D2.1 Initial release of the Digital Twin Platform & Architecture

Preliminary release of the Digital Twin serving as a solid base to expand on further digital twin development.

Project name	BIPED: Building Intelligent Positive Energy Districts
Duration	January 2024 – December 2026
Project ID	101139060
Coordinator	Technical University of Denmark
Type of action	Innovation Action
Call ID	HORIZON-MISS-2023-CIT-01
Website	https://www.bi-ped.eu/
Document name	D2.1. Initial release of the digital twin platform & Architecture: preliminary release of the Digital Twin serving as a solid base to expand on further digital twin development.
Document status	Final
Delivery date	28 June 2024
Dissemination	Public
Authors	Eva Schmitz (DKSR), Lukas Koch (DKSR), Max Serra (DKSR), Martin Traunmueller (AIT), Arne Schilling (VCS), Shahrzad Pour (DTU), Johanne Bräuner Nygaard Hansen (AAKS), Jan Jezek (INNO)

Funded by the European Union

BIPED is funded under the EU Horizon Europe Research and Innovation programme. Grant ID: 101139060. Every effort has been made to ensure the accuracy of information provided in this document, which reflects only the authors' views. The European Commission is not responsible for any use that may be made of the information contained herein. Credits: cover image from Adobe Stock.

Document history

Version	Date	Contributor	Description
0.1	17.04.2024	Martin Traunmueller (AIT)	Table of Contents
0.2	20.05.2024	Shahrzad Pour (DTU), Martin Traunmueller (AIT)	First draft
0.3	29.05.2024	Karel Jedlička (UWB), Tomáš Andrš (RT), Jan Ježek (INNO)	Contribution
0.4	05.06.2024	Max Serra, Eva Schmitz (DKSR)	Contribution; Refinement
0.5	07.06.2024	Lukas Koch (DKSR)	Contribution
0.6	14.06.2024	Reviewers Konstantinos Tsiolis (DTU), Milos Sipetic (AIT)	Review 1
0.7	21.06.2024	Review by WP Lead Martin Traunmueller (AIT)	Review 2
0.9	26.06.2024	Eva Schmitz (DKSR)	Review 3
1.0	28.06.2024	Eva Schmitz (DKSR)	Final version

Table of Contents

List of Figures

List of Tables

List of Abbreviations/Glossary

Abbreviation	Definition
AOI	Area of Interest
AMQP	Advanced Message Queuing Protocol
API	Application Programming Interface
BIPED	Building Intelligent Positive Energy Districts
DMI	Danish Meteorological Institute
DMP	Data Management Plan
GDPR	General Data Protection Regulation
glTF	GL Transmission Format
GPU	Graphics processing unit
HTTPS	Hypertext Transfer Protocol Secure
IAM	Identity and Access Management
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
KTX2	KHRONOS Texture Compression Format
LWJGL	LightWeight Java Game Library
MIMs	Minimal Interoperability Mechanisms
MQTT	Message Queuing Telemetry Transport
OGC	Open Geospatial Consortium
OUP	Open Urban Platform
PED	Positive Energy District
REST	Representational State Transfer
SFTP	Secure File Transfer Protocol
SRS	Software Requirement Specification
SSL	Secure Sockets Layer
WP	Work Package

Executive summary

This report documents the process and outcome of the BIPED digital twin in its initial stage after 6 months of conception and development. BIPED aims to provide a holistic view on the subject of Positive Energy Districts (PED) development for the testbed of Braband, Aarhus, by establishing a digital twin platform that includes energy, mobility and cross-sectoral properties describing the urban landscape and its society. This development describes an organic process that will continue to grow and expand throughout the project's timeline and be presented in upcoming deliverables (D2.2, D2.3, D2.4). The initial release of the BIPED digital twin D2.1 establishes a technologically solid platform serving as base for future extensions and refinements. As part of the initial release of the BIPED digital twin platform demonstrator, this report provides in-depth background information on the development within the first 6 months of the project, outlining the methodological approach and scope as well as detailing technical components and data sets under examination. The report further offers the user a handbook or guide (Chapter 5) on how to use BIPED digital twin in its current state.

The document is structured into 6 chapters:

- **Chapter 1 "Introduction"** provides an outline about the Initial release of the digital twin platform and architecture, serving as a solid base for the BIPED digital twin to expand on further development steps. The chapter includes an overview about the methodological steps that have been taken, describes how this deliverable relates to the other deliverables of this project's WP3 (D3.1) and WP4 (D4.1) and concludes with outlining the scope of this deliverable.
- **Chapter 2 "Requirements"** discusses Technical Requirements, originating from the technical partners, and the Software Requirement Specifications (SRS), defined by Functional and Non-Functional Requirements and related Constraints.
- **Chapter 3 "Digital twin architecture and components"** presents the technical architecture of the BIPED digital twin, which will be continuously expanded over the course of the project. The chapter starts with the General Architecture of the platform, continues with related Backend Components, Models and Algorithms, and Frontend Components.
- **Chapter 4 "Demonstrator of the BIPED Digital Twin Platform (Initial Release)"** offers guidance to the user of BIPED digital twin in its current state, offering an interface description including a 2D and 3D map, a legend and a description of the toolbox. Furthermore, this chapter describes how to access live data from sensors located in Aarhus.
- **Chapter 5 "Data Sets and Data Sources"** is structured by three main topics, covered by BIPED (Energy, Mobility, Cross-sectoral) and describes related datasets that have been included into BIPED digital twin in this stage or have been under examination for inclusion. The chapter continues with the involvement of stakeholders, end-users and current actions taken towards the embedment into the European Data Space landscape, by providing an outline of the current EU-defined priority data sets, including the EC High-value datasets and INSPIRE Geo Data Portal, and discusses how BIPED has been nested deep within the European data space landscape.

- **Chapter 6 “Conclusions and Next Steps”** provides an outlook on how these findings will contribute in next working steps and deliverables, together with findings of other work packages as presented in their respective deliverables, to define a holistic digital twin supporting PED establishment in communities and cities.

The following documents are attached to this document:

- Annex 1: DUET Maturity Model
- Annex 2: BIPED Datasheet
- Annex 3: Data Sources

1. Introduction

This deliverable consists of the conception and description of the software architecture for the BIPED digital twin and an initial release of platform components as a demonstrator. This report was produced to accompany and document the current work process as of M06. The report offers background information and detailed insights on technical components and software development, the methodology used, as well as a handbook as guidance to the user of the demonstrator.

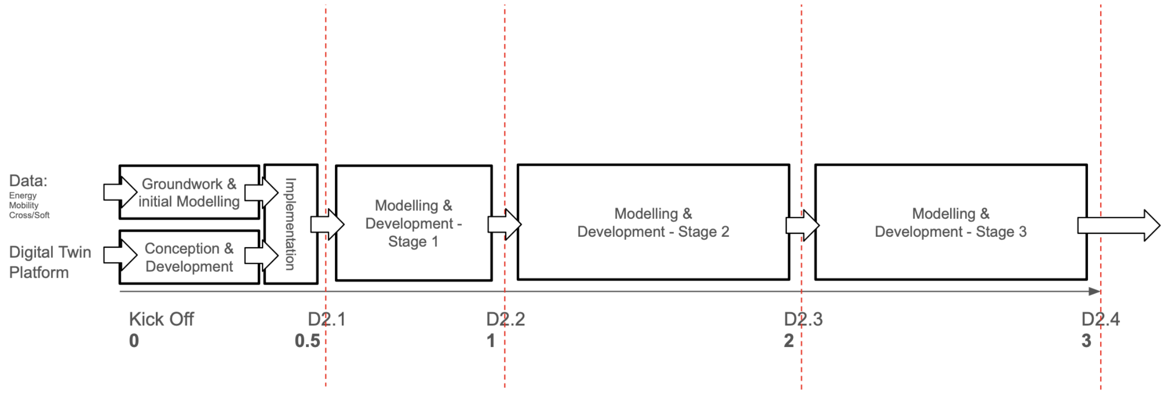


Figure 1. Schematic workplan of BIPED Digital Twin: M01-M36

Developing a holistic digital twin with the aim to support Positive Energy District (PED) establishment in the neighbourhood of Braband in Aarhus, that goes beyond commonly discussed properties, such as energy consumption and mobility, can only be achieved in an iterative process. While the data integration and calculation models have to be built up gradually and are interdependent, a solid technical basis lays the foundation for the software architecture.

The initial release of BIPED Digital Twin (D2.1) aims to provide a technically robust platform as a solid, yet scalable base to start this process. The basic architecture is scalable and designed to flexibly serve new use cases in future releases (D2.2, D2.3, D2.4 - see Figure 1). This comes with the inclusion of more data sources and data models in incremental modelling and development stages to be published. Figure 1 shows a schematic workplan of the overall project (M01-M36).

As outlined in Figure 1, the process in this stage focuses on merging and implementing three parallel development streams – the overall understanding & concept phase, technical development of the digital twin backend (T2.1) as technical basis, and the initial data-related groundwork, including the exploration and evaluation of available data sources (T2.2, T2.3, T2.4).

The Digital Twin Maturity Model (Figure 2) developed in the Horizon 2020 project DUET[1] (Digital Urban European Twins - Grant Agreement No. 870697) helps to validate the organisation and status of the initial work phase. The model describes a high-level pathway for the development of digital twins. The BIPED work phase can be categorised into the Strategy Phase and Exploratory Phase.The Strategy Phase has already been completed for

[1] DUET. https://www.digitalurbantwins.com/

the most part with processing the BIPED project proposal in close collaboration with local stakeholders. The search for suitable data sets is also seen as part of this first phase. This part of the work is still in progress. The current technological development has strong characteristics of the explorative "Exploratory Phase".

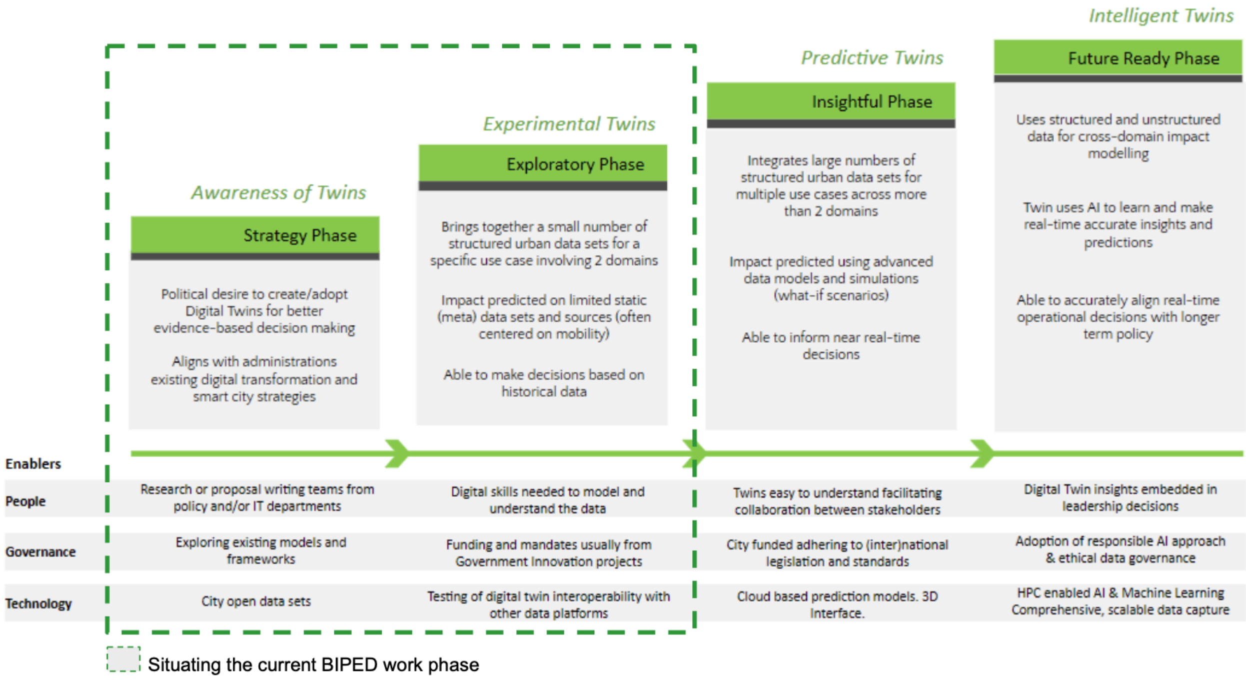


Figure 2. DUET Maturity Model (see also Annex 1)

The BIPED digital twin will serve the city government, decision makers, the local community in Braband/Aarhus, as well as public and private stakeholders in making informed decisions towards PED planning – with the potential to adopt the concept in communities all over Europe. Therefore, the technical development has been aligned with activities in the work packages WP3 and WP4 as well as related work on stakeholder and end-user involvement and the engagement with broader data space communities in T2.5.

The steady communication and feedback from those stakeholder groups ensures a digital twin design serving real-world challenges towards PED establishment for the community by defining specific user requirements, while opening doors to access highly relevant, yet restricted data sets by raising awareness of data owners as extended project partners.

Furthermore, based on the project's aim of establishing BIPED digital twin not only in Braband in Aarhus, Denmark, but in towns and cities all over Europe, the development of BIPED digital twin is deeply grounded in the European data space environment. In doing so, our technical development has been guided by this principle. For instance, the incorporation of open high quality data sources and priority datasets, as defined by the European Commission (EC) has been seen as a priority. This goal is also embedded in the T2.5 activities with key relevant European data space initiatives, raising awareness, exploring opportunities for cooperation and assessing broader applicability and replicability of BIPED solutions beyond the pilot setting.

1.1. Scope & Objectives

To accomplish Deliverable 2.1, three main activities were carried out by aligning and integrating contributions from various partners and work packages (WPs): Conceptual Work, Data Mapping and initial Technical Implementation.
The scope encompasses collaborative efforts of joint knowledge building, architectural design, definition of software requirements, basic technical development, data mapping and handling, and synchronisation with parallel workstreams.

The primary objectives for Deliverable 2.1 are:

1. Establish a joint Knowledge Base between partners:
    - Facilitate effective communication and knowledge sharing among all partners.
    - Understand each partner's capabilities and technical solutions
    - Document and disseminate the collective knowledge to support ongoing and future project phases.
2. Achieve Accurate Data Mapping and Initial Modelling:
    - Conduct thorough data mapping to understand the data landscape and requirements.
    - Develop initial data models that are accurate and align with the architectural design.
3. Define Basic Technical Requirements for the BIPED Twin Development
    - Collect Requirements from potential future end-users as an orientation.
    - Collect Requirements from the Technical partners to ensure a robust base.
    - Specify general Software Requirements.
4. Design a Robust Architecture for Digital Twin (Initial Release):
    - Create a scalable and flexible architecture that can serve as the foundation for the Digital Twin / specify the initial schematic architecture.
    - Ensure the architecture supports integration with various data sources and technical components and models.
5. Develop Core Technical Components:
    - Implement the essential technical components that will form the backbone of the Digital Twin.
    - Ensure these components are modular and can be easily extended or modified.
6. Integrate and Align with Parallel Workstreams:
    - Incorporate stakeholder and end-user feedback from WP3 and KPI requirements from WP4 into the Digital Twin's design and implementation.
    - Ensure the Digital Twin development is in sync with WP3 and WP4 to meet broader project goals.

1.2. Methodological Approach

By achieving the objectives above, the work package aims to deliver a solid foundation for the Digital Twin, ensuring it is well-positioned for future development phases and effectively meets the needs of all stakeholders involved. Figure 3 (below) outlines this process, which will be described in more detail below. In the following months, an agile development approach will be established to incrementally advance the platform and integrate more data based on the refinement of User Requirements.

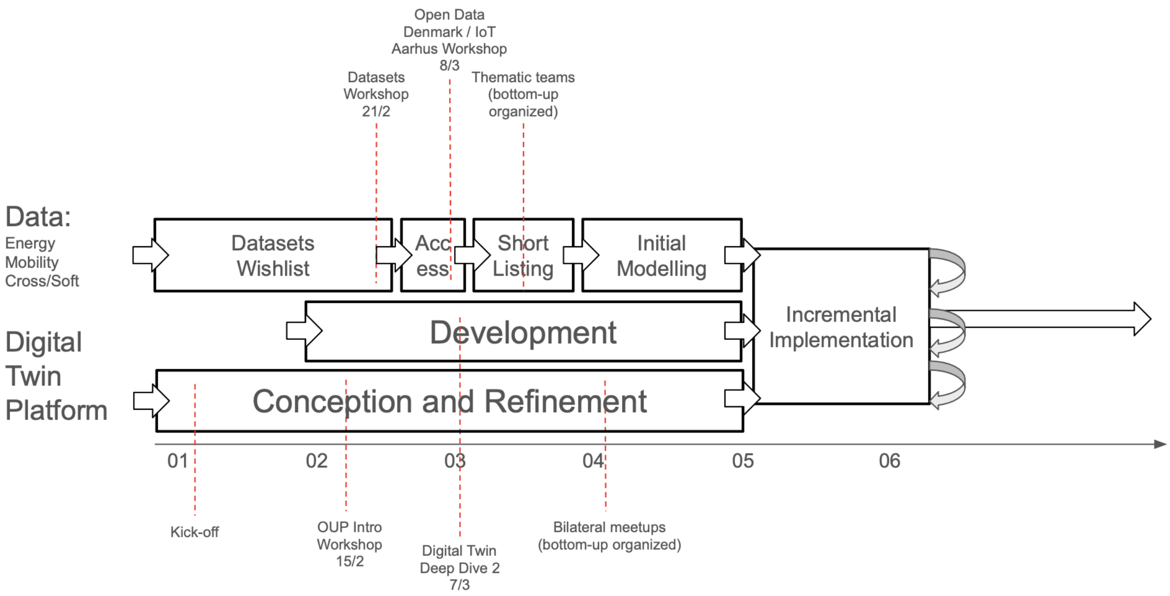


Figure 3. Schematic workplan of BIPED Digital Twin: M01-M06

1.2.1. Data

1.2.1.1. "Dataset Wishlist": BIPED Data Sheet

In a first step, variables describing the urban and social landscape of Brabrand have been discussed and collected in a "dataset wishlist" and linked to representative datasets as found in open data portals, such as Open Data Denmark[2] and Aarhuskommune[3] (see Annex 3 for the complete list of discussed data sources). This data sheet serves as living document and tool for the project team providing an overview on discussed datasets, additional information (e.g. accessibility, level of data granularity, ownership) and the current state of the dataset within the BIPED digital twin project ("available", "owner-contacted", "owner-not-contacted", "not-available") throughout the project duration. (See Annex 3).

While topics related to energy and mobility are firmly established in the scientific communities for modelling PED development, relevant and necessary datasets for modelling purposes could therefore be included in the BIPED Data Sheet rather quickly, driven by the expertise of the consortium partners. However, the inclusion of cross-sectoral data such as datasets describing the social, economic, environmental and physical characteristics of a neighbourhood or city, requires an exploratory phase.This exploration has to be thorough due to the novelty of including such less-discussed, while potentially significant properties in an urban PED-related context. On the team's expertise in urban data science topics, such properties have been discussed with respect to their relevance to energy consumption in an urban neighbourhood. Datasets reflecting selected properties have been identified and added to the data sheet.

[2] Open Data Denmark: https://www.opendata.dk/city-of-aarhus
[3] Aarhuskommune: http://aarhuskommune.dk/

Besides such quantitative measurable data, BIPED will also include the collection and exploration of cross-sectoral soft data in next development. As outlined in D1.3, the European Commission/Eurostat glossary defines soft data as qualitative or quantitative data resulting from an approximation of economic phenomena through surveys and interviews., besides other sources, such as social media scraping and various crowdsourcing methods.

For the purposes of BIPED, soft data are mainly data that attempt to describe and quantify people's perceptions or emotions in order to complement hard data models. According to the proposal, one of the objectives of BIPED is to explore how the quantitative collection of soft data can support progress in the development of digital twins, and to develop novel methods to quantitatively collect such data so that it can be incorporated into the digital twin.

1.2.1.2. Access & Shortlisting

In doing so, a rather excessive and well defined pool of datasets has been established and in a next step, a short-list on prioritized datasets created. The shortlisting process followed fact-based decisions on how properties relate to energy consumption (for instance, as suggested by the literature) as well as their accessibility (as for instance, easy accessible datasets were prioritized). Relevant data owners have been contacted and discussions on data sharing initiated (see a list of prioritized data sets in Table 1, Chapter 4).

Having defined a short-list with necessary datasets allowed the establishment of sub-groups for energy, mobility and cross-sectoral modelling: Lead partners of related tasks (T2.2, T2.3, T2.4) organized task-specific recurrent meetings with involved partners to share insights and development steps as well as to plan next progress steps to be taken.

Due to this process, two primary challenges have become clear:

- Accessibility and Sharing of Datasets by Third-Party Owners: Extensive efforts have been made to gather comprehensive information from various sources. We have established contacts with local energy providers for energy consumption data, the Aarhus Traffic Department for the city's traffic model and sensor data, various Aarhus Municipality departments for demographic and socio-economic data, and the Aarhus GIS Department for geospatial data necessary to build the digital twin. However, finalising agreements and procedures to access this data remains ongoing.
- Existing Language Barriers: Many open datasets are available only in Danish, which complicates data processing and integration. Despite these challenges, we are working diligently on translating and interpreting these datasets to ensure their usability.

As of May 2024, we are in contact with these departments to finalise necessary agreements and procedures to access the data and proceed with our data integration and analysis tasks (see Table 1, Chapter 4).

1.2.1.3. Modelling

Following up on the challenges expressed above, the comprehensive AI-based modelling has not yet fully commenced. However, several preliminary steps have been undertaken, and we have outlined the planned next steps as data becomes available.

With AI techniques, we refer to the AI taxonomy presented in the "AI Watch. Defining Artificial Intelligence 2.0"[4] document by the European Commission. The AI Watch is designed to establish AI from an operational perspective. To achieve this, the core domain of AI and transversal topics are characterized, along with a list of subdomains. Each

[4] https://ai-watch.ec.europa.eu/publications/ai-watch-defining-artificial-intelligence-20_en

subdomain is provided with a list of keywords. AI domains are considered as "Reasoning, Planning, Learning, Communication, Perception, Integration and Interaction, Services, AI Ethics and Philosophy." Part of taxonomy, including AI domains, their subdomains, and a list of keywords, is shown in the table below. The reader is referred to Table 2 of the AI Watch taxonomy for the complete list. To cope with changes in the practicality of methods in different domains, the keywords can be dynamically added to the list. Therefore, with AI, we mean any algorithms that can be used as part of extracting value from data.

Table 1. Part of EU AI Watch Taxonomy

AI domain	AI subdomain	Keyword	
Reasoning	Knowledge representation; Automated reasoning; Common sense reasoning	case-based reasoning	inductive programming
		causal inference	information theory
		causal models	knowledge representation & reasoning
		common-sense reasoning	latent variable models
		expert system	semantic web
		fuzzy logic	uncertainty in artificial intelligence
		graphical models	
Planning	Planning and Scheduling; Searching; Optimisation	Bayesian optimisation	hierarchical task network
		constraint satisfaction	metaheuristic optimisation
		evolutionary algorithm	planning graph
		genetic algorithm	stochastic optimisation
		gradient descent	
Learning	Machine learning	active learning	feature extraction
		adaptive learning	generative adversarial network
		adversarial machine learning	generative model
		adversarial network	multi-task learning
		anomaly detection	neural network
		artificial neural network	pattern recognition
		automated machine learning	probabilistic learning
		automatic classification	probabilistic model
		automatic recognition	recommender system
		bagging	recurrent neural network
		Bayesian modelling	recursive neural network
		boosting	reinforcement learning
		classification	semi-supervised learning
		clustering	statistical learning
		collaborative filtering	statistical relational learning
		content-based filtering	supervised learning
		convolutional neural network	support vector machine
		data mining	transfer learning
		deep learning	unstructured data
		deep neural network	unsupervised learning
		ensemble method	

Currently, we are in the process of data exploration and collection, and in Chapter 4 we will outline potential user scenarios for energy, mobility, and cross-sectoral datasets. Our efforts thus far include:

- **Energy Data**: Continuous efforts are being made to acquire energy data with the assistance of Aarhus municipalities and other partners. These efforts include:
    - Investigating the possibility of obtaining energy consumption data from Kredsløb, a key energy sector data provider.
    - Realising the potential to access data on all public buildings in Aarhus.
    - Exploring additional data sources beyond Kredsløb.
    - Working on acquiring data from utility providers in Aarhus; until then, a "dummy" dataset from another utility is available via Center Denmark Portal[5].

  Once sufficient and necessary data for the energy sector has been connected, we will use a data-driven digital twin methodology to access key parameters of the energy network. This method, developed at DTU, will enable real-time assimilation of sensor data from energy networks for optimal decision-making, accounting for system uncertainties and forecasts, and facilitating reliable scenario generation.

- **Mobility Data**: Following the data-driven approach in the first phase of the digital twin platform creation, we have started to:
    - Identify and store traffic sensor data for analysis.
    - Develop a contemporary traffic model of the Area of Interest (AOI), or acquire datasets from which the model can be created.

  The current status of prioritised datasets and their availability is captured in Table 1.

- **Cross-sectoral Data**: Several activities have been undertaken over the first period of the project, including:
    - Investigating high-value datasets and APIs introduced by the EC.
    - Reviewing the comprehensive list of datasets available for Aarhus Kommune on OpenData.dk.
    - Exploring socio-economic data availability by Aarhus municipality for the years 2021, 2020, 2019, and 2017, and contacting consultants to understand the data and formats.
    - Identifying various data sources in Denmark, including Data Distributor[6], SDFI[7], and Open Data DK[8].
    - Identifying weather data from the Danish Meteorological Institute (DMI) and Open Weather platform, and currently deciding which data source to proceed with.

The next steps involve scoping the boundary of dataset selection in relation to potential user scenarios, validating these scenarios with stakeholders, exploring additional data sources, and selecting appropriate models and techniques for effective data analysis.

1.2.2. DKSR OUP as technical core

The BIPED digital twin is being developed on the basis of the existing urban platform from consortium partner DKSR. The Open Urban Platform (OUP) is an event-based real-time data platform design with a focus on urban IoT and control applications (see section DKSR OUP Platform below). This makes the platform the excellent tool for a PED where real-time data is paramount, and in later stages the control functionality could be enabled to automate certain processes (e.g. energy routing based on dynamic signals).

[5] https://portal.centerdenmark.com/en-US/
[6] https://datafordeler.dk/
[7] https://sdfi.dk/
[8] https://www.opendata.dk/

1.2.2.1. OUP Introduction

In a first step, it was necessary to familiarise consortium partners (in particular those from the technical team involved in data modelling) with the platform and to discuss its implementation and extension throughout the course of the BIPED digital twin development. To this end, a series of workshops organised by DKSR have been organised, such as the OUP introduction and several deep dive sessions (see Figure 4), in which technical details have been discussed and the extension into the digital twin has been planned.

These workshops proved to be very helpful to settle a common ground of understanding on the OUP among the consortium. A deeper description of the platform functions and technical description can be found in chapter 3.2.1

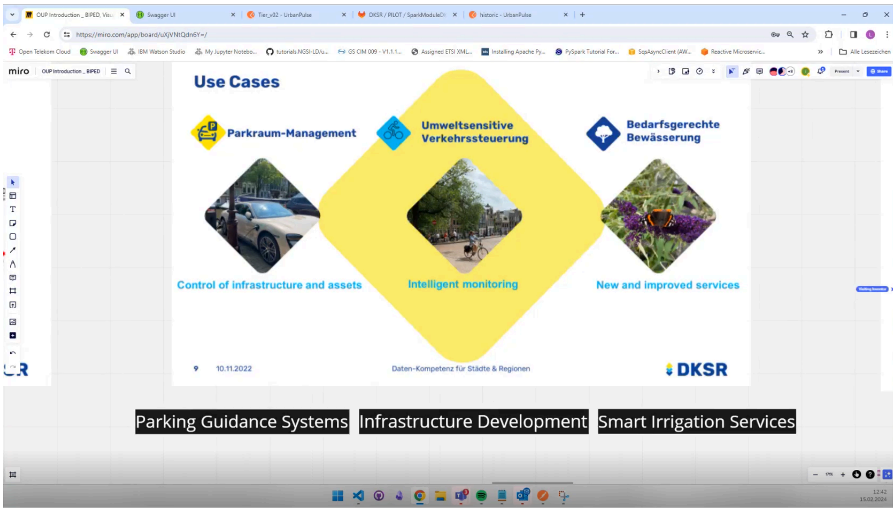


Figure 4. Online Whiteboard used for a workshop on the OUP, led by DKSR.

1.2.2.2. Bilateral meetings

With a common understanding of what the platform can and cannot do as-is, a series of bilateral meetings have been organised in thematic sub-groups to clarify in detail how the OUP will need to be configured and, potentially, extended to incorporate the datasets and data models developed by partners. This has been done on a topic by topic basis and has led to further technical discussions which have shaped a common vision of how all the parts will need to be integrated throughout the course of the BIPED project. This is reflected in chapter 3, in which the digital twin architecture is presented.

1.2.2.3. Refining and designing

All the outputs of the activities described above have been refined internally within DKSR and aggregated into the design of this preliminary release of the BIPED digital twin. Major effort has been dedicated to ensuring that this design is scalable and flexible enough to accommodate for what will be developed within BIPED and beyond, thinking always on system interoperability and connection to other european data efforts.

1.2.2.4 Requirements

Requirements for the platform should be derived from the needs of future end-users and specific Use Cases. As the exploration of stakeholders and users is taking place in parallel in WP3, the consortium partners have decided, in consultation with WP3, to define an initial, experience-based user set and rudimentary, first user stories. This set is methodologically helpful as it can be used as a basis for orientation and discussion for further technical developments. Requirements for the BIPED twin were also discussed and documented from the perspective of the technical system users, i.e. the consortium technical partners. These technical partners' requirements are then structured as Software Requirements Specification (SRS) to serve as a sound basis for the continuing development of the digital twin architecture.

2. Requirements

In the BIPED project, user requirements originate either from the technical partners involved in the project as part of their task obligations or from stakeholders by expressing their expected outcomes and intended use of the tool. While stakeholders have been involved during the initial phase of the project (WP3), the primary focus at this stage is on gathering requirements from the technical partners to develop the core components of the digital twin platform.

The requirements of a software system can be collected in various ways, such as user stories, scenarios, textual format, mock-ups, and more. At this stage of the project, the current technical requirements have been gathered through several formal and informal meetings, which have been documented in textual format, and these constitute the basis for the development of this preliminary release of the BIPED digital twin.

In the following subsections we first explore the general user requirements from the perspective of foreseeable end-users. We then collect the technical user requirements from the point of view of the different consortium members. These technical requirements are then structured as Software Requirements Specification (SRS) to serve as a sound basis for the further development of the digital twin architecture.

2.1. General User Requirements as User Stories

As the definition of the end-users and their needs is still in progress (WP3), more general placeholders have been described to guide the technical development.

Based on the partners' experience in the field of digital solutions for cities and the development of use cases for urban digital twins, we understand the following list of typical users as a guideline for the development - it does not yet represent the defined end-user selection. We have come up with six 'user archetypes' or personas that will typically interact with and use the BIPED twin in the future. They serve as orientation for the conceptual structure. In parallel to the conceptualisation of the twin and the technical development of the basic infrastructure within this work package, WP3 is also investigating the actual user groups on site in the city of Aarhus and the neighbourhood Braband. This process is currently ongoing. The needs of local users will be iteratively translated in user stories and integrated into the development process continuously.

We use the familiar standard format of user stories to describe the requirements of the different user types: *As [User], I want [to perform this action] so that [I can accomplish this goal]*

User Fact Sheet 1: Decision Maker

Purpose: Needs to make decisions at a strategic level in order to drive urban development forward efficiently, meeting the City's objectives.

User Stories:

- UR-1.1DM: As a Decision Maker, I want to know what influence a considered measure will have on the traffic itself, on traffic energy consumption and on the traffic environmental impact so that I can consider alternative measures and select the optimal one.
- UR-1.2DM: As a Decision Maker, I want to evaluate the performance of current heating infrastructure so that I can recommend upgrades and maintenance to improve efficiency and reduce emissions.

- UR-1.3DM: As a Decision Maker, I want to analyse heat demand forecasts so that I can identify areas for expanding the district heating network and to ensure that the future capacity meets the projected growth.
- UR-1.4DM: As a Decision Maker, I want to integrate weather data to forecast demand spikes during extreme weather conditions so that I can implement strategies to balance supply and demand, preventing outages and maintaining grid stability.
- UR-1.5DM: As a Decision Maker, I want to use temperature data to provide drivers with accurate EV range predictions so that I can optimize battery usage and ensure reliable vehicle performance in cold weather.

User Fact Sheet 2: Politician

Purpose: Needs to explain and justify decisions and measures taken in order to get a voters' confirmation of his/her politics.

User Stories:

- UR-2.1PO: As a Politician, I want to know how much closer we are to district energy positivity now than we were before I took charge of the district so that I can present how my political decisions helped the district.
- UR-2.2PO: As a Politician, I want to review the social, economic, and environmental benefits of the proposed twin so that I can advocate for policies and investments that align with my constituents' needs and the city's long-term sustainability goals.

User Fact Sheet 3: Data Analyst

Purpose: Needs a data driven understanding of a real world phenomenon in order to help decision makers to make data driven decisions.

User Stories:

- UR-3.1DA: As a Data Analyst, I want to iteratively explore the history of traffic in the district so that I can look for spatiotemporal traffic patterns.
- UR-3.2DA: As a Data Analyst, I want to integrate various data sources related to urban heating so that I can provide comprehensive insights to decision-makers for strategic planning.

User Fact Sheet 4: Citizen

Purpose: Needs to be informed about conditions which can affect his/her behavior in order to make conscious decisions driven by facts (data) instead of subjective feelings.

User Stories:

- UR-4.1CI: As a Citizen, I want to see the traffic in the district in real-time so that I can choose the appropriate mode of transport
- UR-4.2CI: As a Citizen, I want to understand how district heating will affect my utility costs and comfort levels so that I can support initiatives that improve my quality of life and align with sustainable urban development.

User Fact Sheet 5: Community Engagement Officer of the City

Purpose: Needs visual instruments to foster citizen engagement as part of participatory processes in the context of Urban Planning.

User Stories:

- UR-5.1CE: As a Community Engagement Officer, I need a visualisation tool that I can use in participatory processes with citizens to clarify planning measures, e.g. compare designs.
- UR-5.2CE: As Community Engagement Officer, I need an easy to handle tool that I can interact with on-site and that allows me to quickly draw citizens' ideas onsite into the 3D-Map.

User Fact Sheet 6: Architects from private companies

Purpose: Needs a greater availability of data (e.g. shadow impact, energy consumption, mobility needs) to provide better designs.

User Stories:

- UR-6.1Arch: As an architect in the competition process, I want to have direct access to selected data sets in order to customise my designs.
- UR-6.2Arch: As an architect, I want to be able to test my designs directly in the virtual 3D model and better understand its impact.

2.2. Technical Partners' Requirements

Requirement from Task 2.1 - All Technical partners: In task 2.1, all technical partners work together in defining and developing a software platform that integrates all relevant urban data and models, and thus can be regarded as a Digital Twin solution. This solution provides means to process, analyse and demonstrate all relevant information including the physical urban environment, dynamic real-time information representing the current state in the real world as well as historic information used for data analysis (e.g. for UR-3.1DA). The platform must be designed in a flexible way such that scenarios and use cases modelled in tasks 2.2, 2.3 and 2.4 can be supported. It will be deployed as a system of systems incorporating solutions from all technical partners. Therefore it is mandatory to follow interoperability guidelines and concepts developed by global organisations such as ISO, OASC, and OGC.

User stories related to Data Analyst (UR-3.xDA) and Decision Maker (UR-1.xDM) can be directly covered by task 2.1 by enabling access to all raw data within the Digital Twin.

Requirement from Task 2.2 - Energy Modelling partners: In task 2.2, focusing on energy-related data sources, the technical partners are engaged in developing advanced models and analytics tailored to the district energy system of both demand and supply sides. Particularly important here are the user requirements of the Persona Decision Maker, e.g. UR-1.3DM and Citizens. These models aim to analyse various aspects of energy consumption and leveraging diverse datasets for optimising energy-related processes. Just like in task 2.4, the integration of these energy-related datasets requires the implementation of a data interface within the digital twin framework.

Requirements from Task 2.3 - Mobility and Mobility Environmental Impact Modelling partners: In task 2.3, focusing on gathering mobility-related data to prepare the traffic model of the AOI and integrate it to the BIPED digital twin. It therefore fulfils the User Requirements of the personas Decision Maker, Politician, Data Analyst and Citizen (UR1.1DM, UR2.1PO, UR-3.1DA and UR 4.1CI). GLayer (see chapter 3.2.4) will be used for raw data cleanup and aggregation. RT SW (see Chapter 3.3.1) will be used for calculation of the base traffic model and Traffic Enviro analyst (see Chapter 3.3.2) for traffic energetic and other environmental impact analyses. These analyses should provide both the assumption of the real traffic situation impacts as well as relevant simulations based on what-if scenarios. The topics of the what-if scenarios have to be defined in a closer cooperation with presumed users of the digital twin (e.g. Decision Makers, Citizens, Politicians).

Requirement from Task 2.4 - Cross-sectoral & Soft-Data Modelling partners: In task 2.4, the technical partners are engaged in developing models that analyse cross-sectoral data, particularly focusing on weather data in the first iteration. These models are crucial for providing insights and making informed decisions within the digital twin framework. This relates to UR-1.4DM and UR-1.5DM. As the models will be cross-thematic, the requirements are likely to come from a broad user group, e.g. the strategic users like the Decision Maker or practical users like the Community Engagement Officer. To effectively utilise these models, there's a need for a gateway—a point of integration—that allows incorporation of the diverse data sources into the digital twin platform.

2.3. Software Requirement Specification (SRS)

In software development, system requirements are typically categorised into three main types: functional requirements, non-functional requirements, and constraints. A technical document that translates user requirements into these categories is the Software Requirements Specification (SRS). Particularly, SRS serves as the foundation for software design before the implementation of a software solution.

On a general level, SRS plays a critical role in ensuring the software projects' success by satisfying several goals which are valid for all the tasks along the project including BIPED. SRS clearly defines the requirements by detailing what the digital twin should do, including data collection, processing, visualisation, interaction capabilities, and specifying quality attributes via functional and non-functional requirements, respectively. It aligns stakeholders, ensuring all parties have a shared understanding of the project goals, functionalities, and constraints. It establishes the project scope and boundaries, providing a framework to avoid unnecessary extensions beyond the requirements. Moreover, it serves as a blueprint for the development team, ensuring alignment with BIPED's goals and consistent integration of all components. For the development team, SRS provides a basis for developing test cases to validate and verify that the BIPED digital twin meets the specified requirements. Moreover, it acts as a comprehensive reference document throughout the project lifecycle for both the development team and stakeholders. Finally, it ensures the BIPED meets any relevant industry standards, regulatory requirements, or legal obligations. Accordingly, this section, by documenting the requirements, fulfils the above goals.

2.3.1. Functional Requirement/s

The BIPED digital twin platform is designed around five core functional components:

1. **Data Ingestion:** Essential for acquiring data from various sources through gateways and other collection mechanisms.
2. **Data Streaming:** Facilitates seamless data integration (offline (batch processing), online, or real-time), crucial for timely insights.
3. **Data Modelling:** Enables accurate AI-based modelling, extracting significant value from the data.

4. **Data Storage & Structuring:** Manages and organises large volumes of data securely.
5. **Data Visualization:** Provides intuitive interaction, visualisation, and access to analytical tools.

These components ensure the BIPED digital twin will be a robust platform for simulation, analysis, and optimization across diverse domains.

During this phase, we analysed the technical partners' requirements (chapter 2.2) to derive the essential functional specifications for the project. These specifications are then assigned to the corresponding digital twin core components, ensuring coherent integration that aligns with user needs (technical users or stakeholders). As the project progresses, these requirements will be continuously refined, evolving with the development process and guiding the creation of future deliverables.

2.3.1.1. Data Ingestion

Data is undoubtedly the key asset of a digital twin, and therefore the ingestion of both dynamic and static data into the BIPED platform must be robust and flexible. To enable the integration of energy-related data (task 2.2), mobility-related data (task 2.3), and cross-sectoral data (task 2.4), the digital twin platform must provide a variety of robust and versatile connections.

These connections should support HTTP/HTTPS protocols and adhere to RESTful principles, allowing for secure and efficient communication between the digital twin core components and external models. Additionally, while the API should provide endpoints for data ingestion, retrieval, and manipulation, supporting various data formats such as JSON and XML, it is important to recognize that data ingestion will not always require an API. Data will also be ingested through other means, such as pushing to an FTP server, using SSH, or connecting directly to a database.

For instance, most of the more or less static data for the fundamental 3D model will likely be imported into the BIPED database using well-established interoperable formats, such as the Open Geospatial Consortium (OGC) standards. This approach ensures that even static data is integrated consistently and is interoperable with other systems. Standardised formats facilitate data exchange and integration, making it easier to incorporate diverse datasets into the platform. Moreover, all the technical partners of the BIPED consortium agree on the need to ensure all endpoints have open documentation, e.g. Swagger[9], to facilitate integration and usage by various stakeholders.

By incorporating various data ingestion methods and adhering to standardised formats, the BIPED platform will ensure that it can handle a wide array of data sources and types, thereby enhancing its robustness and flexibility for future applications and analyses. This comprehensive approach to data ingestion underpins the platform's capability to support simulation, analysis, and optimization across diverse domains, ultimately driving the success of the BIPED digital twin.

2.3.1.2. Data Streaming

Streaming technologies are important for making raw data available to end user applications. Web applications for instance must be lightweight in terms of CPU and memory utilisation - especially when targeting mobile platforms (e.g. tablets). Therefore, any data displayed must

[9] https://swagger.io/specification

be made available as data chunks that can be forwarded to the on-screen renderer without heavy data processing, which would stall the entire application.

In the case of urban spatial data displayed as a dynamic map, content is often downloaded from specialised REST services as images from rectangular areas (tiles) and put together on the screen. Similar strategies have been developed for vector data (vector tiles). For showing 3D scenes and models in web applications, the glTF format developed by KHRONOS has been established as the most widely used streaming standard. Furthermore, it is advisable to use compression technologies for reducing the network bandwidth. Typical compression standards are JPEG for images and KTX2 for texture data. Draco compression can be used for reducing mesh data. Data provided for thematic vector overlays can be encoded in GeoJSON, which is a lightweight format describing georeferenced geometries.

2.3.1.3. Digital Twin Model Building

Data modelling in the context of Digital Twin involves the creation of comprehensive and accurate representations of real-world systems. Apart from creation of this representation, it is also essential to use it for advanced simulations, predictive analytics etc. This means that data modelling can be divided into two groups (phases), data building and data modelling.

The first phase, data building, focuses on constructing comprehensive representations of real-world systems. This involves creating a detailed energy, traffic and cross-sectoral model. This phase encompasses the collection and integration of various energy or traffic-related datasets (such as energy consumption or road network data, traffic generators, and information from traffic surveys) as well as relevant cross-sectoral datasets (such as population demographics, socio-economic statistics or weather data). These datasets are then processed into a unified structure, enabling the creation of an accurate model. The large amounts of data available for both traffic and energy offer a great opportunity to use AI-based models here. These AI-based models, mostly referring to deep learning, have, in recent years, shown to perform remarkably well in many fields due to their large number of parameters allowing them to successfully model complex systems. To ensure the models accurately reflect the real-world system, they must be calibrated (or in terms of deep learning, trained or fitted). Calibration (training or fitting) involves aligning the model with real-world measurements, such as energy or sensor traffic data, to ensure its outputs are realistic and reliable. This process is crucial for the model to provide meaningful insights and support effective decision-making.

Once the model is constructed, the subsequent phase involves utilising it to conduct various analyses and extract actionable insights. This phase, data modelling, emphasises the application of the developed model for simulation, prediction, and decision support. A key activity in this phase is the creation of "what-if" scenarios. These scenarios are crucial for predicting how the system will respond to specific changes (for instance, in the context of traffic mobility, a scenario might explore the impact on traffic on Street A if there is a closure on Street B. The outcomes of such scenarios provide critical information to facilitate data-driven decision-making). At this stage the models would bring the Digital Twin into the “Future ready” maturity phase (see Figure 2).

2.3.1.4. Data Storage & Structuring

Strategies for storing data depend on the nature of the data and whether it is static such as data of the urban environment, which does not change regularly, or dynamic such as sensor data. Cadastre data representing the urban environment of the Digital Twin are managed by the municipalities. For making it available for streaming and data analysis, cadastre data must be stored in a spatial database using spatial indexing, which speeds up spatial queries significantly.

Storing sensor or real-time data is done by one or multiple document stores. This provides benefits like high scalability, flexible data models, and fast read/write operations. Document-based databases can handle large volumes of data with varied structures, making them ideal for applications requiring rapid and dynamic data processing. Additionally, they support easy indexing and querying, enhancing performance for real-time data retrieval and updates. They are also good because they store data in JSON, making it easily readable by humans.

2.3.1.5. Data Visualisation

Data visualisation is the last but more user-facing part of the BIPED digital twin platform, enabling users to interact with and interpret complex systems effectively. The platform must provide effective visualisation tools to support diverse user needs, including real-time monitoring, historical analysis, and predictive insights.

The platform should feature concise and interpretable dashboards, essential for summarising the PED key performance indicators (KPIs) and other relevant metrics. The development of these dashboards should follow a user-centred approach, ensuring that communication objectives and end-user needs are clearly defined before design begins. This approach ensures that the dashboards are not only visually appealing but also serve their intended purpose effectively. The platform's architecture should leverage state-of-the-art dashboarding tools.

To enhance spatial data analysis, the BIPED platform should incorporate a 3D map visualisation tool. This tool will allow users to view and interact with geospatial data, providing a detailed and immersive perspective on urban data. The 3D map should support various data layers, enabling users to visualise different aspects of the urban environment, such as infrastructure, energy consumption, and mobility patterns.

Moreover, users should have the ability to customise their data views and interact with visualisations to gain deeper insights. This includes filtering data, zooming in on specific details, and adjusting parameters to explore different scenarios. Such interactivity ensures that users can tailor the visualisation to their specific needs and derive meaningful insights. Here a distinction will need to be made between user roles, e.g. the Decision Makers, who might require more control and flexibility, and Citizens or Engagement Officers, who might prioritise simplicity, easy-to-user interfaces and interpretability of already digested results.

The visualisation tools should be designed with modularity and flexibility in mind, allowing for easy updates and integration of new features as the project evolves. This ensures that the platform can adapt to changing requirements and incorporate the latest advancements in data visualisation technology.

2.3.2. Non-Functional Requirement/s

Modularity is a crucial non-functional requirement in the design of BIPED digital twin, as it ensures the twin's components are independently functional, easily maintainable, and scalable. By structuring BIPED in modular units, developers can isolate and address specific functionalities without impacting the entire twin but also allowing for concurrent development of separate modules. This enhances the ability to troubleshoot, upgrade, and optimise individual modules. Furthermore, modularity facilitates the integration of new features and technologies, allowing BIPED to adapt to evolving technical and stakeholder requirements and remain relevant along the project but also afterwards.

Open licences and of the core components and open standards such as those from OGC ensure the BIPED twin to be scaled up in follower cities across Europe.The digital twin core components, the DKSR OUP and VC Map framework (see chapter 3.4.1) are licensed as open source projects, the code is publicly available in online repositories (GitHub). Open Source fosters collaboration and innovation through community-driven development. This approach also ensures transparency and flexibility, allowing for easier customization and rapid problem-solving. The development projects will be documented in English, so will be the Code. Conformity with the MIMs is an important feature for achieving interoperability of data, systems and services between cities and different suppliers. Especially this refers to MIM1 (Context Information) that focuses on the comprehensive and integrated use, reuse and sharing of data, enabling the bringing together of context information from different systems and sources through a web-based API, thus turning data into a strategic resource, which are valuable to guide the usage of data across different domains in BIPED LDT solution as well; MIM2 (Data Models) which supports cities and communities to use consistent and machine-understandable definitions of all the entities for data and MIM7 (Places) that focuses on geospatial data supporting the ability to be able to integrate and transfer data between internal and external IT systems.

Additionally, to secure data privacy and protect sensitive data, the BIPED twin follows robust security measures to be compliant with EU regulations, such as GDPR will be ensured. This will support MIM4 (Trust) specifically on personal data management.

A further quality feature is reliability, the BIPED digital twin should consistently perform its intended functions under specified conditions without failure, ensuring high availability, speed, responsiveness and minimal downtime.

3. Digital twin architecture and components

The BIPED digital twin architecture has been designed during the first six months of the project through the recurrent and fluid exchange between the consortium partners. At this stage, we focus on the functional modules of the system rather than the specific technological implementation, similarly to how the requirements above have been framed. It is with this mindset that we set off to design an architecture that is both tailored to the demands of PEDs and modular enough for later adaptation. Through this refinement and design phase we have strived to underpin the importance of a robust dataflow and dependency structure between technical components. In the following sections we describe the proposed and partly implemented architecture (see Figure 5), then we provide insights on the different components of the BIPED platform. Note that we separate Models and Algorithms from the Back-end components to emphasise their relevance and to bring up the flexibility of the platform to incorporate a variety of models.

3.1. General architecture

The general architecture we propose contains three distinct components. At the bottom are the data sources, these are in grey to indicate that they are exogenous to the system and not the product of this work.

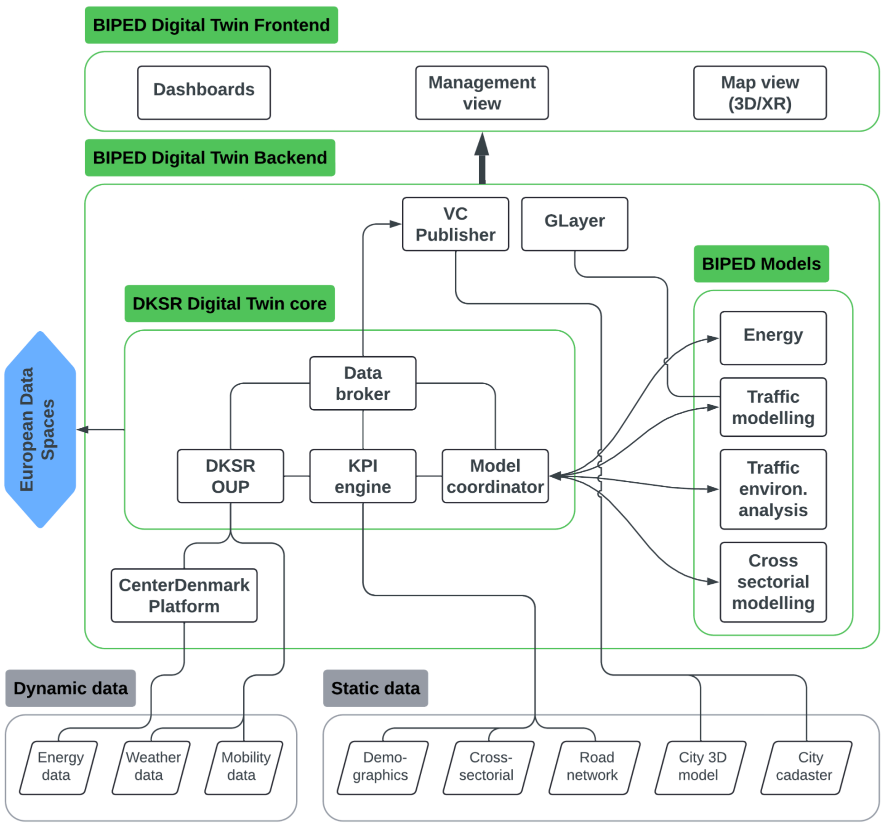


Figure 5. BIPED Digital Twin architecture

In the middle, the BIPED Digital Twin Backend includes the components and modules to ingest, stream, store, model and simulate the data in a consistent and structured way.

The diverse nature of the data leads to three different ingestion paths; the DKSR OUP, the KPI engine and the VC Publisher. It is worth mentioning here that we don't route the "City 3D Model" to the core, but straight to the VC Publisher because the 3D model of a city (here we are referring to satellite imagery and building shapes shown when navigating the map in 3D, refer to Interface description) is relevant for the VC Publisher and doesn't need to be funnelled through the BIPED Digital Twin core, which would add some unnecessary overload. The DKSR OUP is responsible for ingesting and processing the dynamic data and it will potentially also be able to perform some analytics tasks if required. The KPI engine is, additionally to the computation of the key performance indicators (KPI), responsible for ingesting the static data, meaning that it will perform recurrent fetching of data when it changes or at predefined intervals.

The KPI engine together with the Model coordinator are the components where the business logic of the Digital Twin is implemented. These two components are meant to enable the smooth interaction of the platform by orchestrating the different models and providing them with the necessary data, as well as, storing the simulation results to present them to the users to support decision making. These models, grouped under BIPED Models, are able to leverage the input data from the different domains to forecast energy demand, simulate traffic what-if scenarios, perform environmental analysis and more. The specifics of whether these models are hosted by the respective owners, or deployed in the same cluster as the rest of the BIPED components remains an open topic, and we believe that the Model coordinator should be able to handle both scenarios. One consideration in this regard is data volume: if large amounts of data need to be sent back and forth, it might be more efficient to have the models "close" to the data storage. Finally, the Data broker, VC Publisher and GLayer are the backend components responsible for provisioning the data for the Frontend.

The BIPED Digital Twin Frontend includes a 3D map view, a series of dashboards (which might be embedded to the map viewer) and a management view. These different components provide all the necessary user interfaces to explore the data, analyse the results and manage the system.

3.2. Back-end components

3.2.1. DKSR OUP Platform

The DKSR OUP is essentially an event-based big data platform, designed for high-volume, high-speed data processing and a high diversity of urban data sources, developed in the Vert.X[10] framework. The architecture follows a lambda architecture. This means that data is processed directly via a bus system against the moving data stream (speed layer), persisted and, if required, processed as a "batch" (batch layer). Data is stored semi-structured in a NoSQL DB or kept in-memory. This enables horizontal scaling across clusters of computing units and thus the handling of big data in near real time. For processing reactive microservices allow a large number of messages per second and for processing problems with large data streams without blocking other tasks (multi-threading).

The core architecture is based on DIN SPEC 91357 for Open Urban Platforms, the de facto European reference architecture for open urban platforms based on the Smart City Marketplace (former EIP-SCC) lighthouse projects' results. Figure 6 shows the architecture of the OUP at a conceptual level. The data flow can be read from bottom to top.

[10] https://vertx.io/

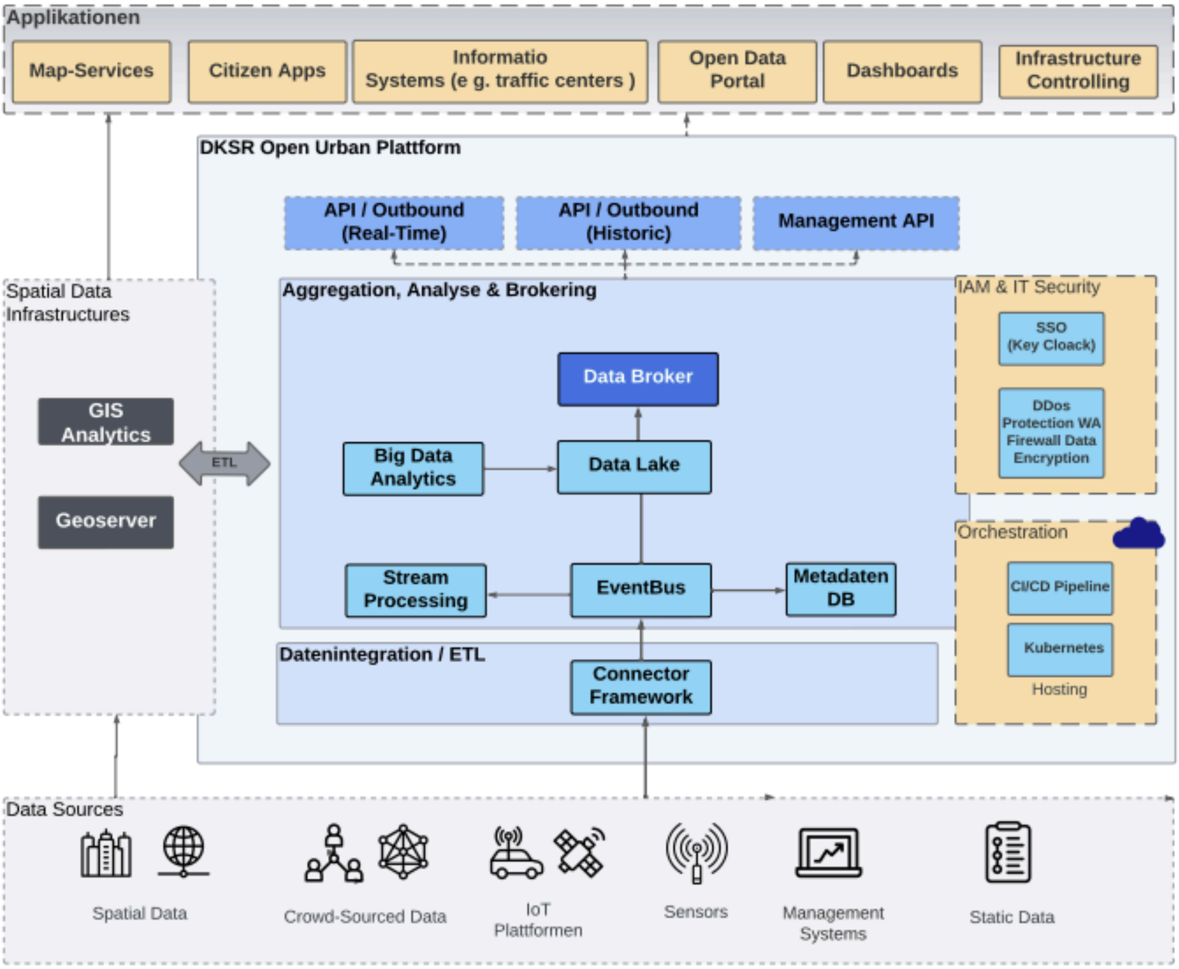


Figure 6. High level architecture of the DKSR OUP

The OUP Core consists of Stream Processing, Data Lake and Big Data Analytic Tool. Stream Processing is event based and is executed by an event processor module. The analyses can be performed in real time against the moving event stream. Analyses can be created automatically from predefined patterns or determined manually. The predefined patterns contain standard operations such as min, max and average of the measured values. The event processor module offers an SQL-like syntax for this.

The Data Lake provides a configuration option for a hierarchical storage implementation. This allows the advantages of different storage systems to be combined and the data to be stored where it is needed. For example, the first level can be configured to use an in-memory storage system to provide fast responses, while the second level can be a long-term storage system. In this way, the module can collect large amounts of data on the one hand, while ensuring a fast response time when historical data is requested on the other.

For Big Data processing a Big Data Analytic Module is provided. The module is based on Apache Livy, which allows seamless communication with Spark within the modular architecture. Apache Livy acts as a bridge between the user applications and the Spark cluster, enabling efficient job submission and monitoring, thus simplifying the interaction and management of Spark jobs.

Each module makes its functionality available to the other modules. The interaction between the modules is carried out by the Vert.X event bus system, which enables asynchronous communication. In combination with REST APIs, the back-end can ultimately communicate both event data for the real-time view and persistent data to the historical view. In addition to the Vert.x event bus (part of the Vert.x framework), various technologies from the streaming and messaging environment can be considered for the event bus system, which offer additional functionalities if required. For the persistence module and the event processor module, the OUP is technology agnostic. Elastic Search is proposed as a best practice persistence module, and Esper Tech Complex Event Processing Engine as a persistence module.

3.2.2. VC Publisher

VC Publisher is an authoring tool for creating and managing geo-web applications, which present a core component of Digital Twin solutions. With VC Publisher, users (e.g. Data Analysts) are able to collect, configure, and process cadastral data that is to be integrated into a special geo portal solution (VC Map) enabling easy browsing and analysis of geospatial data. Due to the focus on 3D data processing, VC Publisher was used for preparing the basic geospatial data of the physical environment of the BIPED digital twin.

Raw data and data from municipal REST services is converted into streaming data sets, which can be easily accessed and processed by lightweight web applications. These converters cover the following data types:

- pointcloud - point cloud conversion
- qmesh - terrain conversion
- tms - Tiled Map Service (raster overlay)
- x3dm - 3D object calculation
- oblique - oblique aerial image calculation

The VC Publisher provides a graphical editor for configuring web applications, which controls the layout and style of the application, available renderers (2D, 3D and oblique), available layers, and available tools and other functionality. BIPED-specific tools, e.g. for performing energy-related analysis of live data, will be implemented as plugins. The web application is pushed on the BIPED platform and connects to other BIPED components through REST services.

The VC publisher backend and the graphical editor are meant to be used by administrators, who are in charge of controlling the contents and behaviour of the web application. A user and project management system can be connected to an IAM system for controlling the privileges of users. Figure 7 shows the administrator rights to components within VC Publisher.

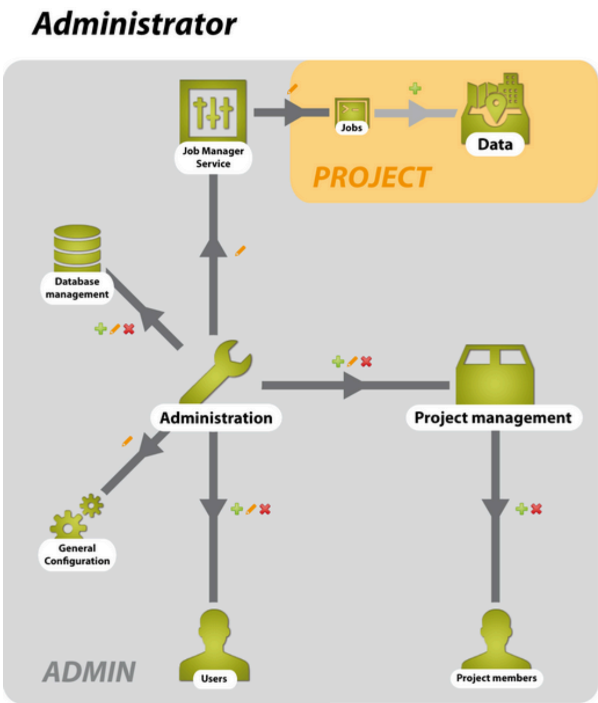


Figure 7. Administrator access rights in VC Publisher (add, remove, edit)

An additional REST API is available for controlling a VC Publisher instance from external modules. The REST API is defined using OpenAPI 3.0.1 technology, which facilitates discovering and understanding the capabilities of a service without requiring access to source code, additional documentation, or inspection of network traffic. The API enables access to almost all functionalities of VC Publisher, depending on the user role.

3.2.3. Center Denmark Energy Platform

The Center Denmark Energy platform serves as a tool for accessing information on energy resources, infrastructure, and other relevant datasets within the energy and utility sector. It aims to provide a user-friendly experience that facilitates easy access to energy-related data in different municipalities of Denmark. Its architecture is organised into distinct layers, each playing a crucial role in data management.

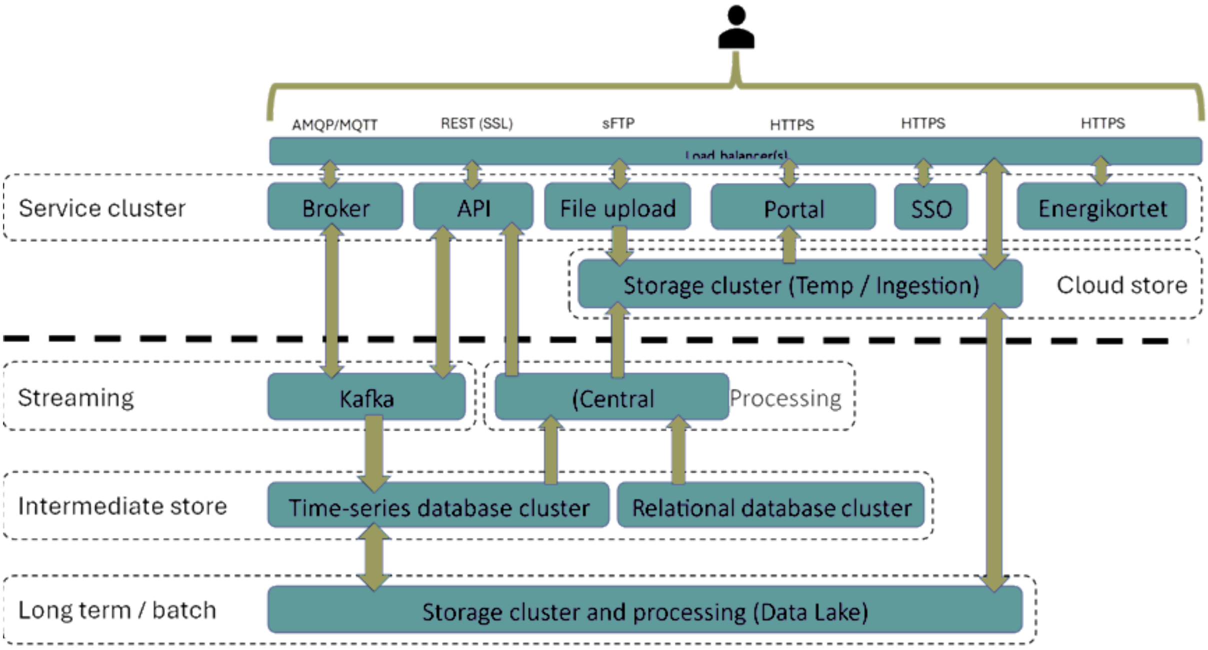


Figure 8. Schematic Architecture CenterDenmark Platform

For long-term storage and batch processing, the platform utilises a data lake, which stores large volumes of raw data in its native format, supporting extensive batch processing operations. The intermediate store layer includes two main clusters: a time-series database cluster for analytics on time-indexed data, and a relational database cluster for structured data, supporting complex queries and transactions. In an Intermediate store Kafka, a distributed streaming platform, manages real-time data streams, directing them to appropriate intermediate storage or processing units. The storage cluster temporarily holds ingested data, organising it for further processing. At the service and user interaction layer, the platform supports AMQP/MQTT, REST (SSL), sFTP, and HTTPS protocols, with load balancers ensuring high availability and performance. Users engage with the system through services like a broker for managing data communication, APIs for programmatic access, a file upload service, a user-friendly portal, SSO for secure authentication, and the specialised Energikortet application.

3.2.4. GLayer Server

Glayer is a GPU-accelerated backend software focused on fast data aggregation, filtering and visualisation. GLayer is capable of performing analytical queries on large datasets in a scope of milliseconds with special emphasis on spatial data and spatial aggregation. GPU uses thousands of lightweight processing cores, leverages data parallelism, and has high memory throughput. This architecture is known to be very effective for computer graphics. In the GLayer concept, the dataset is splitted into smaller parts and each part is processed in parallel by a GPU core which allows the high performance of aggregation-based analytical queries.

The GLayer server is written in Kotlin and uses LWJGL as a GPU interface. GLayer is designed as platform independent, however the target device for testing is NVidia GPUs. The backend provides REST API with swagger / OpenAPI specification.

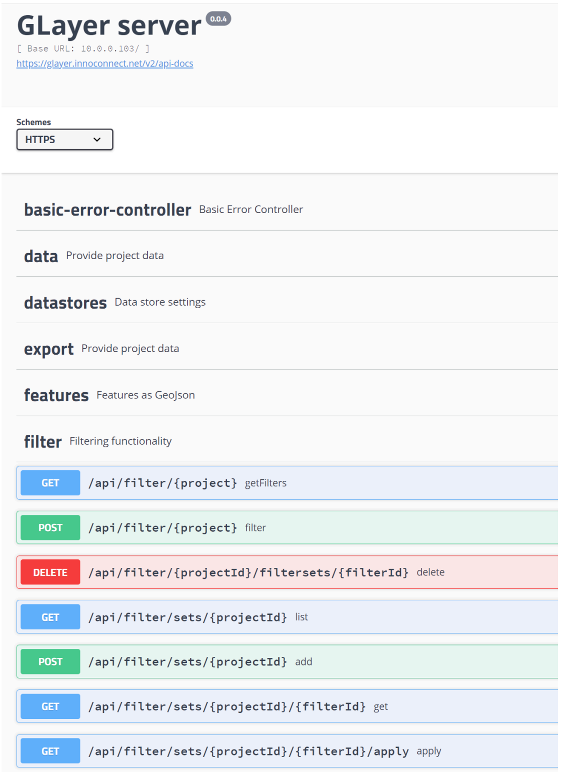


Figure 9. GLayer API Description

On top of the RestAPI there is a frontend part which serves these purposes:

- Datastore configuration: data connectors can access data from permanent storages such as SQL databases or csv files.
- Project configuration: project can define how the dataset is going to be visualised and filtered. This includes data type definitions, aggregation strategies, filtering capabilities and cartographical outputs such as choropleth maps, heat maps, histograms etc.
- Project visualisation: Once the project is configured, it can be accessed through a permanent link via a standard web browser on a desktop or mobile.

3.3. Models and algorithms

As described in above (section 2.3.1.3) we divide data modelling in the two phases: data building and data modelling. This process is currently in an early stage, it is also highly dependent on the availability of local datasets. We are currently working on obtaining domain-specific data from local stakeholders of the city, e.g. utility company.

Advanced modelling software is already in the pipeline in the mobility domain, provided by the consortium partner RoadTwin. The following is intended to show the integration plan of RoadTwin Software into the BIPED twin.

Related to energy consumption, advanced modelling for describing how the district heating load depends on the local meteorological data, has been initiated. This includes methodologies for adjusting the standard meteorological forecasts with potential data in the city, such that the relevant city weather forecasts can be established and used for proving state-of-the-art heat load forecasts. The Energy Map developed by Center Denmark, will be considered for providing a flexible user interface.

In order to operate the local district heating network, the temperatures in the network have to be estimated in real time. Prototype methods for estimating the network temperature using data from individual energy metres in e.g. residential homes, have been developed.

3.3.1. RoadTwin Software

The Roadwin Software (RT SW) application models the impact of planned traffic restrictions (change of speed limit or closures) on traffic flow. Thanks to this SW, it is possible to browse through the planned traffic constructions that will take place in the area of interest. It is also possible to view the traffic volumes and level of traffic in the area under ideal conditions without traffic restrictions.

The application architecture consists of two main components: the frontend (client) and the backend. The frontend is designed for interacting with the traffic model and conducting "what-if" analyses of city traffic. This includes creating traffic scenarios and visualising the results on a map. The frontend provides various ways to view the results, such as comparing two scenarios, aggregating data comparisons, and comparing real-world measurements with the traffic model. It is foreseen that a subset of the RT SW frontend functionality will be integrated into the VC Map component (section 3.4.1) in order to integrate the RT SW results into a comprehensive graphical client for the BIPED Digital Twin Platform.

The backend is responsible for performing the aforementioned analyses. It offers a REST API with Swagger already available[11]. In addition to calculating the basic traffic model and modifications (traffic scenarios), the API supports creating new models, updating existing ones, performing static traffic assignments, calibrating the model, and more. Detailed documentation of the API and its functionalities can be found here[12].

3.3.2. Traffic Enviro Analyst

The Traffic Enviro Analyst prototype aims to leverage traffic modelling to analyse and calculate the impact of traffic on the environment. An accurate traffic model is crucial for calculation of traffic energy consumption and environmental impact.

Generally various environmental impacts can be calculated and visualised from traffic models. The Traffic Enviro Analyst created by University of West Bohemia (UWB) aims to calculate:

- emissions, particularly CO_2, and their dispersion
- energy consumption from traffic
- noise modelling

To achieve accurate results, it is essential to obtain the appropriate data about types of vehicles (passenger/freight, types of fuel, etc.) and traffic share for the particular vehicle types, for which the above mentioned environmental impacts can be derived, based on the traffic flows provided by the traffic model.

The Enviro Analyst can be chained to RoadTwin software (see chapter above) and provide the calculated environmental impacts to the BIPED Digital Twin Platform through its VC Map client (described in 3.4.1). This integration will allow users to model different scenarios and analyse how certain policy decisions impact the environment.

[11] https://bazina.plan4all.eu/swagger-ui/index.html?configUrl=/v3/api-docs/swagger-config#/
[12] https://gitlab.com/tramod/rt_sw-api-documentation-public/-/blob/main/api_new.md?ref_type=heads

3.4. Front-end components

The BIPED Digital Twin Frontend includes a 3D map view, a series of dashboards (which might be embedded to the map viewer) and a management view. The frontends will be specified based on the User Requirements. For example, a more KPI-centred view can be helpful for the DM to create reports, while the Citizen Engagement Officer needs the drawing functions in the 3D map for participation events.

3.4.1. VC Map

VC Map is a high-performance web based front end solution with a focus on streaming 3D city models and other geospatial content. A VC Map can be configured using the VC Publisher backend or alternatively based on a set of configuration files along with data converter tools. The technology stack comprises:

- CesiumJS: an open source library for creating 3D globes, developed for aerospace and smart city solution,
- OpenLayers: an open source library for creating dynamic maps in web pages,
- 3D-Tiles: an open standard for massive, heterogeneous 3D geospatial datasets such as point clouds, buildings, and photogrammetry, Built on glTF.

Various tools and widgets can be added and configured for enabling features such as

- export data from backend databases,
- drawing vector geometries as well creating simple 3D shapes,
- measuring distances and areas in 2D and 3D
- creating height profiles
- shadow and viewshed analysis
- uploading and positioning 3D models
- clipping geometries

Support for OGC compliant services such as WMS, WMTS ensures compatibility with existing Geo-IT systems, e.g. run by municipalities and serves MIMs requirements.

The core VC Map framework has been published as an open source project on GitHub. Extensibility is achieved by providing a plugin mechanism allowing third parties to develop their own tool.

3.4.2. Dashboards

The provision of concise and interpretable dashboards is paramount for the success of the project. These will be the focus of the later stages of the project. The previous experience of multiple consortium partners with providing dashboards has emphasised the need to have clearly defined communication objectives and rey results before starting designing them. For this reason and for the delay on the availability of some easily displayable cross-sectoral datasets, it has been general consensus that the development of the dashboards should await the relevant data sources and most importantly, the end-user needs for them.

It is worth mentioning that, regardless of the communication objective still being defined through interaction with the stakeholders, the platform is being built with the flexibility and modularity necessary to later incorporate a state-of-the art dashboarding tool such as Grafana or Apache Superset. Both these tools are well established and powerful platforms that enable connection to multiple data sources, fulfilling the requirements of the BIPED digital twin.

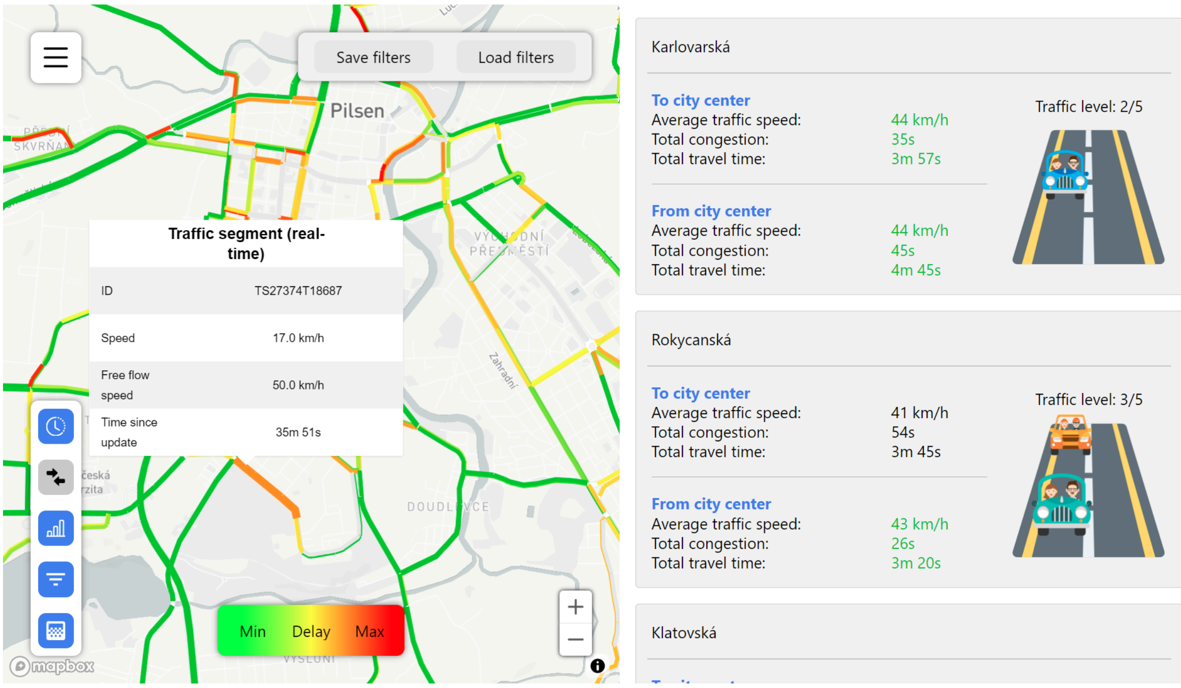


Figure 10. Dashboard showing real-time traffic conditions in Pilsen (from INNO)

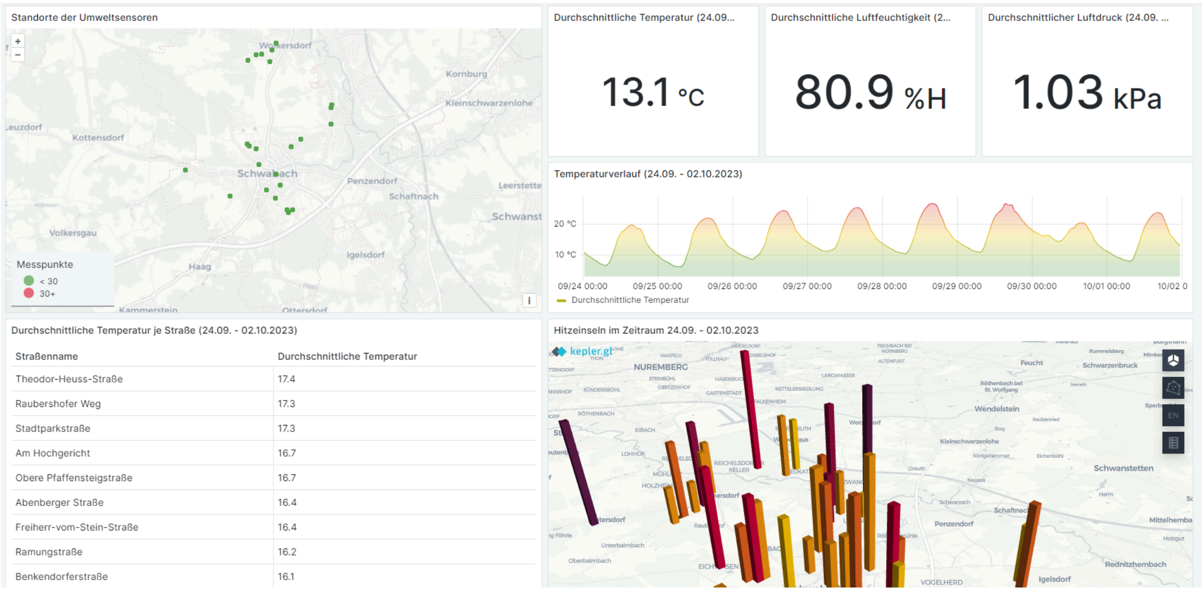


Figure 11. Heat Island Dashboard (from DKSR)

DKSR provides further dashboard options (see Figure 11) based on open source tools such as Grafana or Kepler.gl. In addition, more sophisticated features can be contributed as needed. The example in Figure 12 shows an interface (data centre) that allows data to be displayed and shared between multiple parties, based on user rights and role management. This feature is relevant when a party, e.g. a private architect, wants to access specific data, see user requirement UR-6.1Arch. It can also be used to communicate the development of urban policies based on data. For example, a city can prepare data according to the city's sustainability and climate goals and present the achievement of the goals in key figures (Fig. 12). For example, the dashboard in Figure 12 tracks changes in emissions over time, allowing decision-makers to monitor trends and assess the effectiveness of implemented policies (see UR-1.1DM:). For further technical development, user requirements will determine which front-ends are desired as part of the BIPED twin.

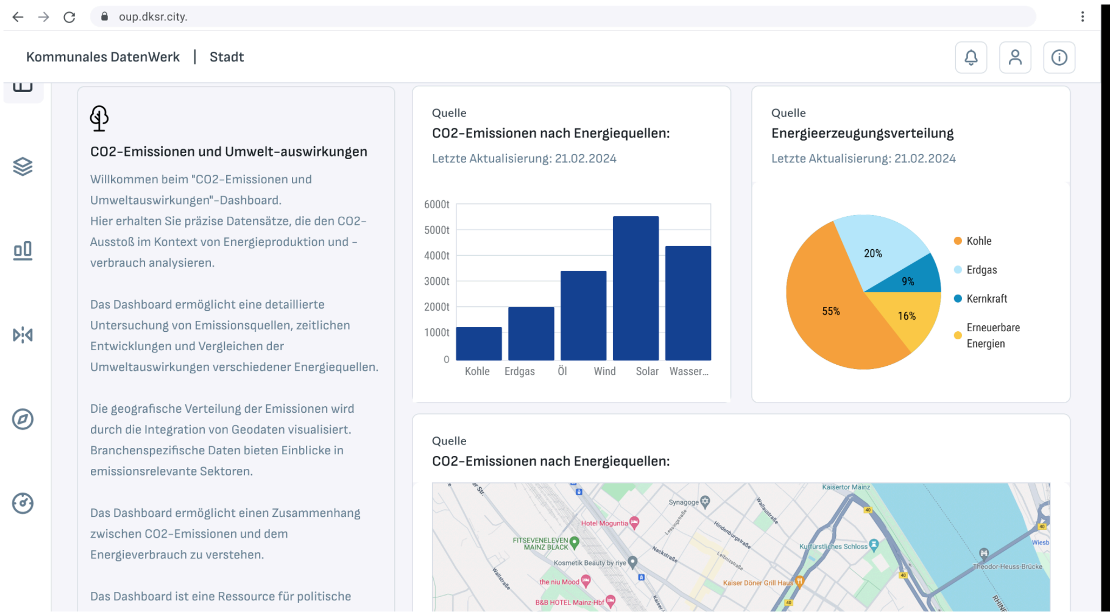


Figure 12. KPI Viewpoint of the municipal data centre tool in German (from DKSR)

4. Demonstrator of the BIPED Digital Twin Platform (Initial Release)

For the initial release, access to the digital twin has been implemented as a browser-based application, which is available online. Please note that it covers only a few of the components shown in Figure 5 with a focus on desktop viewers on the application level and data integration.

The main technologies used are Virtual City Systems’ VC Map framework and a FROST server providing information on air quality sensors. FROST is an open source implementation of the OGC SensorThings API developed by Fraunhofer IOSB.

The application is hosted on https://vcmap.bi-ped.eu/ with an option to be migrated to another location managed by the municipality of Aarhus in the future.

4.1. Interface description

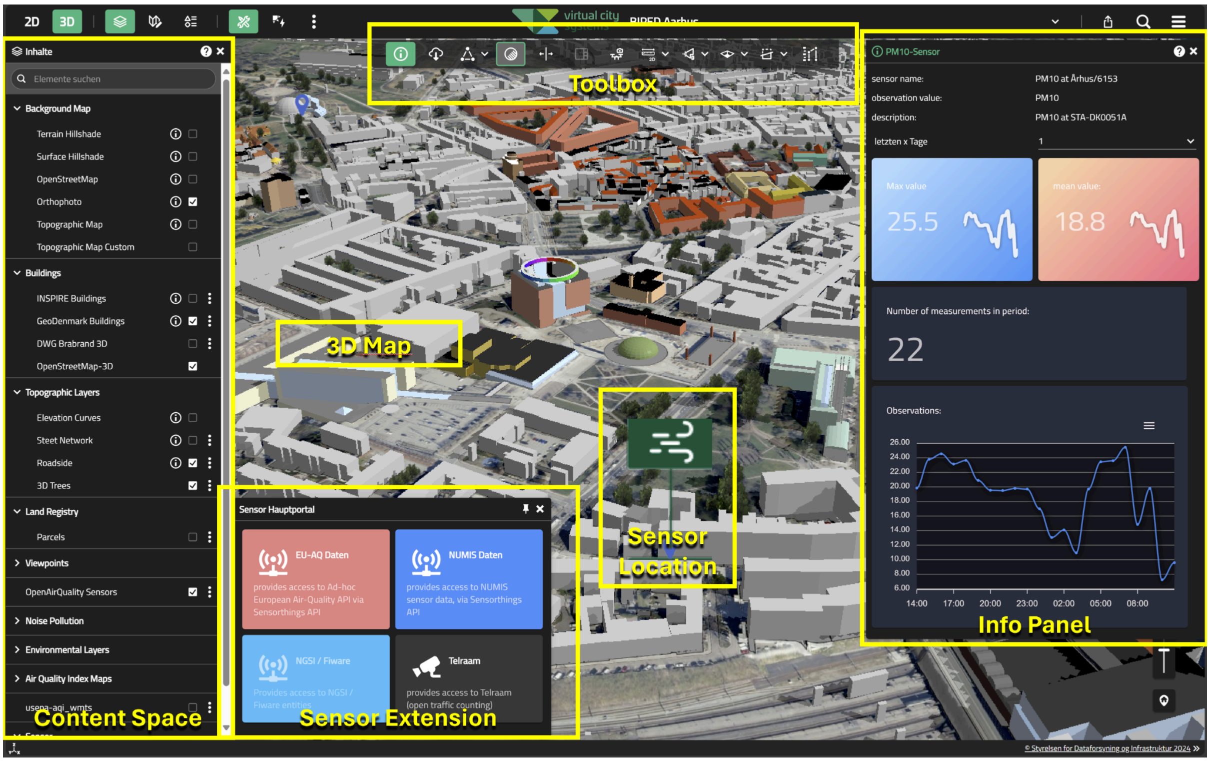


Figure 13. User interface of the online application in desktop mode.

The user interface[13] is divided into the following components:

- the **Map View** as the central element,
- the **Content Space** showing available layers and data sets
- the **Info Panel**, showing details of selected features,
- the **Toolbox** in the header, to access various map functions,
- one or multiple **Extension Panels** for project-specific functionalities, and
- the **Map Navigation Widget**, which provides controls for navigating the current view of the map (hidden under Info Panel box in Figure 13).

[13] https://help.vc.systems/en/vc-map/v5.1/components/index.html#id_image_uiTextoverlay

The general UI concept enables users to open multiple panels and components, which can be arranged freely on the screen. This allows implementing complex workflows with information views which interact with each other. For instance, a Digital Twin solution for PED might include working with planning scenarios while having a dashboard showing the typical energy demand of the district. The Map View will always stay in the background and cannot be moved. The UI implements a responsive design, meaning that depending on the target device and screen size, some user input controls will work differently and be presented in a different way. While on typical tablets most of the panels are still available, the functionality on mobile phones is limited and some complex widgets will not work. Also, the window concept changes to full screen panels on smartphones. However, smartphones have not been identified as a target platform in BIPED, since location based services are not required.

- The **Map View** supports two different modes, which can be toggled at the upper left.
- The **2D Map** displays two-dimensional content only. If the 2D map is activated, 3D objects including the 3D city model of Brabrand cannot be displayed and are thus automatically deactivated in the Content Space. Usually, alternative layers have been configured for the same object type, in this case the building footprints holding the same object properties.
- The **3D Map** displays three-dimensional and two-dimensional content. If the 3D map is activated, all layers and 3D objects are displayed based on the digital terrain model providing the height reference model. 2D layers such as land use areas and orthophotos are projected on the terrain.
- The **Content Space** holds information on the available layers and 3D object types, which can be activated as needed. Additionally, predefined viewpoints for zooming in to areas of interest and custom layers defined by the planning module are available here. Layers are often mutually exclusive, such as noise map overlays for different hours of the day. Other actions provided by the Content Space are access to customizable legends informing users about the layer content or symbols as well as access to declarative styling, which colours objects based on specific properties, for instance applying different colours for private, public and industrial buildings.
- The **Info Panel** shows detailed information about selected objects by clicking on them in the Map View. This typically means showing the available properties in a tabular form or as a simple balloon, which is one of the typical GIS functionalities. It is also possible to provide custom code for enabling a more complex representation such as dashboards. An example is given in Figure 11 with the history of PM10 particle concentrations measured by an air quality sensor in Aarhus displayed as a graph. The sensor locations are shown as a billboard layer in the Map View.
- The **Toolbox** provides access to various additional modules and widgets. These typically come with their own user interface opened in a new panel. Examples are tools for measuring distances and areas (in 2D and 3D), a split screen tool, a tool for generating height profiles, a drawing tool, a tool for setting time and date and enabling shadows, a viewshed tool, and a tool for creating cross sections.
- The **Extension Panels** actually provide project or domain specific functionalities in a separate window. These extensions are implemented as add-ons using the plugin mechanism of the open source core map library. Any custom functionality in BIPED, if not relevant for other applications, needs to be implemented as an extension. The sensor integration extension is a first example for adding project-specific functionalities. It is able to establish a connection to an OGC SensorThings API compliant service, retrieve sensor locations and readings, and create interactive dashboard-like overlays.
- The **Map Navigation Widget** helps navigating in the 3D environment in case that input devices are not available or gestures are not working.

4.2. Accessing Live Data

One of the fundamental concepts of Digital Twins is that it reflects the current conditions of the real world embedded in the virtual model of the city. Users must be enabled to access, analyse and monitor various aspects of public spaces including weather conditions, air quality, traffic conditions, energy flows and the electricity grid. As a technical demonstrator, two types of IoT sensors were integrated in the application of Aarhus. The sensor meta data and real time readings are retrieved on demand through specialised web APIs backed by IoT networks. The functionality was added as an extension as explained in the previous section. Typically, a catalogue of sensor location is downloaded on user request and a spatial point layer is added to the map indicating the location of each sensor and the type of sensor.

The first type of sensor provides information on the current air quality by measuring concentrations of O3, NO, NO2, SO2, CO as well as particles PM10 and PM2,5. The measurements are provided by the OpenAQ initiative[14]. There are currently two OpenAQ sensors located in Aarhus ("Århus/6153" and "Århus/6160"). On selecting the sensor on the map, the latest readings are shown in an interactive panel. The history is displayed as dynamic graphs with options to change the time range and type of measurement.

The second type of sensor collects information on multimodal traffic collected by a network of individual volunteers. It is called Telraam[15] and provides information on the number of cars, trucks, bicycles and pedestrians currently passing by. The idea is to get insights of the current traffic condition and typical densities over the day.

Lastly, data is streamed through the Open Urban Platform (OUP) from Open Weather Map, which includes information on temperature, humidity, wind, rain, and more. This data can be correlated with energy demand and generation. Additionally, it can be analysed in relation to traffic behaviour.

[14] https://openaq.org
[15] https://telraam.net

5. Data Sets & Data Sources

Based on the methodological approach, including data mapping and prioritisation process, as described in Chapter 1.2, this chapter outlines prioritised data sets with the aim of being included in the BIPED baseline platform and describe the current status of availability for each of them, as per submission date of this deliverable D2.1. (see Table 1). Besides "Group", "Variable Name", "Details" and "Format" of the variable, the current "Availability Status" is indicated by colour, from dark green (available) to dark red (currently not-available).

Table 2. Prioritised datasets and their availability status (as per submission date of this deliverable): 1/dark green - available, 2/light green - owner contacted, 3/light red - owner not contacted, 4/dark red - currently not available

Group	Variable Name	Details	Format	Availability Status (1-4)
Digital Twin	Building footprints		GML	1
Digital Twin	3D city model	only Brabrand so far	DWG, CityGML	1
Digital Twin	Orthophotos spring Webmercator	web mercator CRS for easier web integration	WMTS	1
Digital Twin	Terrain and surface raw data	resolution 0.4m	WCS, GeoTiff	1
Cross-sectoral: Environmental	weather data: temperature, wind, precipitation		CSV	1
Cross-sectoral: Social	Demographics: Population age distribution	on household level	CSV	2
Cross-sectoral: Social	Demographics: Population gender distribution	on household level	CSV	2
Cross-sectoral: Social	Demographics: Population ethnicity distribution	on household level	CSV	2
Cross-sectoral: Environmental	Land Use pattern	on parcel level	CSV, JSON, SHP	3
Cross-sectoral: Environmental	green space		CSV, JSON, SHP	3
Cross-sectoral: Environmental	air quality model	continuous spatial dataset covering the AOI	CSV, JSON, SHP	2
Cross-sectoral: Environmental	road network / segments	road segments input data for space syntax analysis (connectivity / accessibility	CSV	3
Mobility	Road network		GeoJSON, JSON, SHP	2
Mobility	Traffic generators (zones)		GeoJSON, JSON, SHP	2
Mobility	OD matrix	OD matrix connecting traffic generators and network	CSV, JSON	2
Mobility	Sensor traffic data		CSV, JSON, SHP, XML, TXT	2
Mobility	Sensor data Calibration of traffic model		CSV, JSON, SHP, XML, TXT	2

Cross-sectoral: Environmental	Noise measurements	on sensor level	CSV, TXT, XML	3
Cross-sectoral: Environmental	Noise model	continuous spatial dataset covering the AOI	CSV, TXT, XML	3
Energy	Enriched Dataset on District heating from Kredsløb	Demand from individual buildings in Brabrand	CSV, TXT, XML	2
Energy	Power Grid: Consumption data based on user		CSV, TXT, XML	2
Energy	Power Grid: Dataset of substations		CSV, TXT, XML	3
Energy	Power Grid: Aggregate consumption data on each building		CSV, TXT, XML	3
Cross-sectoral: Environmental	Altitude model		CSV, TXT, XML	3
Energy	District heating circuit and infrastructure	Location of district heating circuit lines and distribution system	CSV, TXT, XML	3
Energy	Energy infrastructure	Location of power infrastructure and distribution system	CSV, TXT, XML	3
Meteorological forecasts	Energy system	Standard MET forecasts	CSV, TXT, XML	3
Cross-sectoral: Environmental	Identified infrastructure resilience options on energy systems		CSV, TXT, XML	3

Variables and related data sets have been defined by four thematic groups:

- Digital Twin (T2.1): including datasets needed to build the 3D digital twin model (as for instance, a 3D city model, an terrain model or building footprints)
- Energy (T2.2): including energy-related variables, such as district heating, energy consumption per user or energy infrastructure
- Mobility (T2.3): including mobility-related variables, such as road network data, sensor traffic data or an origin-destination (OD) matrix
- Cross-sectoral (T2.4): including cross-sectoral variables, such as population demographic information, green space or weather data. (Properties of quantitatively less tangible data (soft data, such as emotions or perceptions) will be included at a later stage as part of cross-sectoral variables.)

As gaining access to data can be a laborious and time-consuming process and as access rights discussions are at different stages (as indicated Availability Status, Table 1), presented datasets are in progress and will be constantly updated throughout the project.

Already available data sources will be described below in more detail, including meta information and data structure. The full list of mapped data as a living document ("BIPED data sheet") that will support the implementation process can be seen in Annex 2.

Moreover, Data based User Scenarios have been formulated. The question of how to build a data-based solution to a local challenge helps to identify and search for relevant data. The scenarios described here build on the Users already introduced (see Requirements and User Stories). They aim to broaden the perspective on which data is relevant for the BIPED twin depending on the local challenge or question tackled by the user. It is a second methodological step towards driving forward user-orientated development and corresponding data collection.

5.1. Energy

Progress is currently being made to access data from the municipal utility companies. As a work-around, dummy data on a regional level can be accessed via the CenterDenmark platform to examine typical data sets.

4.1.1 Data based user scenarios

In relation to energy, scenarios have been formulated from user perspectives, to gain more insights. As an example based on the user and user story: Politician, UR-2.1PO: District Energy Positivity

- **Challenge:** As a politician, you are an advocate of district energy, a system in which buildings share a central source of heating and cooling. You want to demonstrate the positive impact on energy independence.
- **Solution:** Analyse historical energy consumption data for the district before and after implementing the district energy system. This data can be visualised to show the percentage reduction in overall energy consumption, highlighting the progress towards district energy positivity (T2.2). Additionally, analyse the fuel source mix before and after, demonstrating a shift towards cleaner energy sources if applicable (T2.3).

5.2. Mobility

5.2.1. Description of Data Source

This chapter describes the data that is needed to create a base traffic model of AOI, to calibrate it and also to display traffic sensor data in digital twin.

One of the key datasets is the geometry representation of the Road network. This dataset consists of lines (or segments or edges) and points (or nodes).

Apart from the geometry representation of Road network, the dataset should contain attributes that define the possible traffic on the roads (such as *maximum capacity or average speed)*.

Another key datasource for traffic models is the traffic generators (also called zones), where traffic originates and terminates in the city.

The last part of the traffic data modelling is the Origin-destination matrix (OD matrix), defining the movement of vehicles in the Road network. The OD matrix is usually squared (number of rows = number of columns), representing a zone and related travel activity through it. The OD matrix is usually not easily available, and software like RoadTwin SW can be used to calculate it.

Sensor data (collected by CCTV cameras or road sensors for instance), describing the actual situation of the traffic, can be used for calibration of the traffic model.

The format of all above mentioned data can differ and for the purposes of the BIPED project is the format irrelevant. If we obtain the data from the municipality we need to firstly process them in order to secure the right format that we can work with.

5.2.2. Data based user scenarios

The first example is based on the user and user story: Decision Maker and UR-1.1DM: Traffic Impact Assessment.

Challenge: The city is considering implementing a congestion charge for entering the downtown area. As a decision-maker, you need to understand the potential impact of this measure on traffic flow, energy consumption, and environmental impact.

Solution: Utilising historical traffic data combined with traffic simulation software, the impact of the congestion charge can be modelled. The model can predict changes in traffic patterns (e.g., reduced congestion in downtown, increased traffic on bypass routes), estimate the resulting decrease in energy consumption from reduced traffic, and calculate the potential reduction in air pollution emissions. This data allows you to weigh the potential benefits (reduced congestion, cleaner air) against potential drawbacks (increased travel times on bypass routes) before making a decision.

Another scenario is based on the user and user story: Data Analyst, UR-2.1DA: Spatiotemporal Traffic Patterns.

Challenge: As a data analyst, you need to understand traffic patterns within the district to support decision-making.

Solution: Utilise historical traffic data from various sources (e.g., traffic sensors, GPS data from ride-sharing apps) to identify spatiotemporal traffic patterns. This involves analysing traffic flow variations across different locations and times of the day, week, and month. Visualisations like heatmaps can highlight areas of high congestion or identify peak commute hours. These insights can inform road infrastructure improvements, public transportation scheduling, and traffic management strategies.

Another scenario is based on the potential public user and user story: Citizen, UR-4.1CI: Real-Time Traffic Information.

Challenge: You are a citizen who commutes daily. Real-time traffic information can help you make informed decisions about your travel.

Solution: Develop a mobile application that displays real-time traffic conditions within the district. This application can integrate data from traffic sensors and crowd-sourced information to show congestion levels on various roads. Users can customise the application to receive alerts about accidents, road closures, or unexpected delays on their preferred routes. This empowers citizens to choose the most efficient travel mode (e.g., car, public transport, cycling) based on real-time traffic conditions.

5.3. Cross-sectoral

5.3.1. Description of Data Source

Apart from several parallel dataset exploration, weather data has been the initial focus in cross-sectoral dataset, and consequently cross-sectoral applications of weather data. Accordingly, we have started to understand the weather data from DMI (Danish Meteorological Institute). DMI provides structured and historical data for Denmark, followed by an API that gives us access to values for a 20km area or specific municipalities.We can extract data for over a 30 year period and the dataset is updated hourly.

Key Weather Data attributes from DMI which can be applicable in BIPED are:

- **Temperature**: Daily and hourly temperature readings.
- **Precipitation**: Amount and frequency of rain or snow.
- **Wind**: Speed and direction of wind.

Moreover, the Open Weather Platform is another weather platform which is under investigation by DTU and DKSR. Progress is currently being made in talks with the city to obtain socio-economic data.

5.3.2. Data based user scenarios

In relation to Energy Sector & Weather data, here we have considered two scenarios from the perspective of decision makers in the city, who need to manage the energy grid , based on the user story of a Decision Maker in the city who manages energy supply, UR-1.2DM.

Challenge: During cold temperatures, the demand for energy to heat buildings increases significantly.

Solution: An energy company uses temperature data to predict these peaks in heating demand, ensuring sufficient energy supply and preventing grid overloads.

Challenge: In hot weather, there is a spike in energy usage for air conditioning.

Solution: Utilities use temperature forecasts to anticipate higher electricity consumption and adjust their supply strategies accordingly.

Challenge: The electricity grid needs to handle peak loads during extreme weather conditions like heat waves or cold snaps.

Solution: By integrating weather data (temperature, wind, and precipitation), grid operators can forecast demand spikes and implement strategies to balance supply and demand, preventing outages and maintaining grid stability.

In relation to **Mobility Sector & Weather data,** we have considered below scenarios:

Challenge: In cold weather, EV battery performance can degrade, reducing the vehicle's range.

Solution: An automotive company uses temperature data to provide drivers with accurate range predictions and optimise battery usage.

Challenge: Combining temperature and wind data, a navigation app helps EV drivers plan their routes by predicting changes in range due to weather conditions.

Solution: This ensures drivers are aware of how cold temperatures and strong winds might impact their journey.

Challenge: A city plans the placement and availability of EV charging stations by analysing weather patterns.

Solution: During adverse weather conditions, such as heavy rain or extreme cold, accessibility to charging stations is ensured by strategically placing them in locations less likely to be affected by weather disruptions. This planning helps maintain a reliable charging infrastructure for EV users under various weather conditions.

5.4. European Dataspace

As a European project and to maximise its impact, BIPED digital twin is deeply grounded in the European data space environment. This subchapter outlines actions that have been taken so far to identify relevant data space initiatives the project will reach out in next steps, as well as efforts that have been made in screening priority datasets, as defined by the European Commission, informing our dataset selection (as described in above in this chapter).

The table below provides a list of key dataspace communities that have been identified as being highly relevant to BIPED's LDT solution, led by the work of T2.5. The benefits of active engagement with these communities are primarily twofold. On the one hand, these communities can provide additional technical input to support the dataset usage and management of this project, as outlined in section 3.3 above. On the other hand, working with these broader EU dataspace communities also supports this project's exploration of the applicability and replicability of the BIPED solution beyond the pilot setting in Branband/Aarhus. In particular, they can help to identify the features that are unique to the pilot setting and the gaps that need to be addressed for a broader application. In addition, by involving these data communities throughout the development of the BIPED LDT, this can also help to raise awareness of BIPED solutions, paving the way for a broader adoption of BIPED solutions across Europe in the future, thus generating greater impact. This list will be continuously updated throughout the project. By the time of delivery of this deliverable (M6), these initial groups have been identified. For the next steps, the engagement with these communities will be further synergised with the broader stakeholder/end-user engagement in WP3 and 4. An engagement plan and schedule is currently being developed.

Table 3. Dataspace initiatives relevant to BIPED

Initiative	Why relevant	Link
Datacellar	Sister project of Omega-x that aims to create a federated energy dataspace that can support the creation, development and management of local energy communities in the EU, which can provide inputs on BIPED’s energy data angle.	https://datacellarproject.eu/

Enershare	Aims at supporting the large-scale UE wide replicability and market take-up of energy data sharing technological solutions and data driven services in different socio-economic contexts. This can contribute to not only the BIPED energy data focus but also the cross-sectoral and possibly also the soft data objectives.	https://enershare.eu/
BRIDGE	A European Commission initiative that brings together the Horizon 2020 and Horizon Europe projects on smart grids, energy storage, islands and digitisation to provide a structured view of cross-cutting issues that arise in the demonstration projects and may be a barrier to innovation. This community can provide inputs on data management and regulations.	https://bridge-smart-grid-storage-systems-digital-projects.ec.europa.eu/
European Dataspace for Smart Communities	This project's focus on cross-sectoral data space for governments on all levels to support the Green Deal objectives can support BIPED's efforts on cross-sectoral data	https://www.ds4sscc.eu/
Synergies	Provides a reference energy data space implementation that unleashes data-driven innovation and promotes the creation of an inclusive ecosystem of stakeholders across the energy data value chain, which can provide valuable insights for BIPED's stakeholder engagement.	https://energydataspaces.eu/
European Mobility Dataspace	Brings together diverse stakeholder groups who need data for innovative mobility solutions. This can support BIPED's mobility angle and broader mobility stakeholder engagement.	https://mobility-dataspace.eu/
EDDIE - European Distributed Data Infrastructure	Aims at reducing data integration costs, allowing energy service companies to operate and compete seamlessly in a unified European market, which can support BIPED's activities on data integration and promote replicability across Europe	https://eddie.energy/
Int:Net	Its key objective is to institutionalise an assessment methodology and maturity model to support the harmonisation of energy services, which can not only help with BIPED's energy data integration but also its methodology might also help with cross-sectoral harmonisation,	https://intnet.eu/
Omega-X	There are 4 use case families and 9 pilots within Omega-X: local energy communities, renewables, flexibility, electromobility, several of which can contribute to the BIPED exploration,	https://omega-x.eu/consortium/
GREAT Green Deal Dataspace	Aims to establish the Green Deal Data Space Foundation and its Community of Practice. Its holistic Green Deal focus can help not only the development of BIPED's technical solution but also the stakeholder engagement to facilitate exchange of	https://www.greatproject.eu/

	best practices and knowledge	

5.5. Stakeholders & End-User Involvement as Part of Data Mapping

Stakeholder and End-User involvement (as part of WP3) ensures the alignment between the PED and the technical / data development of the digital twin with local needs. It furthermore fosters community support, and addresses specific challenges within the city of Aarhus, by establishing a BIPED Community and linking stakeholder and end-user involvement to the BIPED digital twin development. Below we provide an overview about past and current activities - see D3.1 for details.

5.5.1. Establishing a BIPED Community

The Deliverable *3.1 BIPED Community* of Work Package 3 focuses on district based but city wide mapping of stakeholders and end-users to secure a proper representation of interests (e.g. social, cultural, economic, environmental). In collaboration with existing actors in the district, the aim is to involve both private and public actors, city developers, energy suppliers, local businesses and NGOs. In correlation to the stakeholder mapping, a framework of engagement will be built to secure a purposeful and fruitful involvement of stakeholders based on principles of people-driven innovation.

5.5.2. Linking Stakeholder and End-User Involvement with the development of the Digital Twin

Mapping, engagement and involvement of stakeholders and end-users provide essential data, context, and validation for an accurate and effective digital twin. Examples can be seen below:

Data Collection and Validation: Stakeholder collected data can provide insights into e.g. real-city processes, strategies, energy consumption patterns, and infrastructure. And End-users' behaviour data enriches the digital twin e.g. their validation ensures accuracy.

Contextualization: Understanding the local context, urban dynamics, and social aspects through the stakeholders can provide context-specific information. End Users can add value giving their preferences, routines, and lifestyle choices and contextualising the digital twin.

Model Calibration and Simulation: Stakeholders' feedback ensures accurate representation. End-Users input on comfort levels, energy-saving behaviours, and usage patterns fine-tunes the models.

Behavioural Modelling: Stakeholders' insights enhance realism. End-Users behavior data informs agent-based models, predicting how people interact with the district.

Scenario Testing and Optimization: Stakeholders can provide test scenarios with their input and optimise PED strategies. End-Users can evaluate scenarios from their perspective e.g. impact on convenience.

Keeping Stakeholders and End-Users in the Loop (feedback): By regularly updating the digital twin based on stakeholder feedback it is possible to address evolving needs. End-Users feedback on the other hand ensures the twin remains relevant and responsive.

6. Conclusion & Next Steps

This report details the preparatory work and the current status of each work stream and work package tasks. During this initial phase, important groundwork has been laid to ensure the success of the project in the coming months.

Based on the three presented workstreams of Data Mapping, Technical Development, and Conception and Refinement the following points and lessons can be highlighted, which will determine the further development towards Release 1 (D2.2) by the end of 2024 (month 12).

6.1. Data Mapping

- In the course of the data mapping exercise, it became clear that the search for and release of data by stakeholders is a lengthy process, but one that is crucial to the development of the BIPED twin. It is therefore a critical dependency and also a risk for the success of the project.
- It is of the utmost urgency that the availability of data is clarified in the next months in order to gain a realistic idea of how models and algorithms can be developed and integrated in a first Release and which user stories can be prioritised for implementation.
- Alternatively, data could be obtained from third-party providers or by the city procuring new sensor technology. Cost-effective options are available and should be discussed with the city's Smart City / IoT department if added value is expected (floating car data, etc.).

6.2. Technical Development

The further technical developments focus on fully demonstrating all functions of the digital twin. This includes:

- All backend components are set-up and first testing data has been incorporated in the twin.
- All data types (energy data, mobility data and cross-sectoral) will be consumed by the digital twin.
- The interfaces between the backend components and the 3D model of the digital twin will be technically advanced.
- The modelling character of the digital twin will be demonstrated.

The further technical development as described above requires a focus on the two BIPED core components (see Figure 5, Chapter 3.1): KPI Engine and Model Coordination.

6.2.1. KPI Engine

A KPI engine is a component designed to enable data analytics for KPIs in a lightweight, flexible, and robust manner using data from diverse sources. It allows the combination of soft, spatial, and temporal data. While the OUP focuses on big data processing of real-time data and the geodigital twin focuses on geodata, the KPI engine integrates these varied data types and enriches them with user defined KPIs. The results from the KPI engine are stored and can be made available on-demand for visualisation or used as input for different models.

6.2.2. Model Coordination

The model coordination component is essential due to the various models that will run in parallel or asynchronously. It centralises communication between the different modelling units and the 3D - and real-time data processing. Furthermore it is facilitating bidirectional communication between different models. Effective communication and data exchange between these units are vital for accurate and comprehensive modelling. Furthermore, the model coordinator enables easy aggregation of the results within the KPI engine.

For example, results of forecasting energy demand and a simulation of current available electric vehicles that potentially can provide energy needs to be aggregated. The model coordinator can handover this data to the KPI engine in a standardised way via a single API.

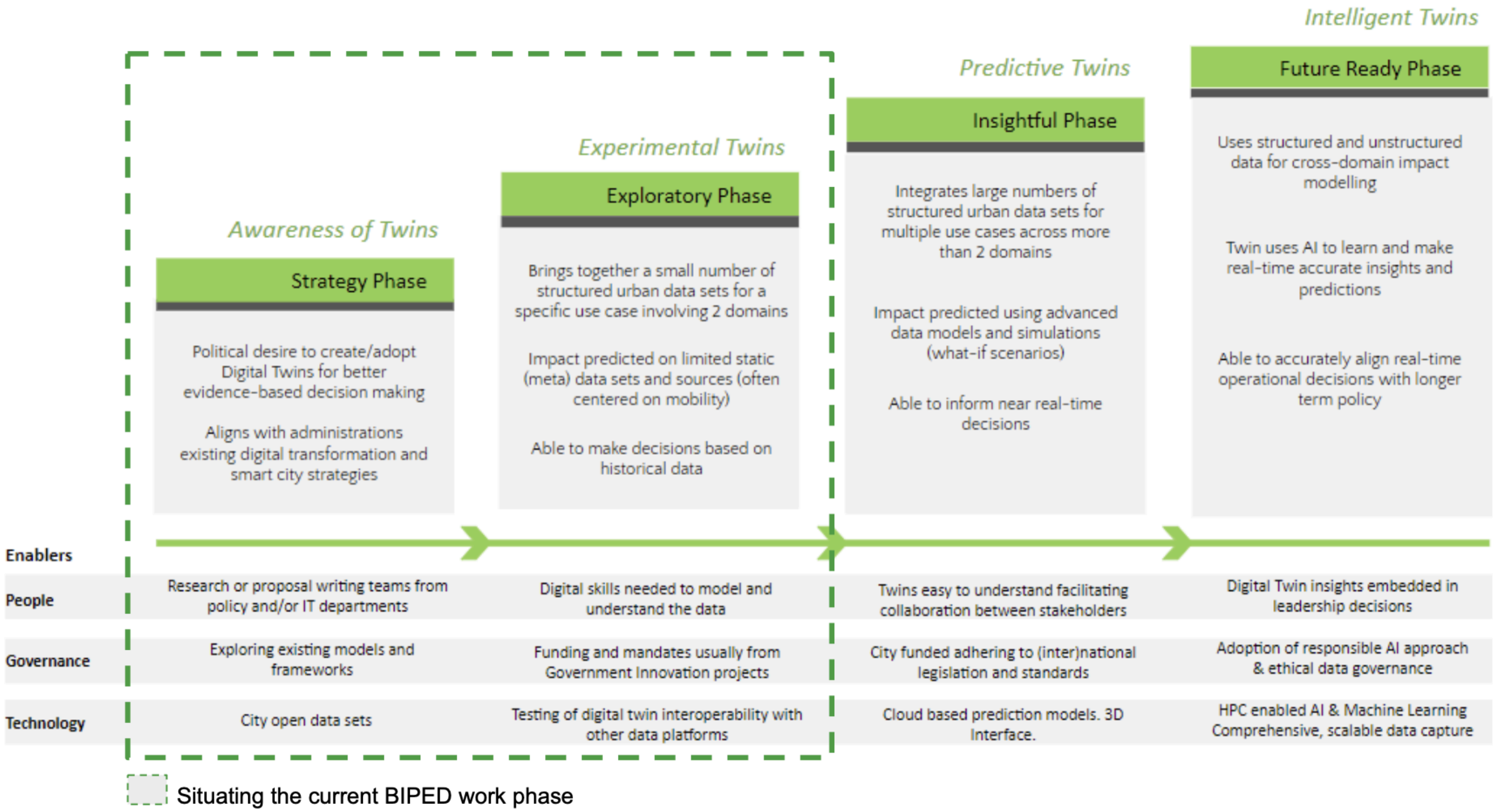
6.3. Conception and Refinement

- The constant exchange between the consortium partners has enabled a common understanding of the components and the future twin architecture. This is an important prerequisite for the architecture description presented here.
- This document, in particular the BIPED Digital Twin Architecture Scheme, can subsequently be considered as a living artefact for the consortium partners to further discuss, refine and develop the BIPED architecture.
- It has been shown that a preliminary definition of user stories was helpful in guiding the conceptual development and design of initial data use scenarios. The added value of the project outcome increases when user stories and KPIs are aligned. The project direction should not be determined solely by the availability of data at this stage.
- In the spirit of user-centred, agile development, work with user stories will be intensified in the future. Close synchronisation with WP3 is important to ensure continuous feedback from local stakeholders. Consideration should be given to who can take on the role of (proxy) product owner in order to prioritise user stories.
- To further focus development, bundle information and avoid friction, the use of a common backlog (based on user stories or epics) should be discussed.

This report, including the initial user stories and the Software Requirements Specification (SRS), will serve as the baseline for further development steps. It represents the accompanying documentation and conceptual work for the technical development, which is visualised in the first demonstrator and made visible to the public. Throughout the BIPED lifecycle, the SRS and a future development backlog will be continuously matured and refined, incorporating evolving stakeholder needs and ensuring that the digital twin solution remains robust, accurate and aligned with the original vision.

7. Annexes

7.1. Annex 1: DUET Maturity Model



7.2. Annex 2: BIPED Data Sheet

baseline dataset	Requestor	Group	Variable Name	Details	Format	Availability Status	Link / Source	AAKS support needed (w details)
x	VCS	Digital Twin	Building footprints		GML	1	Available from GeoDenmark Vector (see other tab)	
	VCS	Digital Twin	Parcels					
	VCS	Digital Twin	Landuse areas					
x	VCS	Digital Twin	3D city model	only Brabrand so far	DWG	1		
	VCS	Digital Twin	Trees		GML	1	Available from GeoDenmark Vector (see other tab)	
	VCS	Digital Twin	Street network		GML	1	Available from GeoDenmark Vector (see other tab)	
	VCS	Digital Twin	Utility networks: district heating			4		
	VCS	Digital Twin	Utility networks: natural gas			4		
	VCS	Digital Twin	Utility networks: electricity			4		

	VCS	Digital Twin	Utility networks: fresh water			4		
	VCS	Digital Twin	Utility networks: waste water			4		
	VCS	Digital Twin	Terrain model hillshade layer		WMTS	1	https://services.datafordeler.dk/DHMSkyggekort/dhm_terraen_skyggekort/1.0.0/WMTS?username=xxx&password=yyy&service=WMTS&request=GetCapabilities	
	VCS	Digital Twin	Surface model hillshade layer		WMTS	1	https://services.datafordeler.dk/DHMSkyggekort/dhm_overflade_skyggekort/1.0.0/WMTS?username=xxx&password=yyy&service=WMTS&request=GetCapabilities	
	VCS	Digital Twin	Orthophotos spring	updated automatically. Last update 2023	WMTS	1	https://services.datafordeler.dk/GeoDanmarkOrto/orto_foraar_wmts/1.0.0/WMTS?username=xxx&password=yyy&service=WMTS&request=GetCapabilities	

x	VCS	Digital Twin	Orthophotos spring Webmercator	same as above, but in web mercator CRS for easier web integration	WMTS	1	https://services.datafordeler.dk/GeoDanmarkOrto/orto_foraar_webm/1.0.0/WMTS?username=xxx&password=yyy&service=WMTS&request=GetCapabilities	
	VCS	Digital Twin	Terrain and surface models as hillshade layers and iso lines		WMS	1	https://services.datafordeler.dk/DHMNedboer/dhm/1.0.0/WMS?username=xxx&password=yyy&service=WMS&request=GetCapabilities	
x	VCS	Digital Twin	Terrain and surface raw data	resolution 0.4m	WCS	1	https://services.datafordeler.dk/DHMNedboer/dhm_wcs/1.0.0/WCS?username=xxx&password=yyy&service=WCS&request=GetCapabilities	
x	AIT	Cross-sectoral: Social	Demographics: Population age distribution	on household level	CSV	2	AAKS	internal data, waiting for connection
x	AIT	Cross-sectoral: Social	Demographics: Population gender distribution	on household level	CSV	2	AAKS	internal data, waiting for connection
x	AIT	Cross-sectoral: Social	Demographics: Population ethnicity distribution	on household level	CSV	2	AAKS	internal data, waiting for connection
	AIT	Cross-sectoral: Social	School locations		CSV	3	AAKS	internal data
	AIT	Cross-sectoral: Social	Health records, health data		CSV	3	AAKS	internal data

	AIT	Cross-sectoral: Economic	Education level	on household level	CSV	3	AAKS	internal data, waiting for connection
	AIT	Cross-sectoral: Economic	Crime Statistics	lon/lat	CSV	3	AAKS	internal data, waiting for connection
	AIT	Cross-sectoral: Economic	Complaint recordings	lon/lat	CSV	3	AAKS	internal data, waiting for connection
	AIT	Cross-sectoral: Economic	Income levels	on household level	CSV	3	AAKS	internal data, waiting for connection
	AIT	Cross-sectoral: Economic	car ownership	on household level	CSV	3	AAKS	internal data, waiting for connection
	AIT	Cross-sectoral: Economic	Land value	on parcel level	CSV	3	AAKS	internal data, waiting for connection
	AIT	Cross-sectoral: Economic	Rent prices	on household level, on parcel level, on street level	CSV	3	AAKS	internal data, waiting for connection
x	AIT	Cross-sectoral: Environmental	Landuse pattern	on parcel level	CSV	3	AAKS	internal data
	AIT	Cross-sectoral: Environmental	POIs: shopping, bars, culture etc	lon/lat	CSV	3	AAKS	internal data
x	AIT	Cross-sectoral: Environmental	green space		CSV	3	AAKS	internal data
	AIT	Cross-sectoral: Environmental	public vs private space		CSV	3	AAKS	internal data
	AIT, UWB	Cross-sectoral: Environmental	air quality	on sensor level	CSV	2	AAKS	internal data

x	UWB	Cross-sectoral: Environmental	air quality model	continuous spatial dataset covering the AOI	CSV	2		
x	AIT / DTU	Cross-sectoral: Environmental	weather data: temperature, wind, precipitation		CSV	1		
	AIT	Cross-sectoral: Environmental	sunrise / sunset		CSV	3		
x	DTU, AIT, INNO	Cross-sectoral: Environmental	road network / segments	road segments input data for space syntax analysis (connectivity / accessibility	CSV	3		
	AIT	Cross-sectoral: Soft Data	perception of safety in an area	to be crowdsourced	CSV	4	BIPED generated	
	AIT	Cross-sectoral: Soft Data	perception of wealth in an area	to be crowdsourced	CSV	4	BIPED generated	
	AIT	Cross-sectoral: Soft Data	comfort level (general, weather related etc)	to be crowdsourced	CSV	4	BIPED generated	
	AIT	Cross-sectoral: Soft Data	sense of community of an area	to be crowdsourced	CSV	4	BIPED generated	
x	RT	Mobility	Road network	Geometry: - **lines** (edges): geographical coordinates of the start/end and break points of the lines, that creates the road network	GeoJson or Json or Shapefile (SHP)	2		

x	RT	Mobility	Traffic generators (zones)	Used to generate traffic and also to attract traffic (vehicles trevel from one generator to another). It is usually a point layer, that is connected to the road network via node (connected only by attribute)	GeoJSON or JSON or Shapefile (SHP)	2		
x	RT	Mobility	OD matrix	OD matrix connecting traffic generators and network	CSV or JSON	2		
x	INNO, RT	Mobility	Sensor traffic data	Data that typically comes from automatic data collectors - sensors (such as road sensors on crossroads, or video traffic cameras).	CSV	2		waiting for AAKS to connect INNO with the person responsible for traffic data in Aarhus
x	RT	Mobility	Sensor data Callibration of traffic model	Data that is used to calibrate the traffic model (get the ifnormation about the precision of the model and make it more precise).	CSV	2		

	UWB offered this dataset in project proposal	Mobility	Open Transport Map	Allows routing and visualization of traffic volumes of the whole EU. The underlying data come from OpenStreetMap and are accessible in a scheme compatible with INSPIRE Transport Network.		1	https://hub.plan4all.eu/otm , https://opentransportmap.info/	
	UWB offered this dataset in project proposal	Mobility	Open Land Use	a composite map that is intended to create detailed land-use maps of various regions based on certain pan-Europen datasets such as CORINE Landcover, UrbanAtlas enriched by available regional data.		1	https://hub.plan4all.eu/olu	
	UWB offered this dataset in project proposal	Mobility	Smart Point of Interest	an open dataset published as Linked Open Data. It contains more than 33 million points of interest over the whole world.		3	https://hub.plan4all.eu/spoi	
x	UWB	Cross-sectoral: Environmental	Noise measurements	on sensor level	CSV	3		internal data

x	UWB	Cross-sectoral: Environmental	Noise model	continuous spatial dataset covering the AOI	CSV	3		internal data
x	CDK	Energy	Enriched Dataset on District heating from Kredsløb	Demand from indivual buildings in Brabrand	CSV	2	To be done together with Kredskøb	
	CDK	Energy	Dataset on individuel flats consumption	Not yet availlable		4		
x	DTU, Center Denmark	Energy	Power Grid: Consumption data based on user		CSV	3		
x	DTU, Center Denmark	Energy	Power Grid: Dataset of substations		CSV	3		
x	DTU, Center Denmark	Energy	Power Grid: Aggrigated consumption data on each building		CSV	3		
	DTU, Center Denmark	Energy	Power Supply: PV systems, EV fleet					
x	DTU	Cross-sectoral: Environmental	Altitude model		CSV	3		
x	DTU	Energy	District heating circuit and infrastructure	Location of district heating circuit lines and distribution system	CSV	3		
x	DTU	Energy	Energy infrastructure	Location of power infrastructure and distribution system	CSV	3		

x	DTU	Cross-sectoral: Environmental	Identified infrastructure resilience options on energy systems		CSV	3		

7.3. Annex 3: Data Sources

Name	Overview / Topics	meta data / variables	Link
Open Data Aarhus	General public data; various topics	Befolkning og samfund energy Energi international-issues Internationale spørgsmål agriculture-fisheries-forestry-food Landbrug, fiskeri,... environment Miljø economy-and-finance Økonomi og finanser government-public-sector Regeringen og den... regions-cities Regioner og byer justice-legal-system-public-safety Retssystem og offentlig... health Sundhed transport Transport education-culture-and-sport Uddannelse, kultur og... science-and-technology Videnskab og teknologi	https://www.opendata.dk/city-of-aarhus
Inspire Geo Portal	Priority Dataset (EC)		https://inspire-geoportal.ec.europa.eu/srv/eng/catalog.search#/home

Open Transport Map	Allows routing and visualization of traffic volumes of the whole EU. The underlying data come from OpenStreetMap and are accessible in a scheme compatible with INSPIRE Transport Network. Open Transport Map can serve as a map itself as well as a layer embedded in your map. OTM is accessible via both GUI and API. Experimental version of OSM is published as Linked Open Data and available through the SPARQL endpoint. OTM is implemented in cultural tourism by many various ways: background data showing traffic infrastructure, input data for analysis or data about traffic volumes, which are essential in cases overcrowded parts of cities or areas struck by air pollution.	https://opentransportmap.info/img/OTM_physicalModelAndCodelists.svg	https://hub.plan4all.eu/otm, https://opentransportmap.info/
Open Land Use	a composite map that is intended to create detailed land-use maps of various regions based on certain pan-Europen datasets such as CORINE Landcover, UrbanAtlas enriched by available regional data.		https://hub.plan4all.eu/olu
Smart Point of Interest	an open dataset published as Linked Open Data. It contains more than 33 million points of interest over the whole world. Data are acquired from public resources (open data) and data from partners of cooperating projects such as SDI4Apps, Peregrinus Silva Bohemica project and others. Data are divided into thematic groups. The structure of SPOI is described in the SPOI Ontology.		https://hub.plan4all.eu/spoi
Enriched Dataset on District heating from Kredsløb	Demand from indivual buildings in Brabrand	Forward and return temperature, M3, kWh	To be done together with Kredskøb
Dataset on individuel flats consumption	Not yet availlable	kWh from Brunata?	
Datafordeler Service by SDFI	PUBLIC BASIC DATA FROM THE DANISH AUTHORITIES		https://datafordeler.dk/
Aarhus Kommune			aarhuskommune.dk
GeoDanmark			Systems and data - geodanmark
Spatial Map - Aarhus Kommune			SpatialMap 4.4.0 (aarhuskommune.dk)
Data Forsyningen			https://dataforsyningen.dk/data?filter=0;format-63&view=gallery&search=