Deliverable

D4.2 DUET Data integration (demonstrator)

Project Acronym:	DUET	
Project title:	Digital Urban European Twins	
Grant Agreement No.	870697	
Website:	www.digitalurbantwins.eu	
Version:	1.0	
Date:	30 November 2020	
Responsible Partner:	TNO	
Contributing Partners:	IMEC, KUL, VCS, UWB	
Reviewers:	**Internal** Philippe Michiels (IMEC) Thomas Adolphi (VCS) **External** Yannis Charalabidis Michiel Van Peteghem Andrew Stott Pieter Morlion	
Dissemination Level:	Public	X
	Confidential – only consortium members and European Commission	

This project has received financial support from the European Union's Horizon 2020 Programme under grant agreement no. 870697

Revision History

Revision	Date	Author	Organization	Description
0.1	19.11.2020	Max Schreuder	TNO	Initial structure
0.2	20.11.2020	Walter Lohman	TNO	Contents
0.3	23.11.2020	Max Schreuder/Walter Lohman	TNO	for internal and external review
0.4	24.11.2020	Expert review	External experts	Expert review
0.5	26.11.2020	Max Schreuder/Philippe Michiels	TNO/imec	refine/update
0.6	27.11.2020	Walter Lohman/Max Schreuder	TNO	provision of final version
0.9	30.11.2020	Expert review	External experts	Expert review
1.0	30.11.2020	Max Schreuder	TNO	Final version

Table of Contents

Tables

Figures

Executive Summary

In the DUET Digital Twin system many different data sources are combined to deliver an integral digital version of the city; varying from simple indicators, 3D information of objects to output of live IoT data streams and output from computational models mapped to elements in the terrain. This document briefly describes the DUET Alpha Version demonstrator and explains the different data types and data sources; e.g. CityGML for 3D Objects, Cityflows IoT data streams, GeoJSON, Terrain data, Pointclouds, etc.

The DUET Alpha Version is able to show integration of the various data types including live data from IoT data streams. This gives a good indication of how different data sources can be integrated and processed not only in an End User Interface but also in a chain of publishers and consumers including models. It helps to understand the difficulties in the data integrating process and detect design problems at an early stage. Work done and lessons learned from the development of the DUET Alpha Version migrates to developing the final DUET system in Work Package 5 System Integration.

1. Introduction

Deliverable D4.2 is the demonstrator; the DUET Alpha Version. This document briefly describes data types and the integration of data in the DUET Alpha Version and gives an outline of data integration in the final DUET system. Data integration is work in progress.

The DUET Alpha Version demonstrator is discussed in the document 'Designing the Alpha Version'. The inner workings of data integration is detailed in the deliverables;

- **D3.1 IoT stack and API specifications v1** describes the components onboarding of data into DUET;
- **D3.5 Data integration** provides the generic description of the minimal API functions for model integration and discusses HPC;
- **D3.8 Digital Twin data broker specification and Tools v1** describes the high level roadmap and technical architecture and implementation tools and frameworks of the digital twin data broker from the viewpoint of DUET;
- **D5.1 System Architecture & Implementation Plan** discusses an overall architecture and implementation or deployment plan.

The DUET system is a Digital Twin system where many different data sources are combined to deliver an integral digital version of the city. As the term 'data' is a generic term, it can describe a wide range of different information types. Data varies from simple numbers, indicators and KPIs to detailed information about the terrain, 3D information of objects (buildings, trees, roads etc.), live IoT data and even output of computational models mapped to elements in the terrain.
The Message Streaming Platform in DUET also connects various data publishers and subscribers to data, making a flow of data though the system possible. IoT data streams for example, flow into the system and can not only be visualized, but also be part of a chain of computational models.

The DUET system aims to integrate the different data types and sources, make them comprehensible and useful for end-users, in e.g. visualisations, but also enables use of the data across the total chain of connected DUET modules. The integration of data across the DUET system is done using the DUET T-Cell with a data catalog, a data broker and (data) gateways. Visualisation systems for example will search for data using the data catalog and stream the data from supplied data links.

As a demonstrator an Alpha Version of DUET is developed. The DUET Alpha Version outlines the possibilities for end users to access various data sources and show how data can be integrated available from various sources such as output of complex computational models or IoT data available from live data streams. The DUET Alpha Version will give an insight in the handling and integration of data. It enlightens future work of the DUET development steps. See also Deliverable D3.8 Digital Twin data broker specification and Tools v1. This document briefly describes different data sources and discusses how the DUET Alpha Version retrieves and processes data.

The development of the DUET Alpha Version as a demonstrator is a valuable step in the project, as it enables detection of design problems in an early stage of the development process. Important aspects of the DUET system are implemented, to prove the architecture and principles work. Simulations and what-if scenarios are out of scope for the DUET Alpha version.

2. Supported use cases

Use cases to be supported in the DUET Alpha Version were selected based on consideration and discussion of the following criteria:

- Availability of the required data, or clear time horizon for availability of certain sets
- Potential legal hurdles
- Support from local policy makers / decision makers (high-level officials)
- Technical complexity and feasibility
- Alignment with the project ambitions (assess the similarities and differences between the cases)

Furthermore the project marked 3 cities; Athens (Greece), Flanders (Belgium) and Pilsen (Czech) for our Pilot studies using the DUET system. The ambitions (epics) for the final DUET system are formulated in detail for each of the pilot cities in **Deliverable D2.3 Final list of user requirements for the DUET solution**.

To fulfil the Pilot ambitions in the end, an early version of the DUET system, the Alpha Version, is developed. With the ambitions in mind, use cases are translated to Epics. Table 1 contains the list of epics that have been agreed with the pilot cities that will be part of the DUET Alpha Version.

Theme	Epic
Basis infrastructure - View current state	As a user of the digital twin, I can browse the **3D model of the area of interest** so I can get a detailed look at the surrounding of areas I want to inspect
Basis infrastructure - View current state	As a user, I can browse the **2D road network** on top of the 3D model so I can see where the roads are and can get extra attributes by clicking on a road
Basis infrastructure - View current state	As a user, I can see the **historical information** being sent out by **sensors** on top of the 3D model, so I get an accurate indication of the current local status
Traffic model	As a user, I see the **prediction of traffic flows** (e.g. the KUL, P4All model) as calculated in the traffic model of the area, so I can correlate what is displayed by the sensor with typical traffic flows
Mobility state estimation	As a user, I see the **measurements done by the sensors interpolated** (by a model that fuses information coming from different sensors from different sensor types) so I can get an approximation of the **density of people** also in places there are no sensors
Air quality model	As a user, I see the **real time air quality model** as calculated (by a model that fuses information coming from different sensors from different sensor types) so I can get an approximation of the **air quality** in places there are no sensors

Table 1: List of epics supported by the DUET Alpha Version

3. Datasets and formats

Data for Digital Twins have wide ranges of types, formats and sources. The data types used in the DUET systems are outlined in this chapter. The DUET Alpha Version demonstrator is able to show integration of these various data types. In the table the Epics are linked to well-known standard data types. In the following sections these data types are highlighted and briefly explained.

Epic	Relevant data types
As a user of the digital twin, I can browse the **3D model of the area of interest** so I can get a detailed look at the surrounding of areas I want to inspect	● CityGML for 3D Objects ● GeoJSON ● Terrain data ● Pointclouds ● Raster images
As a user, I can browse the **2D road network** on top of the 3D model so I can see where the roads are and can get extra attributes by clicking on a road	● GeoJSON ● Terrain data ● Information (Text) linked to geographical location
As a user, I can see the **historical information** being sent out by **sensors** on top of the 3D model, so I get an accurate indication of the current local status	● IoT/Cityflows ● Model data ● CityGML for 3D Objects
As a user, I see the **prediction of traffic flows** (e.g. the KUL, P4All model) as calculated in the traffic model of the area, so I can correlate what is displayed by the sensor with typical traffic flows	● IoT/Cityflows ● Model data ● CityGML for 3D Objects
As a user, I see the **measurements done by the sensors interpolated** (by a model that fuses information coming from different sensors from different sensor types) so I can get an approximation of the **density of people** also in places there are no sensors	● IoT/Cityflows ● Model data ● GeoJSON
As a user, I see the **real time air quality model** as calculated (by a model that fuses information coming from different sensors from different sensor types) so I can get an approximation of the **air quality** in places there are no sensors	● IoT/Cityflows ● GeoJSON ● Model data

Table 2: Epics linked to relevant data types

The following sections highlight and briefly explain the following data types:

- IoT/Cityflows
- CityGML for 3D Objects (buildings, trees, roads, etc.)
- GeoJSON
- Terrain data
- Pointclouds
- Raster images
- Model data
- Information (Text) linked to geographical location

3.1 IoT/Cityflows data streams

For the cityflows data streams we make use of virtual sensors. Data from several providers (Citymesh, Proximus, Telraam, …) is captured using different API/web polling agents that take the data, translate the data into NGSI-v2 TrafficFlowObserved entities and submit the data points as updates for virtual sensors to the imec City of Things NGSI context broker.

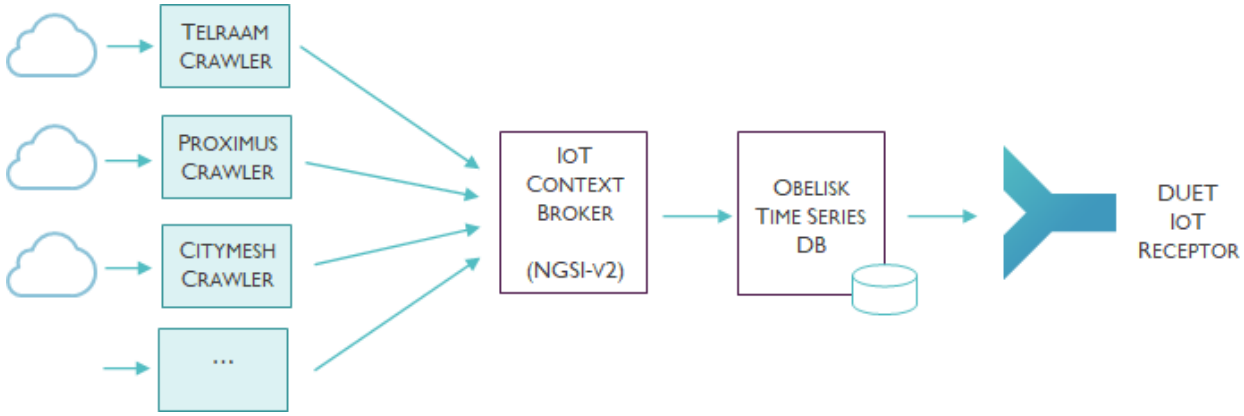

Figure 1: Cityflows data flow

An NGSI compliant time series database Obelisk is used to store the time series data. This is done using the NGSI subscription protocol.

```
{
  "refRoadSegment": 420147,
  "refDevice": "urn:ngsi-v2:cot-imec-be:device:telraam-420147",
  "intensity": 0,
  "source": 420147,
  "dateObservedTo": "2020-11-26T04:59:00Z",
  "dateObservedFrom": "2020-11-26T04:00:00Z",
  "dataSource": "telraam",
  "dateObserved": "2020-11-26T04:59:00Z",
  "areaCovered": {
        "type": "LineString",
        "coordinates": [...]
  },
  "area_covered": {
        "type": "LineString",
        "coordinates": [...]
  },
  "vehicleType": "car",
  "direction": {
        "flowUp": {
          "type": "Number",
          "value": 0
        },
        "flowDown": {
          "type": "Number",
          "value": 0
        },
        "bearing": {
          "type": "Number",
```

```
          "value": 350.22457386529817
        }
    }
}
```


Figure 2: An NGSI-V2 TrafficFlowObserved sample from the Telraam data source

The cityflows model in turn consumes the TrafficFlowObserved data that is published in DUET, processes it and returns a new TrafficFlowObserved stream that covers the full extent of the road network.

3.2 CityGML for 3D Objects

Most of the 3D objects in the DUET system are described using CityGML[1].

CityGML is an open standardised data model and exchange format to store digital 3D models of cities and landscapes. It defines ways to describe most of the common 3D features and objects found in cities (such as buildings, roads, rivers, bridges, vegetation and city furniture) and the relationships between them. It also defines different standard levels of detail (LoDs) for the 3D objects, which allows the representation of objects for different applications and purposes, such as simulations, urban data mining, facility management, and thematic inquiries.

Figure 3: Pilsen CityGML example

[1] http://www.citygml.org/

3.3 GeoJSON

To link information to a geographical location, the GeoJSON (RFC 7946)[2] format is used.

GeoJSON is a geospatial data interchange format based on JavaScript Object Notation (JSON). It defines several types of JSON objects and the manner in which they are combined to represent data about geographic features, their properties, and their spatial extents. GeoJSON uses a geographic coordinate reference system, World Geodetic System 1984, and units of decimal degrees.

3.4 Terrain data

Terrain data is usually represented as DGM files; raster files containing a height map of the terrain. These are grayscale or colored images, where the pixels contain information about the terrain.

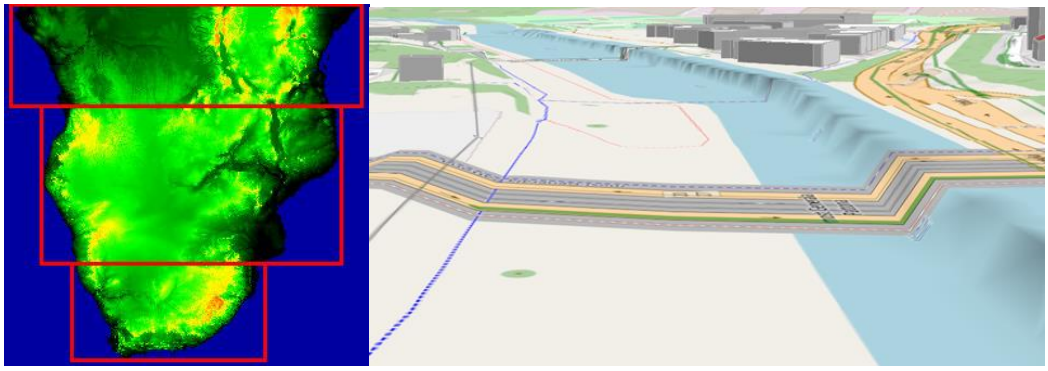

Figure 4: Examples of colour heightmap(left) and mapping of height data in the terrain (right)

Terrain data can also be represented as TIN (Triangular Irregular Network)[3]

A triangulated irregular network (TIN) is a representation of a continuous surface consisting entirely of triangular facets, used mainly as Discrete Global Grid in primary elevation modelling.

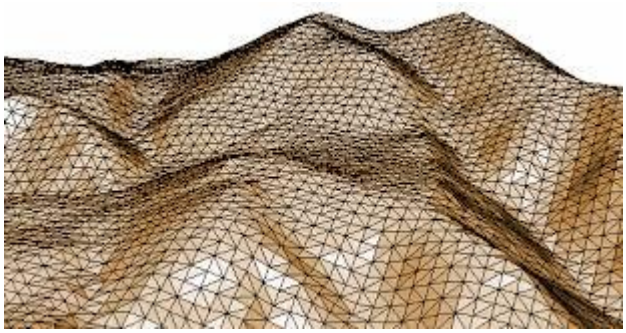

Figure 5: Example of a triangulated irregular network

[2] https://geojson.org/ , https://tools.ietf.org/html/rfc7946
[3] (https://en.wikipedia.org/wiki/Triangulated_irregular_network)

3.5 Point clouds

A bit similar to the DGM files, are point clouds[4]. Where DGM files are mostly bitmaps, where pixel information is used for spatial data, point clouds are just a set of points. The points are geo referenced and have additional information such as distance attached. When a terrain or 3D object is scanned, sets of points can be acquired. This information can be used for example, to analyse roof types or to enhance satellite images with depth information.

A point cloud is a set of data points in space. The points represent a 3D shape or object. Each point has its set of X, Y and Z coordinates. Point clouds are generally produced by 3D scanners or by photogrammetry software, which measure many points on the external surfaces of objects around them. As the output of 3D scanning processes, point clouds are used for many purposes, including to create 3D CAD models for manufactured parts, for metrology and quality inspection, and for a multitude of visualization, animation, rendering and mass customization applications.

Point clouds can be generated from aerial measurements.

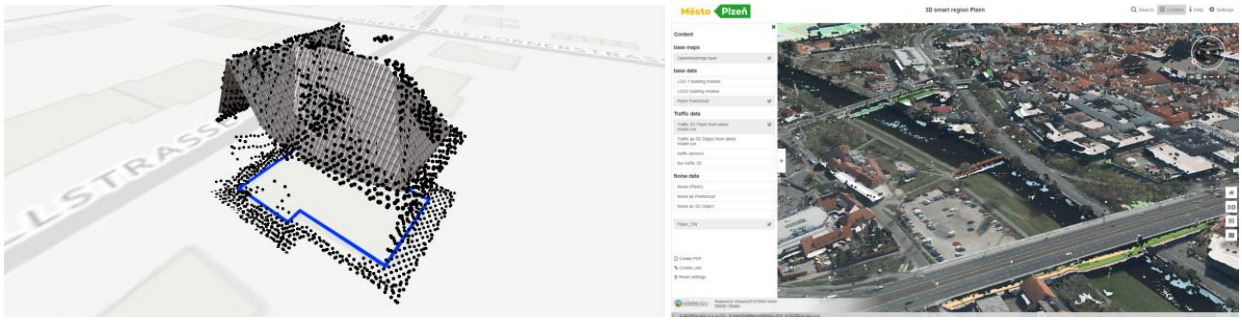

Figure 6: Lidar roof mesh example[5] (left) and integration of point clouds in city data (right)

3.6 Raster images

Raster images in general are digital images . The representation of the data is pixelated forming the image. When data has a geo reference, the data can be placed on a map. In practice, the raster images are most commonly divided into 'tiles' so only the necessary part of the raster is brought into view. Tiles can be requested from 'tile services' using WMTS (Web Map Tile Service) or TMS (Tile Map Service).

[4] https://en.wikipedia.org/wiki/Point_cloud
[5] https://marian42.de/article/roofmeshes/

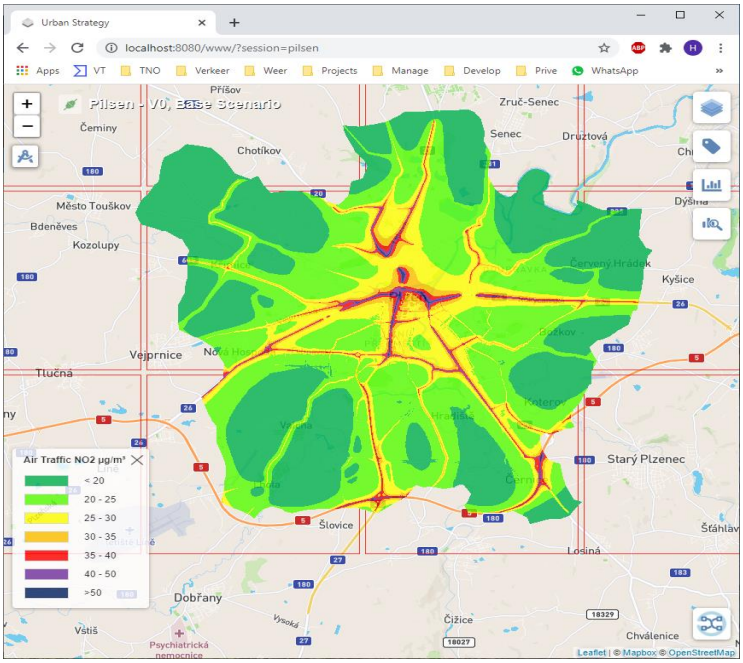

Figure 7: Air pollution data placed in Tiles for the city of Pilsen.

3.7 Model data and information linked to geolocation

When computational models are used, the simulated or modelled output can be placed on a map and also be integrated with other information of the mapped area. Output from Traffic prediction models, Noise Pollution or Air Pollution can be integrated with other data. This can be additional data for buildings (Location type) or additional data for roads (Street names), which can be overlaid on the map, as well as KPI data (graphs), regarding the objects in view that can be integrated.

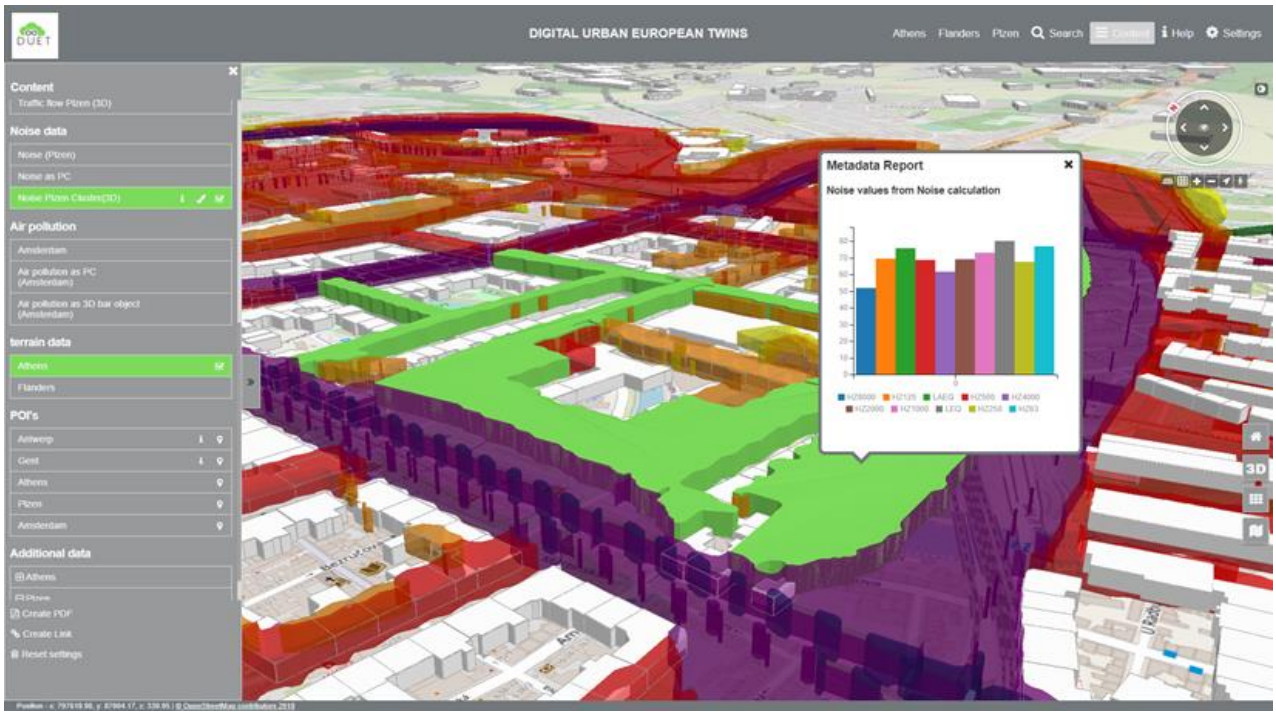

Figure 8: Noise pollution overlay and KPIs in graphs

4. Data integration

In the Alpha version of DUET, data integration is done at the viewer (end-user visualisation) level. This approach enables the project, at this point in the development, to formulate and assess the requirements for data integration, to ensure that the end-users are presented with comprehensible and useful information and visualisations.

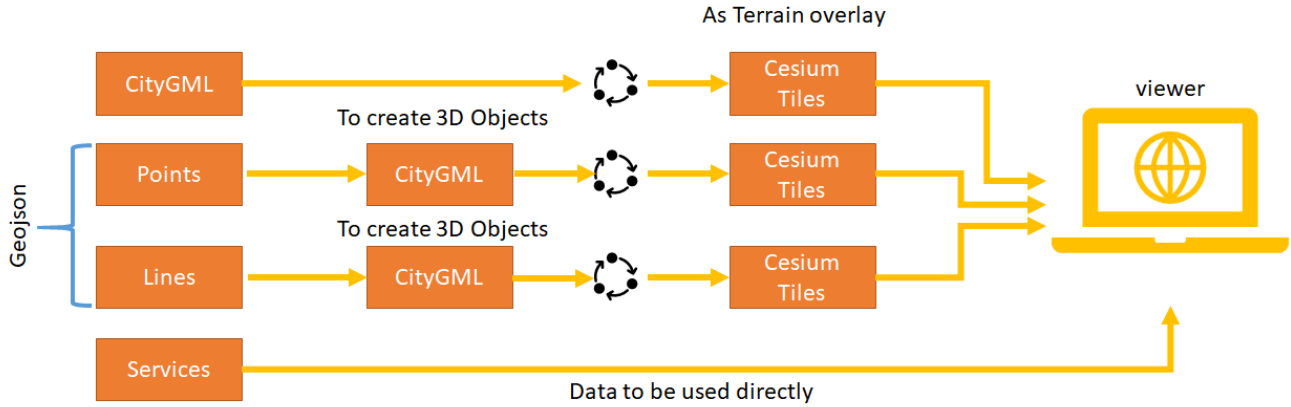

Figure 9: Alpha version workflow to integrate data

The figure outlines 3 major pipelines for data processing. Points and Lines data are converted into CityGML format. CityGML data is converted into Cesium[6] Tiles and rendered by the 3D Cesium Web Based viewer. Services provide data that can be processed directly in the viewer. IoT Cityflow data, for example, is requested through the DUET services and integrated in the viewer. The end-user can view the map area of interest and combine sources of data prerendered to Cesium Tiles or from a direct connected DUET Services.

Figure 10: DUET Alpha version data integration of traffic data and digital terrain (buildings, etc)

[6] Cesium is a platform for 3D geospatial data handling. The basis is 3D tiling, which transforms 3D geospatial data into streamable 3D content. Cesium can layer together data from multiple sources. https://cesium.com/

5. Future work and conclusion

The final version of the DUET system aims to have an advanced method of making data available for integration. The T-Cell design with the data catalog and data broker together with the relevant gateways will interconnect the different sources for streaming toward the visualisation system. See figure 11 and figure 12.

In the DUET architecture from the visualisation perspective, requested data is looked up in the data catalog, providing the links to external data storage. Gateways connected to the T-Cell will provide the necessary conversion and transitions to stream to the viewer.

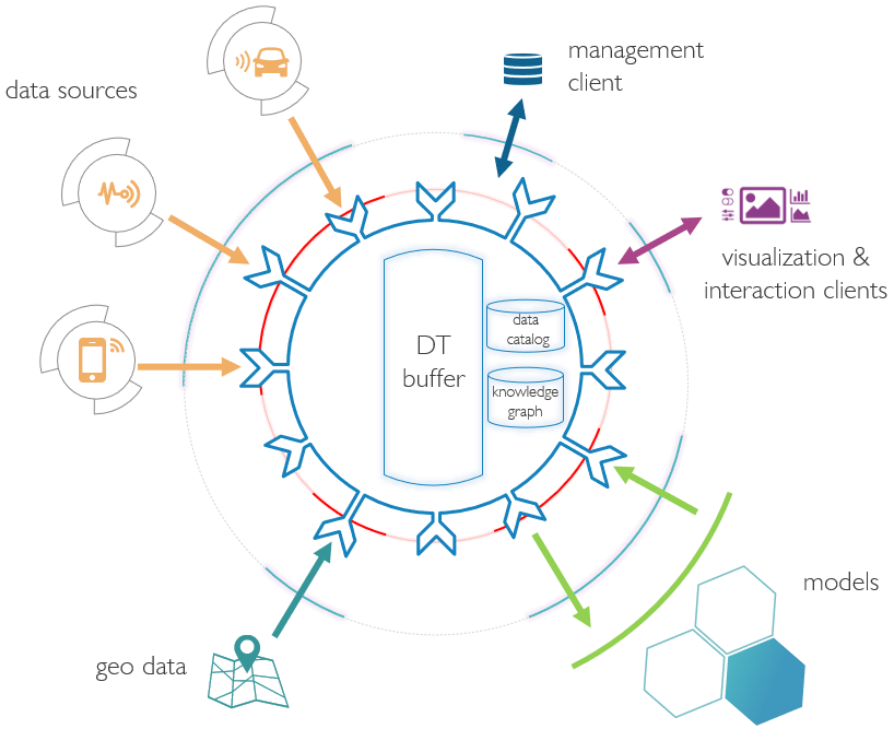

Figure 11: DUET T-Cell & Data[7]

The Alpha version of DUET gives a good indication how different datasources can be integrated and processed in an End-User Interface. It helps to understand the difficulties in the data integrating process and detect design problems at an early stage. Work done for the development of the Alpha version migrates to developing the final DUET system.

[7] source: Deliverable D3.8 Digital Twin data broker specification and Tools v1

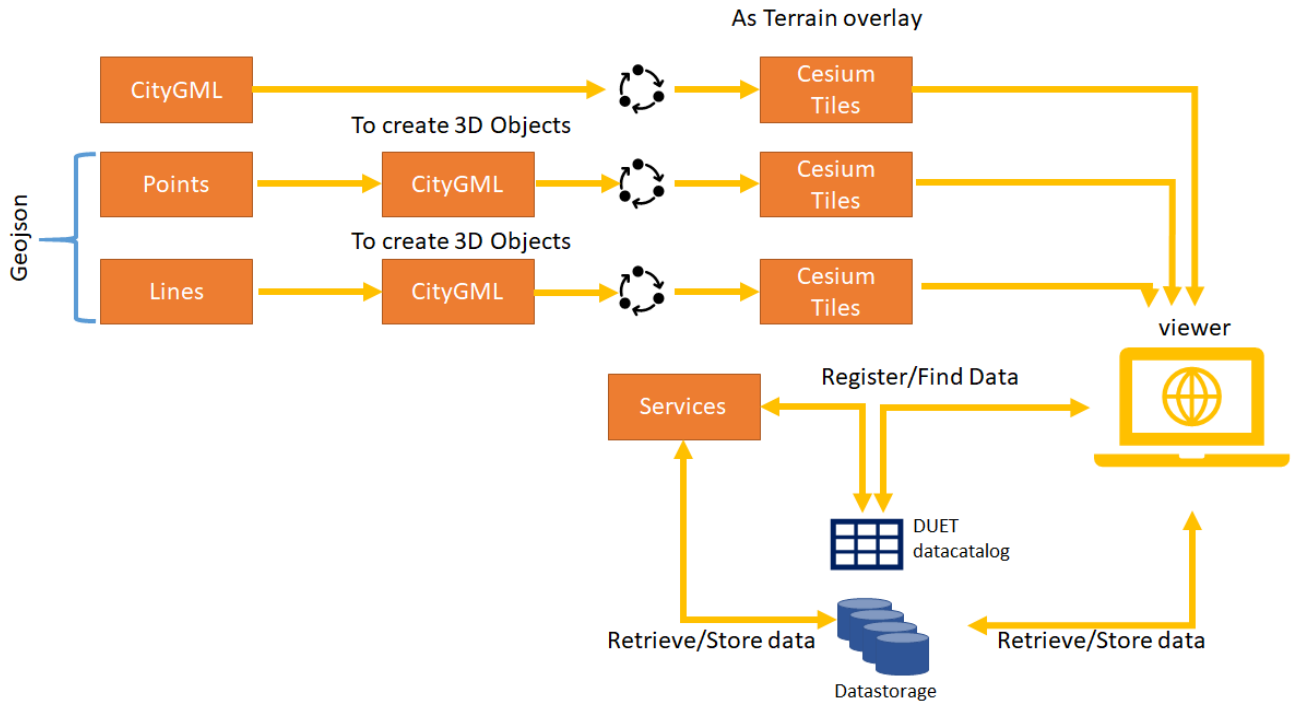

Figure 12: Future workflow for data integration in the viewer (end user interface)

In the final version of the DUET system, data will be located by a query to the DUET data catalog. It will point to the location of data. This can be stored in a Data storage or can be directly available via a Services API (IoT data/Cityflows).

6. References

- Demonstrator DUET Alpha Version
- Document 'Designing the Alpha Version'
- Deliverable D2.3 Final list of user requirements for the DUET solution
- Deliverable D3.8 Digital Twin data broker specification and Tools v1
- Deliverable D5.1 System Architecture & Implementation Plan
- http://www.citygml.org/
- https://geojson.org/
- https://tools.ietf.org/html/rfc7946
- https://en.wikipedia.org/wiki/Triangulated_irregular_network
- https://en.wikipedia.org/wiki/Point_cloud
- https://marian42.de/article/roofmeshes/
- https://cesium.com/