--- Pending EC Approval ---

Open framework for boosting EU High Value Datasets from Public Sector

beopen-dep.eu

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

D5.5 – HVDs catalogue (first version)

Document History: with the use of shared documents, it makes less sense to use versioning, and therefore to track the document history. The document has continuously populated with content until it's release to the consortium for contributions and to the peer reviewers for quality assessment.

Ver.	Date	Description	Author	Partner
0.1	17/11/2024	ToC and preliminary content	V. Venckauskaitė	ID Vilnius
0.2	04/12/2024	Pilot contributions	All partners	All partners
0.3	10/12/2024	HVD scoring update	C. Viceto	Ubiwhere
0.4	11/12/2024	Addition to content, conclusions	V. Venckauskaitė	ID Vilnius
0.5	17/12/2024	Final Review	F. D'Agresti / M. Dias	Engineering / Porto Digital
0.6	18/12/2024	Final version	V. Venckauskaitė	ID Vilnius
1.0	19/12/2024	Submission to EC	A. Filograna	ENG

Disclaimer

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

Publication details
Grant Agreement Number 101069831
Acronym BeOpen

Full Title	Open framework for boosting EU High Value Datasets from Public Sector
Topic	DIGITAL-2022-CLOUD-AI-02-OPEN-AI
Funding scheme	DIGITAL-SIMPLE
Start Date	01/01/2023
Duration	30 Months
Project URL	https://beopen-dep.eu
Project Coordinator	ENGINEERING
Deliverable	D5.5 – HVDs catalogue (first version)
Work Package	WP5 - Communication, dissemination, and exploitation
Delivery Month (DoA)	M24
Version	V01
Actual Delivery Date	19/12/2024
Nature	Report
Dissemination Level	Public – PU
Lead Beneficiary	ID Vilnius
Authors	All Partners
Quality Reviewer(s)	ENG & PD
Keywords	HVD, Catalogue, API, Portal

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

D5.5 – HVDs catalogue (first version)

Advancing Open Data Integration and Accessibility through the BeOpen HVDs Catalogue

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

Table of Contents

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

List of Figures

List of Tables

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

Abbreviations and Acronyms

ACRONYM	Definition
AI	Artificial Intelligence
API	Application Programming Interface
DS	Digital Service
EDP	European Data Portal
HVD	High Value Dataset
UC	Use Case

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

1 Introduction

1.1 Objective of the HVDs Catalogue

The primary objective of the High-Value Datasets (HVDs) Catalogue is to enable the efficient dissemination and reuse of public open data across European Union member states, enhancing accessibility for both governmental bodies and the public. This deliverable aligns with the broader goals of the BeOpen project, aiming to improve the usability and interoperability of public datasets through adherence to standardized metadata models and data formats. The HVDs Catalogue is a foundational step toward achieving Europe-wide data integration and promoting transparency and data-driven decision-making at all levels.

1.2 Overview of BeOpen's Approach to HVDs Harmonization and Dissemination

The BeOpen project adopts a multi-layered approach to harmonizing and disseminating HVDs across various public portals, maximizing their accessibility and usability. This process is designed to ensure that data from diverse sources across pilot cities are standardized, enriched with meaningful metadata, and made available through open APIs.

Key steps in BeOpen's approach include:

- **Data Mapping and Standardization**: The project employs widely adopted European ontologies to map datasets into consistent, machine-readable formats. This step facilitates cross-regional data interoperability, making datasets easier to interpret and integrate with other public data sources.

- **Metadata Harmonization**: BeOpen leverages DCAT-AP and NGSI-LD standards to harmonize dataset metadata. This ensures that datasets published on national and European portals meet established European standards, enhancing their discoverability and trustworthiness.

- **Open Data Provisioning through APIs**: Recognizing the importance of accessible data for third-party applications, BeOpen provisions datasets via the CKAN API as the primary access point, with additional support for other standards when specific data types or use cases require them or see potential for improving accessibility. This approach enables seamless data retrieval and integration by external users, aligning with the official portal for European data requirements for harvestable APIs (European Commission, 2024). Indeed, upon release of the final version of the BeOpen HVDs catalogue, the datasets will also be available on the Idra portal and through its API, the tool that will allow the federation of BeOpen HVDs within the European data portal.

Using tools such as Idra, CKAN, the FIWARE Orion Context Broker, BeOpen ensures that datasets are readily accessible to developers, policymakers, and researchers, thereby promoting data-driven decision-making and public engagement with open data.

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

1.3 Achievability Statement

BeOpen is well-positioned to achieve its objective of harmonizing and disseminating HVDs due to the consortium’s extensive experience and the maturity of its technological assets. The partners bring to the project proven solutions and expertise from prior initiatives, allowing for a robust foundation in data interoperability, metadata harmonization, and API-driven dissemination.

Several key assets contribute to this goal:

- **Idra Open Data Federation Platform** (Engineering Ingegneria Informatica SpA, 2024): Developed by Engineering Ingegneria Informatica (ENG), Idra serves as a central component for open data harmonization, enabling datasets from various pilot cities to be aggregated, standardized, and federated across platforms. Its established use in prior open data projects ensures reliability and efficiency in data processing.

- **FIWARE Orion Context Broker** (FIWARE, 2024): A core part of the FIWARE platform, the Orion Context Broker enables BeOpen to support the NGSI-LD standard, essential for providing context-aware, real-time data to users. This tool is integral for ensuring that BeOpen’s HVDs can be accessed dynamically and are interoperable across systems.

- **Big Data Test Infrastructure (BDTI)** (European Commission, 2024): As part of the European Commission's Digital Europe Programme (DEP), BDTI provides a powerful environment for processing and analyzing large-scale public sector data. This infrastructure enhances BeOpen’s capacity to generate actionable insights and align datasets with the European Data Strategy. For instance, Naples’ pilot explores BDTI capabilities in addressing climate change adaptation by identifying relief hubs during urban heat island events. The integration of BDTI tools enables BeOpen to leverage advanced analytics and big data capabilities, aligning the project with emerging priorities in the digital transformation of Europe.

Together, these assets empower the BeOpen consortium to achieve its objectives, ensuring that HVDs are accessible, standardized, and useful for a wide array of stakeholders. The consortium’s approach is rooted in the use of European standards, ensuring the long-term sustainability and relevance of the datasets provided.

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

2 HVDs Data Catalogue Structure and Content

2.1 Overview of Data Format Standards

Consistency in data formats is essential to ensure interoperability, usability, and scalability across various platforms and portals. The BeOpen project emphasizes adopting widely accepted standards for HVDs to meet the requirements of both local and European open data initiatives. This section outlines the key formats currently in use and explores how they support the project's goals.

Most common data formats used:

1. **GeoJSON** is extensively used for geographic datasets, offering a lightweight and versatile format for encoding spatial data structures such as points, lines, and polygons. It is highly compatible with modern web applications, GIS software, and APIs. The benefits include easy visualization, simplicity to process programmatically, and inherent support of geospatial metadata standards.

2. **JSON** is utilized for non-spatial datasets, particularly those that focus on structured tabular or attribute-based data. It is widely recognized for its simplicity and broad compatibility across platforms. It is lightweight, human-readable, and easily parsed by a variety of programming languages and tools. Some pilot cities use JSON for datasets that do not require geospatial representation but still demand structured and harmonized metadata.

The adoption of GeoJSON and JSON ensures that the project adheres to best practices in data format standardization, promoting **interoperability –** datasets are easily transferable between platforms, including the BeOpen CKAN, national open data portals, and the European Data Portal (EDP). Also, **scalability** and **usability** fore lightweight formats enable efficient data processing, even for large datasets and end-users, from government agencies to the general public, benefit from simplified access and integration with widely used tools.

In addition to adopting standardized data formats, several pilot cities have undertaken significant efforts to improve the quality and usability of their datasets. For instance, certain pilots transitioned their datasets from proprietary formats, such as shapefiles, to open standards like GeoJSON, enhancing compatibility with APIs and simplifying integration with other platforms. This shift also ensures easier dataset visualization and processing in web-based applications.

Other pilots went beyond format standardization by leveraging smart models and automated workflows to enrich their datasets. For example, machine learning techniques were applied to preprocess and classify data, improving accuracy and enabling better insights. These advancements not only align with the goals of the BeOpen project but also demonstrate the pilots' commitment to providing high-quality, reusable data.

By focusing on format improvements and integrating intelligent data processing methods, the pilots have ensured that their datasets are not only aligned with the project's standards but also optimized for diverse end-user needs. These enhancements have resulted in more accessible, interoperable, and future-proof datasets that contribute to the project's vision of seamless open data dissemination across Europe.

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

2.2 Enhancing Data Accessibility via Standardized APIs

To maximize the utility of HVDs, the BeOpen project seeks to provide data in machine-readable formats accessible via standard APIs. This approach simplifies data discovery and access for third-party systems, applications, and services, fostering innovation and encouraging a broader ecosystem of public data utilization. The HVDs Catalogue will support various data attributes by leveraging different types of APIs, ensuring a tailored solution for diverse use cases.

A core component of the BeOpen project is assisting pilots in selecting the most suitable API types for their datasets. To streamline this process, an API typologies mapping was created. Table *1* categorizes APIs based on their capabilities, such as geospatial data handling, contextual relationships, and real-time updates. By providing pilots with clear guidance, this mapping enables them to align their dataset characteristics with the API most suited for their use case.

For example, pilots managing static datasets with basic retrieval needs can leverage the CKAN API, while datasets involving contextual metadata or IoT data streams may benefit from NGSI-LD. This typologies mapping has helped pilots confidently select APIs that improve the accessibility and usability of their datasets, ensuring a balance between technical feasibility and user requirements.

To provide a solid foundation for accessibility, most HVDs are uploaded as static datasets in standard formats such as GeoJSON or JSON. These formats are immediately accessible via the CKAN API, allowing users to perform basic data retrieval and analysis. This is a solid foundation and makes the data immediately accessible via the CKAN API. However, the BeOpen project goes further, enabling the addition of specialized API endpoints tailored to the unique requirements of the data. CKAN's external link resource feature allows to offer multiple API endpoints under the same dataset entry. Potential additional API endpoints include Python API (selected by four use cases) for custom data processing and machine learning workflows. Flexible and easily adaptable for use cases requiring data transformation, automation, or statistical analysis. Another potential API type identified by a use case is NGSI-LD API for a dataset with contextual linked data. The combination of these APIs ensures that datasets are discoverable and accessible in ways that meet diverse use cases, from general open data to real-time and statistical applications.

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

Data type	CKAN API	NGSI-LD API	SDMX API	OGC API	SensorThings API	JSON	SPARQL Endpoint	Python
General Open Data	✔	✔ (if contextual metadata)	✕	✔ (basic metadata)	✕	✔	✔ (if in RDF)	✔ (customizable)
Time Series Data	✔ (limited, metadata)	✔ (contextual & time)	✔	✔ (supports time in geospatial data)	✔	✕	✔ (limited)	✔ (customizable)
Real-Time or Streaming Data	✕	✔	✕	✔ (with standards like WFS or MQTT)	✔	✕	✕	✔ (customizable)
Statistical Data	✕	✔ (if dynamic context)	✔	✕	✕	✕	✔ (if structured in RDF)	✔ (customizable)
Sensor or IoT Data	✕	✔	✕	✕	✔	✕	✕	✔ (customizable)
Geospatial Data	✔ (basic metadata only)	✔	✕	✔ (supports complex geospatial operations)	✔ (supports spatial queries)	✕	✔ (if in RDF with geospatial ontology)	✔ (customizable)
Linked Data or Semantic Queries	✕	✔ (supports linked data relationships)	✕	✕	✕	✕	✔	✔ (customizable)

Table 1: API typologies mapping

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

CKAN API: provides access to data stored in CKAN, suitable for basic data retrieval. Ideal for general open data, accessible in formats like CSV, GeoJSON, and metadata queries.

NGSI-LD API: if a static dataset has dynamic aspects or contextual relationships (like geospatial or IoT-related data), the pilot has the option to transform it into the NGSI-LD format. For example, data about air quality measurements, traffic counts, or sensor data could benefit from NGSI-LD, as this format supports relationships, context, and even real-time updates.

SDMX API: for datasets that contain structured statistical or time-series data (like demographic statistics, monthly environmental data, or energy usage reports), the pilot could reformat them to SDMX. This makes the data more suitable for structured analysis and sharing within statistical platforms.

OGC API: provides access to geospatial data with support for spatial queries. Suitable for geospatial applications, like maps, GIS data, and environmental datasets requiring complex spatial queries.

SensorThings API: designed for IoT and sensor data with temporal support. Best for sensor networks that require real-time updates, like weather, pollution, and traffic monitoring systems.

JSON API: a lightweight data-interchange format, easy for machines to parse and generate. Ideal for basic API responses or when simplicity and cross-platform compatibility are needed for data exchange.

SPARQL Endpoint: a protocol and query language for RDF data, allowing complex queries over linked data. Useful for querying semantic datasets, especially in knowledge graphs, ontologies, and linked data platforms.

By adding these different formats as external resources in CKAN, the pilots:

- Enable **users to choose the format** best suited to their needs (e.g., those needing raw data versus those needing structured statistical data).
- Ensure the data becomes **discoverable and accessible** via multiple APIs (CKAN, NGSI-LD, Python), broadening its usability and accessibility.
- Facilitate **interoperability** across platforms and data consumers, as some users or systems may be configured to work specifically with for example NGSI-LD or SDMX standards.

Idra then ensures that all formats are harmonized and standardized, making them compatible with European Data Portal requirements. By having CKAN, NGSI-LD, and other datasets all accessible within a single system, the portal can harvest a complete set of HVDs for European-wide accessibility.

This approach maximizes the reach and utility of the HVDs, allowing a broader audience to use the data in ways that suit their specific use cases and technical needs. It also makes the data interoperable with various data platforms and applications, ensuring better reuse and integration.

2.3 List of Catalogues and Data Portals

A central tool among the many proposed by the BeOpen framework is the **CKAN portal**. It is the world's leading open-source data management system and specifically CKAN is an open-source platform for powering data hubs and data portals, making it easy to publish, share and use data. It was chosen as portal for adding and publishing the HVDs developed during the BeOpen project.

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

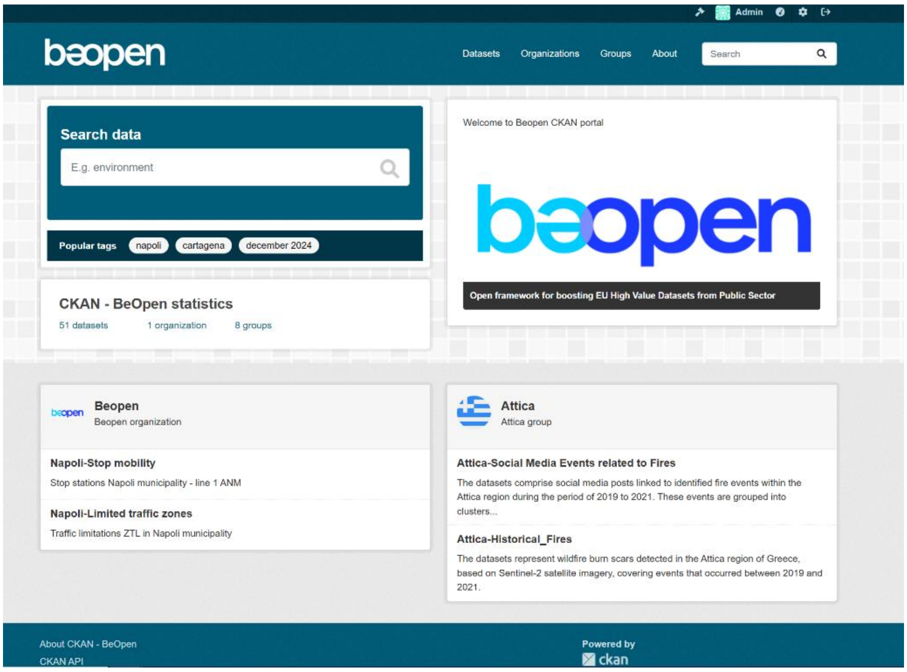


Figure 1: BeOpen CKAN Portal

The BeOpen CKAN platform serves as the central repository for HVDs contributed by the pilot cities. Accessible at https://platform.beopendep.it/ckan/dataset/, the platform allows users to discover, explore, and access datasets through a user-friendly interface and machine-readable APIs. Groups were created within it, one for each city involved in the project, as shown in Figure 2.

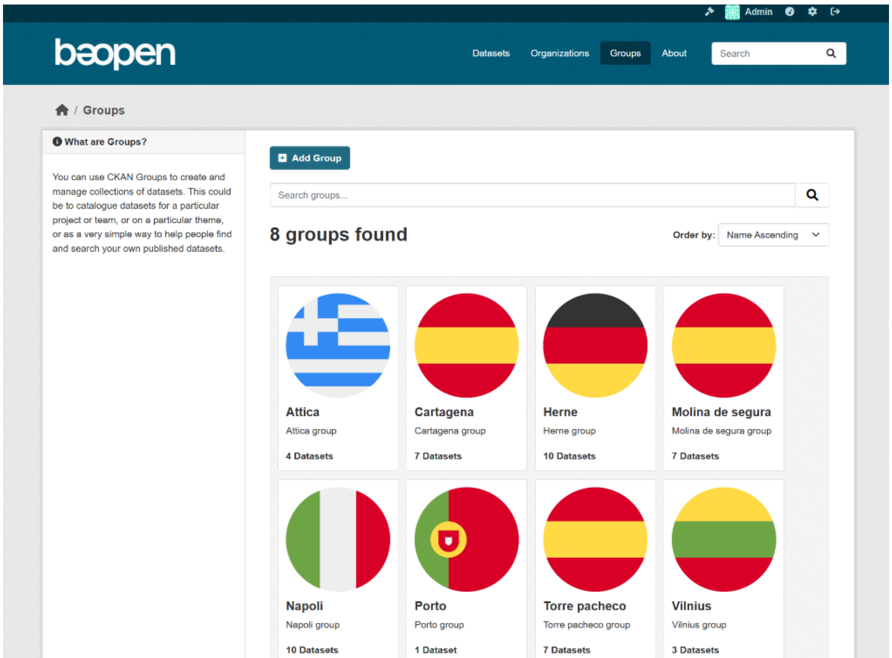


Figure 2: BeOpen CKAN Portal - Groups

Effective dissemination of HVDs is a cornerstone of the BeOpen project. To achieve this, pilot cities are leveraging a combination of local, national, and centralized platforms. These include their respective

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

open data portals and the BeOpen CKAN, which acts as a unifying catalogue for harmonized data publication. This multi-layered approach ensures data accessibility at both regional and European levels, aligning with the Open Data Directive and the European Data Portal (EDP) requirements. Table *2* below provides an overview of the selected catalogues and portals in separate pilot countries, highlighting their role in the dissemination process and their alignment with the project's goals.

Pilot City	Open Data Portal	EDP Harvestability	Use of Idra	Portal URL
Attica region	N/A	N/A	Yes	N/A
Cartagena	Portal Nacional de Datos Abiertos - España	Yes	Yes	https://datos.gob.es/
Herne	Open Data Portal Herne	No	Yes	https://opendata.herne.de/
Molina de Segura	Portal Nacional de Datos Abiertos - España	Yes	Yes	https://datos.gob.es/
Naples	N/A	N/A	Yes	N/A
Porto	Porto Open Data	No	Yes	https://opendata.porto.digital/
Torre Pacheco	Portal Nacional de Datos Abiertos - España	Yes	Yes	https://datos.gob.es/
Vilnius	Vilnius Open Data Portal	No	Yes	https://opendata.vilnius.lt/

Table 2: Selected portals of piloting countries

BeOpen CKAN serves as a central catalogue. In addition to national open data portals, almost all pilot cities will publish their datasets on the BeOpen CKAN platform. This ensures centralized access to HVDs while facilitating further federation with European-level platforms such as the European Data Portal (EDP). An important note to make is if the pilots' selected catalogues are added to Idra, the upload of datasets to the BeOpen CKAN is not compulsory (in case of Porto) as well as if the selected portal already supports EDP harvestability, of which the Spanish use cases are an example of and exemplify a collaborative approach, sharing a common national platform. Figure *3* highlights the importance of using Idra to ensure harmonization and harvestability, aligning datasets with EDP requirements.

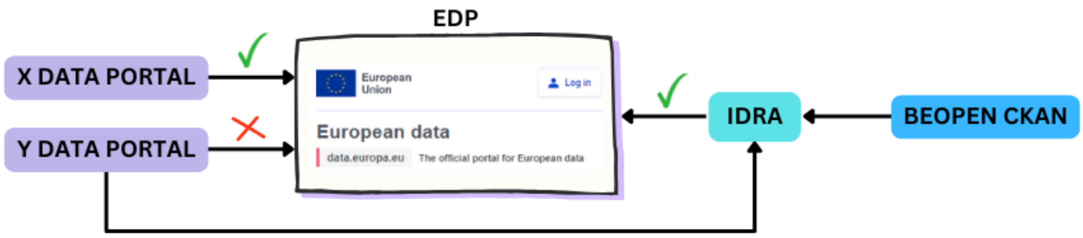


Figure 3: Dataset harvestability by the EDP though Idra

Contribution to dissemination goals. The selected catalogues and portals play a crucial role in the dissemination and accessibility of HVDs. Essentially, the government can officially recognize or endorse any public open data portal that can be considered as national infrastructure. By leveraging both local (e.g., Vilnius Open Data Portal), national (e.g., Portal Nacional de Datos Abiertos - España), and centralized platforms (BeOpen CKAN), the project ensures:

- **Increased accessibility**: Localized portals address regional needs, while national and centralized platforms promote broader data access.

- **Compliance with open data directives**: Use of national portals that are EDP-harvestable ensures compliance with European open data policies.

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

- **Scalability and reusability**: BeOpen CKAN serves as a flexible, centralized repository that complements the capabilities of local and national portals.

Compliance with the European Data Portal Standards. Idra, the data harmonization tool, plays a crucial role in preparing datasets for EDP harvesting by transforming and aligning metadata across formats. Through Idra, each dataset's metadata is mapped to the EDP's DCAT-AP standard, ensuring that the data conforms to EDP guidelines. This mapping process facilitates the EDP's ability to recognize, categorize, and index datasets from BeOpen. Metadata is enriched with necessary attributes (e.g., descriptions, keywords, and geospatial coverage) to enhance discoverability. Regular quality checks ensure that metadata remains accurate and meets EDP standards, which further supports reliable harvesting and cataloguing by the EDP. The BeOpen consortium demonstrates the interconnected processes that support Europe-wide accessibility. By maintaining compliance with EDP requirements, BeOpen maximizes the visibility and impact of its HVDs, promoting transparency, data-driven decision-making, and cross-border collaboration.

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

3 HVD Assessment and Improvement Process

3.1 Overview of Baseline and Scoring System

The BeOpen framework (BeOpen consortium, 2024) provides tools to allow a clear understanding of the requirements to be followed in order to improve your datasets and make them High Value Datasets. HVD means a dataset structured and made available following a certain methodology defined within the Commission Implementing Regulation (EU) 2023/138 (European Union, 2024).This document establishes a list of specific high-value datasets and the arrangements for their publication and re-use. The requirements of this regulation have been summarized and outlined within the BeOpen framework in a specific section called **High Value Datasets Operating Guide** (Figure 4: HVD Operating Guide in the BeOpen framework). The purpose of this guide is to help interested users in correctly understanding the content of the HVD Regulation.

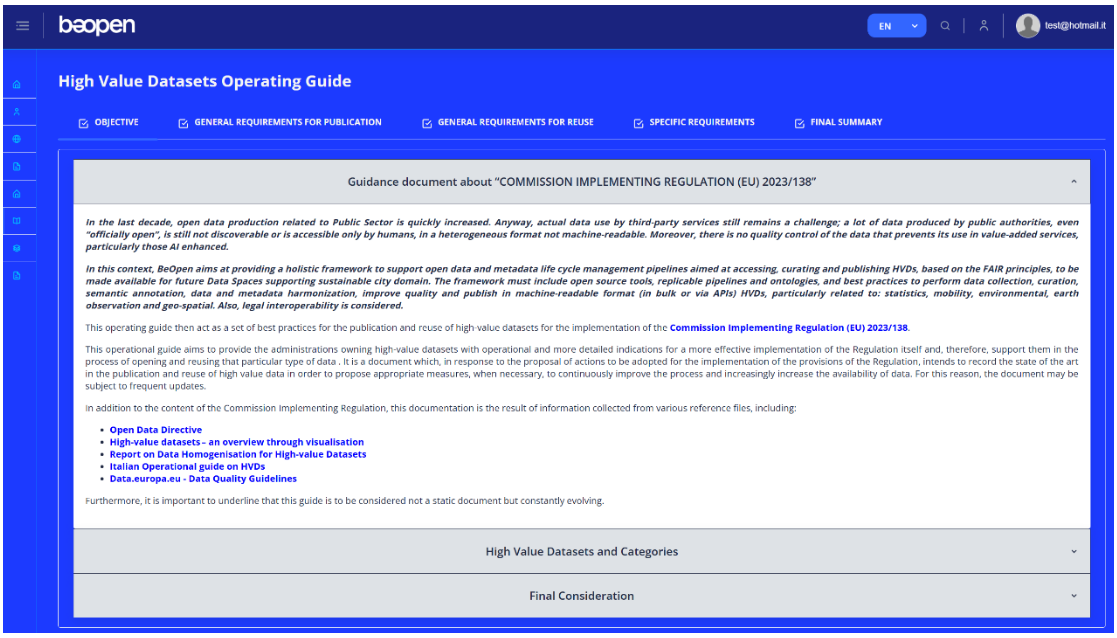


Figure 4: HVD Operating Guide in the BeOpen framework

The BeOpen framework also provides another tool, the **HVD Impact Assessment**, showed in Figure 5, for evaluating datasets against the HVDs compliance criteria. It is a scoring system that ensures that datasets are assessed consistently across partner cities, allowing for a clear understanding of each dataset's readiness for public dissemination and compliance with European open data standards.

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

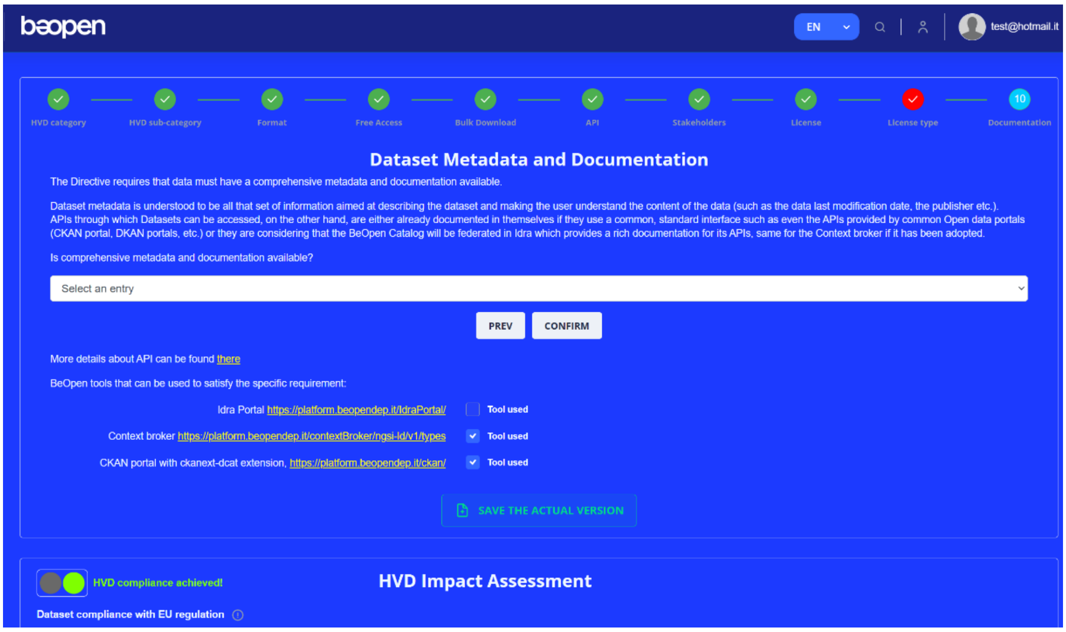
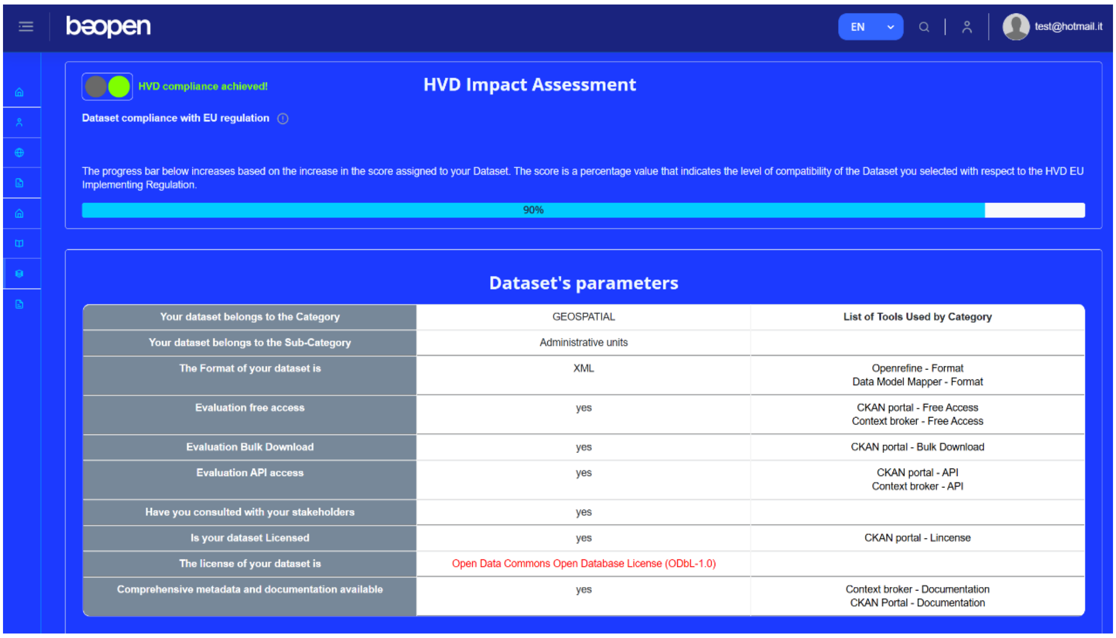


Figure 5: HVD Impact Assessment in the BeOpen Framework

The tool, in addition to providing in each step the reference to the HVD Operating Guide to allow the user to better analyse the request, contains the tools proposed by the BeOpen framework that could be useful to satisfy the requirement. In the second part, as shown by Figure 6, the tool provides the current score, a graphic indicator that shows whether compliance with the regulation has already been achieved or not and, at the end, a report that contains the answers to the previous steps already completed.



Your dataset belongs to the Category	GEOSPATIAL	List of Tools Used by Category
Your dataset belongs to the Sub-Category	Administrative units	
The Format of your dataset is	XML	Openrefine - Format Data Model Mapper - Format
Evaluation free access	yes	CKAN portal - Free Access Context broker - Free Access
Evaluation Bulk Download	yes	CKAN portal - Bulk Download
Evaluation API access	yes	CKAN portal - API Context broker - API
Have you consulted with your stakeholders	yes	
Is your dataset Licensed	yes	CKAN portal - Lincense
The license of your dataset is	Open Data Commons Open Database License (ODbL-1.0)	
Comprehensive metadata and documentation available	yes	Context broker - Documentation CKAN Portal - Documentation

Figure 6: HVD Impact Assessment in the BeOpen Framework (second part)

The first version of the calculation formula involved identifying the relevant criteria for defining HVDs, resulting in eight criteria with no prioritisation. This formulation was shared among the partners for

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

providing information about the existing data and whether each criterion was met for individual datasets. In total, a dataset could achieve a maximum score of 8, being identified as an HVD. The HVDs Impact Assessment Tool evaluates datasets across 10 critical criteria, each designed to assess the dataset's alignment with key open data principles and compliance requirements. The criteria are as follows:

1. **HVD Category Selection**: This initial step involves classifying the dataset under the appropriate HVD category, aligning it with specific high-value thematic areas as defined by the European Union.

2. **HVD Sub-Category Selection**: Further refinement of the dataset's categorization through sub-categories to ensure accuracy and relevance within the chosen thematic area.

3. **Data Format Type**: The format of the dataset is evaluated, with priority given to machine-readable formats (e.g., HTML, CSV, JSON). Machine-readable formats are essential for interoperability and ease of reuse, a critical component of HVD compliance.

4. **Free Access to Dataset**: This criterion verifies whether the dataset can be accessed without any restrictions or financial barriers, ensuring that it meets the fundamental principles of open data accessibility.

5. **Bulk Download Availability**: Bulk download functionality is assessed to determine if users can access the entire dataset in a single download, an essential feature for facilitating large-scale analysis and reuse by external entities.

6. **API Access**: Datasets are evaluated for API availability, which enhances accessibility and enables seamless integration into third-party systems, applications, and services, promoting dynamic data usage.

7. **Stakeholder Consultation**: The assessment tool verifies whether consultations with relevant stakeholders have been conducted. This ensures that the dataset aligns with the needs and expectations of potential users and beneficiaries, thereby enhancing its value and relevance.

8. **Licensing**: A licensing check is conducted to verify the presence of a suitable open license, ensuring legal clarity and promoting reuse.

9. **Type of License**: The type of license is documented, with preference for licenses that align with open standards (e.g., CC BY 4.0 or equivalent), facilitating unrestricted reuse and adaptation.

10. **Comprehensive Metadata and Documentation**: This final criterion ensures that the dataset includes extensive metadata and documentation. High-quality metadata improves discoverability, while comprehensive documentation supports ease of use and understanding.

The first two steps are mandatory to proceed with the assessment; they therefore do not weigh in terms of scores but only place the dataset in a specific category among those proposed by the EU Commission implementing regulation., Then, although the assessment tool encompasses 10 steps, the maximum score attainable is 8 points. This scoring system reflects the core mandatory aspects required for a dataset to be considered HVD-compliant, as outlined by BeOpen standards. Datasets achieve compliance when they fulfil these core requirements, even if all 10 steps are not strictly necessary, such as stakeholder consultations and points 8 and 9 encompass the same requirement yet demands detailed information as to make sure the dataset is licensed with the right license.

The initial baseline assessment serves as a starting point, capturing each dataset's current compliance status and identifying areas for improvement. This baseline score, recorded at the beginning of the

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

project, enables the tracking of progress over time. By year-end, datasets are reassessed, with actions documented to reflect steps taken by pilot cities to elevate datasets closer to full compliance. This iterative scoring process fosters ongoing improvement and enhances the overall quality and accessibility of HVDs

The second version involved a revision of the EU Commission implementing regulation, allowing for the prioritisation of mandatory criteria in identifying a dataset as an HVD. The criteria concerning machine-readability, bulk download, stakeholder consultation, API access, and free access to datasets are considered mandatory, reflecting crucial aspects to ensure the minimum quality of the data and its compatibility with HVD requirements. The three optional criteria, namely licensing/type of license, comprehensive metadata and documentation, allow for an additional evaluation of the data's robustness for their later integration into Digital Services.

Furthermore, the scoring scale was changed from 0 to 8 to a scale in percentage, making the score more intuitive and easier for users to interpret. Thus, the mandatory criteria each carry a weight of 17% (total of 85%), and a dataset must achieve at least this score to be identified as an HVD. The three optional criteria each carry a weight of 5%, contributing to a total of 15%, which is a lower weight compared to each mandatory criterion (17%). Data with a score of 100% fully meet all criteria, both mandatory and optional.

Table 3 shows the logic of the second version of the HVD impact assessment tool algorithm. This improved version provides the dataset score in percentage and the image relates this percentage with its meaning.

< 50%	No HVD	At least 3 mandatory criteria are missing
51% - 67%	No HVD	2 mandatory criteria missing
68% - 85%	No HVD	1 mandatory criterion missing
= 85%	HVD	Meets all mandatory criteria
> 85%	HVD	Meets all mandatory criteria + some optional
= 100%	HVD	All mandatory and optional criteria are met

Table 3: Score in percentage of the second version of the HVD impact assessment tool

3.2 Pilot Contributions on HVD Improvement

The BeOpen project relies on the active participation of its pilot cities to identify, harmonize, and disseminate HVDs in alignment with European open data initiatives. Each pilot city plays a critical role in addressing local challenges while adhering to project-wide standards for data interoperability and accessibility.

In this section, we outline the contributions made by the pilot cities toward achieving the project's objectives. These contributions span various domains, including dataset curation, format standardization, metadata enrichment, and the adoption of innovative technologies for improved usability. The pilots'

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

efforts exemplify diverse approaches to ensuring the availability and quality of HVDs, tailored to their unique geographic, social, and administrative contexts. The subsequent subsections provide a detailed account of each pilot's contributions, highlighting their innovations, challenges, and best practices, along with dataset overviews.

3.2.1 Pilot Attica: Natural Disaster Shield for Citizens and Environment

Attica use case has selected 5 datasets for the purposes of visualisation, one dataset as a result of the creation of the digital service and a historical dataset owned by the National Observatory of Athens for use in the digital service (**Table 3**).

Dataset	Use
Historical Fires Dataset	Validation role for DS, Visualisation
Social Media Posts Related to Fires	Development of DS
Social Media Events related to Fires	Detection, Visualisation
Predicted Fire Danger Levels Using EFFIS Data	Supporting Role
Copernicus Data Space Ecosystem Satellite Data	Development of DS, Verification

Table 4: Attica pilot HVD overview

The Historical Fires Dataset shows the Wildfires occurred between 2019 and 2021 in Attica region, Greece. This geospatial dataset plays a critical role in verifying the accuracy of detected wildfire events, ensuring that identified incidents align with actual fire-affected areas. Burned scars provide a visual representation of affected areas, aiding stakeholders in assessing the severity of past incidents and planning recovery or preventive measures.

The Social Media Posts Related to Fires dataset includes the Fire-related posts from X social media platform including Tweet ID, Tweet URL, metadata produced from CERTH processing (i.e., timestamp, location extracted from text, links to related images, fire danger level, relevance score). The dataset undergoes initial analysis using a clustering technique to identify potential fire incidents, with the DBSCAN method—a density-based clustering algorithm—being employed. It groups closely packed points within a defined space, effectively identifying clusters of fire-related activity.

The Social Media Events Related to Fires dataset is identified using the above outlier detection method and is utilized for visualization within the BeOpen framework, effectively displaying the detected fire events.

The Predicted Fire Danger Levels Using EFFIS dataset shows the Predicted fire danger levels for the areas of Athens. Predictions were made using an MLP model trained on embeddings generated by SatCLIP and based on fire danger data from the European Forest Fire Information System (EFFIS). This dataset is used to support the event detection algorithm by providing predictive insights into areas at high risk of fire.

The Copernicus Data Space Ecosystem provides access to Sentinel-5 satellite measurements from 2019 to 2021 in the Attica region. These measurements are utilized to verify detected events by analyzing air pollutant levels through temporal and spatial outlier detection methods and integrating them with non-EO data.

Dataset	Initial score	End-of-year score
Historical Fires Dataset	2	85%
Social Media Posts Related to Fires	2	90%
Social Media Events related to Fires	2	90%
Predicted Fire Danger Levels Using EFFIS Data	2	90%

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

Copernicus Data Space Ecosystem Satellite Data	2	39%

Table 5: Attica pilot HVD scores

Actions taken to improve the datasets. The datasets converted to JSON data into the NGSI-LD format, enabling semantic interoperability and adherence to linked data standards through **FIWARE Smart Data Models**. Then the datasets uploaded into the **BeOpen CKAN** – Attica group using the BeOpen framework. Additionally, we aim to federate CKAN with ***Idra***, which acts as a unified access point for searching and discovering open datasets from heterogeneous sources, enhancing interoperability and accessibility across platforms. For visualization, we aim to use **BeOpen Framework's capabilities** to display the processed data outputs of the analysis. Finally, we aim to use the **Metadata Quality Validator** to assess the quality of the metadata.

3.2.2 Pilot Cartagena: Data Visualisation Platform

The Cartagena pilot addresses three interconnected use cases leveraging HVDs and digital services to enhance urban decision-making and policy development.

The first use case focuses on urban security in the historic city center, utilizing data on air quality, noise levels, crowd monitoring, and traffic to improve safety and manage critical infrastructure.

The second explores the sustainable introduction of LED lighting, assessing its health and environmental impacts through satellite luminosity maps, manually collected luminosity data, and energy consumption monitoring.

The third targets climate change mitigation by addressing urban heat islands and heatwaves using Copernicus satellite data, environmental metrics like humidity and temperature, and air quality datasets.

Together, these use cases demonstrate how HVDs can transform urban management, optimize resources, and address challenges related to security, sustainability, and climate resilience.

Dataset	Use
Road traffic data	Real road traffic information, characterizing type of emissions, speed, vehicle capacity, type. Generates historical data over time
Crowd Monitoring	It characterizes the pedestrian flow through areas of the city. In real time and historically.
Luminosity indices	Night-time lighting measurements of streets with LED lights. Real-time and historical.
Energy consumption	Historical archive of energy consumption by LED street lightning.
Environmental Data	Measures meteorological conditions, temperature, humidity, precipitation, wind. Characterizes environmental conditions of gaseous pollutants and particles. Real time and generates historical data.
Air Quality	Collects measurements of different pollutants measured by local stations deployed in several points of the city.
Noise	Real-time and historical measurement of noise produced by traffic and pedestrians.

Table 6: Cartagena HVD overview

Actions taken to improve the datasets. The improvement of HVDs in Cartagena follows a structured approach supported by the BeOpen framework tools. Datasets from Libelium IoT devices (including Air Quality, Environmental Data, Crowd Monitoring, and Noise Levels) already adopt a smart data model and contain well-defined metadata. This ensures their interoperability and quality, requiring less extensive improvements compared to other datasets.

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

For datasets from external sources, the BeOpen tools are used to enhance metadata quality and ensure standardization. These datasets are processed through **OpenRefine** for cleaning and mapped to standardized data models using the **Data Model Mapper** (DMM). Metadata is validated using the **Metadata Quality Validator**, ensuring compliance with the DCAT-AP standard before federation into the **Idra** platform. Once improved, all datasets are integrated into the **CKAN**-based platform, enabling API accessibility and visualization via Grafana. The **Impact Assessment Tool** was also used to verifying compatibility with the EU Implementing Regulation of the HVDs. This process ensured their usability in municipal applications as HVDs.

This process guarantees that both IoT-generated and externally sourced datasets meet HVD requirements, facilitating consistent, interoperable, and reusable data to support Cartagena's urban security, sustainable lighting, and climate change mitigation initiatives.

Dataset	Initial score	End-of-year score
Road traffic data	3	100%
Crowd Monitoring	3	100%
Luminosity indices	3	100%
Energy consumption	3	100%
Environmental Data	4	100%
Air Quality	4	100%
Noise	4	100%

Table 7: Cartagena HVD scores

3.2.3 Pilot Torre Pacheco: Data Visualisation Platform

The use cases and datasets applied in Torre Pacheco are identical, further validating the adaptability of the approach to different urban contexts.

This pilot also addresses three interconnected use cases leveraging HVDs and digital services to enhance urban decision-making and policy development. The first use case focuses on urban security in the historic city centre, utilizing data on air quality, noise levels, crowd monitoring, and traffic to improve safety and manage critical infrastructure. The second explores the sustainable introduction of LED lighting, assessing its health and environmental impacts through satellite luminosity maps, manually collected luminosity data, and energy consumption monitoring. The third targets climate change mitigation by addressing urban heat islands and heatwaves using Copernicus satellite data, environmental metrics like humidity and temperature, and air quality datasets. Together, these use cases demonstrate how HVDs can transform urban management, optimize resources, and address challenges related to security, sustainability, and climate resilience.

Dataset	Use
Road traffic data	Real road traffic information, characterizing type of emissions, speed, vehicle capacity, type. Generates historical data over time
Crowd Monitoring	It characterizes the pedestrian flow through areas of the city. In real time and historically.
Luminosity indices	Night-time lighting measurements of streets with LED lights. Real-time and historical.
Energy consumption	Historical archive of energy consumption by LED street lightning.

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

Environmental Data	Measures meteorological conditions, temperature, humidity, precipitation, wind. Characterizes environmental conditions of gaseous pollutants and particles. Real time and generates historical data.
Air Quality	Collects measurements of different pollutants measured by local stations deployed in several points of the city.
Noise	Real-time and historical measurement of noise produced by traffic and pedestrians.

Table 8: Torre Pacheco HVD overview

Actions taken to improve the datasets. The improvement of HVDs in Torre Pacheco follows a structured approach supported by the BeOpen framework tools. Datasets from Libelium IoT devices (including Air Quality, Environmental Data, Crowd Monitoring, and Noise Levels) already adopt a smart data model and contain well-defined metadata. This ensures their interoperability and quality, requiring less extensive improvements compared to other datasets.

For datasets from external sources, the BeOpen tools are used to enhance metadata quality and ensure standardization. These datasets are processed through **OpenRefine** for cleaning and mapped to standardized data models using the **Data Model Mapper** (DMM). Metadata is validated using the **Metadata Quality Validator**, ensuring compliance with the DCAT-AP standard before federation into the Idra platform. Once improved, all datasets are integrated into the **CKAN**-based platform, enabling API accessibility and visualization via Grafana. The **Impact Assessment Tool** was also used to verifying compatibility with the EU Implementing Regulation of the HVDs. This process ensured their usability in municipal applications as HVDs.

This process guarantees that both IoT-generated and externally sourced datasets meet HVD requirements, facilitating consistent, interoperable, and reusable data to support Cartagena's urban security, sustainable lighting, and climate change mitigation initiatives.

Dataset	Initial score	End-of-year score
Road traffic data	3	100%
Crowd Monitoring	3	100%
Luminosity indices	3	100%
Energy consumption	3	100%
Environmental Data	4	100%
Air Quality	4	100%
Noise	4	100%

Table 9: Torre Pacheco HVD scores

3.2.4 Pilot Molina de Segura: Data Visualization Platform

The Molina de Seguro pilot also serves as a replica of the Cartagena pilot, implemented in a medium sized city to demonstrate the solution's effectiveness and replicability. The use cases and datasets applied in Molina are identical, further validating the adaptability of the approach to different urban contexts.

This pilot also addresses three interconnected use cases leveraging HVDs and digital services to enhance urban decision-making and policy development. The first use case focuses on urban security in the historic city center, utilizing data on air quality, noise levels, crowd monitoring, and traffic to improve safety and manage critical infrastructure. The second explores the sustainable introduction of LED lighting, assessing its health and environmental impacts through satellite luminosity maps, manually collected luminosity data, and energy consumption monitoring. The third targets climate change mitigation by addressing urban heat islands and heatwaves using Copernicus satellite data, environmental metrics like humidity and temperature, and air quality datasets. Together, these use cases demonstrate how HVDs can transform urban management, optimize resources, and address challenges related to security, sustainability, and climate resilience.

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

Dataset	Use
Road traffic data	Real road traffic information, characterizing type of emissions, speed, vehicle capacity, type. Generates historical data over time
Crowd Monitoring	It characterizes the pedestrian flow through areas of the city. In real time and historically.
Luminosity indices	Night-time lighting measurements of streets with LED lights. Real-time and historical.
Energy consumption	Historical archive of energy consumption by LED street lightning.
Environmental Data	Measures meteorological conditions, temperature, humidity, precipitation, wind. Characterizes environmental conditions of gaseous pollutants and particles. Real time and generates historical data.
Air Quality	Collects measurements of different pollutants measured by local stations deployed in several points of the city.
Noise	Real-time and historical measurement of noise produced by traffic and pedestrians.

Table 10: Molina de Segura HVD overview

Actions taken to improve the datasets. The improvement of HVDs in Molina de Segura follows a structured approach supported by the BeOpen framework tools. Datasets from Libelium IoT devices (including Air Quality, Environmental Data, Crowd Monitoring, and Noise Levels) already adopt a smart data model and contain well-defined metadata. This ensures their interoperability and quality, requiring less extensive improvements compared to other datasets. For datasets from external sources, the BeOpen tools are used to enhance metadata quality and ensure standardization. These datasets are processed through **OpenRefine** for cleaning and mapped to standardized data models using the **Data Model Mapper** (DMM). Metadata is validated using the **Metadata Quality Validator**, ensuring compliance with the DCAT-AP standard before federation into the Idra platform. Once improved, all datasets are integrated into the **CKAN**-based platform, enabling API accessibility and visualization via Grafana. The **Impact Assessment Tool** was also used to verifying compatibility with the EU Implementing Regulation of the HVDs. This process ensured their usability in municipal applications as HVDs.

This process guarantees that both IoT-generated and externally sourced datasets meet HVD requirements, facilitating consistent, interoperable, and reusable data to support Cartagena's urban security, sustainable lighting, and climate change mitigation initiatives.

Dataset	Initial score	End-of-year score
Road traffic data	3	100%
Crowd Monitoring	3	100%
Luminosity indices	3	100%
Energy consumption	3	100%
Environmental Data	4	100%
Air Quality	4	100%
Noise	4	100%

Table 11: Molina de Segura HVD scores

3.2.5 Pilot Herne: AI-Tools for Street Management Investments

The use case of "AI-Tools for Street Management Investments" focuses on enhancing and integration datasets to address the challenges of managing Herne's road infrastructure. The core dataset, Road Condition, coming from a data hub of assessment information, is continuously updated using edge devices attached to city vehicles, which classify and assess road damages with standardized rating for

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

street segments. The second dataset, Road traffic, represents the traffic flow and integrates sensor data into the service, allowing the analysis of how traffic impacts road condition.

The other datasets are examples of external factors and provide supplementally insights, enabling the analysis of environmental, structural and traffic related influences on the road network.

Dataset	Use
Road condition	Core dataset represents road segments with standardized ratings of streets condition.
Road traffic	Represents traffic flow from one street sensor, based on vehicle count, classification and speed.
City Boundaries	External factor. Provides spatial boundaries.
Construction Sites	External factor. Ongoing or planned construction that may disrupt traffic and affect road condition
Main Traffic Routes	External factor. Highlights key routes in the city.
Speed Zones	External factor, maps area with speed limits influencing road wear.
Areas with High Heat Load and Vulnerability - Climate Change	External factor defines zones with high heat exposure.

Table 12: Herne pilot (UC11) HVD overview

Actions taken to improve the dataset. The "**Road condition**" and "**Road traffic**" were exported from the municipal IoT platform as JSON files and uploaded to the BeOpen **CKAN** portal. Metadata for each dataset was generated using CKAN and linked to **Idra**, where RDF files for both individual datasets (via CKAN) and the entire catalogue (via Idra) were created. These were evaluated using the **Metadata Quality Validator**, ensuring compliance with DCAT-AP standards and achieving sufficient quality scores.

The **Impact Assessment Tool** is also used to verifying compatibility with the EU Implementing Regulation of the HVDs. This process ensured their usability in municipal applications as HVDs.

The additional datasets from the city's open data portal require only minor metadata preparation regarding the DCAT-AP, as they already fulfilled most HVD criteria.

The harmonized datasets resulting from the digital service are used in municipal applications, such as planning construction projects, optimizing maintenance schedules, and managing road infrastructure investments. These datasets enable data-driven decisions by integrating road condition and traffic data with other contextual information, such as environmental factors and construction activities.

Dataset	Initial score	End-of-year score
Road condition	1	78%
Road traffic	1	78%
City Boundaries	7	78%
Construction Sites	7	78%
Main Traffic Routes	7	78%
Speed Zones	7	78%
Areas with High Heat Load and Vulnerability - Climate Change	7	78%

Table 13: Herne pilot (UC11) pilot HVD scores

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

3.2.6 Pilot Herne: Management of Large-Scale Events and Civil Protection

This use case focuses on improving real-time data collection and forecasting to enhance safety and mobility during large scale events like the "Crange Kirmes" fair in the city. The used datasets are "**Parking space occupancy**", reported by parking lor operators through a simple app; the "**Road traffic**" which tracks vehicles count and speed on specific street using sensors; and "**People density**", which monitors crowd density at specific hotspots using edge cameras.

Dataset	Use
Parking space occupancy	Reports the percentage of occupied parking spaces, enabling proactive parking management
Road traffic	Represents traffic flow from one street sensor, based on vehicle count, classification and speed
People density	Track crowed density at hotspots on the fair using edge cameras

Table 14: Herne pilot (UC12) HVD overview

Actions taken to improve the dataset. The datasets Parking Space Occupancy, Road Traffic, and People Density were harmonized into time-series formats in the city's IoT data platform. Semantic harmonization was performed to align the datasets with predefined data models, ensuring compatibility and interoperability within municipal applications.

The datasets were exported to the BeOpen **CKAN** portal, where Metadata for each dataset was generated using and linked to **Idra.**

The **Metadata Quality Validator** was used to ensure compliance with DCAT-AP standards and achieve sufficient metadata quality scores. The **Impact Assessment Tool** was also used to confirm that the datasets met the HVDs EU Implementing Regulation requirements.

These datasets provide real-time overview of event dynamics and support forecasting to prevent overcrowding and traffic jams.

Dataset	Initial score	End-of-year score
Parking space occupancy	1	78%
Road traffic	1	78%
People density	1	78%

Table 15: Herne pilot (UC12) HVD scores

3.2.7 Pilot Porto: Machine learning for forecast urban floods

The BeOpen Porto Pilot aims to address urban flood challenge by developing data-driven solutions to improve flood prediction, response, and recovery. By integrating historical flood data, weather forecasts, and real-time sensor information, the digital services will provide accurate, localized predictions of when flooding is most likely to occur in Porto. This allows emergency teams to act proactively by preparing resources and issuing warnings ahead of time.

Dataset	Use
Firefighters Occurrences	Real-time occurrences provided by firefighters, including location, status of resolution and number of human resources and vehicles.
Observed Meteorology – Porto Digital sensory network	Observation of weather conditions in a given place and time (meteorological stations).

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

Drainage network sensing - LACROIX	Real-time information related to water quality (e.g. pH) and flow conditions (e.g. speed).
Porto Water and Energy Floodings - TAGO	Sensorisation system of water level to detect flood situations, in real time, in riverside, rainwater and Douro River infrastructures.
H2Porto - NAVIA	Platform for recording all requests and interventions, carried out by the company, in the public drainage and water supply networks in the city of Porto, with georeferencing service of the intervened locations by the teams and type of action carried out.

Table 16: Porto pilot (DS 1) HVD overview

Actions taken to improve the datasets. Through the capabilities of **Orion Context Broker**, Porto uses NGSI-LD format to standardize and enrich datasets, ensuring semantic consistency and enabling real-time updates for dynamic data.

CKAN Portal is leveraged to enhance dataset discoverability with comprehensive metadata and make datasets accessible via APIs or downloads. The **Metadata Quality Validator (MQV)** methodology was also applied to improve metadata quality and accessibility. Lastly, Porto's open data **CKAN** portal is federated with **Idra** to facilitate seamless data sharing.

The Porto municipality datasets have been published in the Porto CKAN portal, but it is planned to include them in the BeOpen CKAN as well, to have all the HVDs produced within the projects in one place.

Dataset	Initial score	End-of-year score
Firefighters Occurrences	2	100%
Observed Meteorology – Porto Digital sensory network	8	100%
Drainage network sensing - LACROIX	2	100%
Porto Water and Energy Floodings - TAGO	2	100%
H2Porto - NAVIA	2	100%

Table 17: Porto pilot (DS 1) HVD scores

3.2.8 Pilot Porto: Dashboard for real time visualization of occurrences and emergency teams

The Porto pilot which focuses on a dashboard for real time visualisation of occurrences and emergency teams is planned to integrate eleven datasets, which will have been improved by using the BeOpen Framework. The dashboard will serve as a centralised platform to aggregate, process, and display information, with a particular focus on enhancing emergency response times and optimising route planning.

Dataset	Use
Firefighters Occurrences	Represents events in Porto where the firefighters were called.
Observed Meteorology – Porto Digital sensory network	Represents the observations of observed meteorological parameters by the sensors of Porto Digital sensory network
Drainage network sensing – LACROIX	Sensorisation system of Porto's drainage network (sanitation + rainwater + streams), which allows, in real time, get information related to water quality (e.g. pH) and flow conditions (e.g. speed).
Porto Water and Energy Floodings – TAGO	Sensorisation system of water level to detect flood situations, in real time, in riverside, rainwater and Douro River infrastructures.

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

H2PORTO - NAVIA	Platform for recording all requests and interventions, carried out by the company, in the public drainage and water supply networks in the city of Porto, with georeferencing service of the intervened locations by the teams and type of action carried out.
Police and firefighters geolocation buildings	Store the location of first responders and emergency teams buildings.
Watersheds boundaries	Spatial information defining the geographical boundaries of watersheds in Porto.
Roadway and street dimensions	Detailed measurements of the width of roads in Porto.
Fire hydrants	Geolocation of fire hydrants within Porto.
Trees	Location of trees in Porto, including their geolocation, scientific names, and estimated ages.
Traffic Management System – CCTV	Location of CCTV video cameras, associated with vertical light signs in the Municipality of Porto.

Table 18: Porto pilot (DS 2) HVD overview

Actions similar to "Pilot Porto: machine learning for forecast urban floods" were undertaken to improve the quality and scope of the datasets. The initial score, before the improvement of the datasets and the expected score after the enhancement is shown in the table below.

Dataset	Initial score	End-of-year score
Firefighters Occurrences	2	100%
Observed Meteorology – Porto Digital sensory network	8	100%
Drainage network sensing – LACROIX	2	100%
Porto Water and Energy Floodings – TAGO	2	100%
H2PORTO – NAVIA	2	100%
Police and firefighters geolocation buildings	8	100%
Watersheds boundaries	2	100%
Roadway and street dimensions	2	100%
Fire hydrants	2	100%
Trees	2	8100%
Traffic Management System – CCTV	8	8100%

Table 19: Porto pilot (DS 2) HVD scores

3.2.9 Pilot Naples: Integrating Mobility and Environment Data for Metropolitan Transport and public space management

Naples pilot is creating two digital services for its use case in the form of maps. The first will host the data on Sustainable Mobility improved in the BeOpen project, allowing easy consultation, while the second map will offer a description of public space in the urban context, with datasets containing relevant information on the state of the environment.

The Datasets in the Naples pilot that have been assessed belong to different domains:

- Mobility Data describe the elements of the infrastructure of public transport in the city (traffic restriction areas, public transport routes, cycle paths, railway stations and mobility nodes).

- Public space data, assets through which the Municipality provides schooling, welfare and eco-systemic contribution to its citizens, are integrated with information on environmental features: quality of green features (expressed with a green Index assessment) and Urban temperature assessment, as a differential with rural areas temperature.

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

Actions taken to improve the datasets have focused on data standardization through BeOpen **Data Model Mapper** tool, integration of additional information through the use of GIS software (QGIS) and accessibility improvement. The datasets are uploaded on **BeOpen data space** that will host the data to improve accessibility and metadata quality. At present the Municipality offers access to open data only through a web page where data are documented and can be singularly downloaded. Existing Metadata can be valued as poor-quality information. Naples is currently building a CKAN portal for Open Data, where Open data will be collected in a unique sharepoint by the Administration, but datasets of the geographic and earth observation domain, due to their specific features, can conveniently use multiple tools for data sharing.

Dataset	Initial score	End-of-year score
Municipal recreation parks	6	100%
Heat exposure evaluation on public parks	4	100%
Electrical charging stations street level	5	100%
Parking areas	5	100%
Municipalities (administrative units)	6	100%
Traffic restrictions (zone 30, pedestrian areas)	4	100%
Mobility nodes	6	100%
Bus routes	5	100%
Spotted urban heat exposure datasets	7	100%
Cycle paths	3	100%
UrbanGreenIndex_Napoli	7	100%

Table 20: Naples pilot HVD scores

3.2.10 Pilot Vilnius: Detection and Projection of Invasive Species Spread

The digital service for the Vilnius pilot aims to tackle the ecological threat posed by the invasive species *Heracleum sosnowskyi*. The service focuses on identifying the current locations of this invasive plant and projecting its potential spread if left untreated.

The primary objective is to create a digital map of detected locations and predict future spread scenarios, enabling the city to devise informed strategies for limiting and eradicating the plant. Vilnius use case has selected 3 contextual datasets for the purposes of visualisation (with the prospect of utilising those datasets for priority drone flight planning), one dataset as a result of the creation of the digital service and an open access dataset owned by the Lithuanian hydrometeorological service for use in the digital service with projected expansion.

Dataset	Use
Natural framework	Contextual, supporting role
Extensively and intensively used green spaces	Contextual, supporting role
Green area accessibility zone	Contextual, supporting role
Invasive species	DS labels collected into a comfortable geospatial format for visualization to stakeholders
Wind data Meteo.lt API	Modelling of invasive species expansion.

Table 21: Vilnius pilot HVD overview

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

Actions taken to improve the datasets. Supporting datasets provided by ID Vilnius were standardized to the Dublin Core metadata standard, ensuring consistent and interoperable metadata across the Vilnius open data portal. Given that these datasets contained no sensitive information, they were assigned a CC BY-SA 4.0 license, allowing for unrestricted reuse, and were made freely accessible. By integrating these datasets into the Vilnius open data portal, which operates on CKAN, they automatically gained API accessibility through the CKAN platform. Additionally, the datasets are translated to English and uploaded into the **BeOpen CKAN** – using the **BeOpen framework**. The federation of all catalogue's datasets to **Idra** ensures the DCAT-AP standardization. The framework's **Metadata Quality Validator** acts as a guide to ensure the sufficiency of the metadata of improvable datasets.

The invasive species dataset is essentially an output of the DS creation process. It contains Sosnowski hogweed's location information and will be available along supporting datasets in the visualisation map, that will have gone through the same improvement processes as the supporting datasets.

The meteorological data was already compliant with the mandatory requirements of an HVD. ID Vilnius has bought the service while it was still restricted and integrated it into their system, but as of today, the data is freely accessible from the owner. Access to this dataset will be described next to the invasive species dataset in the catalogue.

Dataset	Initial score	End-of-year score
Natural framework	2	100%
Extensively and intensively used green spaces	2	100%
Green area accessibility zone	2	100%
Invasive species	0	51%
Wind data Meteo.lt API	8	100%

Table 22: Vilnius pilot HVD scores

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

4 API Access and Integration

4.1 Overview of Possible Different API Types Available in Catalogue

In this initial phase of the BeOpen HVDs Catalogue, not all datasets have been fully integrated, and there is ongoing work to ensure the inclusion of datasets from all pilot cities. As such, this section outlines the potential for incorporating various API types in the Catalogue, focusing on future standardization and API accessibility plans.

For each dataset, the suitability of standardization to either the NGSI-LD, python or other API format is being assessed, based on its data type, structure, and intended use. This approach will facilitate greater data interoperability and allow for more dynamic data integration, particularly for datasets that benefit from real-time or spatial data updates (NGSI-LD) or statistical aggregation (SDMX).

The table below (*Table 23*) provides an overview of the datasets contributed by each pilot, detailing their current data format, data type, and the potential for API standardization. By identifying these opportunities, the BeOpen project lays the groundwork for a scalable API structure that supports both real-time data exchange and statistical data access across European platforms. This ongoing task will be completed by the project's end, ensuring that datasets are accessible through API types most appropriate for their characteristics and usage contexts.

Use case	Dataset name	Data format	Data type	Potential additional API endpoint
UC 15 - Vilnius	Invasive species	JSON	Geospatial	Python
UC 13 - Porto	Fire hydrants	CSV	Geospatial	Python
UC 13 - Porto	Trees	CSV	Geospatial	Python
UC11 - Herne	Road Traffic	JSON	Geospatial	Python
UC12 - Herne	People density	JSON	Geospatial	Python
UC1- Attica	Social Media Events related to fires	JSON	Linked data	NGSI-LD API

Table 23: Selection of HVDs for potential additional API endpoints

The advantages of a multi-API approach include flexibility for end-users, interoperability and scalability, and efficient data delivery. By supporting diverse APIs, the catalogue accommodates varied user needs, from basic metadata retrieval to advanced data queries and real-time updates. APIs like NGSI-LD promote standardized data exchange across platforms, ensuring datasets are easily integrated into local, national, and European infrastructures. The APIs enable real-time or near-real-time access to datasets, making the catalogue valuable for applications such as urban planning, environmental monitoring, and statistical analysis. By offering a broader spectrum of API types, the BeOpen catalogue caters to the diverse needs of stakeholders while promoting openness, usability, and innovation in the use of HVDs.

4.2 Linking Data Assets with APIs for Easy Access

To facilitate the seamless access and integration of HVDs the BeOpen platform connects each data asset with relevant APIs, allowing users to retrieve, query, and interact with data efficiently. Linking data assets to APIs enables flexible access to datasets, whether users need raw data for analysis, real-time data

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

updates, or structured statistical data for specific applications. An example of the varying data flows could be seen in Figure 2.

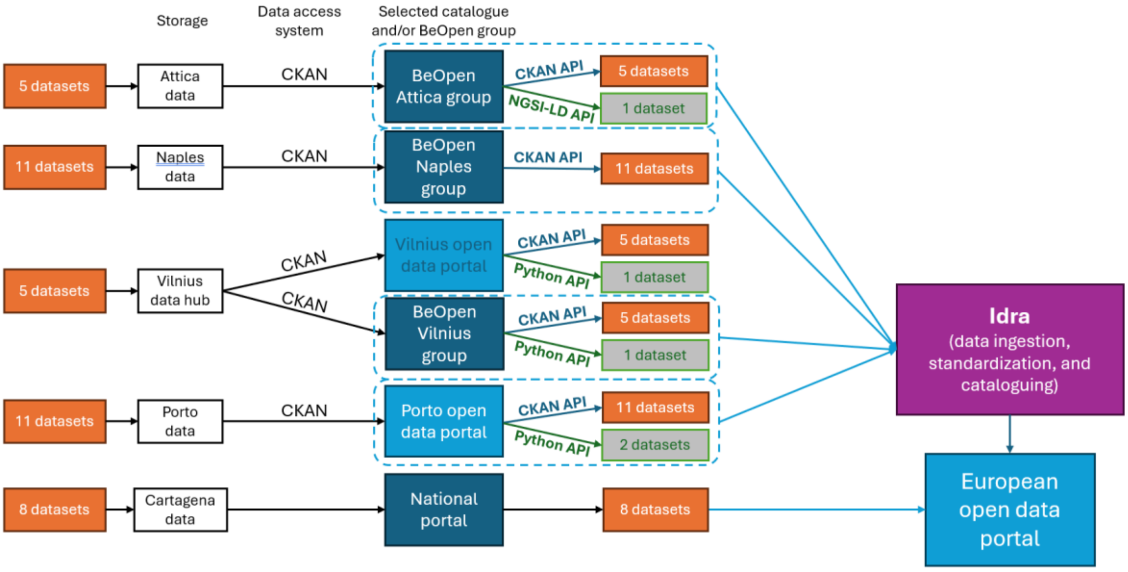


Figure 7: Data flow and API typologies for European open data integration

For each dataset uploaded to CKAN, the BeOpen platform would be able to support multiple API endpoints, tailored to different data access needs. This multi-API approach provides a layer of flexibility, ensuring that datasets are accessible to diverse audiences with varying technical requirements. By linking datasets to APIs directly within CKAN, users can seamlessly transition from data discovery to data integration without needing extensive setup or technical expertise.

Additionally, this strategy allows each pilot city or data provider to select the most suitable API based on the dataset's characteristics, thereby maximizing the dataset's usability across applications. Ultimately, this approach broadens accessibility, promotes interoperability, and supports the broader objectives of the BeOpen project by making HVDs more accessible and actionable.

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

5 Conclusion and Next Steps

5.1 Summary of HVD Catalogue's First Version

The **first version of the HVDs Catalogue** represents a significant milestone in the BeOpen project's efforts to harmonize, standardize, and disseminate public datasets across the European Union. It provides an initial framework for making data accessible to both governmental institutions and the public through modern open data portals, ensuring compliance with European data interoperability and accessibility standards.

This version of the catalogue includes contributions from pilot cities, showcasing their efforts to improve and align datasets with project goals. Each city focused on addressing specific local challenges while adhering to the shared objectives of standardization and harmonization. Through their work, pilot cities have demonstrated innovative approaches, such as utilizing new metadata models, implementing smart APIs, and adopting lightweight data formats like GeoJSON and JSON to enhance usability and scalability.

The catalogue reflects a strong commitment to the European Data Portal (EDP - https://data.europa.eu/) requirements, integrating metadata schemas and facilitating harvesting through platforms such as CKAN and Idra. By ensuring compliance with these standards, the HVD Catalogue lays the groundwork for seamless data sharing and interoperability between local, national, and European portals.

Key features of this version include:

1. **Data Standardization**: The adoption of universally recognized data formats and metadata structures to ensure compatibility with a wide range of platforms and tools.

2. **API Integration planning**: Implementation of versatile APIs to facilitate efficient access and interaction with datasets, ensuring usability across different user groups.

3. **Pilot Contributions**: Highlights from each pilot city, demonstrating the potential of harmonized datasets to address diverse urban and environmental challenges.

Pilot cities participating in the BeOpen project have played a pivotal role in the development of the first version of the HVD Catalogue. Each city contributed datasets tailored to address their specific local challenges while working toward harmonization with European standards. The main existing themes of these use cases are urban monitoring and management, digital platforms for data integration and visualization, environmental impact and sustainability, public health and safety. While the categorization provides a helpful framework, it is important to note that the use cases presented span multiple categories, reflecting their multifaceted nature that can be seen in the pilot contribution sections. These use cases along with their respective HVDs address interconnected challenges that extend beyond the scope of a single category, demonstrating their broad applicability across various aspects of urban management and governance.

While this version provides a robust foundation, it also identifies areas for future work. These include enhancing metadata quality, expanding the list of available datasets, and strengthening the connection between local initiatives and the broader European open data ecosystem. The BeOpen project team will continue refining the HVD Catalogue, leveraging the feedback and insights gained from this initial iteration to ensure its relevance and utility in supporting data-driven decision-making at all levels.

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

5.2 Plans for Updates and Future Revisions

As the HVD Catalogue continues to evolve, several updates and revisions are planned to ensure that the datasets meet the highest standards of accessibility, usability, and compliance with EU directives. While significant progress has been achieved, ongoing efforts remain focused on:

Onboarding of New Datasets. Some pilot cities, such as Vilnius, are in the process of developing datasets tied to their specific use cases, such as outputs from new digital services. These datasets are expected to reach HVD status upon completion, further enriching the catalogue.

API Implementation. While the foundational data-sharing mechanisms are in place, additional APIs tailored to specific data types and user needs are planned for development. These APIs will enhance interoperability and provide new ways to access and utilize the datasets.

Iterative Improvements. Pilots are working diligently to refine their datasets, metadata, and dissemination strategies. While challenges exist, all efforts are being made to ensure the catalogue continues to grow in quality and functionality. These updates reflect the collaborative nature of the project and a commitment to gradual and meaningful progress.

5.3 Anticipated Improvements, Challenges and Expansion

Looking ahead, the HVD Catalogue is envisioned as a cornerstone for fostering open data reuse across the EU. Its future development aligns with both the BeOpen project’s long-term goals and broader EU directives for data harmonization and accessibility. Expanding Metadata Linkages is a critical next step involving incorporating metadata links to datasets that are not directly owned by the pilot cities but were instrumental in creating those within the catalogue. This approach will enhance transparency, traceability, and context for users, showcasing the collaborative value of HVDs.

Community involvement will play a pivotal role in driving the success of the HVD Catalogue. Efforts to engage developers, researchers, and citizens should be in place to demonstrate the practical benefits of open data. These activities align with the project's KPIs, aiming to foster a thriving ecosystem of data users and contributors. Over time, the project anticipates integrating advanced tools for data visualization, analytics, and cross-regional comparisons. This would empower stakeholders to derive actionable insights and further utilize the datasets in decision-making processes.

A critical challenge for the project is ensuring seamless integration and interoperability across different data portals at local, national, and centralized levels. While some pilot cities, such as Naples, are actively developing new open data platforms, others face challenges with existing infrastructures. For example, Porto’s open data portal currently lacks integration with Portugal's national portal, and Vilnius similarly does not yet connect with Lithuania’s national platform. This diversity in portal usage highlights the multifaceted nature of open data dissemination. Different user groups often prioritize platforms that align with their thematic or geographic interests. For instance, stakeholders seeking localized data on Vilnius are likely to consult the city’s open data portal first, while broader national or regional users may rely on centralized or national platforms. This user-driven variability presents both an opportunity and a challenge: ensuring that data is accessible through multiple avenues without fragmenting the user experience or duplicating efforts. Through these expansions and ongoing efforts, the HVD Catalogue aims to grow beyond its current scope, becoming a dynamic and indispensable resource for open data collaboration across the EU.

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

6 Appendices

Figure 8: Impact Assessment Tool on the BeOpen framework showing the dataset scores and history (example of Vilnius pilot)

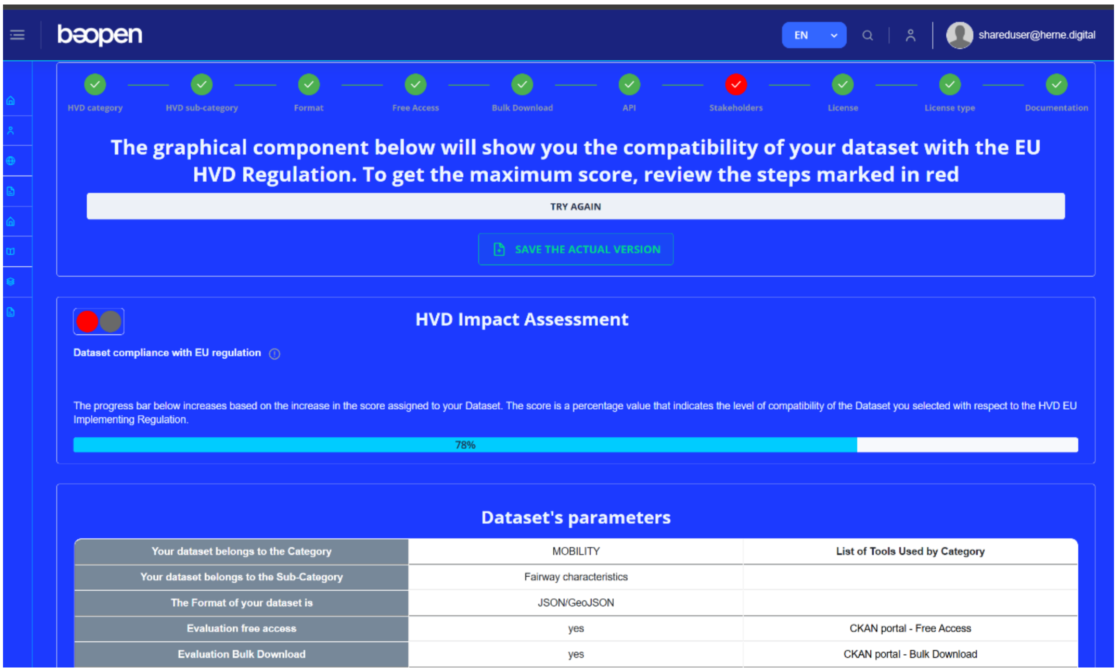


Figure 9: Impact Assessment Tool on the BeOpen framework in case a mandatory criteria has not been met

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

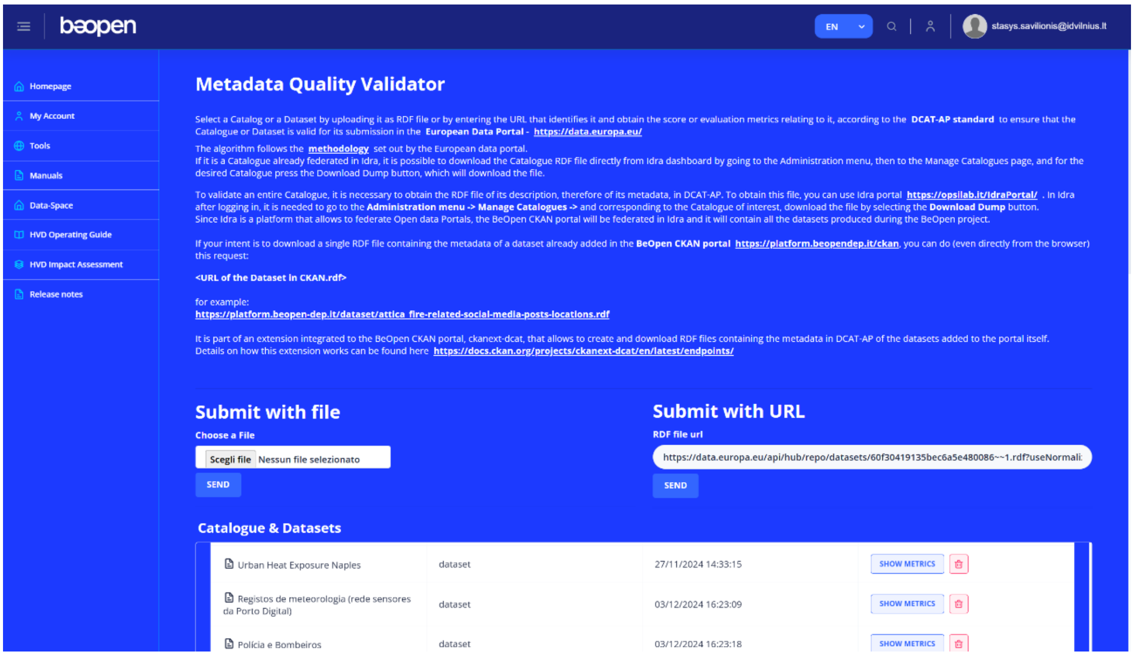


Figure 10: Pilots checking the metadata quality in the BeOpen Framework using the Metadata Quality Validator tool

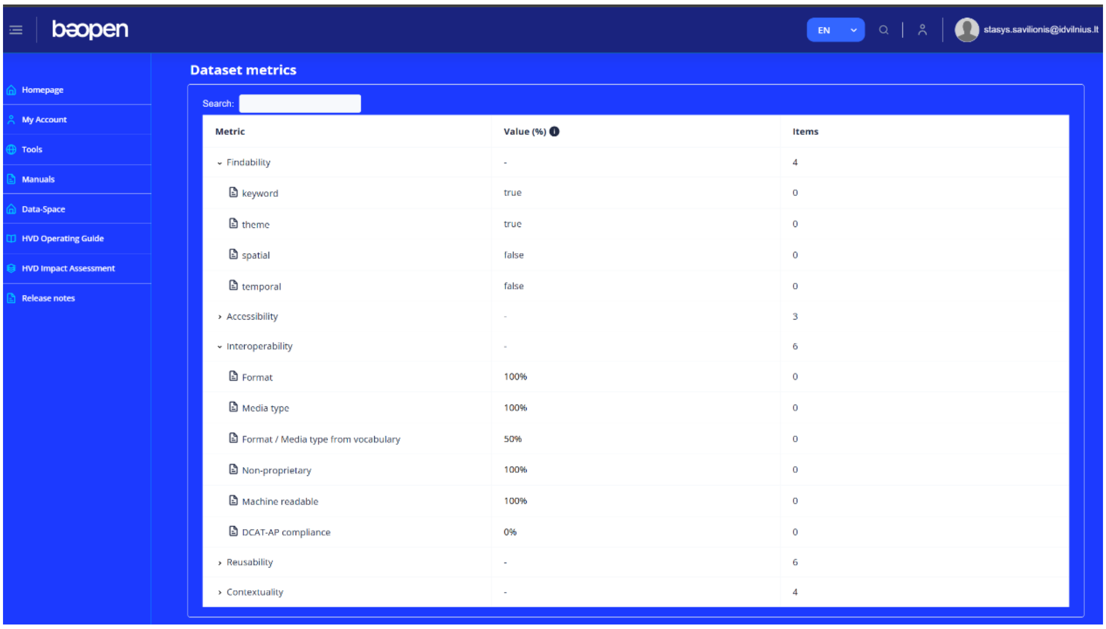


Figure 11: Pilots checking the metadata quality in the BeOpen Framework using the Metadata Quality Validator tool - Results

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807

7 References

BeOpen consortium. (2024). *BeOpen Framework*. Retrieved from https://platform.beopendep.it/

Engineering Ingegneria Informatica SpA. (2024). *Idra - Open Data Federation Platform*. Retrieved from https://idra.readthedocs.io/en/latest/

European Commission. (2024). *Big Data Test Infrastructure (BDTI)*. Retrieved from https://big-data-test-infrastructure.ec.europa.eu/index_en

European Commission. (2024). *European data portal*. Retrieved from https://data.europa.eu/en

European Union. (2024). *Commission Implementing Regulation (EU) 2023/138*. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32023R0138

FIWARE. (2024). *FIWARE Orion Context Broker*. Retrieved from https://fiware-orion.readthedocs.io/en/master/

BeOpen has received funding from European Union's Horizon Europe Research and Innovation programme under the Grant Agreement No 101100807