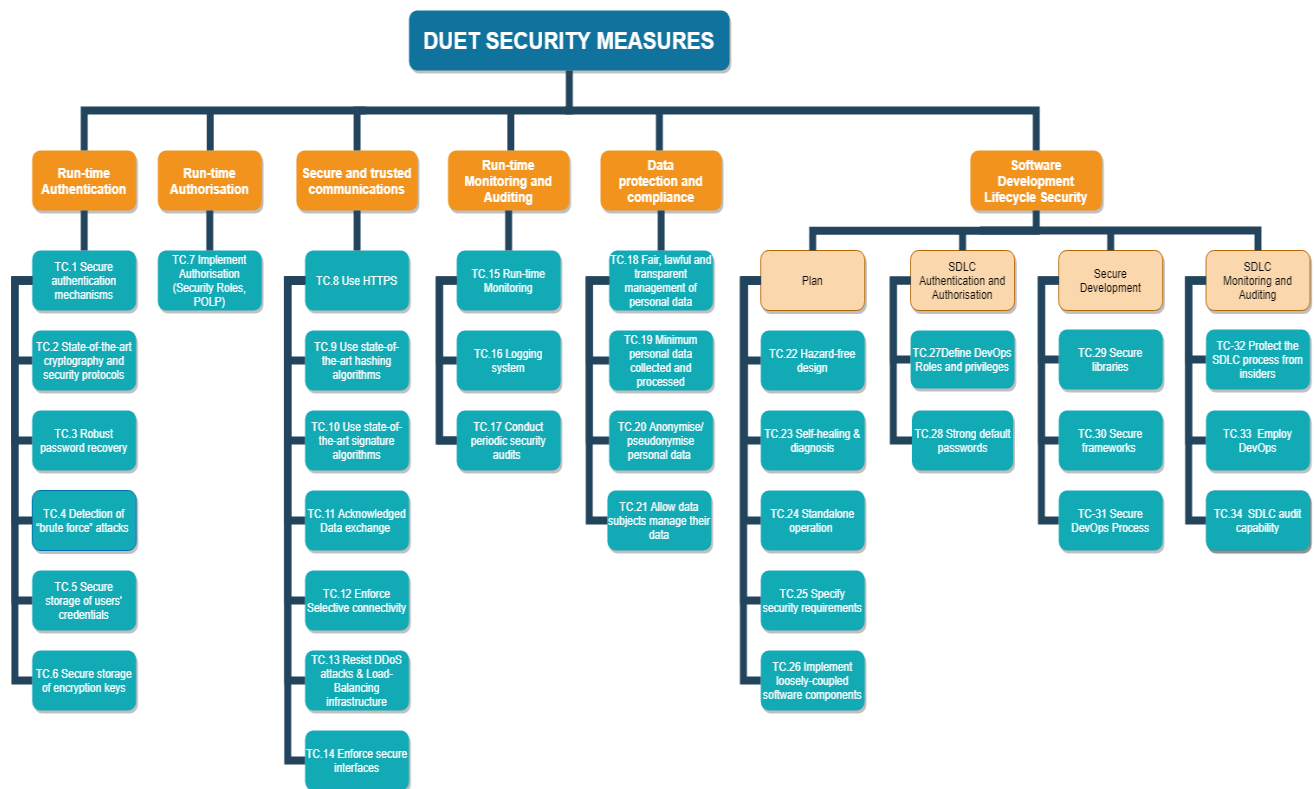


Figure 1: Taxonomy of DUET security measures and respective TCs

5. Ethical aspects

The ethical and legal aspects were scrutinised extensively during the project. Work package one, “Ethics, privacy and legal requirements for a European Cloud Infrastructure”, together with work packages nine and one, contains the legal and ethical requirements. The ethical aspects are about identifying legal and ethical considerations and providing an overview of ethical aspects. Next to that, measures and procedures are described to ensure legal compliance. Each of the pilots has been scrutinised on the impact of ethics and legal issues. In WP 9, procedures and criteria have been described and dealt with elements like recruiting and managing personal contact data. Especially D1.2 and D1.3 Cities Guide to Legal Compliance for Data-Driven Decision Making Iterations 1 and 2 deal with ethical aspects applied to each of the pilots.

In general, the DUET Digital Twins is not meant for personal advice and don't store or unlock personal data. Only the necessary (proportionality principle) log-in data is stored in a secured way. The used sensor data is an excellent example of privacy by design since no pictures are visible, recorded or sent over the network. Another example where personal data has been processed at the source are the anonymised census data used to create the origin/destination patterns of the traffic models.

The DUET Digital Urban Twin doesn't use any personal data defined in the GDPR. The anonymised sensor data, background data like OSM, 3D BIM models, and the used simulation models contain legal or ethical issues regarding the DUET internal data processing or visualisation output.

Nevertheless, the way data and model results are visualised has an ethical dimension. The DUET consortium is aware of the impact of the way data is visualised and results. The DUET landing pages provide detailed context information about each Digital Twin case, the approach, expected outcome (explaining what can be expected as an outcome), how to use information, and an overview of the used datasets and simulation models. This context information is essential to create the proper context for using the Digital Twin and interpreting results and outcomes.

The consortium has adhered to the H2020 ORDP (Open Research Data Pilot) convention with the EC, which explicitly caters to delivering a DMP (Data Management Plan).

6. References

- European Commission, Directorate-General for Research & Innovation (2016). *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020*, Version 3.0, July 26th.
- European Commission, Directorate-General for Research & Innovation (2017). *H2020 Programme Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020*, Version 3.2, March 21st.
- DUET Consortium Agreement (2019). Innovation action No. 870697. Project Title: Digital Urban European Twin for smarter decision making.
- DUET DoA (2019). Grant agreement No. 870697. Digital Urban European Twin for smarter decision making.
- Raes, L., Michiels, P., Adolphi, T., Tampere, C., Dalianis, T., Mcaleer, S., & Kogut, P. (2021). DUET: A Framework for Building Secure and Trusted Digital Twins of Smart Cities. *IEEE Internet Computing*. <https://doi.org/10.1109/mic.2021.3060962>

7. Annex 1: Datasets & models overview

The following running list of data sources will be used during the project to keep track. Further updates will appear in the forthcoming editions of this Deliverable.

Flanders datasets and models overview related to the epics realised in the open beta release:

Dataset name	Type	Theme	Description	Format	EPIC 1	EPIC 2	EPIC 3
	Dataset typology	Which of the DUET themes does it touch?	Brief text description in EN.	In which format(s) are the data available?	Difference in traffic density of traffic-closed roads vs base density	Impact on air pollution during road closures	Impact on noise pollution during road closures
GRB LOD 1 (Spatial reference database 3D LOD 1)	Geospatial dataset	Basemap	The Large-scale Reference File (GRB) is a geographic information system that serves as a topographical reference for Flanders. It is a common geographical basis on which all users can graft their own data. The GRB only contains geographic and characteristic information of well-defined, conventionally accepted reference data: buildings, plots, roads and their layout, watercourses, railways and the road network. LOD 1 is basically a 3D presentation based on buildings footprint.	Shape	Yes	Yes	Yes
GRB 2D Base layer	Geospatial dataset, geospatial webservice	Basemap	Idem. 2D representation.	Shape, WMS, WFS	Yes	Yes	Yes
Digital height model Flanders (1 meter, 5 meter, 25 meter, 100 meter)	Geospatial dataset	Basemap	Digital surface model of the ground level including objects in raster format with a ground resolution of 1 meter. This DSM was derived from LIDAR height data that was collected within the framework of the Digital Height Model Flanders II (DHMV II).	WMS		Yes	Yes
Orthofoto high scale (10cm)	Geospatial dataset, geospatial webservice		Covering medium-scale orthophoto coverage of the Flemish Region, including the Brussels-Capital Region. This assignment includes the realization of digital photographic aerial shots in the winter flying season with a ground resolution of 17 cm and, subsequently, the production of an ortho photo mosaic with a ground resolution of 25 cm.	WMS	Yes	Yes	Yes
Antwerp LOD 2 model	Geospatial dataset	Basemap	Idem GRB LOD 1 description, but also the roof shapes are represented.	ESRI GeoDB	Yes	Yes	Yes
Gent LOD 2 model	Geospatial dataset	Basemap	Idem GRB LOD 1 description, but also the roof shapes are represented.	DWG, DXF, CityGML (future)	Yes	Yes	Yes
Gent Trees	Geospatial dataset	Spatial planning	Trees on public domain (width & height)	GeoJSON, Shape	Yes	Yes	Yes
Road works (ongoing and planned) - GIPOD	Service	Mobility	Overview of ongoing and planned road occupations (road-works and events).	API	Yes	Yes	Yes
Air quality model - QUARK (VITO)						Yes	
Flemish multi modal traffic model	Simulation model	Mobility	Strategic multi modal models of persons- and freight transport in Flanders. Based on the BASMAT, MM and RMM instruments build with Cube Voyager.	Cube Voyager	Yes	Yes	Yes
MatLabTrafficToolbox	Simulation model	Spatial planning			Yes	Yes	Yes
Traffic model Gent	Simulation model	Mobility			Yes	Yes	Yes

Athens datasets and models overview related to the epics realised in the open beta release:

<<<Insert Table>>>

datasets and models overview related to the epics realised in the open beta release:

<<<Insert Table>>>

European datasets and models overview related to the epics realised in the open beta release:

Name	Type	Theme	Description	Case 1	Case 2	Case 3	Case 6	Case 7	Case 8	Case 13
Dataset name	Dataset typology			Difference in traffic density of traffic - closed roads vs base density	Impact on air pollution during road closures	Impact on noise pollution during road closures	Seeing current traffic flow based on the model and available sensors	Seeing current noise pollution based on the model and available sensors	Seeing current air pollution based on the model and available sensors	Testing green routes proposed by the city
Shops	Dataset	Spatial planning	Large world dataset containing the information of the shops https://wiki.openstreetmap.org/wiki/Map_Features#Shop	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Amenities	Dataset	Spatial planning	Large world dataset containing the information of the amenities https://wiki.openstreetmap.org/wiki/Map_Features#Amenities	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Public Transport	Dataset	Spatial planning, Mobility	Large world dataset containing the information of Public Transport https://wiki.openstreetmap.org/wiki/Map_Features#Public_Transport	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Building	Dataset	Spatial planning	Large world dataset containing the information of buildings https://wiki.openstreetmap.org/wiki/Map_Features#Building	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Roads	Dataset	Spatial planning	Large world dataset containing the information of the roads https://wiki.openstreetmap.org/wiki/Map_Features#Highway	Yes	Yes	Yes	Yes	Yes	Yes	Yes
OpenAQ	Sensor	Health and environment	Open air quality sensor data in Europe						Yes	