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

D2.1 BeOpen technical framework design and technical specifications (First version)

Document History

Ver.	Date	Description	Author	Partner
0.1	05/06/2023	Table of Contents	Martino Maggio	ENG
0.2	30/06/2023	Initial structure of the chapters	Martino Maggio	ENG
0.3	21/07/2023	First release of chapter 1 and 3	Martino Maggio, Francesca D'Agresti	ENG
0.4	04/09/2023	First release of chapter 2	Martino Maggio, Francesca D'Agresti	ENG
0.5	13/10/2023	Second release of all chapters	Martino Maggio, Francesca D'Agresti	ENG
0.6	27/10/2023	Final draft for internal review	Martino Maggio, Francesca D'Agresti	ENG
0.7	31/10/2023	Reviewed version with comments	Stefano De Panfilis Pedro Sánchez Romero Anna Ida Hudig	FIWARE HOPU ARTHUR
1.0	3/11/2023	Final version	Martino Maggio	ENG

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	D2.1 BeOpen technical framework design and technical specifications
Work Package	WP2
Delivery Month (DoA)	M10
Version	V01
Actual Delivery Date	3 November 2023
Nature	Report
Dissemination Level	Public – PU
Lead Beneficiary	ENG
Authors	Martino Maggio, Francesca D'agresti (ENG)
Quality Reviewer(s)	Stefano De Panfilis (FIWARE), Pedro Sánchez Romero (HOPU), Anna Ida Hudig (ARTHUR)
Keywords	Framework, Architecture, Interoperability, HVDs, Open Data

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

D2.1 BeOpen technical framework design and technical specifications (First version)

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

Table of Contents

Introduction 9

1 BeOpen framework technical requirements 10

1.1 Requirements categories 10

1.1.1 Data management 10

1.1.2 Data Publication 10

1.1.3 Data Interoperability 11

1.1.4 System performance & security 11

1.1.5 System interoperability 11

1.1.6 User experience 11

1.2 BeOpen framework general requirements 12

1.2.1 Requirements from the pilots for the BeOpen framework 16

2 Framework architecture 18

2.1 Open Data management phases and related functionalities 19

2.1.1 Data Collection 19

2.1.2 Data Quality Improvement 21

2.1.3 Semantic Harmonization and Data Integration 22

2.1.4 Data Publication 24

2.1.5 Data Access and Visualization 25

2.1.6 Cross-phase capabilities 28

2.1.6.1 Metadata Management 28

2.1.6.2 Privacy and Security 29

2.1.6.3 Data Storage 30

2.1.7 BeOpen Framework and AI-related capabilities 31

2.2 BeOpen framework architecture overview 33

2.3 Reusable tools for BeOpen framework 35

2.3.1 Mapping between tools and framework functionalities 39

3 Standards and interoperability specifications 41

3.1 European Technical Standards 41

3.1.1 Commission Implementing Regulation (EU) on HVDs 41

3.1.2 DCAT Application profile for data portals in Europe (DCAT-AP) 45

3.1.2.1 DCAT-AP annex for High Value Datasets 47

3.1.3 ETSI NGSI-LD API for Context Information Management 48

3.1.4 INSPIRE - Infrastructure for spatial information in Europe 50

3.2 European Key Initiatives 51

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

3.2.1 European Data Spaces 51
3.2.2 Smart Data Models initiative 52
3.2.3 Minimal Interoperability Mechanisms – MIMs 53
Conclusions 55
References 56

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

List of Figures

Figure 1: BeOpen framework high level logical architecture 34
Figure 2: Features of HVDs according to the Implementing Regulation [25] 42
Figure 3 - BeOpen framework architecture and HVDs regulation high level requirements 44
Figure 4: Benefits of DCAT-AP for Data providers and Data reusers [30] 46
Figure 5: Data linking with DCAT-AP [30] 47
Figure 6: NGSI-LD information model as UML [43] 49
Figure 7: Relationship between INSPIRE implementing Rules and Technical Guidance [46] 51
Figure 8: Common European data spaces [48] 52
Figure 9: State of play of MIMs in June 2023 [54] 54

List of Tables

Table 1: General requirements of BeOpen framework 16
Table 2: Requirements from the pilots for the BeOpen framework 17
Table 3: Framework functionalities of Data Collection 20
Table 4: Framework functionalities of Data Quality Improvement 22
Table 5: Framework functionalities of Semantic Harmonization and Data Integration 23
Table 6: Framework functionalities of Data Publication 25
Table 7: Framework functionalities of Data Access and Visualization 27
Table 8 : Reusable tools for the BeOpen framework 38
Table 9: Properties introduced with the creation of the DCAT-AP annex for HVDs 48

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
AWS	Amazon Web Services
CC	Creative Commons
CEF	Connecting Europe Facility
CKAN	Comprehensive Knowledge Archive Network
CLI	Command Line Interface
CSV	Comma-Separated Values
DBMS	Database Management System
DCAT-AP	DCAT Application Profile
DT	Data
ESRI	Environmental Systems Research Institute
ETSI	European Telecommunications Standards Institute
EU	European Unit
FAIR	Findable Accessible Interoperable Reusable
GDPR	General Data Protection Regulation
GIS	Geographic Information Systems
GTFS	General Transit Feed Specification
GUI	Graphical User Interface
HDFS	Hadoop Distributed File System
HTTP	Hypertext Transfer Protocol
HVD	Hign Value Dataset
ID	Identification code
IR	Implementing Rules
ISA	Interoperability Solutions for European Public Administrations
ISO	International Organization for Standardization
IT	Information Technology
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation for Linked Data
NGSI-LD	Next Generation Service Interfaces for Linked Data
OASC	Open & Agile Smart Cities
ODMS	Open Data Management System
OLAP	Online Analytical Processing
OSM	OpenStreetMap
PA	Public Administration
POI	Point Of Interest
RDF	Resource Description Framework
REST	REpresentational State Transfer
SEMIC	SEMantic Interoperability Community
SI	System Interoperability
SP	System Performance and security
SPARQL	SPARQL Protocol and RDF Query Language
SQL	Structured Query Language
UML	Unified Modeling Language
UX	User eXperience
WMS	Warehouse Management System
XML	eXtensible Markup Language

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

Introduction

This document, Deliverable D2.1 "BeOpen technical framework design and technical specifications (First version)" reports information about requirements, logical architecture and high-level specification of the BeOpen technical framework. The BeOpen framework represents the core of the project providing the capabilities to "*improve technical data interoperability, semantics and quality of the public sector dataset".* The framework, that will consist in a set of technical tools, will be validated by the different Public Administration involved in the project that will use it to manage and improve their High Value Datasets with the final aim of publishing them via standard and interoperable formats and interfaces.
This document, that will be followed by final version at the end of the project, (June 2025) is composed by three main chapters.

Chapter 1 describe the technical requirements of the framework identified during the first months of the project. Some of the requirements are related to the main objectives of BeOpen project as stated in the Description of Actions, others are, instead, related to the domains and pilots specific needs identified during the technical meetings/calls of the first project months. This chapter reported the requirements in tables, classified based on typologies and categories.

Chapter 2 presents the logical structure and functionalities related to the BeOpen framework based on the needs and requirements reported in chapter 1. In this first version of the deliverable, the attention was paid to the definition of the logical architecture rather than the technological one which is in the process of being defined together with the framework implementation. This chapter is structured following a set of typical phases of the open data lifecycle, identifying for each phase a set of functionalities to be provided by the BeOpen framework in relation to the requirements identified in chapter 1. In this chapter is also presented a list of tools, that are taken in consideration for a possible reuse in the implementation of the framework, in connection with the logical functionalities previously presented.

The last chapter has been divided into two parts: the first resumes the scope of key official European standards and norms that the BeOpen framework will support to guarantee interoperability and standardisation of the HVDs, while the second describes the scope of some key European initiatives related to the BeOpen objectives, making a reference to the official documents.

As previously mentioned, the Deliverable 2.1 will be updated by Deliverable D2.4 "BeOpen technical framework design and technical specifications (Final version)" that will include the final version of the architecture and specification of BeOpen framework.

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

1 BeOpen framework technical requirements

This chapter reports the BeOpen framework technical requirements elicited during the first part of the project (by M10). The requirements relate to different aspects of the framework and they have been identified by different categories stakeholders contributing to the project. A set of requirements is related to the objectives of the BeOpen project as defined in the project Grant Agreement: these requirements are fundamental to achieve the overall scope of the project. A specific set of requirements is identified in order to provide essential capabilities to allow a correct framework functionality in the context of the project needs that emerged during the technical discussion among the partners. Another typology of requirements is directly expressed by the cities/regions involved in the project in relation to specific needs for the implementation of the pilot application and services.
Each requirement has been classified based on typology (functional/non-functional) and a category to simplify the following analysis, presented in chapter 2 aimed to extract the framework logical functionalities. These categories are described in the following section.

1.1 Requirements categories

1.1.1 Data management

Data management includes the requirements related to activities that aim at collecting, improving and preparing data for the final stage of High Value Datasets publication. In particular, this category covers all the requirements focused on the data (and metadata) acquisition from heterogeneous sources, the validation and reduction of errors, enrichment and transformation in machine readable formats for their future use by third party application and services. In this context a key aspect is data quality requirements: high-quality data will be essential for the effective adoption of artificial intelligence and automation technologies in workflows. Inaccurate results can be expected when machine learning algorithms are trained using low-quality data.

1.1.2 Data Publication

In BeOpen project one of the key objectives is the publication of the HVD produced during the project via APIs and other interoperable data provisioning interfaces. The Data Publication category includes the requirements related to the final process of publication of datasets in a specific type of output file, via a platform/interface adopted to share the files in a public way. Requirements included in this category refer to functionalities and approaches to make the HVDs available to multiple applications, users, or organizations maximizing their reuse; it this category the aspect of license management is also included. It is fundamental to ensure that open data can be reused in a proper way, in compliance with the FAIR principles [1].

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

1.1.3 Data Interoperability

Data interoperability can be considered the core of the BeOpen project and its technical framework. Without interoperable data it is not possible to guarantee the reuse of the HVDs across multiple cities, domains and applications, which, instead, represents one of the key objectives of BeOpen. In order to reach data interoperability, the use of standards is mandatory: the requirements in this category focus on the reuse of common data models, ontologies and, in general, on the semantical approach to produce Linked Open Datasets. Linked Data's ability to create a web of interconnected data provides valuable context and insights that are not always apparent in traditional isolated data silos, fostering HVDs reusability in cross-domains contexts.

1.1.4 System performance & security

System performance and security are crucial aspects of any IT system and therefore for the BeOpen framework. It is essential that the system runs smoothly, efficiently, and securely. Any IT system must have security as a critical aspect, and it requires constant vigilance and attention. Regularly monitoring and assessing the system is necessary to ensure its performance and security are at their optimal level. This category includes the requirements covering the above-mentioned aspects but data protection and security (e.g. anonymisation) are also capabilities that are considered.

1.1.5 System interoperability

This category includes the requirements related to the capability to enable information systems to communicate and understand information exchanged between them. This facet bears significant relevance within the context of BeOpen because the framework has to be able to interact with different technologies and external systems already present in the in the pilot sites foreseen by the project. Moreover, in this category we refer also to the adoption of standards, protocols, technologies, and mechanisms that allow data to flow between different systems without any human intervention.

1.1.6 User experience

In the User experience category are included the requirements to ensure that users can easily and effectively interact with open data at every stage of its lifecycle. Whether they are seeking data, analysing it, developing applications, or using it for research, a positive user experience promotes transparency, accessibility, and the overall effectiveness of open data initiatives, fostering innovation and informed decision-making.

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

1.2 BeOpen framework general requirements

ID	Name	Description	Type	Category	Rationale
DT-01	**Data import, ingestion and extraction**	The framework shall involve the capability to handle the collection and extraction of data from various sources. It also shall include functionalities like data acquisition from external systems, data import/export mechanisms, data streaming, and real-time data ingestion.	Functional	Data management	Data ingestion is a crucial step that helps in extracting and transferring data in an automated way. With data ingestion pipelines established, it allows to focus on extracting value from data and finding new insights.
DT-02	**Data available in machine-readable formats**	The datasets shall be made available in an open and machine-readable format, such as CSV, JSON, XML, or RDF, to provide a digital representation of data in a file that can be automatically imported, readed and processed by a computer system for further processing.	Functional	Data management - Data Interoperability	BeOpen framework, as stated in the project objectives, must include open-source tools that can make public open data available in machine-readable format.
DT-03	**Data validation and quality control**	The framework shall involve the capability to validate the integrity, accuracy, and consistency of the open data. It shall also involve performing automated checks, enforcing data validation rules, identifying, and handling data anomalies or errors, and ensuring data quality control measures are in place. It involves tasks such as data cleansing (removing duplicates, and correcting errors).	Functional	Data management	BeOpen framework, among the objectives, must validate and assess the HVDs accessibility and reusability through the execution of EU wide pilot services AI applications. The high quality of input data of an AI application is an essential requirement.
DT-04	**Data enrichment and augmentation**	The framework shall involve the capability to enhance the open data by enriching it with additional relevant information or augmenting it with complementary data from other sources. It shall include the capability to perform techniques such as data linking, data integration, data enrichment through APIs or external datasets, and data augmentation through data mining or machine learning algorithms.	Functional	Data management	Data enhancement is a crucial step that ensures the data is up-to-date and accurate in combination with data cleansing.
DT-05	**Data integration and aggregation**	The framework shall involve the capability to integrate and aggregate data from multiple sources to provide a unified and comprehensive view of the open data. It shall also involve tasks such as data fusion, data integration across different systems or databases, and data aggregation to create meaningful summaries or higher-level insights.	Functional	Data management	Some pilots present the need to integrate data belonging to different domains (such as mobility and urban data) to provide citizens data which usually are available from different sources.
DT-06	**Data format conversion and transformation**	The framework shall involve the capability to support the conversion and transformation of open data into different formats. It shall include functionalities such as converting data between CSV, JSON, XML, or other standard formats, as well as performing data transformations or	Functional	Data management	The requirement addresses the specific needs of the pilots of having tools in the BeOpen framework able to transform input file in specific machine-readable formats.

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

		manipulations to meet specific requirements.			
DT-07	**Data standardization**	The framework shall involve the capability to standardize the open data to ensure consistency and conformity to predefined formats or standards. It shall involve tasks such as data normalization and transforming data into a common format or structure.	Functional	Data interoperability	The requirement addresses the need for BeOpen project to aim at relying on, and be aligned with, key relevant initiatives, specifications and standards related to open data.
DT-08	**Metadata management and discovery**	Metadata management and discovery: the framework shall involve the capability to manage and facilitate the discovery of metadata associated with the open data. It shall involve capturing and organizing metadata attributes such as data source, data description, data licensing, and other relevant information to enhance searchability and understandability of the open data.	Functional	Data interoperability	The metadata management is an essential requirement: it empowers data publishers to describe and organize their data, making it more accessible, usable, and trustworthy.
DT-09	**Compliance with EU HVDs Implementing Regulation**	The BeOpen framework shall implement the Commission Implementing Regulation (EU) of 21.12.2022 [2] laying down a list of specific high-value datasets and the arrangements for their publication and re-use. The directives listed in the regulation have the objective of ensuring that public data of highest socio-economic potential are made available for re-use with minimal legal and technical restriction and free of charge.	Non-functional	Data interoperability – Data Publication	BeOpen, among the objectives, must prioritize the use and provisioning of the Datasets categories identified in the regulation and be compliant with the minimum requirements for reuse and dissemination.
DT-10	**Compliance with DCAT-AP standard**	The BeOpen framework shall support the adoption of the DCAT Application profile for the metadata definition of the datasets produced in the project.	Functional	Data interoperability	DCAT-AP is the key specification to describe open dataset in Europe. Its compliance is mandatory for the BeOpen framework.
DT-11	**Compliance with ETSI NGSI-LD standard**	The framework should provide tools to support the generation and management of data and metadata following ETSI NGSI-LD standard [3].	Functional	Data interoperability	NGSI-LD is one of the key standards to be reused in BeOpen project. Moreover, this standard is already adopted by different pilot cities of the project.
DT-12	**Data mapping and schema matching**	The framework must provide capabilities to simplify the access and reuse of common and standard data models for the data description of HVD (e.g. FIWARE data models)	Functional	Data interoperability	The way in which data and metadata are represented is important. It is crucial to use a common data structure for data that does not change using different platforms.
DT-13	**Compliance with INSPIRE directive**	The framework has to provide capabilities compliant with INSPIRE directive [4] that aims to create a European Union spatial data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment.	Functional	Data interoperability	INSPIRE is a key directive in the context of HVDs and BeOpen framework has to be compliant with it.
DT-14	**Compliance with "Portal for**	Public open data provided by the project, must be discoverable by the Portal for European Data [5] and	Functional	Data publication	BeOpen framework, among the objectives, must include open-source

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

	European data" requirements	possibly other national open data portals.			tools that can make public open data available in compliance with the Portal for European Data [5] allowing it to automatically harvest the datasets.
UX-01	**Data visualization and exploration**	This requirement focuses on the framework's ability to provide visualizations and exploration tools to analyse and present the open data in a meaningful way. It includes features such as charts, graphs, maps, and interactive visualizations to facilitate data exploration, discovery of patterns, and generation of insights.	Functional	User Experience	The requirement answers to the specific needs of the pilots of having tools in the BeOpen framework able to provide data visualization and analysis experiences offering real-time monitoring of the data obtained.
DT-15	**API and web services for data access**	The framework shall involve the capability to provide application programming interfaces (APIs) and web services for easy and programmatic access to the open data.	Functional	Data publication	BeOpen framework, as key objective, must include the datasets publication through APIs to boost the re-use, combination, and integration of open public data across the EU.
DT-16	**Data catalogue and search capabilities**	The framework shall involve the capability to maintain a comprehensive catalogue of available open datasets and provide robust search functionalities to help users discover and locate specific datasets of interest.	Functional	Data management	The requirement answers to specific needs of the pilots of having tools in the BeOpen framework able to provide search engines that ensure that data relevant to citizens is easily accessible and searchable through a user-friendly interactive way.
DT-17	**Open data licensing and attribution**	The framework shall involve the capability to support open data licensing and attribution mechanisms. It shall include features such as assigning appropriate licenses (e.g., Creative Commons) to the open data, providing clear guidelines for data usage, and facilitating proper attribution of the data to ensure compliance with legal and ethical requirements.	Functional	Data publication	The requirement addresses the objective for BeOpen project to aim at publishing, datasets publication under open licenses in particular the Creative Commons BY 4.0 license or any equivalent or less restrictive open license.
DT-18	**Compliance with FAIR data principles**	The data management methodology, metadata standard, and standard data format adhere to FAIR data guidelines, ensuring that the same data models are applicable to various data requirements.	Non-functional	Data publication	To address legal interoperability in the context of public Open Data dissemination and reuse, BeOpen framework must allow adherence to the FAIR principles.
SP-01	**User identity management, authentication and access control**	The framework shall involve the capability to provide mechanisms for user authentication, ensuring that only authorized individuals or entities can access and interact with the open data managed by the public administrations.	Functional	System performance & security	This requirement is considered one of the minimum system specifications that the platform must meet in order to ensure to provide essential access and security features.
SP-02	**Performance monitoring and optimization**	The framework shall involve the capability to monitor and optimize the performance of data management and processing operations. It shall involve collecting and analysing performance metrics, identifying bottlenecks,	Non-functional	System performance & security	The requirement is considered one of the minimum system specifications that the platform must meet in

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

		optimizing resource utilization, and ensuring efficient handling of large-scale data processing tasks.			order to ensure to work properly.
SP-03	**Scalability and high availability**	The framework shall involve the capability to scale and handle increasing volumes of open data while maintaining high availability and performance. It shall include features such as horizontal scalability, load balancing, fault tolerance, and redundancy to ensure continuous availability and responsiveness of the system even under high loads or in the event of failures.	Non-functional	System performance & security	The requirement is considered one of the minimum system specifications that the framework must meet in order to ensure to work properly.
SP-04	**Data privacy and security measures**	The framework shall involve the capability to define an implementable ethical and legal framework, with the aim of ensuring full compliance of the project with relevant normative provisions and assessing the project's activities from an ethical and legal standpoint focusing on issues of privacy and data protection.	Non-functional	System performance & security	Privacy and security of the open data is a key aspect that the framework must support. It shall encompass features such as data anonymization, encryption, access control mechanisms, audit trails, and compliance with data protection regulations.
SI-01	**High modularity**	The framework must be based on modules that can be easily integrated, but at the same time can be used as separated tools allowing the users to use only the ones that are needed in their data management process	Non-functional	System interoperability	The requirement is considered one of the minimum system specifications that the platform must meet in order to allows the software to be manageable for the purpose of implementation and maintenance.
SI-02	**Cloud deployment**	The framework must be designed and implemented to run on the most relevant cloud environments.	Non-functional	System interoperability	The requirement is considered one of the system specifications that the platform must meet in order to achieve numerous benefits, including faster and simplified deployments, to saving costs and providing an agile and scalable platform.
UX-02	**Web based user interfaces**	The framework must provide web-based user interfaces to allow an easy access of the capabilities via widely used web browsers (e.g. Chrome and Safari).	Functional	User experience	The adoption of a web interfaces will allow a simple access by the different stakeholders using a simple web browser overcoming technological barriers due to different devices and operating systems.
UX-03	**Intuitive and user-friendly interface**	The framework shall involve the capability to provide an intuitive and user-friendly interface for users to interact with the framework. It shall include features such as clear navigation, organized layouts, and user-friendly design principles to ensure ease of use and a positive user experience.	Functional	User Experience	The requirement is necessary to simplify the usage of the framework for all the typologies of users including the non-technical ones.

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

UX-04	**Customizable dashboards and reports**	The framework shall involve the capability to allow users to customize dashboards and generate reports tailored to their specific needs. It shall include features such as drag-and-drop widgets, configurable visualizations, and flexible report generation options to provide users with personalized insights and analytics.	Functional	User Experience	Requirement added as it is considered useful for final pilot applications to better visualize and analyse data.
UX-05	**Collaborative data annotation and feedback**	The framework shall involve the capability to facilitate collaboration and user participation in the open data management process. It shall include features such as allowing users to provide annotations, comments, or feedback on the data	Functional	User Experience	This requirement aims at fostering collaboration among users and enabling community-driven data improvement.
UX-06	**Responsive and mobile-friendly design**	The framework shall involve the capability to provide responsive interfaces also optimized for mobile devices. It shall include features such as adaptive layouts, mobile-friendly interactions, and responsive design principles to ensure a seamless user experience across different devices and screen sizes.	Functional	User Experience	BeOpen framework must offer its services through a web environment that must also be properly usable from mobile devices.
UX-07	**Contextual help and documentation**	The framework shall involve the capability to provide contextual help and comprehensive documentation within the framework in order to understand the functionalities, features, and best practices of the framework.	Functional	User Experience	The requirement is necessary to simplify the usage of the framework for all the typologies of users including the non-technical ones.
UX-08	**Multilingual support**	The framework shall involve the capability to provide multilingual support within the framework to cater to users from different language backgrounds.	Functional	User Experience	The BeOpen framework has to provide language support at least for the countries involved in the project
UX-09	**Real-time notifications and alerts**	The framework shall involve the capability to provide real-time notifications and alerts to users. It shall include, for instance, push notifications, email alerts, or in-app notifications to keep users informed about updates, changes, or important events related to the open data they are interested in.	Functional	User Experience	Requirement to satisfy specific pilot applications scenarios
UX-10	**Personalized user preferences and profiles**	The framework shall involve the capability to allow users to set personalized preferences and maintain user profiles within the framework.	Functional	User Experience	Requirement added as the functionality that is intended to be provided, can be useful for final pilot applications.

Table 1: General requirements of BeOpen framework

1.2.1 Requirements from the pilots for the BeOpen framework

In the following section are presented the most relevant pilot specific requirements identified by the cities/regions involved in the BeOpen project. Even though the pilots contributed to the

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

identification of the above reported general requirements, some specific framework requirements, have been collected and included in Table 2. This is not an exhaustive list that will be detailed and improved during the following months in relation to the pilot application design phase.

Requirement name	Requirement description	Requirement type	Requirement category	Pilot city/region
Social network data integration	Provide APIs that serves the X (formerly Twitter) metadata (or even other social networks) to NOA to offer information for fire prevention in the Attica region.	Functional	System interoperability	**Attica region**
Required data formats	Needs to interface with different formats such as CSV, GeoJSON (open machine-readable formats), ESRI Shapefiles, WFS, WMS, also with datasets of static pictures of streets, and historical status of streets in addition with temperature in specific areas of the city (both historical and dynamic data).	Functional	Data Management	**Herne**
Required existing datasets	Use existing data from PLINIUS research center and cooperate with them to achieve more information on models about climate risk assessment used in the Urban plan PUC.	Functional	System interoperability	**Naples**
Required integration with existing systems	Consider integration with systems already in use for different goals: IPMA, APA, SVARH. Work on supra municipal scale, needs to scale down at city level with real time monitoring access.	Functional	System interoperability	**Porto**
Required technologies in data collection phase	Compliance with specific technologies used in data collection: Apache NiFi, a powerful and reliable system to process and distribute data and Apache Airflow, a platform to programmatically author, schedule and monitor workflows.	Functional	System interoperability	**Porto**
Required protocol for data optimization	Increase data quality using the IEEE P2510 protocol for environmental datasets such as datasets related to gases.	Functional	Data Management	**Cartagena**
System integration and communication	The company that installs the sensors and maintains the infrastructure, which specializes in the field of sensors, has precise information on the equipment and the data collected. The system must be able to interface with the infrastructure of that company during the data acquisition phase.	Functional	System interoperability	**Vilnius**

Table 2: Requirements from the pilots for the BeOpen framework

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

2 Framework architecture

This chapter presents the logical structure and functionalities related to the BeOpen framework based on the needs and requirements reported in chapter 1. The aim of the BeOpen framework is to support public administration in the management of HVDs offering technical tools that can help in specific activities related to the life cycle of the open datasets. The BeOpen framework will specifically supports the public administration in the processes that start from the collection of (raw) data to their final publication via API in machine readable formats, helping them in the implementation of the relevant national and European regulation and in particular the HVD implementing act [2]. It is possible to identify different phases in the management of open datasets: in BeOpen, based on the project objectives and requirements, it has been decided to take in consideration the following ones, that will be better described and detailed in the next sections:

- **Data Collection**: Data collection is the process of systematically gathering information from diverse sources. Data can be collected, for instance, by devices (e.g. sensors), other platforms, o via data entry, aiming to acquire raw data that will later be processed and made available. A key aspect of this phase in the support and interoperability with different data protocols and standards.
- **Data Quality Improvement**: this phase encompasses techniques and procedures to enhance the accuracy, consistency, and reliability of collected data, ensuring it meets specific standards and criteria. The tasks performed in this phase allow to produce high quality data improving the row ones, that can be affected by different types of issues.
- **Semantic Harmonization and Data Integration**: this phase involves aligning data from various sources with consistent formats and meanings, enabling seamless integration and interoperability among disparate datasets. The use of common data models and ontologies is fundamental in this phase.
- **Data Publication**: it includes all the activities and functionalities that aim at making collected and processed data accessible through different public channels such as, open API, open data portals and platforms, ensuring transparency and ease of access.
- **Data Access Visualization**: this phase includes the capabilities of providing the final endpoints to access and query open datasets, but also functionalities to visualise them using graphical techniques to represent data in a comprehensible and informative manner, making it easier for users to understand, analyse, and derive insights from the data.

In the whole open dataset life cycle can be identified other phases (e.g. Data Archiving and Preservation), but they are considered out of the scope of the project so they are not covered by the BeOpen framework.

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

2.1 Open Data management phases and related functionalities

In the next sections, for each of the above-mentioned phases, it will be presented a more detailed description with a set of specific functionalities, to be provided by the framework to fulfil the phase objectives, in relation with the requirements identified in chapter 1.

2.1.1 Data Collection

Data Collection is the foundational phase in the open data lifecycle, serving as the point of origin for the wealth of information that will become publicly available. This phase encompasses the processes involved in gathering raw data from diverse sources, such as sensors, government records, external platforms and proprietary databases. Data generation can be an automated process in the case of sensor data, where information is continuously collected, or a more sporadic endeavour, like conducting a survey to gather specific data points. Data collection, on the other hand, involves the systematic acquisition of data through methodologies designed to capture information relevant to a particular purpose or research objective.

The methods employed during data generation and collection are crucial in determining the quality, accuracy, and relevance of the data. Data may be collected using a variety of techniques, including manual data entry, automated data recording from instruments or devices, web scraping, and more. The types of data collected can range from structured data, which is highly organized and often found in databases, to unstructured data, such as text documents or multimedia content. Regardless of the source or format, data must be collected in a manner that minimizes errors and bias. This often involves adhering to best practices, establishing data collection protocols, and ensuring data quality assurance measures are in place.

One critical consideration during data collection is the preservation of data integrity. Data integrity ensures that the data accurately represents the information it is meant to capture. To maintain data integrity, proper data collection procedures should be followed. This includes documenting the data source, ensuring data is collected at the right time and place, and maintaining data security and privacy standards.

Furthermore, metadata, or data about data, plays a pivotal role in this phase. Metadata includes information about the data source, the methods used for data collection, the date and time of collection, and any relevant contextual details. Properly documented metadata is essential for later stages of the open data lifecycle, as it helps data users understand the origin and characteristics of the data they encounter, facilitating its discoverability and usability.

The Data Collection, in some cases, is managed by a different entity from the ones that instead manages the following phases in the open data life cycle: for instance, in some public administration (PA) this activity is performed by a controlled company or an external vendor that can provides the devices (e.g. sensors) and a proprietary platform to collect the data. This means that, in this specific case, the PA has no direct control over the way in which raw data are collected and represented (e.g specific standards) so additional activities have to be performed to achieve a specific data compliance (see Data Quality Improvement phase)

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

Functionality name	Functionality description	Requirements mapping
Data Source connection	The framework provides the seamless connection and integration of data supporting various sources and standards, including IoT devices, databases, and external APIs, ensuring data can be collected from a wide range of origins efficiently and in real-time. The data collection process is done automatically avoiding manual data collection efforts through scheduling, scripting, and monitoring to increase efficiency and reduce human errors.	**DT-01**
Data Sampling and Aggregation	It is a crucial function that aids in the efficient management, analysis, and utilization of large datasets across various phases of the open data lifecycle. Data sampling involves selecting a representative subset of data for analysis, aiding in initial exploration, quality assurance, and performance optimization. Aggregation, on the other hand, consolidates, summarizes, or groups data to provide higher-level insights. Key functionalities include various sampling methods, aggregation techniques, granularity selection, and performance optimization. When used judiciously, data sampling and aggregation enhance data accessibility, interpretation, and performance, delivering valuable insights and safeguarding privacy across diverse domains.	**DT-04** **DT-05**
Data Streaming and Real-time Processing	It involves the continuous capture and processing of data as it's generated, enabling immediate insights and responses. By leveraging specific real-time technologies or cloud-based services, PA can collect and process data from various sources, such as IoT devices, social media, or web applications, in real time. This instantaneous data flow empowers organizations to make informed decisions, detect anomalies, and respond swiftly to changing conditions. These capabilities are necessary to manage open datasets coming from (near) real time devices providing data streams, strictly related to "smart city" applicative scenarios.	**DT-01** **UX-01**

Table 3: Framework functionalities of Data Collection

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

2.1.2 Data Quality Improvement

Data Quality Improvement focuses on enhancing the accuracy, reliability, and overall quality of collected data. This phase is vital because the value and utility of open data depend on the trustworthiness and completeness of the information it contains.

At its core, data quality improvement involves a series of processes and strategies to address issues that may compromise data integrity. These issues can arise during the data collection phase or result from inconsistencies and inaccuracies in the original data sources. Therefore, this phase plays a crucial role in transforming raw data into reliable and usable assets.

One of the primary aspects of data quality improvement is data validation. This process includes a set of procedures and rules designed to check data for errors, inconsistencies, and completeness. Data validation helps identify anomalies, such as missing values, outliers, and incorrect entries, ensuring that the data is accurate and reliable. Common validation methods include format checks, range checks, and logical consistency checks. For example, if data represents temperature values, a validation rule could ensure that the values fall within a plausible temperature range.

Data cleansing is another critical component of data quality improvement. It involves the identification and correction of errors and inconsistencies in the data. Data cleansing methods may include removing duplicates, standardizing data formats, and correcting typographical errors. For instance, if a dataset contains multiple entries for the same individual with slight variations in spelling, data cleansing processes would consolidate and standardize those entries to eliminate redundancy and improve accuracy.

Additionally, data enrichment can be part of the data quality improvement process. Data enrichment involves supplementing existing data with additional information to enhance its quality and usefulness. For instance, geographic data may be enriched with coordinates and demographic data to provide more context and value to the original dataset.

To ensure the ongoing quality of open data, the data quality improvement phase often includes establishing and adhering to data quality standards. These standards define the criteria and rules that data must meet to be considered of high quality. They may encompass aspects such as accuracy, completeness, timeliness, and consistency. Adherence to these standards helps maintain the long-term quality and reliability of open data resources.

Effective data quality improvement, as the following phases in the open data life cycle also requires maintaining metadata. Metadata includes information about the data, such as the data source, collection date, and transformations applied during the quality improvement process. Properly documented metadata ensures transparency and allows data users to understand how the data has been enhanced and validated.

Moreover, it's essential to maintain data lineage, which tracks the history and transformations applied to the data. This information is crucial for auditing and tracing any changes or enhancements to the data over time, ensuring transparency and accountability.

Data quality improvement is an iterative and continuous process, as data quality can degrade over time or with changes in data sources. Regular monitoring, validation, and cleansing of data are essential to uphold data quality standards. The quality of open data directly impacts its usability, reliability, and the trust users place in it. It is, therefore, a fundamental phase that underpins the broader goals of transparency, informed decision-making, and innovation driven by open data initiatives.

Some technical tasks related to this phase are in some cases performed together with the collection phase. For instance, a middleware that oversees collecting information from sensors could also perform most of the functionalities related to the data quality improvement.

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

Functionality name	Functionality description	Requirements mapping
Data Validation and Cleansing	Data validation involves scrutinizing data for adherence to predefined rules and standards. Inconsistencies, errors, or missing values are identified and flagged. Data cleansing, on the other hand, rectifies these issues through techniques such as imputation, outlier removal, and standardization. This crucial functionality guarantees that data is error-free and conforms to predefined quality standards, ensuring the integrity of analyses, reporting, and decision-making.	**DT-03**
Data Enrichment	It involves augmenting existing data with additional, relevant information from external sources. This process can include, for instance, appending geographical data, socio-demographic details, or financial metrics to enhance the context and depth of the dataset. Data enrichment not only refines the quality of data but also expands its utility, enabling organizations to make more informed decisions and extract richer insights.	**DT-04** **UX-01** **UX-04**
Data Monitoring and Alerting	It is a key process that ensures data remains accurate and reliable over time. It involves continuous surveillance of data, looking for anomalies, inconsistencies, or deviations from defined quality standards. Automated alerting systems are put in place to notify data stewards and administrators when issues are detected, enabling them to take corrective actions promptly. This functionality is indispensable for upholding data integrity, especially in environments where data is frequently updated or sourced from diverse origins. By proactively monitoring and alerting, organizations can maintain data quality, trustworthiness, and the effectiveness of data-driven operations.	**DT-03** **UX-09**

Table 4: Framework functionalities of Data Quality Improvement

2.1.3 Semantic Harmonization and Data Integration

Semantic Harmonization and Data Integration is a critical phase in the open data lifecycle, tasked with the challenging goal of combining data from diverse sources while ensuring that it is not only technically compatible but also semantically meaningful and contextually relevant. This phase plays a pivotal role in unlocking the full potential of open data, as it addresses the common challenge of data heterogeneity, data that may be represented in various formats, use different schemas, and exhibit discrepancies in terminology and meaning.

At the core of this phase is the process of semantic harmonization, which involves aligning the terminologies, definitions, and structures used across different datasets. It is essential to ensure that when data from multiple sources is integrated, it can be understood, compared, and analysed cohesively. This process might require mapping terms, concepts, and classifications to create a common data vocabulary. For instance, when integrating health data from different hospitals, the phase of semantic harmonization would ensure that "patient records," "medical cases," and "diagnosis codes" are understood consistently across datasets.

To facilitate such harmonization, it's common to adopt open standards, ontologies, and controlled vocabularies. These standardization measures help bridge the semantic gaps and facilitate interoperability, making data integration feasible. For instance, adopting widely recognized

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

ontologies like RDF (Resource Description Framework) or using standards like JSON-LD (Linked Data) enables datasets to express their semantics in a machine-readable and interoperable manner. Beyond harmonizing semantics, the phase of data integration deals with the technical aspects of combining datasets. This includes addressing data format differences (e.g., converting data from CSV to JSON), handling data structures (e.g., flattening hierarchical data), and ensuring data consistency and integrity. The process may also involve the mapping of data fields and attributes to enable data aggregation and correlation. For example, when integrating transportation data from multiple sources, this phase ensures that location data from various sources can be combined accurately, regardless of different coordinate systems or data encoding.

Another crucial component is data transformation, which involves the conversion of data into a common format or structure, enabling seamless integration. Data transformation can also involve the derivation of new data attributes or the generation of composite datasets from the integrated sources.

The result of successful semantic harmonization and data integration is a coherent and unified dataset that can be queried, analysed, and visualized as a single resource, even though it may originate from multiple sources. Users can access and interact with the integrated data without needing to understand the intricacies of the source datasets. This cohesiveness enhances the usability and effectiveness of open data, making it more accessible and valuable for a wide range of applications, from research and policy development to business intelligence and data-driven decision-making.

Functionality name	Functionality description	Requirements mapping
Data Mapping	It involves the creation of correspondence between data elements, attributes, or fields in different datasets, ensuring a common vocabulary and structure. This process is often guided by schema mappings or transformation rules that enable the seamless integration of data from disparate origins. Data mapping is crucial for establishing data compatibility, enabling efficient data aggregation, and ensuring that integrated datasets can be queried and analysed cohesively. It serves as the bridge that harmonizes and unifies data, laying the foundation for comprehensive, accurate, and insightful data integration across various industries and domains.	**DT-07** **DT-12**
Data Formatting and Structuring	This functionality involves converting data from its original format to a standardized format that ensures compatibility during integration. For example, it may convert data from CSV to JSON or XML to enable data from various sources to be combined seamlessly. It ensures that data is presented in standardized, machine-readable formats supporting data providers in structuring data efficiently for human and software consumption. The output of this functionality enables seamless integration by ensuring that data from different sources adheres to a unified format, making it easier to aggregate, query, and analyse.	**DT-02** **DT-06**
Data Transformation	Data transformation encompasses the conversion of data into a common structure, which could involve flattening hierarchical data, changing data types, or generating new attributes. It also includes the generation of composite datasets that merge data from multiple sources.	**DT-10** **DT-11** **DT-13**

Table 5: Framework functionalities of Semantic Harmonization and Data Integration

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

2.1.4 Data Publication

Data Publication is a key phase marking the transition from internal data management in order to make data accessible to the public and relevant stakeholders. In this phase, organizations and data providers take the data they have collected, cleaned, harmonized, and integrated, and make it available through open data portals, platforms, or other distribution methods. One prominent aspect is the provision of data via Application Programming Interfaces (APIs), which enables seamless programmatic access to the datasets, promoting easy integration into various applications and services. The primary goal is to enhance transparency, accountability, and accessibility, allowing data to be harnessed for a wide range of purposes, including research, policy development, business intelligence, and innovation.

Data publication typically involves several key steps and considerations. First, data providers need to choose the appropriate data distribution platform, often open data portals or websites, where data is not only accessible through download but also available via APIs for programmatic access. This allows developers to build applications and services that can dynamically fetch and utilize the data. Once the platform is selected, data must be formatted and structured in a way that is user-friendly and aligns with open data standards. This includes selecting a common file format such as CSV, JSON, or XML, and structuring the data to ensure it is easily machine-readable and interpretable by both humans and software applications. Standardized metadata, which describes the dataset's content and context, is crucial for helping users and developers understand the data.

Data access control and licensing are also important considerations in the data publication phase. Open data is typically made available to the public without discriminatory barriers, but data providers may implement access controls to protect sensitive information. Licensing is used to communicate the terms under which the data can be used, ensuring that users understand their rights and responsibilities. Common open data licenses, such as Creative Commons, are often applied to allow for data reuse while specifying attribution and usage restrictions.

To maximize data discoverability, metadata and documentation are essential. Each dataset should be accompanied by detailed metadata that describes its attributes, source, collection methods, and relevant usage restrictions. Documentation may also include data dictionaries, codebooks, and examples of how to use the data, making it easier for users and developers to understand and leverage the information.

Data publication should also prioritize data versioning and updates. Data is dynamic, and it's important to maintain historical records and provide users and developers with access to the most current information through APIs. Versioning ensures that previous iterations of data remain accessible, allowing for trend analysis and research.

Lastly, data publication involves creating a collaborative environment that encourages user feedback and contributions. This can involve mechanisms for reporting data issues, suggesting improvements, and even contributing to data updates or extensions. Open data thrives when the community of users and developers actively engages with the data, providing valuable input and helping to keep it accurate and relevant. APIs play a crucial role in enabling this collaborative ecosystem by simplifying data access and integration for developers.

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

Functionality name	Functionality description	Requirements mapping
Licensing Management	The framework will provide specific functionalities to support the establishment of clear, legally binding terms that define how data can be accessed, distributed, and utilized. Open data licenses (e.g., Creative Commons) outline usage rights, attribution requirements, and any restrictions, ensuring that users understand the boundaries and responsibilities associated with the data. By providing a legal framework for open data usage, licensing promotes transparency, protects intellectual property, and enables responsible and ethical data utilization, fostering trust and cooperation among data publishers and users.	**DT-09** **DT-17**
Data Versioning and Updates	It ensures that published datasets remain current, reliable, and aligned with changing data sources. It involves the systematic recording of dataset versions and updates, enabling users to access the most recent data while retaining access to prior iterations. This function is essential for maintaining data integrity and auditability. Users can track changes, understand dataset evolution, and trust that they have access to the latest information. It fosters transparency, accountability, and data quality, offering users a dynamic, up-to-date resource for informed decision-making and analysis in various fields and applications.	**DT-03** **DT-04**
Data publication workflow management	This functionality involves last-mile efforts to ensure that the data is ready for public consumption. It includes the final quality checks to confirm that the data is accurate, up-to-date, and compliant with relevant standards. Metadata is crafted, providing essential context and documentation for the dataset. Data formatting is performed to ensure consistency and usability, and any necessary data transformation or aggregation is completed. Licensing terms are also established to guide users in the responsible use of the data. Finally, the dataset is made accessible through an open data portal or distribution platform, making it available to a wide audience of stakeholders.	**DT-14** **DT-15** **DT-16** **DT-18** **UX-07**
Dataset public documentation	The framework will provide functionalities dedicated to create and manage comprehensive, user-friendly documentation for open datasets. This documentation serves as a vital companion to the data, providing context, usage instructions, and metadata that enable users to understand and utilize the dataset effectively. It typically includes data descriptions, schema explanations, information on data sources, and details on how to interpret and navigate the data. Moreover the documentation covers all the aspects related to the Open APIs structures and usage (e.g. endpoints, parameters, security etc.)	**UX-07**

Table 6: Framework functionalities of Data Publication

2.1.5 Data Access and Visualization

In the "Data Access and Visualization" phase the open datasets become accessible and through different channels, in particular via Open APIs and Open Data Platform, they are presented into a visual form that is not only more comprehensible but also rich with insights. This phase is instrumental in conveying patterns, trends, and anomalies hidden within the data, making it a crucial step in data analysis, interpretation, and communication. Data visualization bridges the gap between data and human understanding, as it leverages the human brain's ability to process visual information quickly and effectively. It goes beyond merely presenting numbers and statistics; it

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

paints a vivid picture of data that facilitates decision-making, supports research, and enhances communication.

Key functionalities within the Data Access and Visualization phase enable users to explore, understand, and communicate insights from data effectively. Interactive visualizations are at the core of this process, allowing users to delve into data dynamically. These visualizations often incorporate features like zooming, panning, and tooltips, providing users with the ability to interact with the data on a deeper level. Diverse chart types are offered to accommodate different data types and help users present their findings in the most suitable way.

Customization options are a valuable feature, enabling users to tailor the appearance of their visualizations to their preferences or to match their organization's branding. Real-time data updates are supported, a crucial feature for monitoring rapidly changing data patterns and trends. Users can query filter and select specific data subsets, focusing their attention on areas of particular interest.

Export and sharing options facilitate wider dissemination of insights, whether as images, PDFs, or shared links within reports and presentations. Data drill-down functionality empowers users to explore underlying details within the visualization, going beyond the surface to gain deeper insights. The ability to create interactive dashboards, comprising multiple visualizations, provides a holistic view of data from various angles. Real-time collaboration features like commenting and sharing within the visualization platform encourage discussions and collaboration among teams.

Performance optimization, especially when handling large datasets, guarantees a smooth user experience. Drill-through links enable users to access additional contextual information or source data, adding depth to their exploration. For users seeking automated insights, some frameworks offer suggestions based on the data being visualized, making pattern and trend discovery more accessible.

Annotations and storytelling features allow users to provide context and narrate data insights effectively, transforming the visualization into a powerful storytelling tool. By combining these key functionalities, users can unlock the full potential of data visualization, transforming raw data into actionable insights, and bridging the gap between data and decision-making across various domains, including business, research, and public communication.

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

Functionality name	Functionality description	Requirements mapping
Datasets' Bulk Download	This functionality allows users to download complete datasets rather than individual data points, enabling them to work with data on their own terms and within their preferred tools or environments. It is particularly beneficial for researchers, analysts, and developers who require comprehensive access to data for in-depth analysis or integration into their applications. This functionality can be accessible from both APIs and Open Data Portal solutions.	**DT-09** **DT-15**
Dataset's Open APIs	It consists of the key open data dissemination channel offering a structured and programmatic way for users to interact with open datasets. The Open APIs provide standardized methods for querying, retrieving, and even updating data, enabling developers and applications to seamlessly integrate open data into their workflows. Dataset Open APIs enhance data accessibility and usability, making it easy for a wide range of users to extract specific information or develop custom applications that rely on the dataset's content. Moreover, Open APIs facilitate seamless data integration, enabling AI systems to access and analyze data efficiently. This not only expedites AI development but also ensures that AI models are trained on up-to-date, high-quality data, enhancing their accuracy and performance.	**DT-09** **DT-15**
Data query	It empowers users to extract specific subsets of data from a larger dataset based on their unique criteria. It offers an interactive and dynamic way to explore data, enabling users to filter, sort, and retrieve the information they need in real-time. Data query tools provide a user-friendly interface for constructing and executing complex queries, facilitating data exploration and analysis. This functionality supports the usage of different query languages tailored for linked open datasets (e.g. SPARQL) and can be accessible from both APIs and Open Data Portal solutions.	**DT-08** **DT-16**
Export and Sharing	Enabling users to export visualizations as images, PDFs, or share them on social media or within reports promotes wider dissemination of insights.	**UX-01**
User Feedback and Contributions	Encouraging user feedback and contributions enhances data quality. The framework should allow users to report issues, suggest improvements, and contribute to data updates and extensions, fostering a collaborative data ecosystem.	**UX-05**
Interactive Chart and Dashboards	Providing a variety of chart types (e.g., bar charts, scatter plots, heatmaps) and interactive dashboards that allow users to explore data dynamically and to gain a holistic view of data from various angles.	**UX-01**
Data Filtering and Selection	Offering data filtering and selection capabilities allows users to focus on specific data subsets, enhancing the exploration of data-driven insights.	**UX-01**

Table 7: Framework functionalities of Data Access and Visualization

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

2.1.6 Cross-phase capabilities

2.1.6.1 Metadata Management

Metadata management is a fundamental functionality that underpins the entire open data lifecycle, ensuring data is not just accessible but also understandable and valuable to users. This function plays a central role in each phase, from data collection to data access and visualization, making data more transparent and meaningful.

In the Data Collection stage, metadata management begins with the creation of metadata tags, which describe the data's source, structure, and characteristics. These tags provide context for the data being collected, detailing where it originates, how it's generated, and any relevant timestamps. By documenting metadata at this early stage, data collectors and organizations can lay the foundation for consistent and comprehensible data, contributing to its reliability and usefulness in subsequent phases.

Metadata in the Data Quality Improvement phase acts as a compass for data stewards and quality improvement teams. Detailed metadata descriptions allow them to understand the data's intended purpose and make informed decisions about data cleansing, validation, and transformation. Metadata tags serve as a reference, helping quality assurance processes by providing insights into the expected data quality standards and criteria.

In the Semantic Harmonization and Data Integration phase is Metadata is a critical element: as data from various sources is integrated, metadata schemas guide the alignment of diverse data elements. By mapping metadata from multiple sources to a common structure, harmonization becomes more efficient. Metadata standards like schema.org and JSON-LD enable organizations to structure data in a semantically consistent way, ensuring that data from disparate origins can be integrated cohesively.

Detailed metadata descriptions, during Data Publication, provide essential context and accessibility for open datasets. Metadata tags include information about data ownership, licensing, usage restrictions, and data format. These descriptions are integral to the organization's data catalog or portal, making it easier for users to discover, comprehend, and effectively use the data. Open data standards like DCAT-AP guide metadata creation and publication.

During the Data Access and Visualization phase metadata acts as a guide for data users, offering insights into dataset structure and content. Users can rely on metadata descriptions to understand what data is available and how to retrieve it. APIs often include metadata descriptions, allowing users to programmatically query data. This phase also involves the creation of visualizations and dashboards. Metadata helps users understand what data elements are represented in visualizations and how they should be interpreted.

Metadata descriptions are also a crucial companion to AI applications. AI algorithms rely on comprehensive data to train and operate effectively. Metadata tags indicate data source, quality, and potential biases, helping AI practitioners understand the data's fitness for a particular purpose. For AI model deployment, metadata may guide how data is pre-processed and integrated.

Metadata management, as demonstrated across the open data lifecycle, is the backbone of data accessibility, understanding, and usability. It provides the context that transforms data from a mass of numbers and characters into a valuable resource for research, innovation, and informed decision-making. By documenting the data's origin, structure, ownership, and licensing, metadata tags ensure that open data remains a trusted and effective asset throughout its lifecycle, benefiting a wide range of users and stakeholders.

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

2.1.6.2 Privacy and Security

Even though the core of the BeOpen project is Open Data, and information that by definition is publicly available Data Privacy and Security represent a fundamental and pervasive functionality that permeates every stage of the open data lifecycle. As data flows through its various phases, safeguarding the integrity and confidentiality of data is of paramount importance to assure that, at the end of the process, the data that we are going to share is correct and safe respecting all the privacy requirements and regulations.

In the initial phases of data generation and collection, data privacy and security measures are enacted at the very point of data creation. Whether it is personal information collected from individuals or sensitive business data, mechanisms such as encryption and access controls help protect the data from unauthorized access. In some cases, data anonymization techniques are employed to de-identify data, ensuring that the privacy of individuals or organizations is maintained.

During the data quality improvement phase, data privacy and security play a multifaceted role. Version control mechanisms ensure that data alterations are traceable, safeguarding data integrity and also serving as a crucial tool for auditing any changes to the dataset. This traceability is indispensable for data stewards and quality improvement teams to identify issues, correct inaccuracies, and maintain data quality.

In the semantic harmonization and data integration phase, data from disparate sources is consolidated and aligned. Privacy and security concerns escalate as sharing of data, but also attributes or inferences of it needed for specific purpose, increases. Advanced access control mechanisms, along with encryption, become necessary to safeguard data repositories and ensure that sensitive information remains protected. Additionally, data mapping and transformation processes are secured to prevent data breaches during the integration process.

As data moves into the data publication phase, privacy and security are still paramount. Data repositories must implement access controls, allowing for controlled, authenticated, and authorized access to the workflow of the open data publication. Open data licensing terms and usage policies are closely aligned with data privacy. These legal frameworks outline how data can be used and shared, ensuring that data publishers and users are clear about their rights and responsibilities. Data visualization, which serves as a gateway for users to access and comprehend data, also involves privacy and security considerations. Visualization tools must provide secure access, with authentication and authorization mechanisms to protect sensitive data from unauthorized viewers. Privacy settings should also allow users to control the visibility of data, balancing openness with data confidentiality. Finally it is important to underline that the different phases must adhere to privacy regulations, such as GDPR, for datasets that contain personal or sensitive information, so the BeOpen framework has to support the users in reaching this compliance.

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

2.1.6.3 Data Storage

Data storage serves as the bedrock of the open data lifecycle, seamlessly connecting each phase of data management. This functionality begins its role with data generation and collection, ensuring that data from diverse sources is securely organized, structured, and ready for subsequent steps. As data quality improvement commences, storage systems preserve multiple data versions, offering a comprehensive history of changes for auditing and troubleshooting purposes. During semantic harmonization and data integration, these systems support the standardized formatting and harmonization of data from various sources, maintaining data integrity. In the data publication phase, data storage repositories play a pivotal role, implementing access controls to safeguard sensitive information and housing data in formats suitable for open distribution.

As the open data journey extends to the data visualization phase, efficient and secure data storage remains at the core. Data visualization tools rely on the structured, accessible data stored within these systems to generate dynamic and interactive charts, empowering users to extract insights.

Throughout this data management process, several key considerations come to the forefront. Scalability is crucial, and cloud-based storage and distributed NoSQL databases provide the capacity needed to accommodate burgeoning datasets. These technologies facilitate horizontal scalability and improved performance, enabling the storage and retrieval of large datasets. Big data technologies (e.g.Hadoop and Apache Spark) have also transformed data storage by providing distributed file systems and data processing frameworks that manage and analyze vast amounts of data efficiently.

Data versioning is indispensable, ensuring that data revisions are meticulously recorded, preserving historical records and tracking changes. Data storage systems also must incorporate robust backup and recovery mechanisms to mitigate data loss risks and facilitate disaster recovery. Furthermore, data accessibility is optimized, supporting swift data retrieval and diverse access methods.

Lastly, data archiving strategies come into play, ensuring the enduring accessibility of data for future use. In essence, data storage is the unifying thread in the open data lifecycle, underpinning every facet of data management, from initial creation to its ultimate utilization. Its multifaceted functionalities, from security and scalability to versioning and metadata management, assure data's protection, availability, and continued relevance, making it the cornerstone of successful open data initiatives in a multitude of domains.

For each of the above mentioned phases of the open data management life cycle different technologies and standards can be applied, following as an example are reported some of the possible approaches to be adopted:

- **Data Collection**: In this phase, data is often stored using traditional database management systems (DBMS), such as MySQL, PostgreSQL, and MongoDB. For sensor data or real-time collection, NoSQL databases like Apache Cassandra are employed. Data storage standards, such as SQL and NoSQL, guide the organization and retrieval of data.

- **Data Quality Improvement**: During this phase, data storage often relies on version control systems like Git or databases with built-in versioning features. These technologies maintain historical records of data changes. Data Quality standards like ISO 8000 provide guidelines for data quality management.

- **Semantic Harmonization and Data Integration**: Data lakes, implemented using cloud-based services like AWS S3, Azure Data Lake Storage, and Google Cloud Storage, are

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

common for storing raw and harmonized data. JSON-LD and schema.org are standards for semantic harmonization and structuring data.

- **Data Publication**: For open data publication, data is typically stored in repositories managed by open data platforms (e.g. CKAN, DKAN) is often published in formats compliant with standards such as CSV, JSON, and XML. Open data standards like DCAT (Data Catalog Vocabulary) guide how data should be described and published.

- **Data Access and Visualization**: In this phase, data storage focuses on providing efficient data retrieval. Technologies like distributed databases (e.g., Hadoop HDFS) and in-memory databases (e.g., Apache Kafka) are used to support real-time access to data. Web standards like RESTful APIs are employed for data access. For data visualization, interactive data storage, and retrieval technologies like Elasticsearch are utilized to enable fast search and visualization capabilities.

2.1.7 BeOpen Framework and AI-related capabilities

One of the objectives of the BeOpen project is to provide a framework that can "*improve technical data interoperability, semantics and quality of the public sector datasets*" in order to be used in Artificial Intelligence applications: the BeOpen framework must encompass a wide array of capabilities to ensure that datasets are prepared for AI-driven tasks. In this context the technical functionalities presented in the previous sections play an essential role in guaranteeing not only accessibility but also the seamless integration of data with AI models and applications.

Data pre-processing tools constitute a foundational component, offering capabilities for data cleaning, feature engineering, and transformation. These tools are instrumental in readying data for machine learning models by addressing common issues like missing values and standardizing data formats (e.g. CSV, JSON, and Parquet). Compatibility with these formats simplifies data ingestion and integration with popular AI libraries and frameworks. Additionally, data encoding and character set considerations are crucial, especially for internationalization.

Data version control is a critical necessity for AI practitioners as it enables access to specific dataset versions for reproducibility in research and model development. Version control mechanisms maintain data consistency and traceability, ensuring that users work with precise dataset versions.

API integration plays an essential role, allowing programmatic access to datasets. A well-documented set of APIs exposes data for real-time analysis and model training, making it accessible to developers who are integrating data into their AI applications.

Comprehensive metadata and a clearly defined data schema are essential for a comprehensive understanding of datasets: metadata describes attributes, source information, collection methods, and contextual details, facilitating data exploration and informed model feature selection.

Data quality assessment tools are indispensable in the AI context, automatically identifying and addressing data quality issues. These tools encompass functionalities for outlier detection, data imputation, and statistical tests to ensure the reliability of data for AI modelling.

AI applications often involve large datasets, necessitating scalability and performance optimization. The framework should be designed to handle high data volumes efficiently, utilizing techniques like data sharding, parallel processing, and distributed computing to enhance performance.

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

Data privacy and security features, including data anonymization and encryption, are crucial for protecting sensitive information while still allowing data to be used for AI applications. Access controls and encryption mechanisms are essential for safeguarding data at rest and in transit.
Data linking and joining capabilities are paramount for creating comprehensive training datasets in AI. These functionalities permit the integration of data from various sources, align data using common identifiers, and merge datasets to create richer training data.
A user-friendly data catalogue simplifies dataset discovery for AI tasks. Advanced search and discovery features, such as tagging, filtering, and integration with search engines, enhance dataset discoverability.
Clear data licensing terms and usage policies are essential for AI practitioners to comprehend how they can use, redistribute, and share data. Licensing metadata, combined with embedded licensing information in data files, ensures users comply with licensing requirements.
Data visualization tools, offering the generation of plots, charts, and graphs, are valuable for AI users during the data exploration phase. Visualization aids in identifying patterns and potential features for AI modelling.
Data transformation tools streamline the process of reshaping data, handling missing values, and integrating data from different sources, meeting the requirements of AI models. The framework should offer a variety of transformation functions and scripts.
Finally, community support features, including forums, documentation, and user support channels, foster collaboration and knowledge sharing among AI practitioners. These resources are instrumental in addressing data-related challenges while leveraging open datasets for AI applications, promoting informed decision-making across diverse domains.

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

2.2 BeOpen framework architecture overview

The different functionalities described in the previous sections represent the core logical elements of the BeOpen framework architecture that are depicted in Figure 1. The architecture is structured following the phases of the open data management presented earlier in this chapter. The lower part of the picture represents the collection of data from the different heterogenous data sources to be used in the project: these are mostly related to the data sources identified in city pilots including, for instance, data from internal PA systems, IoT devices, Open data portals, but also datasets from external public entities and organisation.

Data collected and pre-processed through the Data Collection functionalities can be later validated and improved using the capabilities of the Data Quality improvement. The Semantic Harmonisation and Data Integration phase will take care of the data interoperability and harmonisation of the datasets. At the end of these stages, validated and interoperable High Value Datasets will be published moving across a data publication workflow that will assure their availability, completeness and documentation.

The last layer of the architecture, related to Data Access and Visualisations, shows the different functionalities that can be accessed by both final users and the data consumer applications that will reuse/process the HVDs. The access will be granted via multiple channels, that will include open data systems (e.g. open data portals) and APIs. The final stakeholders, that will benefit from the HVDs, are depicted in the top part of the figure: they will access to the HVDs directly (e.g. using an open data portal) or through the data consumer applications (e.g. the pilot applications to be developed in the project) that will provide specific (AI) data processing and analysis capabilities. Thanks to the use of standards and interoperable interfaces and models, HVDs made available through the BeOpen framework can be also accessible, by external third-party systems or National and European data portals (data.europa.eu) for their further sharing and reuse.

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

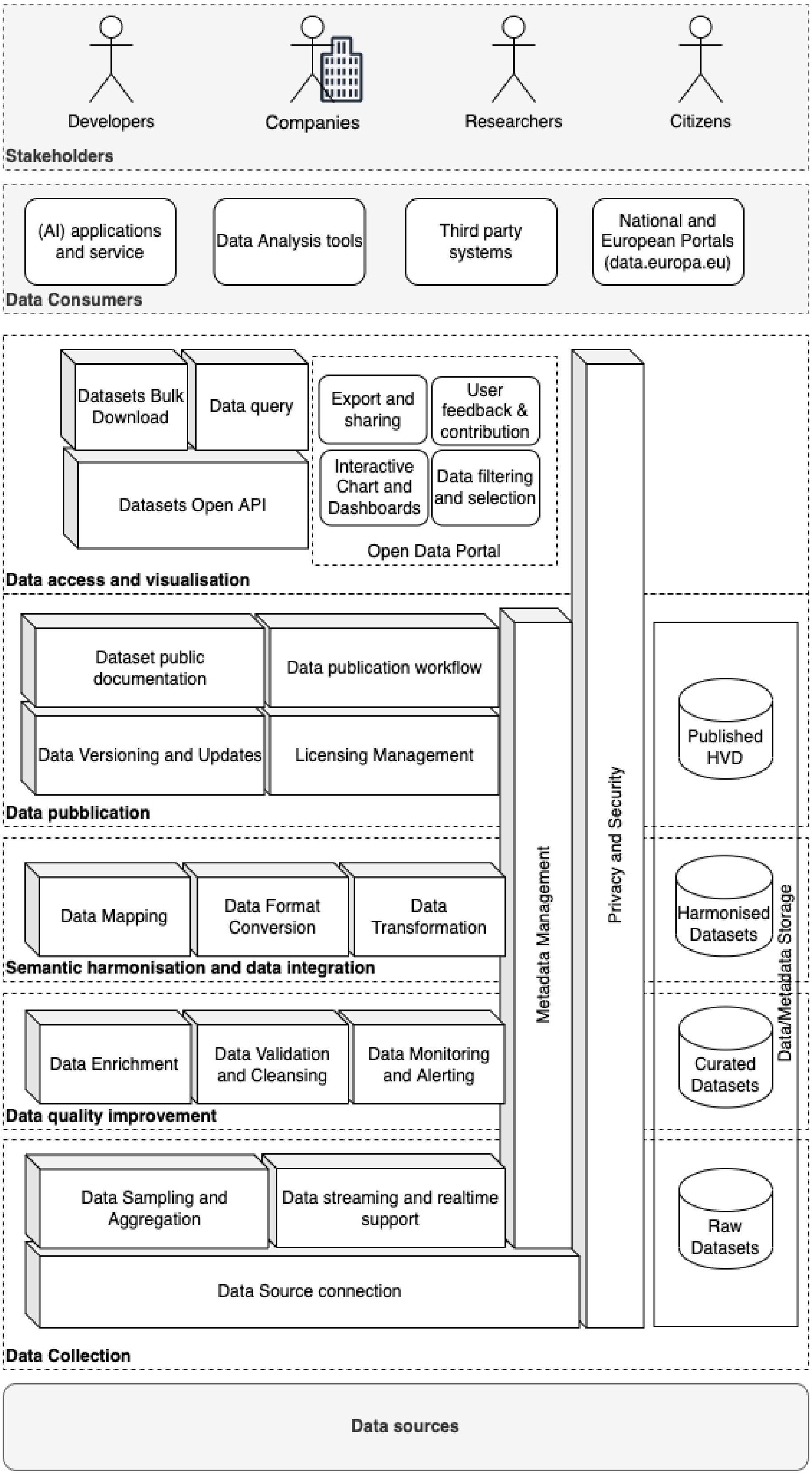

Figure 1: BeOpen framework high level logical architecture

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

2.3 Reusable tools for BeOpen framework

During the initial analysis phases of the BeOpen framework a set of existing data related tools have been studied in order to identify potential capabilities to be reused in the framework implementation; indeed the approach that will be adopted in BeOpen is to maximise the reuse of existing open source tools to speed up the release of the framework and, at the same time, to rely on mature software that, in some cases is already used inside the Public Administrations participating in the project. Table 8 is a tentative list of such possible reusable tools that will be eventually extended during the framework implementation phase where a final choice about their specific usage will be taken.

Name	Description	Functionalities	Main related standards	License
CEF Orion Context Broker [6]	Orion-LD is an NGSI-LD Context Broker implementation written in C/C++. It is a standalone executable and therefore small, fast, lightweight, and easy to handle. Context brokers allow for the management and requesting context of information in a structured manner based on linked data standards following the ETSI NGSI-LD specification. Orion-LD is more suitable for smaller installations or possibly in embedded environments. Orion-LD is not fully compliant with the version 1.3.1 of the ETSI NGSI-LD API specification. The ETSI NGSI-LD specification is a living, changing document (on version 1.5 as of March 2021), with features added in a pace that is just not possible to keep up for an NGSI-LD implementation.	Creation, modification and deletion of Entities and their attributes Batch operations to create/update a set of Entities in a single request. Query/Retrieve Entities, with a rich set of filters and a powerful query language. Query over existing Entity types and Attributes. Subscriptions, to get notified on changes in Entities, instead of actively polling the broker Registrations, to extend the broker with entities that live inside external context sources or brokers Temporal evolution of entities, with corresponding services to query GeoJSON render for Query/Retrieve Entities, to view in existing GIS tools	ETSI NGSI-LD	AGPLv3 [7]
Idra [8]	Idra is a web application able to federate existing Open Data Management Systems (ODMS) based on different technologies providing a unique access point to search and discover open datasets coming from heterogeneous sources. Idra uniforms representation of collected open datasets, thanks to the adoption of international standards (DCAT-AP) and provides a set of RESTful APIs to be used by third party applications. The datasets remain in the original portals, and the platform imports and manages only the metadata, updating it periodically in order to ensure that the information provided is up to date	Web interface searches for open data in a federated and multi-language way Unique access to open data regardless of the heterogeneity of the underlying technologies Public and open APIs to build smart and innovative applications Compliance with DCAT-AP standard and European Data Portal specifications Web scraping technology to federate open data portals that do not provide APIs	DCAT-AP SPARQL CKAN	AGPLv3

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

		Ease of performing SPARQL queries on 5 stars RDF linked open data		
GraphDB [9]	Importing Linked-Data (RDF or turtle files) it is possible to export in different formats and query and explore data. Highly efficient and robust graph database with RDF and SPARQL support. GeoSPARQL support (standard for geographically linked data). SPARQL Federated Query support: it is possible to combine a query on a repository with a remote call to a different SPARQL endpoint. It is an HTTP service endpoint: that can receive API requests and process SPARQL queries queried directly by client applications.	High performance: it allow users to query linked data quickly and efficiently, making them much faster than traditional relational databases. Data integrity: designed to maintain the integrity of data. Organizations can ensure that their data remains consistent and accurate throughout their applications. Provides powerful search capabilities and enable superior performance when performing complex queries on a large dataset. Possibility of exporting a graph in a vast list of formats as well as exporting the results of a query in a different format. GeoSPARQL support for representing and querying geospatial linked data may be of particular interest.	SPARQL GeoSPARQL RDF	Freeware
Open StreetMap [10]	OpenStreetMap (OSM) is a free, open geographic database updated and maintained by a community of volunteers via open collaboration. Contributors collect data from surveys, trace from aerial imagery and also import from other freely licensed geodata sources. OpenStreetMap is freely licensed under the Open Database License and as a result commonly used to make electronic maps, inform turn-by-turn navigation, assist in humanitarian aid and data visualisation. OpenStreetMap uses its own topology to store geographical features which can then be exported into other GIS file formats. The OpenStreetMap website itself is an online map, geodata search engine and editor.	Provides information regarding POI, locations, streets, roads, and other relevant geographical locations, as well as, be used as a database to store newly created geographical open data.	JSON GTFS	ODbL [11]

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

CKAN Portal and extensions [12]	**CKAN Portal** is the world's leading open-source data management system. It is an open-source DMS (data management system) for powering data hubs and data portals. CKAN makes it easy to publish, share and use data.	It is an open-source tool for making open data portals, helping them to manage and publish data collections, which means opening information to the public that fosters transparency and knowledge. Advanced geospatial features, covering data preview, search and discovery. CKAN among its features provides numerous extensions to take advantage of, among which we find extensions to customise layout of CKAN open data portals [13] and extensions to create surveys on CKAN open data portals and store responses as open data [14].	DCAT-AP JSON	AGPLv3
Data Model Mapper [15]	Data model mapper is a tool used to convert a CSV, JSON, or GeoJSON source to a JSON output. The structure of the output file will not be inherited from the structure of the source file but from a related schema provided by the user, the Mapping schema file. This Mapping file determines the way the input values will be set to the output values. The files in input can contain either rows, JSON objects or GeoJSON Features, each of them representing an object to be mapped to an Entity, according to the selected Data Model.	Multiple ways to access and use Data model mapper, as: ○ Command line tool ○ REST API service ○ Frontend tool editor. It is a tool that promotes data interoperability as the initial structure of input data is transformed: it basically allows to conversion of several file types to different FIWARE Smart Data Models. Final output validated objects can be sent directly to a configured Orion Context Broker.	JSON ETSI NGSI-LD	AGPLv3
MinIO Object Storage [16]	It is a High-Performance Object Storage for Virtualized Environments This Object Storage could be the component in charge of managing the persistence. It can handle unstructured data such as log files, backups, tabular data, and container images with a current maximum supported object size of 50TB. It is API compatible with the Amazon S3 cloud storage service. A stable and reliable storage platform for data analytics workloads.	During the first Data collection phase, raw data but also open datasets created after processing, could be stored into the Object Storage component in order to be an input for the following phases. MinIO offers a suite of options to access, such as: ○ graphical user interfaces (GUI) ○ command line interfaces (CLI) ○ application programming interfaces (APIs). It is a high-performance AWS S3 compatible object store: S3 compatibility is a hard requirement: as the RESTful API standard, S3 has revolutionized the interaction between applications, data and infrastructure. S3 compatible object storage and applications can run anywhere.	AWS S3	AGPLv3 MinIO Commercial License [17]

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

SparQLing [18]	Graphical editor for building SPARQL queries, usable via web application.	Graphic tool to allow the creation of queries in the SPARQL language starting from an ontology in Graphol format [19], for users who are not familiar and experienced in using SPARQL language.	SPARQL	MIT [20]
Olap Browser [21]	Web based tool that enables data analysts to reuse existing data stored as RDF Data cubes by presenting a two-dimensional slice of the cube as a table and enabling OLAP operations (roll-up, drill-down, pivot, dimension reduction etc.).	The OLAP Browser enables the exploration of RDF data cubes represented according to the JSON-qb API specification [22], by presenting each time a two-dimensional slice of the cube as a table.	RDF JSON	CC by 4.0 [23]
Cube Visualizer [24]	Web application for graphical representations of an RDF data cubes.	The Cube Visualizer is a web application that creates and presents to the user graphical representations of an RDF data cube's one-dimensional slices. It is built as a client of the JSON QB REST API implementation according to the JSON-qb API specification [22]. User's choices (measure, free dimension and dimension values) are translated to appropriate API calls. The returned data are presented to the user in the form of a chart, the type of which (bar chart, pie chart, sorted pie chart, area chart) can be also selected.	RDF JSON	CC by 4.0

Table 8 : Reusable tools for the BeOpen framework

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

2.3.1 Mapping between tools and framework functionalities

The following table presents the relationship between the framework functionalities, previously described, and the tools proposed to be reused in the BeOpen framework. It is important to remark that a tool, that in the table provides one or more framework functionalities, could cover them only partially and for specific aspects, this means that the framework implementation should also be based on additional software to be identified or to be developed from scratch. Details about degree of coverage of the functionalities by the different tools will be investigated in the next phase during the concrete framework technical design and implementation.

	Data collection				Data quality improvement			Semantic harmonization and data integration		
	Data source connection	Data synchronization and versioning	Data sampling and aggregation	Data streaming and realtime support	Data enrichment	Data validation and cleansing	Data monitoring and alerting	Data mapping	Data formatting and structuring	Data transformation
CEF Orion Context Broker		✓		✓	✓			✓		✓
Idra	✓	✓		✓	✓			✓		✓
GraphDB			✓	✓	✓				✓	
OpenStreetMap										
CKAN Portal and extensions										
Data Model Mapper								✓	✓	✓
MinIO Object Storage		✓		✓	✓		✓			
SparQLing									✓	✓
OLAP Browser										
Cube Visualizer										

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

	Data publication				Data access and visualization						
	Dataset public documentation	Data versioning and updates	Lincensing management	Data publication workflow	Dataset Bulk download	Datasets Open API	Data Query	Export and sharing	User feedback and contribution	Interactive charts and dashboards	Data filtering and selection
CEF Orion Context Broker	✓	✓				✓	✓				
Idra	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓
GraphDB	✓				✓	✓	✓	✓		✓	✓
OpenStreetMap						✓	✓	✓			
CKAN Portal and extensions	✓			✓	✓	✓	✓	✓			✓
Data Model Mapper											
MinIO Object Storage	✓	✓			✓	✓	✓	✓			
SparQLing										✓	✓
OLAP Browser										✓	
Cube Visualizer										✓	

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

3 Standards and interoperability specifications

In this section are presented, in relation with BeOpen framework requirements, architectures and international norms which are the key standards and interoperability specifications that has to be taken in account in the process of open data collection, improvement and publication. This section mainly refers to existing European technical standards and norms including. as primary input, the Commission Implementing Regulation (EU) on HVDs.
The chapter has been divided into specific sections: the first resumes the scope of crucial official European standards, while the second describes the scope of some key European initiatives, both in the context of BeOpen requirements, making a reference to the official documents.

3.1 European Technical Standards

3.1.1 Commission Implementing Regulation (EU) on HVDs

On December 21st 2022 has been released an implementation act focused on High-Value Datasets due to the growing importance of data [2]. The Portal for European Data has published an easy-to-read overview [25]. Commission Implementing Regulation (EU) is a document laying down a list of specific high-value datasets and the arrangements for their publication and re-use. The main objective of establishing the list of high-value datasets is to ensure that public data of highest socio-economic potential are made available for re-use with minimal legal and technical restriction and free of charge. Harmonising the implementation of the re-use conditions of high-value datasets entails the technical specification for making the datasets available in a machine-readable format and via application programming interfaces (APIs). The open data policies in the Member States are strengthened by making high-value datasets available under optimal conditions, which builds on Findability, Accessibility, Interoperability, and Reusability, known as **FAIR principles** [1].
The regulation lays down the themes of high-value datasets by listing six thematic data categories:

1. **Geospatial**
2. **Earth observation and environment**
3. **Meteorological**
4. **Statistics**
5. **Companies and company ownership**
6. **Mobility**

With such regulations the European Commission identified several high-value datasets within each of the six data categories and the arrangements for their publication and re-use following a comprehensive stakeholder consultation and impact assessment for this Implementing Regulation. The provisions of Union and Member State legislation that go beyond the minimum requirements set out in this Implementing Regulation, especially in cases of sectoral law, will remain in effect.

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

Where making high-value datasets available for re-use entails the processing of personal data, such processing should be carried out in accordance with Union law on the protection of individuals with regard to the processing of personal data, including any provisions of national law further specifying the application of the **GDPR** [26]. In order to make as much data as possible available for reuse, member states should use appropriate methods and techniques such as generalization, aggregation, suppression, anonymization, differential privacy, or randomization.
In addition, an objective of such regulation is to promote the use of standard public licenses available online for re-using public sector information. The guidelines set forth in the regulation identify Creative Commons licenses (CC) as a valid example of recommended standard public licenses. Creative Commons licenses have been developed by non-profit organizations and have quickly become a leading licensing solution for public sector information, research results, and cultural domain material worldwide. In such Implementing Regulation it is also explicitly indicated to necessarily refer to the most recent version of the CC license suite, namely. CC 4.0 [23]. The main features of HVDs according to the directives issued by the Implementing Regulation are schematically represented in Figure 2.

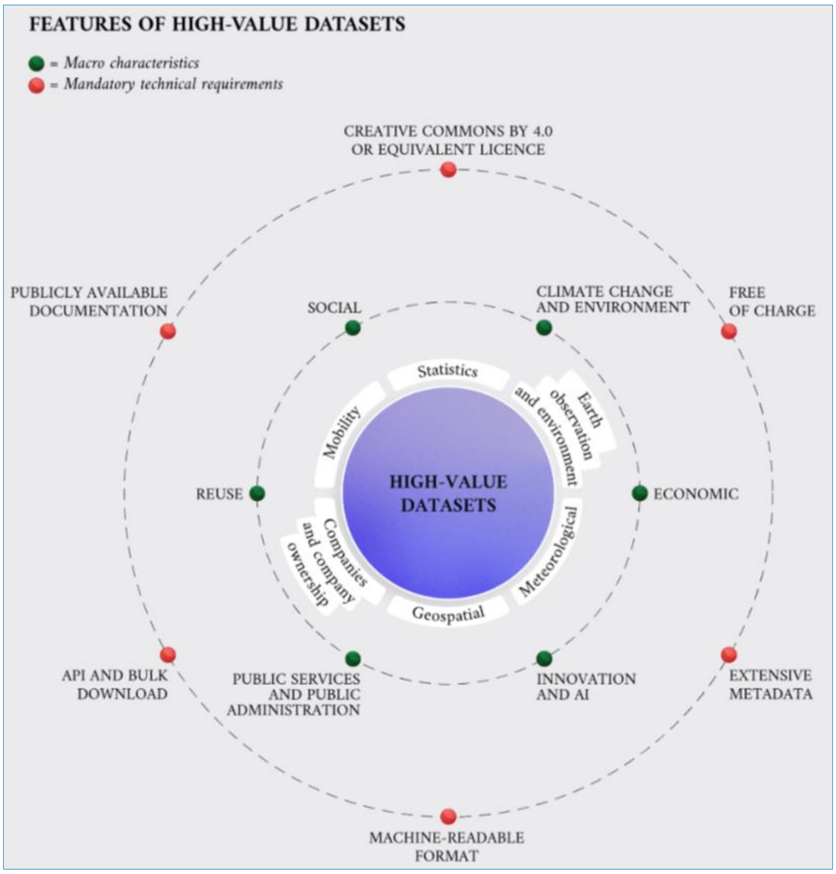

Figure 2: Features of HVDs according to the Implementing Regulation [25]

This Implementing Regulation also lays down the arrangements for **publishing and reusing high-value datasets**, in particular the applicable conditions for re-use and the minimum requirements for disseminating data.
In relation to them, datasets shall be made available according to the following ways:

- Under the conditions of the *Creative Commons BY 4.0 license* [23] or any equivalent or less restrictive open license.
- In a publicly documented, Union or internationally recognized open, machine-readable format.
- Through *application programming interface* (API) [27] that means a set of functions, procedures, definitions and protocols for machine-to-machine communication and the seamless exchange of data.
- Through *bulk download* that means a function that enables a download of an entire dataset in one or several packages.

In addition to these general guidelines common to all six categories of HVDs, the regulation also states more specific and precise indications different for each individual category. These indications are described in detail in the Annex to the Regulation.

The BeOpen framework must adhere to the directives set out in this EU Regulation. BeOpen framework's priority must be to use and provide the dataset categories identified in the Regulation and comply with the minimum requirements for reuse and dissemination. The following picture depict the BeOpen framework logical architecture in relation with some high-level requirements of the HVD regulation for each of the open data management phase.

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

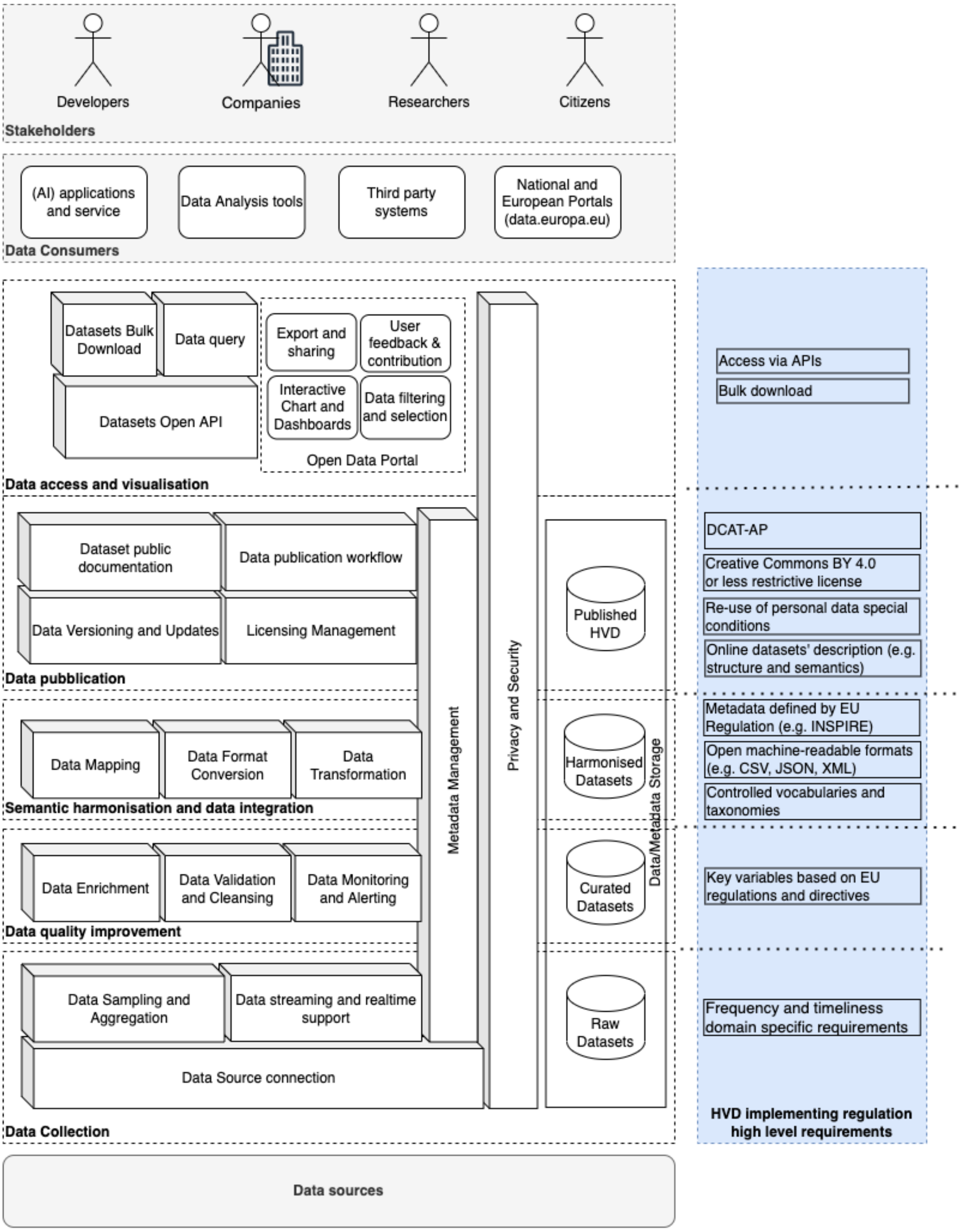

Figure 3 - BeOpen framework architecture and HVDs regulation high level requirements

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

3.1.2 DCAT Application profile for data portals in Europe (DCAT-AP)

The Data Catalogue vocabulary (DCAT [28]) serves as the basis for the DCAT Application profile for data portals in Europe (DCAT-AP) which describes public sector datasets in Europe. The main purpose is to enable cross-data portal search for data sets and increase the searchability of public sector data across borders and sectors. By exchanging descriptions of data sets between data portals, this can be accomplished.

DG CONNECT, the EU Publications Office, and the ISA Programme worked together to define the DCAT-AP. A group of experts from 16 European member states, some European institutions, and the US developed the specification.

At the time, the latest version available of the specification is the DCAT-AP 3.0.0 [29].

By utilizing terms from one or more base standards, DCAT-AP creates specificity by identifying mandatory, recommended, and optional elements for a specific application, as well as recommendations on the use of controlled vocabularies. The two main elements in the DCAT-AP can be considered the following:

- A **Dataset** that is a collection of data published or curated by a single source and can be accessed or downloaded in one or more formats.
- A **Data Portal** that is a system running on a web-based platform to maintain a data catalogue with the descriptions of datasets, offering services that enable the discovery and reuse of datasets.

Ensuring consistent description metadata published by data portals across Europe is a crucial goal. The scenario described below is an example of this and involves two entities, **Data providers** and **Data reusers**:

- By making their datasets searchable and accessible, data providers aim to encourage their reusability. It can be more important to publish description metadata of datasets online than to make the actual data available. By listing their data sets on one or more data portals, they encourage the reuse of their data sets, which can significantly reduce costs.
- Obtaining an overview of datasets and the respective public administrations that are maintaining them, can be challenging for data reusers, particularly when the datasets are located in a Member State where language barriers and unfamiliar government structures are present. Data publishers and portals keep catalogues of datasets made available by public administrations on their websites to address this problem. The quality of the description metadata in these catalogues has a direct impact on the ease of finding datasets.

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

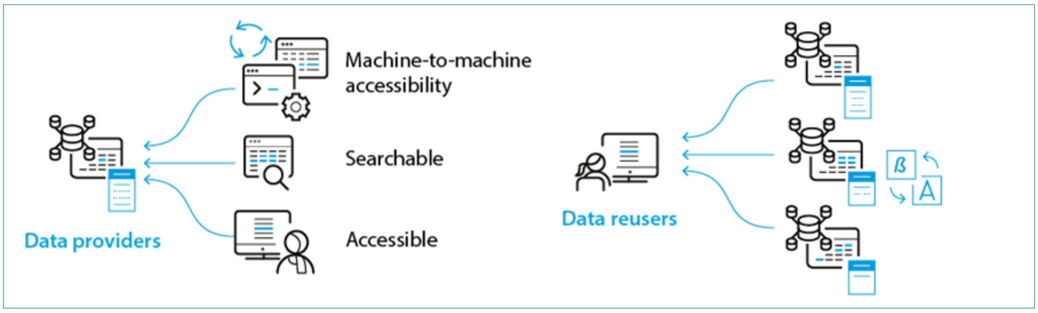

Figure 4: Benefits of DCAT-AP for Data providers and Data reusers [30]

DCAT-AP intends to enable cross-data portal searches for data sets and enhance access to public sector data across borders and sectors. Sharing dataset descriptions among data portals can lead to this. Since the beginning, the DCAT Application Profile was designed to make data reuse easier by adapting DCAT. In concrete terms:

- It proposes mandatory, recommended or optional classes and properties to be used for a particular application;
- It identifies requirements to control vocabularies for this particular application;
- The collection includes other elements that should be prioritized or required for an application, like conformance statement, agent roles, or cardinalities.

DCAT-AP has been implemented by over 15 open data portals in the European Union, including the **European Data Portal** [31]. Moreover, some EU Member States have created extensions for the DCAT-AP, such as the **DCATAP_IT** [32]. To meet the information requirements of the statistical and geospatial domains, the following are some of the main extensions of the DCAT-AP:

- **StatDCAT-AP** [33], created in partnership with Eurostat, brings the statistical and open data communities closer by increasing awareness and making it easier to integrate statistical data sets into open government data portals.
- **GeoDCAT-AP** [34] describes geospatial datasets, dataset series, and services. Its primary objective is to make spatial datasets, data series, and services accessible on general data portals, enabling geospatial information to be more easily searched across borders and sectors.

By using a homogeneous data model, datasets can be retrieved more easily (e.g. via SPARQL queries) including by machine-to-machine access.

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

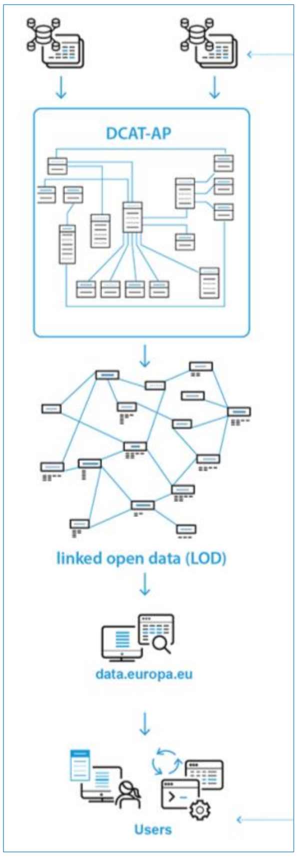

Figure 5: Data linking with DCAT-AP [30]

It was determined that more guidance could be provided for applying the profile in practical situations, such as identification of existing practices or formulation of advice for implementers who need to map their local metadata to DCAT-AP-compliant metadata. Therefore, a DCAT-AP working group of the European Commission developed **implementation guidelines** [35] based on contributions from the implementer community.

As previously stated, DCAT-AP is an essential specification for describing open datasets in Europe, therefore its compliance is mandatory for the BeOpen framework.

3.1.2.1 DCAT-AP annex for High Value Datasets

SEMantic Interoperability Community (SEMIC) Support Centre [36] is responsible for developing solutions that assist European public administrations in performing seamless and meaningful data exchanges across borders and domains. Such exchanges are fundamentally enabled by semantic interoperability. To ensure semantic interoperability solutions are visible and reused, it's crucial to agree on using common semantic standards, promote transparent and well-documented metadata policies, and increase visibility and reuse. SEMIC has been a key player in simplifying the environment where EU countries exchange data for the delivery of electronic public services, while addressing semantic interoperability issues and barriers in the EU.

The Semantic Interoperability Community's top priority is to encourage the use of EU specifications, such as DCAT-AP, at the European, national, and local levels, which will result in an increase in awareness of currently used data standards.

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

SEMIC provides a document containing the guidelines on how to use DCAT-AP for a dataset that is subject to the requirements imposed by the High-Value dataset implementing regulation. The document is called the *"usage guidelines of DCAT-AP for High-Value Datasets"*, in short *"DCAT-AP High Value Datasets"* [37]. Conforming to these guidelines ensures that a dataset within the regulation's scope meets the minimum metadata requirements required for mandatory reporting. However, it does not necessarily mean automated compliance with the regulation because certain aspects are beyond the scope of DCAT-AP. DCAT-AP High Value Datasets is an annex to DCAT-AP that describes additional usage of the DCAT-AP to satisfy the High-Value Dataset implementing regulation. The creation of this annex of DCAT-AP for High Value Datasets, introduced **new properties for DCAT-AP Entities** (such as Agent, Catalogue, Dataset and Distribution). The description and usage of these additional properties is shown in Table 9.

Property	Cardinality	Definition	Usage
applicable legislation [38]	1	The legislation that mandates the creation or management of the Catalogue. It defines the legislation that is applicable to the resource.	For HVD the value must include the ELI [2]. As multiple legislations may apply to the resource the maximum cardinality is not limited.
HVD category [39]	1.. *	A data category defined in the High Value Dataset Implementing Regulation (the six thematic data categories defined are *Geospatial, earth observation and environment, meteorological, statistics, companies and company ownership, mobility*). This property defines the HDV category to which this resource belongs. The creation and maintenance of the code list is the responsibility of the Publications Office. A resource may belong to more than one data category.	For the possible values consult the regulation [2] or consult the controlled vocabulary derived from it.

Table 9: Properties introduced with the creation of the DCAT-AP annex for HVDs

3.1.3 ETSI NGSI-LD API for Context Information Management

NGSI-LD (Next Generation Service Interface with Linked Data) is the API exported by FIWARE Context Brokers and used for the integration of platform components within a "Powered by FIWARE" platform and by applications to update or consume context information. The concept of Linked Data based upon JSON-LD payloads is fundamental to effectively passing data between participating organisations [3].
The NGSI-LD Specification is regularly published and updated by ETSI [40]. At the time, the latest specification is version 1.7.1 which was published in June 2023 [40].

The FIWARE CEF Context Broker [6] is a component compliant with NGSI-LD API developed by FIWARE Foundation [41]. To create Smart Applications in multiple sectors, the FIWARE Foundation aims to develop an open, sustainable ecosystem that focuses on public, royalty-free, and implementation-driven software platform standards. As a consequence, the FIWARE Foundation would suggest a context broker that is in compliance with the ESTI NSGI-LD specifications. The

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

precise context broker to choose depends upon the specific use case; different scenarios may call for a faster context broker, one with a smaller footprint or one which is more secure. NGSI-LD enables linked data (entity relationships), property graphs, and semantics (using JSON-LD [42]'s capabilities). The main constructs of NGSI-LD are **Entity**, **Property** and **Relationship**. NGSI-LD Entities (instances) can be the subject of Properties or Relationships. Properties consist of the combination of an attribute and its value. Linked data can be used to establish associations between instances through relationships. Properties and Relationships can be the subject of other Properties or Relationships. There is no expectation of infinite graphs, and in reality, only one or two levels of property or relationship chaining will take place. The UML representation of the main elements of the NGSI-LD information model and the relationships between them is shown in Figure 6.

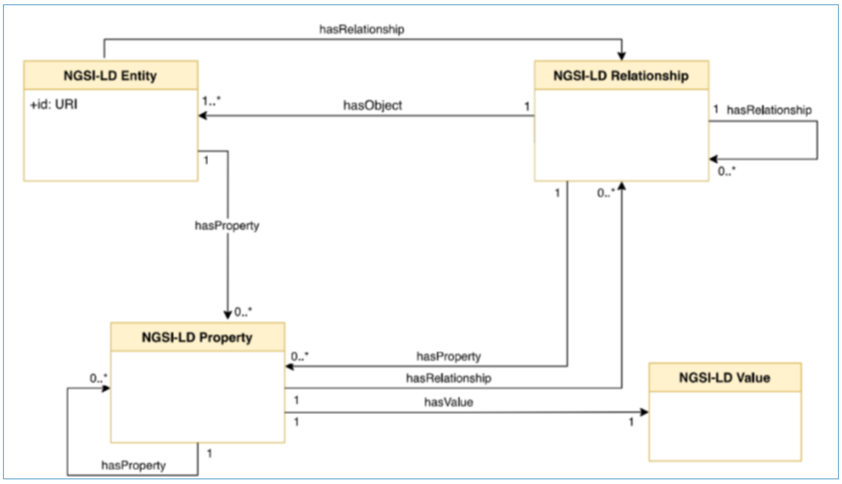

Figure 6: NGSI-LD information model as UML [43]

JSON-LD, developed by the W3C JSON-LD Community Group, is a JSON-based serialization format for Linked Data and is the format used to represent NGSI-LD Entities. It is a W3C Recommendation as of 16 January 2014 [42]. The main advantage of JSON-LD is that it offers the capability of expanding JSON terms to URIs, so that vocabularies can be used to define terms unambiguously.

NGSI-LD is one of key standards to be reused in BeOpen project. Furthermore, many pilot cities have already adopted this standard.

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

3.1.4 INSPIRE - Infrastructure for spatial information in Europe

The European INSPIRE Directive [4] aims to establish a spatial data infrastructure within the European Union to support EU environmental policies and policies or activities that potentially affect the environment. Through this European Spatial Data Infrastructure, public sector organizations can share environmental spatial information, the public can access spatial information across Europe, and policy-making across borders can be facilitated. The infrastructures established and operated by the Member States of the European Union are what INSPIRE is founded on. A total of 34 spatial data themes [44] needed for environmental applications, are addressed by the Directive.

INSPIRE is based on several **common principles**:

- It is important to collect data once and keep it where it can be maintained most effectively.
- The possibility of combining spatial information from different sources across Europe and sharing it with many users and applications should be achievable.
- Sharing information collected at one level or scale with all levels or scales should be feasible, detailed for thorough investigations, general for strategic purposes.
- It is necessary to have geographical information readily and transparently available for good governance at all levels.
- Finding out what geographic information is accessible, how it can be utilized to meet a specific need, and how it can be acquired and utilized are simple tasks.

Common **Implementing Rules (IR)** [45] must be adopted in several specific areas by Member States to ensure that their spatial data infrastructures are compatible and usable in both community and transboundary context as required by the INSPIRE Directive. The specific areas are the following:

- Metadata
- Data Specifications
- Network Services
- Data and Service Sharing
- Spatial Data Services
- Monitoring and Reporting

Adopted as Commission Decisions or Regulations, these Implementing Rules (IR) are binding in their entirety. The Commission received assistance from a regulatory committee made up of representatives of the Member States, headed by a representative of the Commission, in the process of adopting such rules.

In addition to the Implementing Rules, non-binding **Technical Guidance documents** [46] describes the specific implementation aspects and how they are linked to current standards, technologies, and practices. The relationship between the INSPIRE implementation rules and the technical guide is schematically illustrated in Figure 7.

INSPIRE is then a crucial directive for HVDs and therefore the BeOpen framework must be in compliant with it.

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

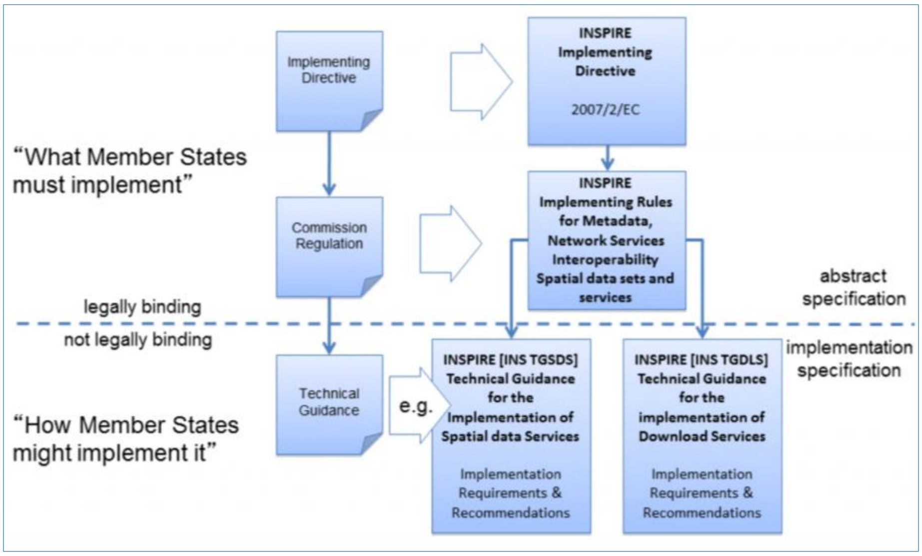

Figure 7: Relationship between INSPIRE implementing Rules and Technical Guidance [46]

3.2 European Key Initiatives

This chapter outlines some European Key Initiatives that appear to be well connected with the objectives and requirements of the BeOpen project. In particular, the following initiatives are described: European Data Spaces, Smart Data models initiative and Minimal Interoperability Mechanisms (MIMs). These initiatives present approaches that will be taken into consideration during the implementation phase of the BeOpen framework.

3.2.1 European Data Spaces

"The European Commission adopted a European strategy for data in 2020 which aims at creating a single market for data that will ensure Europe's global competitiveness and data sovereignty" [47]. The creation of common European data spaces guarantees that more data is available for use in the economy, society, and research, while maintaining control over the companies and individuals who generate it. The Commission endorses the advancement of shared European data spaces in crucial economic areas and areas that are of public concern. According to the Commission Staff Working Document on Common European Data Spaces, "*a common European data space brings together relevant data infrastructures and governance frameworks in order to facilitate data pooling and sharing*" [47].

The use of European data space can be viewed as a system that ensures security and privacy in the process of pooling, accessing, sharing, processing, and using data. A structure that adheres to European rules and values, particularly personal data protection, consumer protection law, and competition law, by accessing and using data in a fair, transparent, proportionate manner and using clear and trustworthy data governance mechanisms.

In the data space, data holders have the option to give access to or share certain personal or non-personal data that they own. To ensure common European data spaces, specific design principles

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

should be followed, including a shared technical infrastructure and building blocks, as well as interconnection and interoperability.

Using data spaces has the objective of speeding up digital transformation in both internal and external sectors and supporting economic recovery plans. The aim is to build a European data space that connects different data sources and ensures that data is shared and utilized extensively while adhering to EU values and regulations in the future. The Common European data spaces initially defined are schematically described in Figure 8

Considering the objectives of the European Data Spaces initiative, the BeOpen project's outcomes can be a significant contribution to the initiative.

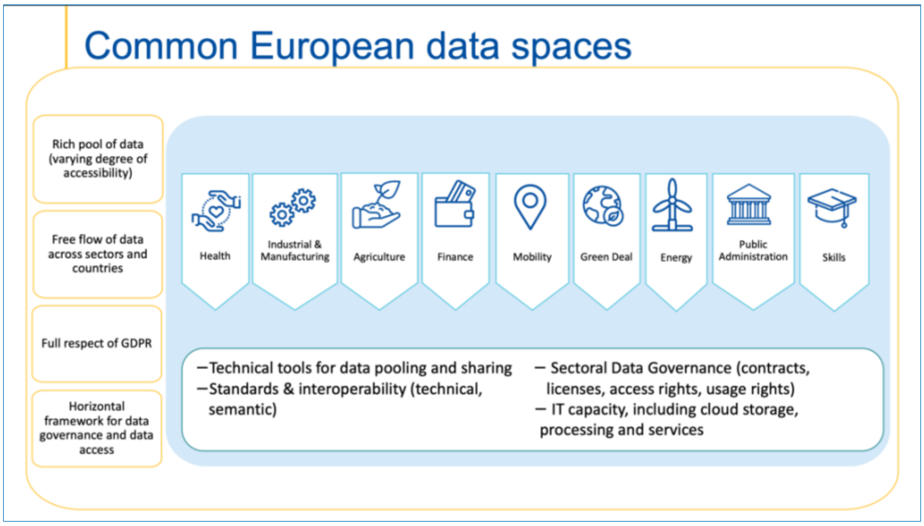

Figure 8: Common European data spaces [48]

3.2.2 Smart Data Models initiative

The FIWARE Foundation [41], TM Forum [49], IUDX [50] and OASC [51] are in charge of a joint collaboration initiative to support the implementation of a reference architecture and compatible common data models that support a digital marketplace of interoperable and replicable smart solutions across several sectors, starting from Smart Cities.

A smart data model includes three elements [52]:

- the **schema**, or technical representation of the model defining the technical data types and structure;
- the **specification** of a written document for human readers;
- the **examples** of the payloads for NGSIv2 and NGSI-LD versions.

All data models are freely accessible and non-proprietary. The licensing mode gives users different rights, including free use, modification, and sharing of modifications. The project is fully published on GitHub [53]; data models are grouped into subjects, and each subject has its own Git repository. The subjects can belong to one or several domains and the domains represent industrial sectors.

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

Data Models can be in one of three stages of their lifecycle [52]:

- **Official**: The data models are accepted and are fully documented and with schema. And with examples available but for specific exceptions.
- **Harmonization**: The data models are accepted but they are in progress to complete some of the mandatory elements.
- **Incubated**: The data models are elaborated by the users and the organization provides support to achieve an official data model.

Examples of domains for which data models are currently available are the following: Smart Cities, Smart Agrifood, Smart Water, Smart Energy, Smart Logistics, Smart Robotics, Smart Sensoring, Cross sector, Smart Health, Smart Destination, Smart Environment, Smart Aeronautics, Smart Manufacturing, Incubated, Harmonization.
The role of data models is crucial as they define the **standardized representation formats and semantics** that applications will use to both consume and publish data. One of the BeOpen project's requirements is to provide capabilities to simplify the access and reuse of common and standard data models for the data description of HVDs. Therefore, the adoption and usage of Smart Data Models can help achieve the goal in order to ensure to use a common data structure for data which is known and well documented and that does not change.

3.2.3 Minimal Interoperability Mechanisms – MIMs

The Minimal Interoperability Mechanisms (MIMs) [54] initiative is being led by OASC [51], with the primary aim of bringing together cities and communities around the world to create a global market for solutions, services, and data that meets the needs of cities and communities. A set of practical capabilities has been developed using open technical specifications to facilitate cities and communities in replicating and scaling solutions globally. Urban data platforms and end-to-end solutions are procured and deployed in cities and communities worldwide with the technical foundation provided by MIMs.

The MIMs that have been formally adopted by all OASC members are schematically described in Figure 9.

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

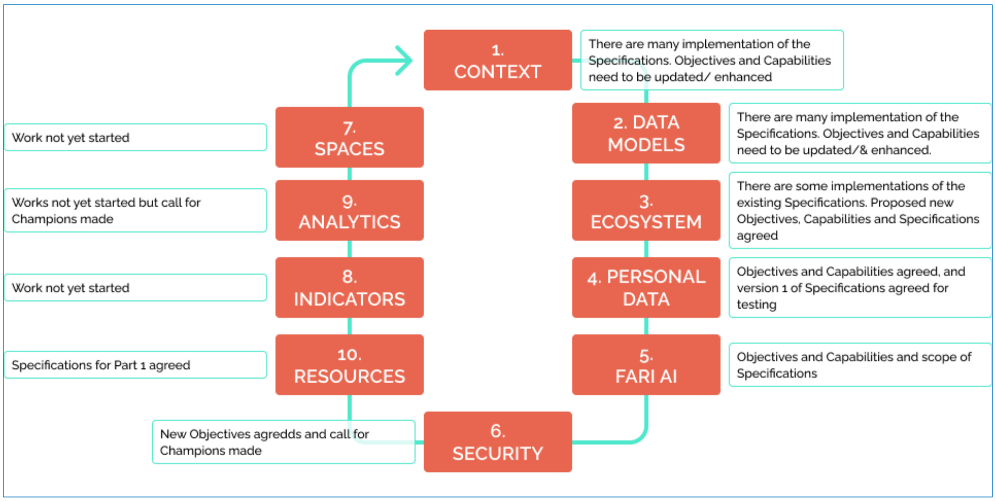

Figure 9: State of play of MIMs in June 2023 [54]

To ensure that data, systems, services, and regulators at various governance levels worldwide are interoperable, MIMs are necessary but not enough capabilities. Cities and communities can achieve interoperability by relying on an inclusive list of baselines and references, which takes into account their different backgrounds and enables them to achieve interoperability with a minimum common ground. No matter the implementation, the same interoperability mechanisms can be employed for critical interoperability points in any given technical architecture. Due to their vendor and underlying technology neutrality, MIMs are a good fit for existing systems and offerings, integrating existing standards.

Considering the objectives of the OASC MIMs, the BeOpen project's outcomes can be a significant contribution to the initiative and at the same time the development of the BeOpen framework will take into account and benefit from the results of the MIMs initiative to promote interoperability between the different European cities adhering to the project.

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

Conclusions

This deliverable 2.1 “BeOpen technical framework design and technical specifications (Final version)” reported a description of technical requirements and the logical architecture of the BeOpen technical framework, together with a set of interoperable standards and norms to be supported. The document focused on the logical description of the different High Value Datasets management phases and related functionalities to be supported by the BeOpen framework and also described a set of existing software tools that will be taken in consideration as baseline for the framework implementation.

This deliverable provided initial information that will be updated and evolved during the next project months: requirements and functionalities are still under revision and in particular the next detailed pilot applications design phase, to be performed during 2024, will be an impact on the framework architecture. Moreover, the possible evolution of standards and related European initiative connected with the BeOpen framework will be followed and eventually implemented as framework changes during the project execution.

This document is the baseline for the ongoing implementation of the framework, that will be released, in its initial version, by the end of 2023.

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

References

[1] GO FAIR, “FAIR Principles,” 2023. [Online]. Available: https://www.go-fair.org/fair-principles/.

[2] European Commission, “COMMISSION IMPLEMENTING REGULATION (EU) 2023/138,” Dicembre 2022. [Online]. Available: https://eur-lex.europa.eu/eli/reg_impl/2023/138/oj.

[3] FIWARE, “NGSI-LD Step-By-Step,” 2023. [Online]. Available: https://ngsi-ld-tutorials.readthedocs.io/en/latest/.

[4] European Commission, “About INSPIRE,” 2023. [Online]. Available: https://inspire.ec.europa.eu/about-inspire/563.

[5] data.europa.eu, “European Data Portal,” 2023. [Online]. Available: https://data.europa.eu/), .

[6] FIWARE, “Orion Context Broker,” 2023. [Online]. Available: https://fiware-orion.readthedocs.io/en/master/.

[7] GNU Operating System, “GNU Affero General Public License,” 2023. [Online]. Available: https://www.gnu.org/licenses/agpl-3.0.en.html.

[8] Engineering Ingegneria Informatica SpA, “Idra - Open Data Federation Platform,” Readthedocs, 2023. [Online]. Available: https://idra.readthedocs.io/en/latest/.

[9] Ontotext, “GraphDB,” Ontotext, 2023. [Online]. Available: https://graphdb.ontotext.com/documentation/10.0/index.html.

[10] OpenStreetMap, “OpenStreetMap,” OpenStreetMap, 2023. [Online]. Available: https://www.openstreetmap.org/#map=6/42.088/12.564.

[11] Open Data Commons, “Open Database License (ODbL),” 2023. [Online]. Available: https://opendatacommons.org/licenses/odbl/.

[12] CKAN, “CKAN User Guide,” 2023. [Online]. Available: https://docs.ckan.org/en/2.9/user-guide.html.

[13] Ubiwhere, “CKAN custom styling,” 2023. [Online]. Available: https://github.com/urbanplatform/ckan-custom-styling/tree/master.

[14] Ubiwhere, “CKAN custom forms,” 2023. [Online]. Available: https://github.com/urbanplatform/ckan-custom-forms/blob/master/README.md.

[15] Engineering Ingegneria Informatica SpA, “Data Model Mapper,” 2023. [Online]. Available: https://github.com/OPSILab/data-model-mapper/tree/frontend_dev.

[16] MinIO, “MinIO Object Storage,” 2023. [Online]. Available: https://min.io/docs/.

[17] MinIO, “MinIO Commercial License,” 2023. [Online]. Available: https://min.io/pricing?jmp=docs.

[18] OBDA systems, “SparQLing,” 2023. [Online]. Available: https://github.com/picorana/sparqling.

[19] D. V. S. D. F. S. a. G. D. G. Lembo, “Graphol: A Graphical Language for Ontology Modeling Equivalent to OWL 2,” 2022.

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

[20] MIT, "Open Sourse Licensing," 2023. [Online]. Available: https://tlo.mit.edu/learn-about-intellectual-property/software-and-open-source-licensing/open-source-licensing.

[21] OpenGovIntelligence, "OLAP browser," 2023. [Online]. Available: https://github.com/OpenGovIntelligence/qb-olap-browser.

[22] OpenGovIntelligence, "JSON-qb API specification," 2023. [Online]. Available: https://github.com/OpenGovIntelligence/json-qb.

[23] Creative Commons, "CC BY 4.0," 2023. [Online]. Available: https://creativecommons.org/licenses/by/4.0/.

[24] OpenGovIntelligence, "Cube Visualizer," 2023. [Online]. Available: https://github.com/OpenGovIntelligence/CubeVisualizer.

[25] JoinUp - European Commission, "High-value datasets – an overview through visualisation," 2023. [Online]. Available: High-value datasets – an overview through visualisation.

[26] GDPR.EU, "What is GDPR, the EU's new data protection law?," 2023. [Online]. Available: https://gdpr.eu/what-is-gdpr/.

[27] Red Hat, "What is a REST API?," 2023. [Online]. Available: https://www.redhat.com/en/topics/api/what-is-a-rest-api.

[28] W3C, "Data Catalog Vocabulary (DCAT)," 2023. [Online]. Available: https://www.w3.org/TR/vocab-dcat/.

[29] European Commission, "DCAT-AP 3.0.0," 2023. [Online]. Available: https://joinup.ec.europa.eu/collection/semic-support-centre/solution/dcat-application-profile-data-portals-europe/release/300.

[30] data.europa.eu, "Linking data: Data Catalogue Vocabulary Application Profile," 2023. [Online]. Available: https://data.europa.eu/en/publications/datastories/linking-data-data-catalogue-vocabulary-application-profile.

[31] W3C, "DCAT Application Profile for Data Portals in Europe," [Online]. Available: https://www.w3.org/2016/11/sdsvoc/SDSVoc16_paper_30. [Accessed 2023].

[32] AGID, "DCAT-AP_IT," 2023. [Online]. Available: https://www.dati.gov.it/content/dcat-ap-it-v10-profilo-italiano-dcat-ap-0.

[33] JoinUp - European Commission, "StatDCAT Application Profile for data portals in Europe," 2023. [Online]. Available: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/statdcat-application-profile-data-portals-europe.

[34] "GeoDCAT Application Profile for data portals in Europe," 2023. [Online]. Available: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/geodcat-application-profile-data-portals-europe/release/101.

[35] European Commission, "DCAT application profile implementation guidelines," 2023. [Online]. Available: https://joinup.ec.europa.eu/collection/semantic-interoperability-community-semic/solution/dcat-application-profile-implementation-guidelines.

[36] JoinUp - European Commission, "About SEMIC Support Centre," 2023. [Online]. Available: https://joinup.ec.europa.eu/collection/semic-support-centre/about.

[37] SEMIC, "DCAT-AP High Value Datasets," 2023. [Online]. Available: https://semiceu.github.io/DCAT-AP/releases/2.2.0-hvd/.

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

[38] SEMIC, "Applicable Legislation Property," 2023. [Online]. Available: https://semiceu.github.io/uri.semic.eu-generated/r5r/release/3.0.0/#applicableLegislation.

[39] SEMIC, "HVD category Property," 2023. [Online]. Available: https://semiceu.github.io/uri.semic.eu-generated/r5r/release/3.0.0/#hvdCategory.

[40] ETSI, "ETSI GS CIM 009 V1.7.1 (2023-06)," 2023. [Online]. Available: https://www.etsi.org/deliver/etsi_gs/CIM/001_099/009/01.07.01_60/gs_cim009v010701p.pdf.

[41] FIWARE, "FIWARE foundation," 2023. [Online]. Available: https://fiware.org/.

[42] json-ld, "JSON for Linking Data," 2023. [Online]. Available: https://json-ld.org/.

[43] FIWARE, "Data Modeling with NGSI, NGSI-LD," 2023. [Online]. Available: https://www.slideshare.net/flopezaguilar/data-modeling-with-ngsi-ngsild.

[44] inspire.ec.europa.eu, "INSPIRE Data Themes," 2023. [Online]. Available: https://inspire.ec.europa.eu/Themes/Data-Specifications/2892.

[45] inspire.ec.europa.eu, "INSPIRE Implementing Rules," 2023. [Online]. Available: https://inspire.ec.europa.eu/inspire-implementing-rules/51763.

[46] inspire.ec.europa.eu, "INSPIRE Technical Guidance," 2023. [Online]. Available: https://inspire.ec.europa.eu/inspire-technical-guidance/57753.

[47] JoinUp - European Commission, "Data Spaces," 2023. [Online]. Available: https://joinup.ec.europa.eu/collection/semic-support-centre/data-spaces.

[48] "Common European data spaces," 2023. [Online]. Available: https://dataspaces.info/common-european-data-spaces/#page-content.

[49] "TM Forum," 2023. [Online]. Available: https://www.tmforum.org/.

[50] IUDX, "India Urban Data Exchange," 2023. [Online]. Available: https://iudx.org.in/.

[51] OASC, "Open & Agile Smart Cities," 2023. [Online]. Available: https://oascities.org/.

[52] FIWARE, "Smart Data Models," 2023. [Online]. Available: https://www.fiware.org/smart-data-models/.

[53] GitHub, "Smart Data Models - repository," 2023. [Online]. Available: https://github.com/smart-data-models.

[54] OASC, "Minimal Interoperability Mechanisms – MIMs," 2023. [Online]. Available: https://oascities.org/minimal-interoperability-mechanisms/.

[55] European Commission, "INSPIRE Technical Guidelines," 2023. [Online]. Available: https://inspire.ec.europa.eu/Technical-guidelines3.

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