Open framework for boosting EU High Value Datasets from Public Sector

beopen-dep.eu

Project co-funded by the European Union under the Digital Europe Programme. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

D2.3 BeOpen Trust Framework Report

Document History

Ver.	Date	Description	Author	Partner
0.1	11/10/2023	Draft Table of Contents	Anna Ida Hudig	ARTHUR
0.2	18/10/2023	Initial Table of Contents	Anna Ida Hudig	ARTHUR
0.2	06/11/2023	Discussion ToC with coordinator and WP2 partners	Anna Ida Hudig	ARTHUR
0.2	07/11/2023	Sharing request for input for chapter 4 of the deliverable with Use Case leaders	Anna Ida Hudig	ARTHUR
0.3	14/11/2023	Finalising chapter 2	Anna Ida Hudig, Arthur van der Wees	ARTHUR
0.4	17/11/2023	Finalising chapter 3	Anna Ida Hudig , Arthur van der Wees	ARTHUR
0.5	22/11/2023	Finalising chapter 4	Anna Ida Hudig, Arthur van der Wees	ARTHUR
0.6	27/11/2023	Integration of input by Use Case Leaders	Anna Ida Hudig, Arthur van der Wees	ARTHUR
0.7	29/11/2023	Updating chapter 1, 4 and 5	Anna Ida Hudig, Arthur van der Wees	ARTHUR
0.8	07/12/2023	Further updates chapter 4 and 5, including integration of input by HERNE	Arthur van der Wees, Celine Prins	ARTHUR
0.9	13/12/2023	Refinements across the document, submission for internal review	Celine Prins, Dimitra Stefanatou	ARTHUR
1.0	21/12/2023	Addressed reviewers' comments and added clarifications	Giacomo Delinavelli	ARTHUR

Disclaimer

The sole responsibility for the content of this publication lies with the authors. It does not necessarily reflect the opinion of the European Union. The European Commission is not responsible for any use that may be made of the information contained therein.

Co-financed by the Connecting Europe Facility of the European Union

Publication details
Grant Agreement Number 101069831
Acronym BeOpen

Full Title	BeOpen
Topic	Open framework for boosting EU High Value Datasets from Public Sector
Funding scheme	Horizon Europe
Start Date	01/01/2023
Duration	30 Months
Project URL	https://beopen-dep.eu
Project Coordinator	ENGINEERING
Deliverable	D2.3: BeOpen trust framework report
Work Package	WP2
Delivery Month (DoA)	M12
Version	1.0
Actual Delivery Date	23/12/2023
Nature	Report
Dissemination Level	PU – Public
Lead Beneficiary	ARTHUR
Authors	Anna Ida Hudig, Arthur van der Wees, Celine Prins
Quality Reviewer(s)	ENG, DPOW
Keywords	Data sharing, trust components, Digital Decade 2030, Open Data, High Value Datasets, public sector, availability, accessibility, quality, reliability, interoperability, licensing and reuse, transparency, accountability

Co-financed by the Connecting Europe Facility of the European Union

D2.3 BeOpen Trust Framework report

Table of Contents

List of Figures

List of Tables

Abbreviations and Acronyms

ACRONYM	Definition
AI	Artificial Intelligence
API	Application Programming Interfaces
EC	European Commission
EAB	Ethics Advisory Board
EU	European Union
DCAT	Data Catalogue Vocabulary
DCAT-AP	DCAT Application Profile for data portals in Europe
DEM	Data and Ethics Manager
DPO	Data Protection Officer
DPIA	Data Protection Impact Assessment
DIGITAL	Digital Europe Programme
CA	Consortium Agreement
CEF	Connecting Europe Facility
GA	Grant Agreement
GDPR	General Data Protection Regulation
HVD	High Value Dataset
IP	Intellectual Property
IPR	Intellectual Property Rights
M	Month
NIS2	The Directive on measures for a high common level of cybersecurity across the Union
ODD	Open Data Directive
WP	Work Package

Executive Summary

This Deliverable provides a “Trust Framework” aiming to increase the actual availability of the data held by the public administration, which are subject to the EU Open Data Directive and the High-value Datasets Implementing Regulation. To this end, the Deliverable takes into account the societal challenges pertaining to the creation of open data ecosystems, as well as the most relevant for the scope of BeOpen project European policies, regulations and best practices. The Deliverable provides a high-level overview of the requirements concerning data sharing, which include open data obligations for public entities, while, also, considering personal data protection, IP and licensing, as well as cybersecurity aspects. Both public policy ambitions and legal requirements are intended as critical aspects of any data sharing ecosystem, hence the policies and obligations are outlined throughout Chapter 2.

The Deliverable identifies certain key components, namely, accessibility, licensing, quality, control, interoperability, transparency and accountability aiming to contribute to the creation of a trust framework allowing for trustworthy data sharing. Each one of these components is described at high-level, as they all should be adapted to the specific Use Case. Based on the information made available by the Use Case leaders upon request, Chapter 4 discusses the BeOpen trust components identified in the context of each BeOpen Use Case. Furthermore, Chapter 5 produces a set of usable Guidelines and Best Practises aiming to support pilot leaders and other relevant stakeholders in undertaking the most appropriate steps towards making their data available in a trustworthy fashion.

Finally, the Deliverable points at specific considerations relating to the necessary steps for the creation of a trustworthy data sharing ecosystem, which include, among other, legal and regulatory aspects, skills development, and other functional and non-functional requirements, such as interoperability, transparency and accountability.

1 Introduction

This chapter elaborates on the current societal challenges society how open data policies and high value datasets in EU may play a role in solving such challenges. Moreover, the scope and the aims of the present deliverable within the context of BeOpen project activities are oulined.

1.1 Data & Societal Challenges

The 21st century presents a broad spectrum of societal challenges, such as climate change, loss of biodiversity, mobility, resilience, and demographic shifts. Public sector data can play an instrumental role in gaining a better understanding of these challenges and in developing effective strategies to address them. In the BeOpen Use Cases, European cities and municipalities aim to address certain societal challenges by making data publicly available. As such, the underlying *aim* is addressing societal challenges and digital data are merely a *means* to this end.

Data is the common denominator across all activities within the BeOpen project. Data sharing is an integral part of BeOpen activities; information is not only likely to be shared *across* Use Cases, but also *within* each Use Case. For instance, different departments within the same organisation, such as a offices with different roles and responsibilities within a municipality or a city council, need to proactively adopt coordinate practises in order to ultimately make data available to the public. In this context, citizens and society, more broadly, may benefit in several ways through access to data, information and knowledge. The need for information sharing has, thus, increased and making the information trustworthy and available appears, hence, to be a prerequisite.

In this context, a Trust Framework, based on current legislation and ethical principles, would facilitate the trustworthy adoption of digital technologies in general and, in this case, of the BeOpen platform, thus to increase access and interoperability of datasets.

1.2 Setting the scene

The mission of BeOpen is to provide a comprehensive framework supporting Open Data and metadata life cycle management pipelines in the public sector. This framework is intended to facilitate accessing, curating, and publishing high-value datasets (HVDs) to be made available for future Data Spaces supporting the sustainable city domain. A central piece of legislation within BeOpen is the Open Data Directive[1], which defines Open Data as follows:

> *'Open data as a concept is generally understood to denote data in an open format that can be freely used, re-used and shared by anyone for any purpose.'*

From a technical and organisational point of view, the BeOpen strategy involves creating open-source tools, replicable processes, ontologies, and best practices. These are designed to carry out tasks such as collecting and curating data, adding semantic annotations, harmonising data and metadata, improving data quality, and

[1] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast)): https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32019L1024

publishing High-Value Datasets (HVDs). These datasets will be published in machine-readable formats, either in bulk or through APIs. With regards to HVDs, the European Commission (2022) states that:

> *'(…), high-value datasets are characterised by specific technical and legal requirements. The open data licence, the availability of public documentation and ensuring machine readability are all requirements applicable to these datasets. Moreover, high-value datasets are required to be downloadable in bulk (where relevant) and through APIs, free of charge, while also providing extensive documentation for their metadata.'*[2]

Figure 1 shows the features of HVDs. Specifically, this diagram shows that HVDs extend multiple domains and affect different stakeholder groups, including public services and public administration, the economy, climate, social groups etc. It also shows the mandatory technical requirements which are shown in the outer layer in red (Figure 1). The BeOpen Use Cases specifically focus on HVDs related to statistics, mobility, environmental data, earth observation, and geo-spatial information. However, public bodies might face challenges in collecting the appropriate data. One of such challenges is determining what type of data is required to address a certain problem. Furthermore, data sharing barriers can arise, for example when accessing data from local or national authorities. The *BeOpen Trust Framework* is intended to serve as a guiding instrument for public bodies that need to work with open data yet are facing uncertainties such as where to start, what considerations to take into account, and how to ensure data sharing occurs in a trusted, ethical, and interoperable manner.

[2] COMMISSION IMPLEMENTING REGULATION (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use. Available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2023.019.01.0043.01.ENG

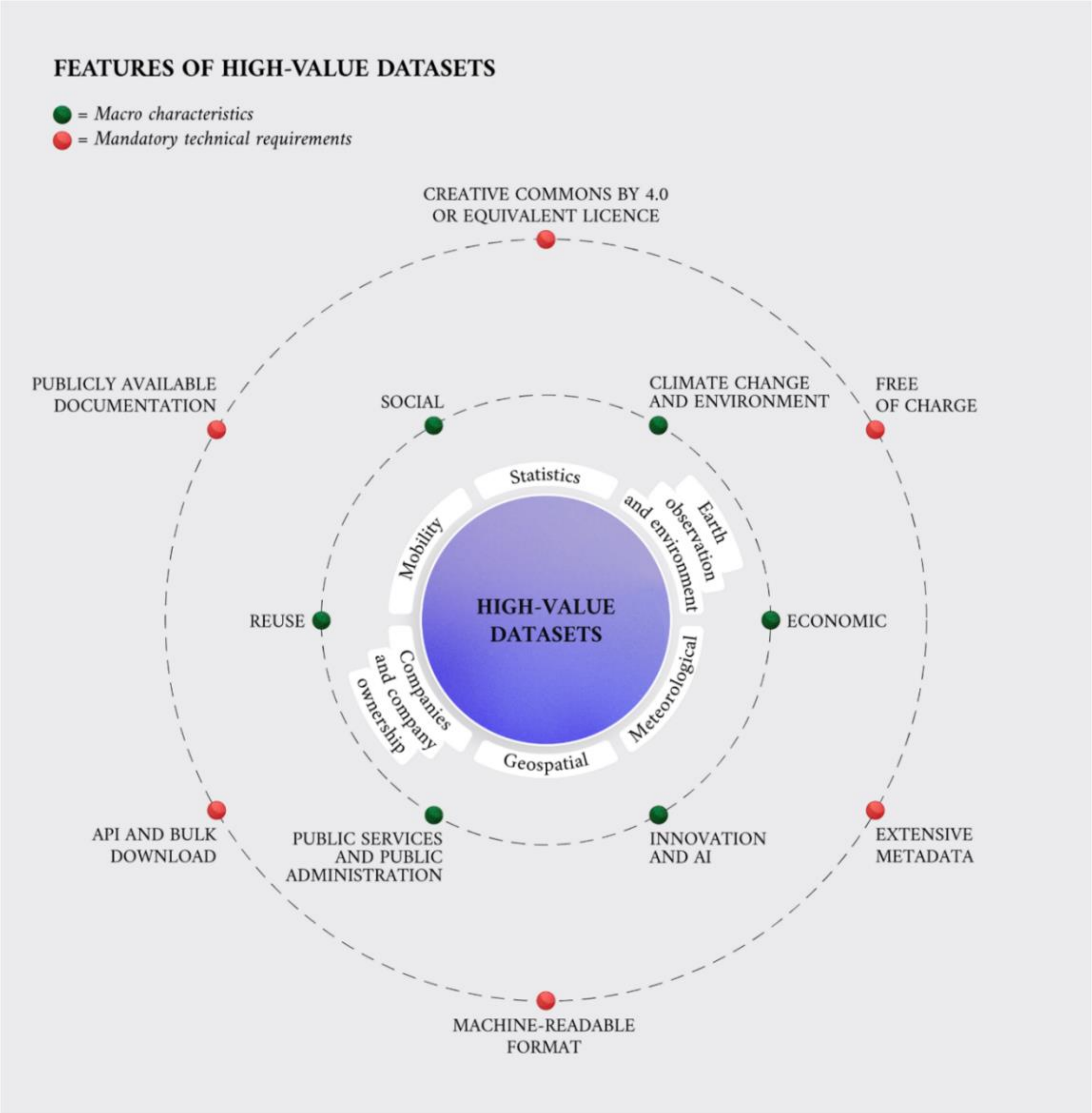
FEATURES OF HIGH-VALUE DATASETS
🟢 = *Macro characteristics*
🔴 = *Mandatory technical requirements*
CREATIVE COMMONS BY 4.0 OR EQUIVALENT LICENCE
PUBLICLY AVAILABLE DOCUMENTATION
FREE OF CHARGE
SOCIAL
CLIMATE CHANGE AND ENVIRONMENT
Statistics
Mobility
Earth observation and environment
REUSE
HIGH-VALUE DATASETS
ECONOMIC
Companies and company ownership
Meteorological
Geospatial
API AND BULK DOWNLOAD
PUBLIC SERVICES AND PUBLIC ADMINISTRATION
INNOVATION AND AI
EXTENSIVE METADATA
MACHINE-READABLE FORMAT

Figure 1: Diagram of features of high-value datasets (HVDs)[3]

1.3 Scope

The BeOpen project encourages active participation from citizens, public, and private organizations by facilitating the sharing of non-personal, reliable data. It empowers individuals and organisations to make informed decisions and exercise their rights through access to public open data. This enhances scrutiny of public administrations and fosters the development of valuable public and private services, like improving mobility or optimising energy systems. Additionally, the project provides best practices, replicable scenarios,

[3] Visual Diagram of Features of High-Value Datasets (HVDs) available at https://gitlab.com/Giuseppeascone/data-provider-repository/-/blob/master/Data%20stories/1.0_HVD_overview.png

lessons learned, and recommendations to assist public bodies in adopting the BeOpen framework and methodologies.

Work Package 2 (WP2) serves as the primary technical work package dedicated to formulating and integrating the BeOpen technical and organisational framework, as outlined in D2.2, and addressing interoperability. The key goals of WP2 involve designing the technical and organisational structure for High-Value Datasets (HVDs), providing the IT components to implement this framework, and ensuring their integration.

This document and the orchestration of the related works pursues its objectives by building ongoing dialogue, both structured and informal, among participants across different Member States working on eight (8) pilots active in eight (8) different regions, namely, Germany, Italy, Spain, Lithuania, Greece, Portugal in the public environment. Deliverable D2.3 is part of Task 2.4 BeOpen trust framework (interoperability, ethical, & other trust components). It builds on to D1.5 (Data Protection & Ethics) and D1.3 (the Data Management Plan), and is, as such, a continuation of these. The discussion below expands on trust components, producing recommendations such as guidelines and good practices. Notably, the present document paves the way for Deliverable D5.6 on Sustainability Plan and Business Model due in M30.

1.4 Methodology

The methodology involves both desk research and the distribution of questionnaires among Use Case leaders. The desk research specifically concentrated on identifying and discussing the most relevant policy frameworks in EU for the scope of BeOpen project. Based on this discussion, a set of trust components was identified. Building on top of the latter, a questionnaire was developed (Appendix 1). This questionnaire was made available on the project repository and, in that format, further communicated to the Use Case leaders, who were queried about the applicability and implementation of the identified trust components accordingly.

More specially, the said questions were formulated as follows:

1. Did you consider this trust component in your piloting activities? *(yes/no/not applicable)*
2. Please specify the aspects or activities, and the specific functionalities of the Use Case where the implementation of this trust component may be relevant.
3. If any, please provide a short description of any practices, approaches or efforts employed to implement this trust component.
4. If any, please provide a short description of any challenges you experienced or foresee in the consideration & implementation of this trust component.
5. Other remarks (e.g. if you gained any learnings or takeaways of your relevant efforts so far).

Following the feedback received from the Use Case leaders, a set of guidelines and recommendations was put forward.

Notably, the present deliverable was the outcome of an interdisciplinary collaboration; particular emphasis has been put on the accessibility of the content by non-experts.

1.5 Target Audience

In thebigger scheme of things, the present document is relevant for both the public sector and private sector, as well for NGOs, academia, organisations and, last, but not least, individuals that are interested in these critical matters. However, considering the scope and objectives of BeOpen project, the main focus is on public sector organisations as defined in Art. 2 ‘Open Data Directive’, namely, public sector bodies, bodies governed by public law, public undertakings and universities. In particular, among the public sector organisations, a specific reference is made to the BeOpen pilot partners and other municipalities and authorities, such as local authorities, regional authorities and national authorities.

1.6 Structure

This deliverable builds on the following structure. chapter 1 provides an introduction and elaborates on the context of BeOpen project, as well as on the scope of the present deliverable. Chapter 2 provides an overview of the most relevant legal and policy frameworks that provide the basis for BeOpen Trust Components to be discussed under chapter 3. Following the identification of those components and building on input collected directly from Use Case Leaders, chapter 4 expands on how the above-mentioned components are reflected within BeOpen Use Case Activities. Finally, Chapter 5 outlines a set of guidelines and lessons learnt pertaining to the scope of this deliverable by end of Year 1 of the project, while chapter 6 provides for the Concluding Remarks. The input request addressed to all BeOpen Use Case Leaders can be found under the Appendices, while the respective input received is available on the project's repository.

2 Trust Frameworks

BeOpen aims to establish a trust framework that will set guidelines to increase the availability, quality and usability of HVDs held by the public sector actors, in compliance with the legal requirements set out in the relevant European legislation and provide public and private sector actors, including ordinary citizens, data owners, data users or developers who might be interested to produce other open HVDs, with a trustworthy, robust, and citizen-centric data-sharing ecosystem.

This chapter presents an overview of the reference trust framework developed for the BeOpen project. It does so by analysing the relevant European policy, ethical and legislative frameworks and by identifying the most common requirements, standards and conditions from the relevant policy and legislative documents, whilst recognising the primarily B2G nature of the data sharing activities to be facilitated in the context of BeOpen pilot activities.

2.1 Digital Decade & Policy Frameworks

Section 2.1 provides an overview of relevant policy frameworks developed by the European Commission to make the EU '*Fit for a Digital Age*', namely, the Declaration on Digital Rights and Principles, the EU Digital Decade 2030, the European Data Strategy and the Digital Europe Programme. Policies are very important tool for the European community as certain core values and principles, including freedom, equality, the rule of law, justice, and non-discrimination forming the basis of a number of key European legislation are initially recognised in the policies.[4]

2.1.1 Declaration on Digital Rights and Principles for the Digital Decade

The European Declaration on Digital Rights and Principles[5], published in December 2022, is often referred to as the EU's digital DNA, serves as a **noteworthy and easily digestible guide for ethical and accountable behaviour in the digital decade**. This Declaration, rooted in key EU values and freedoms, provides guidance for individuals and organisations navigating the ethical and accountability dimensions of technology. The Declaration focuses on six critical areas: prioritising individuals in the digital transformation, fostering solidarity and inclusion, upholding freedom of choice, encouraging participation in digital life, ensuring safety and security, and promoting sustainability.

As articulated by the Executive Vice-President for a Europe Fit for the Digital Age, the essence of the digital transformation lies in ensuring that technologies are not only safe but also aligned with our interests while respecting our rights and values. The principles enshrined in the Declaration will persistently be supported by new EU legislations.

[4] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) (Text with EEA relevance) available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32022R0868

[5] European Declaration on Digital Rights and Principles for the Digital Decade COM/2022/28 final available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2022:28:FIN

The Declaration also draws from established legal document, including general data protection regulation, ePrivacy directive, and the relevant case law of the Court of Justice of European Union. It further complements the European Pillar of Social Rights, embodying a people-centric design by default. The Declaration empowers not only the European public sector, encompassing municipalities, regions, and national entities, but also individuals and society at large. In the context of the European data society and economy, it stands as an essential framework.

2.1.2 EU Digital Decade 2030

The 2030 digital decade policy programme in the European Union (EU)[6] has the aim of promoting innovation and investment. It sets **clear digital transformation targets for 2030 and enhances cooperation between EU institutions and Member States.** The program aims for an inclusive, open, and secure digital environment, bridging the digital divide, and fostering digital skills in the Union. It emphasises the importance of secure digital infrastructure, online democratic participation, sustainability, and resilience to cyberattacks. Key digital targets for 2030 include creating basic digital skills for 80% of the population, promoting ICT specialists and graduates, widespread high-speed networks, quantum computing capabilities, and digital service accessibility. The Commission and Member States work together to define routes to achieve these targets. The EU Digital Decade has been in effect since January 8, 2023.[7]

The EU Digital Decade underscores the European Union's dedication to upholding fundamental rights, the rule of law, and democracy, including demonstrating a commitment to safeguarding citizens' privacy and data rights. The focus on improving the quality of life and service availability suggests an emphasis on data accessibility, while the commitment to digital sovereignty implies increased control over the EU's digital infrastructure. In addressing cybersecurity, the document highlights the importance of security, resilience, and protecting critical technologies to safeguard digital assets and infrastructure. It is also important to noted that the notions of transparency and accountability are emphasised through references to ‘transparent partnerships’ and ‘strengthening the Union’s digital sovereignty in an open manner’. Furthermore, the importance of digital skills and education is highlighted, with recognition of the need for an informed and skilled population to address privacy concerns and understand digital rights.

[6] Decision (EU) 2022/2481 of the European Parliament and of the Council of 14 December 2022 establishing the Digital Decade Policy Programme 2030 (Text with EEA relevance) available at: https://eur-lex.europa.eu/eli/dec/2022/2481/oj .

[7] COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS 2030 Digital Compass: the European way for the Digital Decade COM/2021/118 final, available at: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:52021DC0118
.

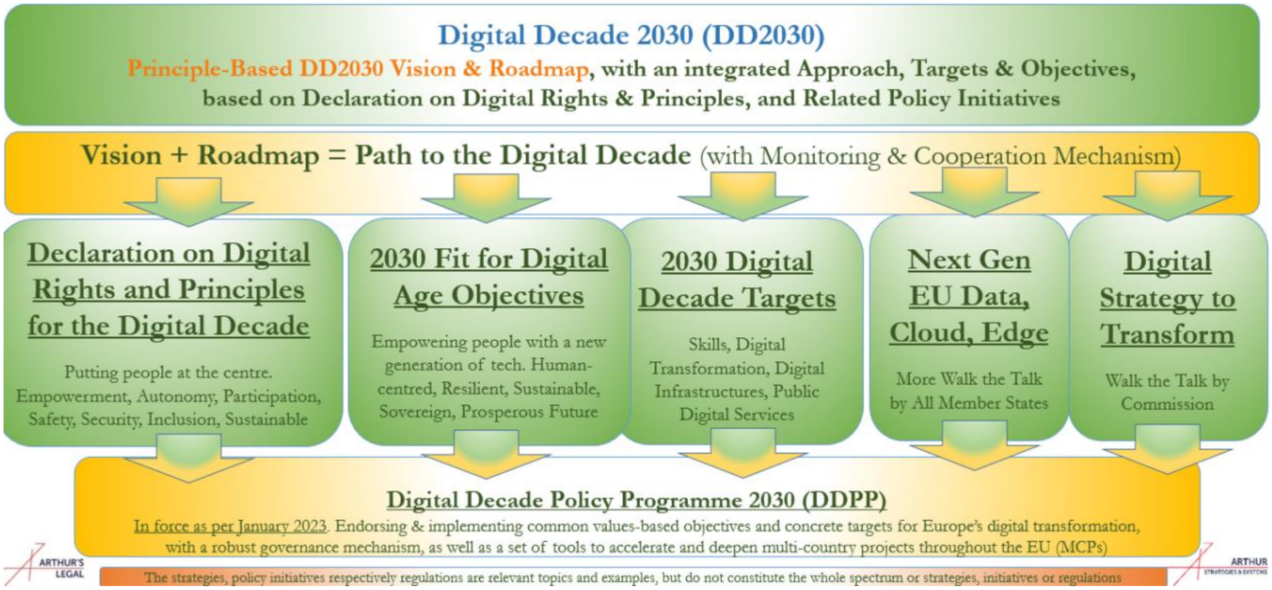
Digital Decade 2030 (DD2030)
Principle-Based DD2030 Vision & Roadmap, with an integrated Approach, Targets & Objectives, based on Declaration on Digital Rights & Principles, and Related Policy Initiatives
Vision + Roadmap = Path to the Digital Decade (with Monitoring & Cooperation Mechanism)
Declaration on Digital Rights and Principles for the Digital Decade
Putting people at the centre. Empowerment, Autonomy, Participation, Safety, Security, Inclusion, Sustainable
2030 Fit for Digital Age Objectives
Empowering people with a new generation of tech. Human-centred, Resilient, Sustainable, Sovereign, Prosperous Future
2030 Digital Decade Targets
Skills, Digital Transformation, Digital Infrastructures, Public Digital Services
Next Gen EU Data, Cloud, Edge
More Walk the Talk by All Member States
Digital Strategy to Transform
Walk the Talk by Commission
Digital Decade Policy Programme 2030 (DDPP)
In force as per January 2023. Endorsing & implementing common values-based objectives and concrete targets for Europe's digital transformation, with a robust governance mechanism, as well as a set of tools to accelerate and deepen multi-country projects throughout the EU (MCPs)
ARTHUR'S LEGAL
ARTHUR STRATEGIES & SYSTEMS
The strategies, policy initiatives respectively regulations are relevant topics and examples, but do not constitute the whole spectrum or strategies, initiatives or regulations

Figure 2: Europe's Digital Decade

2.1.3 A European Data Strategy

The European Commission's strategy for data[8] proposes an **approach for an open, diverse, and democratic digital European environment, with a strong focus on Europe's competitive position in the global data economy**. It highlights the importance of data in the digital era, as it facilitates data-driven innovation, protection of individual interests, and accessibility of non-personal data. The strategy aims to promotes a genuine single market for data by addressing several challenges and building on the following four pillars:

1. **Cross-Sectoral Governance Framework**: this framework aims to create a structure for the data economy whilst considering sector-specific needs. It emphasises enabling legislation for the governance of common European data spaces to support cross-border data use, interoperability, and standards.
2. **Opening High-Quality Public Sector Data**: the strategy plans to make public sector data sets available for innovation, particularly for SMEs. This involves adopting implementing acts on HVDs under the Open Data Directive to make these datasets available for free of charge, in machine-readable formats with standardized APIs.
3. **Legislative Action on Data Sharing**: the Data Strategy enabled the drafting of the Data Act, addressing issues in context of business-to-government and business-to-business data sharing including defining access conditions, and rewriting the intellectual property rights framework.

[8] COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS A European strategy for data COM/2020/66 final available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0066

4. **Compliance and Enforcement**: the Commission will provide guidance on the compliance of data sharing and pooling arrangements with EU competition law. It will also examine state aid guidelines to minimize competition distortions through data-sharing requirements for beneficiaries.

In addition, the Commission aims to improve its own data management and to make data derived from its research programs more accessible to the public. The strategy aims to lead by example and foster an open and data-driven European digital ecosystem.

2.1.4 Digital Europe Programme

To operationalise the objectives set out in the EU Data Strategy, The Digital Europe Programme[9] provides a **funding aimed at bringing digital technology to businesses, citizens, and public bodies**. Some of its key objectives are to enhance the EU's competitiveness in the global digital economy, reduce the digital divide within the EU, and strengthen Europe's capabilities in key digital technology areas. The program runs from 2021 to 2027 with a budget of over €7.5 billion. It focuses on topics such as high-performance computing, artificial intelligence, cybersecurity, advanced digital skills, and the deployment of digital capacity. Additionally, it promotes collaboration between various stakeholders and is open to the participation of non-EU countries. The programme has started on 1 January 2021.

The Digital Europe Programme encompasses various work programmes outlining the objectives, scope, outcomes, and budget allocations for specific topics. The current work programmes include the Digital Europe Work Programme for 2023-2024, European Digital Innovation Hubs Work Programme (2021-2023), the Cybersecurity Work Programme (2023-2024), and High-Performance Computing (extended to 2023).

The Digital Europe Work Programme[10] considers cybersecurity as a crucial component of the digital transformation within the European Union and aims at strengthening the Union's cybersecurity capabilities to protect its citizens and organisations. As the availability and quality of data relies on a high level of cybersecurity, this program aims to enhance the security of digital products and services throughout the entire supply chain. In particular, the activities funded in 2023 and 2024 are directed at:

- Supporting the deployment of cybersecurity infrastructure.
- Strengthening cybersecurity adoption.
- Implementing relevant EU legislation and political initiatives.

These cybersecurity activities are specified in a dedicated work plan for 2023-2024, which will be executed by the Cybersecurity Industrial, Technology, and Research Competence Centre (ECCC) and the Network of National Coordination Centres (NCCs).

[9] Regulation (EU) 2021/694 of the European Parliament and of the Council of 29 April 2021 establishing the Digital Europe Programme and repealing Decision (EU) 2015/2240 (Text with EEA relevance) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32021R0694

[10] Digital Europe Cybersecurity Work Programme 2023-2022

2.2 Other Relevant Regulatory European Frameworks

This Section 2.2 discusses a range of other influential frameworks set under the key European legislation highly relevant to BeOpen project including the Open Data Directive, the Data Act, the Data Governance Act, the INSPIRE Directive, the AI Act, the GDPR, the NIS2 Directive, and the Cyber Resilience Act.

2.2.1 Directive on open data and the re-use of public sector information (Open Data Directive)

Directive (EU) 2019/1024, also known as the "Open Data Directive"[11] is at the heart of the BeOpen activities . This Directive aims to establish a policy framework for the reuse of public-sector information, including geographical, land registry, statistical, and legal data, as well as research data coming from publicly funded research. The Open Data Directive enables and facilitates the availability of public sector information for re-use with minimal legal restrictions and free of charge and in machine-readable data. It also emphasises fair trading and non-discrimination of open data, and release of HVDs. The Open Data Directive does not apply to documents containing third-party intellectual property rights, sensitive information, or created as results of activities unrelated to the public task of a public-sector body. The directive is part of a broader effort to strengthen the European Union's data economy.

The Open Data Directive touches upon different trust components.[12] Primarily, the Directive encourages the availability and accessibility of data including promoting suitable APIs for sharing dynamic data, ensuring easy and immediate access. Furthermore, the Directive states that public sector bodies should retain the rights to exploit documents and to make them available for re-use by others, indicating a level of control that the public sector bodies have over the re-use of these documents. In addition, the Directive encourages the use of open licenses providing minimum restrictions on re-use of data. Additionally, the Directive incorporates transparency, accountability, and the protection of personal data principles, which are important elements in building trust in data management and access.

2.2.2High-Value Datasets Implementing Regulation

Stemming form the EU Open Data Directive, the High-value dataset implementing regulation issued by the European Commission sets out a “list of datasets with a particular potential to generate socio-economic benefits and with harmonised re-use conditions is a significant enabler of cross-border data applications and services.” 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.

Annex I to Directive (EU) 2019/1024 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; and 6) mobility. the Commission identified, within each of the six data categories, several datasets of particularly high value and the arrangements for their publication and re-use. The provisions of Union and Member State legislation that go beyond the minimum requirements set out in this Implementing Regulation, in particular in cases of sectoral law, are to continue to apply.

[11] See FN 1.

[12] Trust components are explained in more details in Chapter 3 (‘BeOpen Trust Components’).

A few limitations to the requirement of making high-value datasets available free of charge are also established: (1) libraries, including university libraries, museums and archives are exempted; (2) data held by public undertakings are not included in the scope of this Implementing Regulation; (3) when high-value datasets include personal data, the safeguards – and limitations – established in the GDPR apply.

The Commission's Guidelines on recommended standard licences, datasets and charging for the re-use of documents[13] identify Creative Commons ('CC') licences as an example of recommended standard public licences. CC licences are developed by a non-profit organisation and have become a leading licensing solution for public sector information, research results and cultural domain material across the world. It is therefore necessary to refer in this Implementing Regulation to the most recent version of the CC licence suite, namely CC 4.0. A licence equivalent to the CC licence suite may include additional arrangements, such as the obligation on the re-user to include updates provided by the data holder and to specify when the data were last updated, as long as they do not restrict the possibilities for re-using the data.

This Regulation entered into force in February 2023 and is binding in its entirety and directly applicable in all Member States. It will be applicable from June 2024.

2.2.3 The Proposal for Data Act

As mentioned, the proposal for Data Act,[14] published by the Commission in February 2022, and recently agreed upon the EU co-legislators,[15] originates from the European Data Strategy. This proposal complements the Data Governance Regulation (see 2.2.3) and aligns with the digital transformation goals of the Digital Decade. The proposal for Data Act is expected to stimulate innovation, create jobs, and provide individuals with greater control over their data by clarifying the conditions for the use and control over data generated by the use of connected products. It empowers users of connected products to easily transfer their data, enhancing control for natural and legal person users by stipulating requirements to level the playing field in the digital environment and improving user access to their data.

The proposal is expected to foster competition, lower repair costs, and contribute to environmental goals by extending product lifespans. It also facilitates data-driven optimisation in certain industries. As such, it addresses several key areas:

a. Increasing legal certainty for data generating companies and consumers;
b. Preventing abuse of contractual imbalances to enhance fair data sharing;
c. Promoting accessibility and use of data by public sector bodies;
d. Allowing consumers to switch between different data processing services.

[13] COMMISSION NOTICE Guidelines on recommended standard licences, datasets and charging for the reuse of documents 2014/C 240/01.

[14] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on harmonised rules on fair access to and use of data (Data Act) COM/2022/68 final https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2022:68:FIN

[15] Data Act: Commission welcomes political agreement on rules for a fair and innovative data economy.

2.2.4 The Regulation on European Data Governance (Data Governance Act)

The Regulation on European data governance[16] ("Data Governance Act") is introduced to facilitate data sharing and reuse, promoting innovation and investment in the EU. Similar to the proposal for Data Act, it stems from the European Data Strategy. The Data Governance Act sets conditions for reusing protected data, regulates data intermediation services, encourages data altruism, and ensures the secure flow of non-personal data outside the EU, with a focus on transparency, neutrality, and accessibility. The regulation includes targets for digital skills, data infrastructure, and data sharing by 2030. It became effective on September 24, 2023.

The Data Governance Act facilitates data sharing across sectors, fostering developments in health, mobility, environment, agriculture, and the public sector. The EU plans to achieve this through trustworthy data intermediaries, citizen and business voluntarily data contribution, i.e. data altruism. The Data Governance Act is expected to drive innovation, create jobs, improve policymaking, and reduce business costs, leading to societal and economic benefits across the EU.

2.2.5 The Directive establishing an Infrastructure for Spatial Information in the European Community (INSPIRE Directive)

The Directive establishing an Infrastructure for Spatial Information in the European Community ("INSPIRE Directive". Directive 2007/2/EC)[17] provides a framework for spatial information in the European Community, aiming to support EU environmental policies and related activities. It applies to spatial data held by public authorities, emphasises the creation of metadata, and encourages interoperability of spatial datasets. The Directive stipulates that EU countries must provide services for discovery, viewing, downloading, and transformation of spatial data. The directive has been in effect since 2007. An evaluation in 2016 highlighted the need for further improvements in implementation and data policy provisions, leading to recommendations and actions by the Commission.

The INSPIRE directive caters for several trust principles across various domains. Although the Directive does not necessarily mention data protection and privacy, it recognises the need to restrict public access to spatial datasets where it might adversely affect international relations, public security, national defence, and other grounds. Further, the Directive promotes data sharing between public authorities and other entities and provides for measures that enable public authorities to share, exchange and use spatial datasets and services. Given that the Directive provides for availability of spatial data, which are largely public sector data, it emphasises the need to minimise licencing fees and promote open access, where possible. In addition, the Directive encourages the use of international standards for interoperability and harmonization of spatial datasets. It also mentions the integration of relevant standards and technical means for achieving interoperability.

[16] See FN 4.

[17] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) available at: https://eur-lex.europa.eu/eli/dir/2007/2/oj

2.2.6 The Proposal for Artificial Intelligence Act

The Proposal for Artificial Intelligence Act, or "AI Act"[18], is a comprehensive legislation proposal by the European Commission to regulate AI systems in the EU. It **aims to ensure responsible AI development and use while upholding fundamental rights**. The proposal defines AI systems and categorises them into four risk tiers. Deployer or developer of High-risk AI systems, described as AI systems used in certain sensitive areas, subject to strict regulatory requirements, in relation to conformity assessments, risk management, data protection and transparency. Certain AI practices, like social credit scoring should be prohibited outright. As enforcement will be delegated to national authorities, either within existing authorities (e.g. Data Protection Authorities, as it seems will happen in France) or with the creation of ad hoc entities (for instance in Spain).

It is important to note that the proposal puts a strong emphasis on transparency. High-risk AI systems must be designed to be transparent to enable users to interpret the system's outcomes. Additionally, the deployers of these systems are required to provide end-users with clear information about the system's characteristics, limitations, performance, and potential risks. The proposal sets out requirements for disclosure of AI system interaction, operation of emotion recognition and biometric categorization systems, and disclosure of deep fake content generation. Accountability is reinforced by the introduction of record-keeping and reporting obligations and the requirements for the ex-ante testing in AI regulatory sandboxes. The proposal underscores the importance of the quality, and reliability of AI systems. The proposed AI Act outlines provisions for technical documentation and testing to ensure high-risk AI systems consistently meet their intended purpose. Developers are required to have a quality management system in place, ensuring compliance with the regulation and covering certain aspects related to the design, development, and quality control of high-risk AI systems. Furthermore, the regulation emphasises developers' responsibility for the quality, relevance, and completeness of training, validation, and testing data sets.

The proposal refers to conformity assessment, CE marking, and harmonized standards, indicating a concern for establishing common European standards and certification processes for AI systems. Another trust principle discussed in the proposal is enforcement by national supervisory authorities. The proposal stipulates that developers shall submit technical documentation to these authorities which will assess compliance of the systems.

2.2.7 The General Data Protection Regulation (GDPR)

The General Data Protection Regulation (GDPR), also known as Regulation (EU) 2016/679, is a European Union regulation that aims to protect individuals when their personal data is processed, both by the private sector and most of the public sector. It grants individuals greater control over their personal data, offers easier access to their data, and establishes certain data rights. Additionally, it requires companies to report data breaches and institutes rules for businesses, creating a level playing field for companies in the EU market. The GDPR encourages innovation and privacy-friendly practices, promotes the free flow of personal data, and provides a modern toolbox for international data transfers. The regulation has been in effect since May 25, 2018 and plays a crucial role in safeguarding personal data in the digital age.

[18] Proposal for a Regulation Of The European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. On 9 December 2023, the European co-legislators reached a political agreement for the AI Act. As soon as a final text will be made available, additional information and analysis will be provided to BeOpen.

The GDPR, designed to safeguard fundamental rights, particularly the right to privacy, ensures that processing of personal data respects individuals' privacy rights. The GDPR permits data sharing but under strict conditions. Organisations must have lawful bases for processing data, and data subjects should be informed about how their data will be shared and with whom. It is important to emphasise thatGDPR (Art. 32) requires organisations to implement appropriate security measures to protect personal data from unauthorized access, disclosure, alteration, and destruction. Moreover, the GDPR gives prominence to transparency through the requirement of clear and concise privacy notices and the right of individuals to access data. Organizations are accountable for complying with data protection principles and must demonstrate their compliance. More detailed information on the data protection principles and requirements set out in the GDPR can be found in Deliverable D1.5 – Data Protection and Ethics submitted in February 2023.

2.2.8 The Directive on measures for a high common level of cybersecurity across the Union (NIS2 Directive)

The Directive on measures for a high common level of cybersecurity across the Union (EU) 2022/2555, also known as “NIS2 Directive”, establishes a common cybersecurity regulatory framework with the aim of enhancing cybersecurity in the EU. The Directive applies to legal entities and organisations operating in various sectors, including energy, transport, finance, healthcare, and digital infrastructure, among others. It requires Member States to adopt national cybersecurity strategies, create computer security incident response teams (CSIRTs), and coordinate vulnerability disclosure. The NIS2 establishes a network of national CSIRTs, a cooperation group, and the European cyber crisis liaison organisation network (EU-CyCLONe) to facilitate strategic cooperation, information exchange, and crisis management. Entities in these sectors must report incidents that can cause severe disruption or damage. The directive also includes provisions for supervision, enforcement, and peer reviews. It must be transposed into national law by October 17, 2024.

The core focus of the NIS 2 Directive is to have European entities implement appropriate measures to manage cybersecurity risks and prevent and minimize the impact of cyber incidents. In that regard, the NIS 2 Directive recognises the importance of safeguarding sensitive information held by critical infrastructure operators and digital service providers. Although it does not explicitly detail data protection and privacy measures, the nature of the directive aligns with the broader principles of protecting data against unauthorised access and ensuring privacy is considered in the context of cybersecurity. While the directive emphasises the need for cooperation and information exchange among Member States, it does not specifically address data sharing and accessibility in the context of principles. However, the collaborative nature of the directive implies a certain level of information sharing to strengthen overall cybersecurity. The Directive promotes transparency among Member States regarding their national capabilities and strategy for ensuring the security of network and information systems. The Directive encourages the use of European and international standards to enhance the overall level of cybersecurity. However, it does not delve into specific standards for data management. Finally, the Directive emphasises the role of competent authorities and requires Member States to designate national competent authorities responsible for the enforcement of the Directive, providing a form of human oversight.

2.2.9 The Proposal for Regulation on horizontal cybersecurity requirements for products with digital elements (Cyber Resilience Act)

In the dynamic landscape of our interconnected world, the proposal for the regulation on horizontal cybersecurity requirements for products with digital elements[19], known as the Cyber Resilience Act (CRA), stands for enhanced cybersecurity. This initiative is not merely a set of rules, but a holistic approach woven into the very fabric of digital products. The main objective of the CRA is to install a holistic cyber security ethos by design, by default and to reduce cyber incidents by 20-30%. It seeks to streamline cybersecurity requirements applicable to a broad spectrum of digital products and associated services, encompassing both tangible and intangible elements, embedded and non-embedded alike. The regulation recognizes the dynamic nature of these products by considering their entire life cycle, currently capped at a maximum of five years.

Central to the CRA is the integration of a product life cycle approach into a regulatory framework, together with responsibilities distributed across the supply and value chain. It emphasizes the need for a demonstrable capability to implement prompt fixes throughout the lifespan of both products and software. Notably, the regulation excludes free, not-for-profit open-source software from its current scope. As of the proposal's official date in September 2022, it is currently under development, with the final Act anticipated to be unveiled in the first half of 2024. Upon entry into force, it is expected to be applicable within 24 months, with a 12-month transition period for reporting obligations. The CRA signals a forward-looking initiative, acknowledging the evolving digital landscape and aiming to fortify cybersecurity in an era of constant connectivity.

[19] Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on horizontal cybersecurity requirements for products with digital elements and amending Regulation (EU) 2019/1020 COM/2022/454 final available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022PC0454 .

3 BeOpen Trust Components

This chapter identifies five (5) key components of the Trust Framework for the BeOpen project. These trust components are extracted from the legal and policy frameworks discussed in chapter 2. It should be emphasised that these trust components are selected on the basis of BeOpen's scope, as stated in Section 1.3. Furthermore, the trust components focus primarily on the data, as other perspectives, such as stakeholders or markets are connected to that. The trust components are:

A. Availability and Accessibility
B. Quality and Reliability
C. Control
D. Interoperability
E. Transparency and Accountability

3.1 Availability and Accessibility

Ensuring data is both available and accessible is pivotal for trust in the digital landscape, and the first trust component to consider when working with Open Data. As mentioned in Section 2.7, the Open Data Directive refers to a set of regulations within the European Union aimed at promoting the availability and accessibility of public sector information. The Directive encourages public entities to make certain categories of public sector data available for reuse.

This includes information in the form of documents, datasets, and other types of information produced, collected, or held by public sector bodies. Similarly, the Directive promotes the accessibility of public sector information by encouraging member states to take measures to make it easily accessible to the widest range of users. Accessibility involves providing the information in formats that are machine-readable and can be easily reused. Machine-readable formats are meant to enhance its usability for various applications and services.

In addition, the European Commission's Data Act and Data Governance Act play a role in fostering data accessibility practices. These acts focus on establishing a single market for data, promoting data-driven innovation, and ensuring accessibility of non-personal data. They address challenges through cross-sectoral governance frameworks, opening high-quality public sector data, legislative action on data sharing, and compliance and enforcement measures.

3.1.1 Licensing and Reuse

Data licensing and reuse are elements that can ensure availability and accessibility of data. Licensing establishes a legal framework, ensuring compliance with regulations. It can also mitigate the risk of misuse and provide protection against unauthorised access. Additionally, allowing data reuse promotes interoperability, enabling the integration of diverse datasets. Incorporating ethical considerations into licensing can contribute to accountability, demonstrating a commitment to responsible data practices. As such, trust in open data relies on clear licensing terms.

The Open Data Directive provides guidelines on open licensing, requiring public sector bodies to use licenses that allow for free and open reuse of data, with minimal restrictions. This ensures that users can trust that they have the right to use and build upon the data. Specifically, the Open Data Directive suggests that public sector bodies should use licenses that are compatible with the principles of openness and allow for the widest possible reuse of the information. This is meant to promote the free flow of data for various purposes, including commercial and non-commercial uses. The Directive emphasises the importance of adopting open licenses that align with recognised principles of openness. Common open licenses include Creative Commons licenses, which allow users to reuse the data under certain conditions, often requiring attribution and prohibiting the imposition of additional restrictions on reuse.

3.2 Quality and Reliability

Quality and reliability of data can enhance trust in a data ecosystem. Poor data quality, for instance, data that is not accurate or consistent, can lead to misinformation or incorrect decision-making, eroding trust and undermining the effectiveness of data-driven initiatives. Quality and reliability of data can depend on the freshness, the trustworthiness of the data source, the accuracy or other indicators for data quality.

The Open Data Directive encourages public sector organisations to maintain high data quality standards, including accuracy, timeliness, and completeness. Guidelines include mechanisms for data validation and data quality assurance. Trust can be strengthened through user feedback mechanisms. Public sector organisations are encouraged by the Open Data Directive to establish channels for users to provide feedback on the quality, relevance, and usefulness of open data, allowing for continuous improvement.

Quality and reliability of data may have implications also for the trustworthiness of AI systems. Specifically, the upcoming AI Act emphasises data quality, fairness, and transparency, requiring organisations to use representative and unbiased data while minimizing data bias. This shows that data quality and reliability have implications that may go beyond their primary use, as the use of poor-quality dataset to train AI algorithms may result in poor outcomes.

3.3 Data Control

Data control is crucial for trust as it empowers individuals and organisations to determine how their data is used, shared, and protected. Knowing that they have control and ensuring that data is handled responsibly, ethically, and in accordance with agreed-upon terms, fosters a sense of trust among stakeholders.

The Data Governance Act and the Data Act also prioritize data control. Specifically, the Data Governance Act underscores the importance of implementing safeguards to control strategic data, by implementing technical measures, like encryption, to prevent unlawful access. While outlining conditions for the re-use of public sector data, the Data Governance Act recognises a role for data intermediation services in the data economy, emphasising voluntary data sharing. It defines these services, excluding certain activities, and calls for a Union-level framework to promote trust, control, and interoperability. Special attention is given to services enhancing individuals' data control, and data cooperatives are acknowledged for their potential benefits. Similarly, in the context of data altruism, the Data Governance Act emphasises individual control of personal data over its usage. The voluntary nature of data contribution is highlighted, with data subjects empowered to make informed decisions.

The Data Act emphasises several key aspects related to data control. Firstly, it introduces a new right for consumers to access and control user-generated data, particularly from Internet of Things (IoT) devices. This is designed to enhance consumer protection and empower users to benefit from a wider range of services. Secondly, the Data Act addresses the issue of unfair contractual terms related to data access and use, specifically protecting SMEs. By restricting such terms imposed by larger entities, the regulation aims to ensure a fairer allocation of value in the data economy. Thirdly, it establishes provisions for switching data processing service providers, aiming to empower business users and individuals by enhancing their control over data processing services. This address concerns related to lock-in effects in the cloud and edge market. Furthermore, the Data Act introduces safeguards against unlawful third-party access to non-personal data held by data processing services in the Union market. Service providers are required to take measures to prevent unauthorized access, reinforcing data control mechanisms. It establishes obligations for data holders, particularly in cases involving personal data, to provide access to users and, when requested, make data available to third parties chosen by the user. The regulation does not grant the data holder a legal basis to provide access to personal data to third parties without the user's consent.

3.3.1 Data Classification

As an element of data control, the classification process primary involves determining what *type* of data are being processed. A correct classification of the data may be necessary in order to allow the lawful availability of data. Within this context, there are a number of key aspects to consider:

Firstly, it is necessary to consider whether the datasets may contain: (1) personal data as defined in point (1) of Article 4 of Regulation (EU) 2016/679 (GDPR); (2) sensitive information, including trade secrets or other IP protected information. The right classification of the data may be contextual, i.e. changing for organisations and countries, as well as depend on the specific status of the body holding the data.

Secondly, it is important to consider whether the data is included in the scope of the *High Value Datasets* Implementing regulation. [20] As described in Section 2.2.2 above, these datasets have a particular potential to generate socio-economic benefits, and made available for re-use with minimal legal and technical restriction and free of charge.

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, in particular Regulation (EU) 2016/679 (GDPR) and including any provisions of national law further specifying the application of the GDPR. Appropriate methods and techniques (such as generalisation, aggregation, suppression, anonymisation, differential privacy or randomisation) should be used, thus making as much data as possible available for re-use.

High-value datasets shall be made available for re-use under the conditions of the Creative Commons Public Domain Dedication (CC0) or, alternatively, the Creative Commons BY 4.0 licence, or any equivalent or less restrictive open licence, as set out in the Annex, allowing for unrestricted re-use. A requirement of attribution, giving the credit to the licensor, can additionally be required by the licensor.

Finally, it is important to consider whether there are any other sensitive or confidential data involved in the datasets. The sharing and processing of certain data types might not be easy to classify, therefore require caution or deliberation before making it available.

[20] COMMISSION IMPLEMENTING REGULATION (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use.

3.4 Interoperability

The interoperability of high value datasets is one of the key focus areas of the BeOpen project. While D2.1 focuses on technical interoperability, this deliverable emphasises the *legal* interoperability in the context of Open Data dissemination and reuse. This refers to policy guidelines or requirements that provide the conditions for the sharing, accessing and using of open data. In the context of the FAIR-principle, interoperability means that data and metadata should be structured in a way that allows for integration with other data or systems. In that way, interoperability ensures that data can be used in combination with other datasets.

Interoperability is an essential aspect of the European Union's approach to data management, and this is reflected in various directives and initiatives, including the Open Data Directive. While the Open Data Directive may not explicitly use the term "interoperability," it does promote the use of common standards and formats to enhance the ability of different systems to work together. Key points related to interoperability in the context of the Open Data Directive may include:

1. Machine Readability: Emphasizing the importance of making data machine-readable to enable easy integration with various applications and tools.
2. Technical Guidelines: Providing technical guidelines to public authorities to ensure that the data they release adheres to certain interoperability standards.
3. Metadata Standards: Addressing the need for standardized metadata to enhance the discoverability and usability of open data.

Furthermore, the INSPIRE Directive emphasises the importance of interoperability, which is the ability of spatial data sets and services to work together seamlessly. The goal is to enable the sharing and use of spatial information across different domains and borders. The directive outlines specific technical guidelines and standards to achieve interoperability, including the use of common data models and encoding specifications. These guidelines cover metadata, data specifications, network services, and data and service sharing. The INSPIRE Directive is designed to facilitate the discovery, viewing, and access to spatial data across various applications and organizations.

3.5 Transparency and Accountability

Transparency in digital environments provides visibility into data flows, data use and decisions, fostering accountability. When individuals or organisations are held accountable for their actions, their stakeholders have a solid ground to trust them, catering for reliability, regulatory compliance, ethical conduct, and a commitment to openness. As such, transparency and accountability work hand-in-hand to establish confidence and integrity in digital interactions, forming the foundation for trust. Essentially, transparency and accountability provide a foundation for other trust components, such as data protection, data sharing, data control, data quality, data ethics and human oversight. Therefore, maintaining this trust component is of key important for the BeOpen project.

Transparency is at the core of various of the mentioned policy and legal frameworks. The Digital Europe Programme, EU Digital Decade 2030, and the Data Act/Data Governance Act all highlight the importance of an open, collaborative, and transparent digital environment. The Open Data Directive encourages the availability and accessibility of data with minimal restrictions, promoting transparency through open formats and machine-readable data. The directive emphasises transparency in the public sector's data practices. Public sector organisations are encouraged to make their data easily accessible to the public, ensuring that citizens and businesses can find and access government data with ease. Furter, the Directive includes guidance on data governance practices within public sector organizations, ensuring that data management processes are transparent, accountable, and trustworthy. Trust guidelines further establish rules for accountability and liability in cases where AI systems cause harm or make erroneous decisions. Similarly, the INSPIRE Directive seeks transparency through metadata and view services, emphasizing the need to minimize licensing charges and promote open access to public sector data.

Accountability is integral in these frameworks as well. The EU Digital Decade 2030 underscores accountability by committing to safeguarding citizens' privacy and data rights and emphasizing "strengthening the Union's digital sovereignty in an open manner." The Open Data Directive incorporates transparency and accountability as crucial elements in building trust in data management and access. Human oversight can contribute to providing transparency and accountability. In complex digital systems and processes, the same human oversight is the safeguard of compliance with ethical standards and regulations. Additionally, human oversight fosters responsibility in technology development and deployment, demonstrating a commitment to ethical practices.

Furthermore, the assurance and monitoring of compliance to applicable regulations contributes to accountability. To ensure that public bodies adhere to the trust-related guidelines, the Open Data Directive includes mechanisms for compliance monitoring and enforcement, such as reporting requirements and supervision. The Data Act and the Data Governance Act include mechanisms for compliance monitoring and enforcement to ensure that organizations adhere to the trust-related guidelines and regulations.

4 Pilots State of Play

This chapter discusses the implementation, approaches, challenges and results of the Trust Components in the Use Cases. This is a preliminary analysis, acknowledging that not all Trust Components are implemented at this project stage. In any event, this chapter paves the way for the Guidelines and Best Practices to be outlined under subsequent chapter.

4.1 BeOpen Use Cases at glance

To facilitate understandability of the next sections, the table below provides a brief overview of the BeOpen Use Cases.[21]

Use Cases (UCs) at glance		Pilot sites
Development of a “shield” protecting Attika region form natural disasters (UC1)	The main objective is to conceive a “shield " from natural disasters for Attica region’s citizens and environment, identifying hazards, assessing risks and mitigation measures. The Attica Region will be the theatre of the Use Case and the BeOpen partners in charge are NOA, CERTH, the Attica Regional Authority. The Use Case builds on socioeconomic and statistical datasets from networks operated by Statistical Authorities and Public/Private Sector; on meteorological datasets; Big Earth Observation, environmental and climatic datasets; Land Use/Land Cover datasets; Open Street map. The expected benefits include, in the long-term, the HVDs will allow to prepare more detailed risk assessment studies for natural disasters.	Attika region (GR)
Management of traffic limitations for improving urban security in historical city (UC2; UC3; UC4)	Urban security in historical city with key critical infrastructure to properly manage traffic limitations and to fully understand the traffic and its impacts. Responsible partners are Cartagena, HOPU, FIWARE and will be replicated in Molina de Segura and Torre Pacheco. The HVDs are traffic datasets including video streams with meta-data; licence plates, vehicles count; environment monitoring, people counting and noise. The expected benefits are the implementation of solutions to analyse the traffic and the environment and decision support tools that allow for better monitoring and planning of interventions.	Cartagena, Molina de Segura, Torre Pacheco (ES)
Assessment of health impact of blue-LED streetlights (UC5; UC6; UC7))	The Use Case focuses on the health impact of installing blue-based LED streetlights and is implemented by the BeOpen partners Cartagena, Molina de Segura and Torre Pacheco together with HOPU. Available HVDs include Geospatial luminosity maps; Energy consumption datasets; historical health statistics about the	Cartagena, Molina de Segura, Torre Pacheco (ES)

[21] See, also, BeOpen D3.1 Pilot requirements and HVDs, submitted in M9.

Use Cases (UCs) at glance		Pilot sites
	urban health impact; historical mobility data about incidents in the 3 cities. The expected benefits include of the lighting in streets; street lighting could be adjusted to avoid the impact on different mental and physical illnesses; reduction of potential incidents.	
Development of mitigation actions for reducing risk of urban health islands and heatwaves due to climate change (UC8; UC9; UC10)	The Use Case focuses on the risk mitigation of urban health islands and heatwaves due to climate changes. Partners involved are Cartagena, Molina de Segura and Torre Pacheco aided by HOPU. HVDs available are green zones dataset including water consumption. Environmental datasets with historical data concerning temperature, humidity, air quality and people data. Copernicus data with land surface temperature. Expected benefits includes more accurate datasets concerning the environment will support activities against climate change.	Cartagena, Molina de Segura, Torre Pacheco (ES)
Use digital tool for improving and optimising the road infrastructure management and maintenance processes (UC11)	Improving and optimising the road infrastructure management and maintenance through AI-tools and HVDs. In relation to this, the datasets are also used to manage mobility and the use of the road infrastructure. This Use Case uses different HVDs around the road infrastructure and involves the city of Herne, HDG and HOPU. The HVDs that will be used for the Use Case include: the over 100 HVDs available at the cities' open data portal; a dataset with static pictures of streets, historical status of streets; climate data; materials databases. The expected benefits are to leverage HVDs to advance AI-based algorithms. More responsibility and sustainability of investments for street management; improvement of datasets would allow for a better city management.	Herne (BE)
Improve information sharing and decision-making processes of large-scale events management (UC12)	The Use Case aims at improving the handling, information sharing and decision-making and management of large-scale events and the associated civil protection. Handling of traffic flows, logistics and regulation, on occasion of the "Cranger Kirmes" in August each year, which attracts around four million visitors over the ten days of duration. The expected benefits are that the application of HVDs shall support private mobility, sustainable mobility, parking, public transport organization and the relevant information systems. In addition, the Use Case will target civil protection issues such as people crowding and crowd management, security and safety service, police intervention.	Herne (D)
Support decision makers and first-aid responders in assessing the impacts on urban mobility of emergencies and natural disasters (UC13)	The Use Case aims at supporting decision makers and first responders assessing the impact of incidents with urban mobility data. The responsible partner is Porto Digital, representing the city of Porto. The Use Case aims to assist in managing and decision-making in cases of emergencies such as wildfires, floods, and earthquakes, by leveraging location datasets to add context about the most affected places and identifying consequences.	Porto (P)

Use Cases (UCs) at glance		Pilot sites
	Main available HVDs are: Geospatial datasets on a) streets and squares, road topology, bridges and points of interest; b) critical infrastructure; c) administrative divisions, requalification of Public Spaces (Squares and Streets); Urban Regeneration Action Plan (PARU); location of incidents and accidents and the traffic flow observations; dynamic datasets on the positioning of municipal vehicles; dynamic meteorological datasets. Expected benefits for citizens include enhanced access to and improved quality of open data. Decision-makers stand to benefit from increased efficiency when addressing transportation incidents and problem-solving. First-aid responders can expect improved capabilities, enabling them to offer faster and more effective assistance and facilitating more strategic incident response planning.	
Enhance decision-making in mobility management (UC14)	The scope of the Use Case is the improvement of mobility management and citizens' information, integrating mobility data and Urban Environmental Information to support decision making. This Use Case will focus on mobility management and involve the city of Naples and MTECH as the technology partner. It uses several HVDs such as: Geospatial datasets; street connected infrastructures; mobility management data; Earth observation data and geospatial datasets from Spotted project; open datasets belonging to different domains. The expected benefits include the implementation of new tools and methodologies for the management and publication of mobility-related data. HVDs will underpin the development of mobility services by third parties;	Naples (I)
Detect and counter the expansion of invading vegetal species due to climate changes (UC15)	The Use Case aims at enabling the detection and countering the expansion of Sosnowsky's hogweed and other invading species in relation to climate change and warming. The partners responsible for the pilot are Vplanas and LAT40. The HVDs include Geospatial datasets on the vegetation landscape; 3D geospatial information of the area; BIM models of Vilnius City buildings. Expected benefits include solutions for monitoring and exterminating invading species of plants. Identification of areas most affected by invading species for citizens' security.	Vilnius (LT)

Table 1 - BeOpen Use Cases summary

4.2 Data availability and accessibility

Use Cases can access open data from institutions. For instance, one Use Case [**NOA/CERTH**] acquires population and building related data for vulnerability, exposure and risk assessment from the national statistical authority, as well supplementary data acquired from various sources (e.g. power networks from the Public national Power Corporation, Digital Elevation and Surface Models from the National Cadastre, etc.). For hazard modelling, they retrieve satellite and ground station data (e.g. meteorological, DEM, technical works, land use, land cover, etc.). Another Use Case [**VPLANAS**] retrieved mostly geospatial data. In the case of [**HERNE**] two different use-cases are being identified: in the case of the crowd management pilot,the data is made available to regulatory authorities, not to the public. Data is owned by the city and access is generated through personal accounts. The data is generated during the time that the fair is taking place, which accounts for ten days and is being monitored continuously during this period. Several control mechanisms which monitor the proper function of the system are being requested to be implemented in this Trust Framework. In terms of measures against cyberthreats and data breaches, this Use Case makes use of end-to-end encryption in which data transmission with access on their servers, is only possible with ssh keys. In terms of transparency and countability with respect to the Use Case of road conditions [**Herne**], so far no thought has been given to liability for harms. This could therefore be taken into account or included in the further course of the project since this Use Case is centred on specific planning activities, therefore concrete and specific actions on the principles design have to be discussed and carried out. All of Herne's city officials and project participants expressed that data management is based on transparency goals. In terms of security, Herne's security officers are not directly involved in BeOpen but their involvement will be necessary the moment that Herne's Open Data Portal will be defined.

In the case of another Use Case [**HERNE**] with regard to the road conditions, geospatial data concerning road conditions and traffic data can be provided only in accordance with national rules or laws. In this case geographical data concerning road conditions and traffic data at Herne Smart City Platform will be shared during Summer 2024. This platform will entail shared editing of geographical feature on single territories.

One of the approaches for data access is the use of *data sharing agreements*. One Use Case leader [**NAPOLI**] is experimenting sharing agreements with other municipalities, to share geographical data on a local mobility project [Napoli PON Metro project PPAT]. They use a platform for shared editing of geographical features of single territories. Furthermore, another Use Case leader [**PORTODIGI**] also makes use data sharing agreements, in addition to service-level agreements. For this, the Use Case leader uses technologies such as MinIO for internal data sharing and PostgreSQL databases.

Another means for data access and availability is open data publication. For this, open data portals have proven to be useful. An example of this is provided by the **Cartagena visualizer**[22].

Data availability and accessibility can come with challenges. One of the Use Cases [**NAPOLI**] experienced a data sharing issue with the public company that manages public transport in the city. In this specific case, the data are owned by the municipality (EU funding) but the dashboard is managed by a public transport provider. Another challenge is a lack of human resources. For example, one Use Case leader indicated that the local administration lacks the people with the adequate skills to address issues like road maintenance, urban planning, energy strategies etc as well as human resources that possess the capability needed to address specific digital and geographical. Finally, there is a risk of data suppliers that withhold data. In one of the Use Cases [**PORTODIGI**], one data supplier left the agreement during piloting activities. Another challenge for data access can be the duration of data acquisition processes [**NAPOLI**].

4.3 Data classification

Data classification refers to the understanding of the *type* of data included in the datasets. First of all, in several Use Cases the availability of *Open Data* and *HVDs* is considered, though it has not become explicit from the

[22] https://pma.ayto-cartagena.es/visualizador/

responses whether the Use Case leaders have verified whether the datasets meet the legal requirements to be classified as Open Data or HVDs.

A number of Use Case leaders have considered the presence of *personal data*. For instance, in mobility and urban planning datasets, the mobility options and preferences are analysed through generalised information, where the data relate to *categories* of users. Where data has been collected from individuals about mobility preferences, this has been done with the use of anonymous questionnaires.

A key approach for dealing with personal data in the context of Open Data includes anonymisation. At least five Use Cases make use of anonymisation tools, before publishing data on open platforms. For instance, in the case of mobility datasets that include traffic plates, vehicle counting and video files, in this information is anonymised to avoid tracking and maintain the citizens' privacy. In case of the crowd management and the road condition Use Cases, the sensors used are either unable to generate personal data or they anonymize it internally to generate information that has not data reference. Furthermore, where data and the metadata is derived by the crawling social media accounts, the Use Case leader indicates that any personal information will be properly erased, and an API will be used to anonymise the final data sets.

Personal data protection brings about certain challenges. For instance, in one of the Use Cases, all data that could reveal the identity of individuals and their movements is considered not to be shared. Data holders (i.e. the city officers acting on mobility strategies) have underlined this issue from the very beginning.

4.4 Data quality and reliability

There are a range of practices and approaches taken by Use Cases to ensure data quality and reliability. One such approach suggested by a Use Case leader is the implementation of a *data quality plan* that stipulates certain requisites. Furthermore, in the context of geo-localisation or relevance scoring of social media data, a mechanism for addressing data quality and reliability is the use of indicators such as *accuracy, precision, and F1.* Where sensor data are used, accuracy can be measured by means of looking at correlation between datasets as an indicator for data reliability [**CARTAGENA** and **TPACHECO**]. In such contexts, it is also suggested to make use of a threshold for number of measurements before using the data, to avoid a sample size that is too low, as well as cross-checking data by reference stations. The sensor data are grouped in time intervals, such that, if not enough instances have been recorded within that interval, the data to be displayed may not be reliable and is therefore discarded. Both measures ensure both the quality and reliability of the data provided.

In the case of the crowd management Use Case [**HERNE**], the data is incomplete and or faulty, therefore the sensor system is susceptible to disruptive influences such as for instance thunderstorms. A challenge is here is that if counting values are being reset, the data will be lost for one day. In order to prevent this from happening, as a effort to implement this trust component, recalibrating of sensors, manual correction or resetting of counting values is being employed. In the case of the road conditions Use Case [**HERNE**], a data quality plan must be developed and implemented since human resources are of crucial importance within the administration. Therefore, work processes that concern the maintenance and updating of data must be carefully prepared and carried out.

Another Use Case [**PORTODIGI**] focuses on the *completeness* as indicator for quality and reliability. They mention: ‘For our automated pipelines we have partially implemented some data quality mechanisms that validate the completeness of datasets and that all expected data arrived at the expected times. If not, then we have recuperation mechanisms in place’. This should ensure the quality and homogeneity of datasets.

4.5 Data control

Data control is about who is authorised to access that data or to process the data. One of the Use Case leaders responded that data control is important during data storage and sharing 'for instance to ensure that only the data providers and the data stewards have access to specific data, and that data shared in open data is never mixed with private data.' In the case of the road conditions Use Case it has been stated that there is a significant importance which lies in the activities regarding the updating time and exporting for open data to be imported and defined in the trust component. The mobility infrastructure data (geographical data concerning road conditions) are produced in the city offices and the challenge here is to analyse and process data from the services on road conditions within acceptable time frames. In case of the crowd management Use Case, the ownership and control of the data lies within the city of Herne. Therefore, access to the data is being granted through personal user accounts. One of the efforts to be employed and implemented in this trust component is that this data will be made available to the public in the near future as a feasible outcome of the BeOpen Trust Framework.

However, in certain Use Cases, the original data is controlled by an external service (e.g. health services or the national fire service) and is not accessible. In such cases, only inferences from datasets can be made publicly available. For instance, where social media data are collected, the partner will not make available the original data set as Open Data, but instead only the outcomes of visual and semantic data processing. Data access is facilitated through a service (REST API), without direct interaction with the database. Data cannot be altered in the database in any way.

One of the Use Cases uses a specific *dashboard* for data control, specifically for data management such as time updates and exporting datasets to Open Data. In this case, mobility data collection is financed by public funds and shared between the mobility municipal agency and the city officers by means of such a dashboard.

4.6 Interoperability

Some issues have already surfaced around the integration of systems and datasets. In the PORTO Use Cases, the solution platforms are only accessible to local technicians, impeding access by external parties. A formal request is required to acquire access. To address this, the Use Case leader put a process in place for formal approval for data to be published in Open Data. However, it appeared to be difficult to implement such an access control system that can be integrated in existing data architectures and allows for data to be made open source.

4.7 Licensing and reuse

There are different means and measures taken for licencing and reuse. Certain data sets are *free access*, such as the planning data [**NAPOLI**]. Furthermore [**CARTAGENA**] data will be available to download via open portals to be reused for various organisations. In the case of health datasets, reuse could be limited due to belonging to another entity. In the case of the road conditions Use Case [**HERNE**] the services enabling data downloading, are to be expanded and improved for users. Additionally, the data will be made available via open portals, also for reuse purposes. With regard to standards to apply, the NGSIv2 is being used, since all data is stored according to the Smart Data models. The challenge within this Use Case is to fulfil the requirements to implement a data standard and to organize a data structure that is compliant to EU and national standards and regulations.

Other datasets are owned by the Use Case leaders. For instance, one Use Case leader [**NOA**] owns the fire database and this partner's agreement cannot be used for commercial exploitation. The data will be made available for downloads via portals. The datasets can be reused by various organisations. Use Cases generally make use of creative commons, while it should be noted that certain respondents acknowledge the need for adopt other open data protocols. For social media data storage, an academic license is used.

Challenges for licensing and reuse could involve that datasets belong to another entity, that limits data access [**NOA**].

4.8 Transparency and accountability

Use Case leaders acknowledge that transparency is important, and that organisation of responsibility should be considered. However, the implementation of transparency and accountability mechanisms has proven to be challenging for the Use Case leaders. One approach taken in the Use Cases is making all data processing public and describe these in the open platforms, yet it remains unclear how this will be executed. In case of the Use Case Road Conditions [**HERNE**], there are a series of alerts which inform any type of failures in the platform on top of reviewing possibilities by its therefore pinpointed and qualified personnel. As it is a complex system with different parts involved, the main challenge is to maintain the correct functioning of the service, minimising the time for fault reporting and troubleshooting. In this particular case, thanks to the emergence of bugs ad their solution, the staff became more experienced and trained capacity and speed while dealing with its system.

5 Guidelines and Best Practices

Based on the discussion in the previous chapters providing for the components of BeOpen Trust Framework and the respective practices within each BeOpen Use Case, this chapter provides for an **initial set of usable Guidelines and Best Practices** aiming to foster the increase of the availability, the quality and the usability of High Value Datasets.

5.1 STEP 1: Definition of Purpose(s) For Data Processing

For BeOpen Trust Framework to be implemented, the initial phase involves considering *why* we need to collect, store, structure or share data in the first place. A clear and unambiguous understanding of the purposes of data collection and transmission is the first stepping stone for developing a technical and organisational system for trusted data sharing. Furthermore, consideration of the "why" extends to both legal and societal perspectives. Legally, the purposes of data processing must be clearly outlined, as this is a prerequisite for assessing the legitimacy and of the data processing in the data processing in the first place. Going beyond legal requirements, from a societal standpoint, it is recommended to gain a sound understanding of the specific societal challenges or sustainable development goals that it is aspired to be served by the specific datasets. Once a comprehensive understanding of the "why" is achieved, the nature of the data to be shared (''what'') and the methodologies for sharing (''how'') should be considered. Notably, in addressing the "what" and the "how," we focus on the input, throughput, and output aiming to provide for a systematic and well-informed approach to data sharing in line with the Trust framework of BeOpen.

5.2 STEP 2: Identification of Data Input

Before sharing data with the data recipient, the data holder needs to apply certain configurations to the data sets to prepare it to be shared. Before we share the data, we might need to *reconfigure* the datasets such that we can share data, while minimizing the risk that issues arise at a later stage. To achieve this, we suggest a 'relative impact-based' approach. This implies that we can share certain attributes or knowledge of information, without revealing personal data or other sensitive information. Once the reasons and the motivation for the sharing of open data are formulated in a clear and unambiguous manner, it is necessary to define the input data.

In light of the Trust Components discussed under chapter 2, a set of relevant questions in this respect is formulated as follows:

A. **Data availability and accessibility**
   a. What data are available?
   b. Which entity has access to the data?
   c. What is the identity of the data holder?
   d. Under what circumstances and conditions can the data be accessed?
   e. Are there any other relevant data availability and accessibility issues to consider?

B. **Data classification**
   a. Does the dataset meet the criteria of Open Data under the Open Data Directive?
   b. Does the dataset meet the criteria of High Value data under the Open Data Directive?
   c. What is the level of granularity of the dataset?
   d. To what extent does the dataset contain personal data?

  e. To what extent does the dataset contain proprietary data?
  f. To what extent does the dataset contain other sensitive data types?
  g. Are there any other relevant data classification issues to consider?

C. Data quality and reliability
  a. How confident are we about the quality of the data?
  b. Are the data 'fresh' and up to date?
  c. How confident are we about the reliability of the data source?
  d. Are there any other relevant data quality and reliability issues to consider?

D. Data control
  a. Who will have access to the data?
  b. What systems are to be used to control the data?
  c. How can we implement secure access management (transit or interface)?
  d. Are there any other relevant data control issues to consider?

E. Interoperability
  a. What are the formats of the datasets (e.g., json/csv)?
  b. Is there a data holder responsible for preventing data leakage?
  c. Is any software necessary to run the dataset?
  d. What are the skills and competences and authorisations of the people at the departments responsible for data input?
  e. Are there any other relevant interoperability issues to consider?

F. Licensing and reuse
  a. If any, under what licenses are the data to be made available?
  b. Are there any other relevant licencing and reuse issues to consider?

G. Transparency and accountability
  a. To what extend to we have insight in data provenance (what is the original source of the data, and what does the 'data supply chain' look like)?
  b. Who can be held responsible for the data?
  c. Are there any other relevant transparency and reuse issues to consider?

5.3 STEP 3: Identification of Data Throughput

In the *throughput* stage, data is transmitted from the data holder to the data recipient. This way, the data recipient becomes the new data holder. Before sharing data, all relevant stakeholders need to have enough *certainty and comfort* in revealing the information received. It is important to check if the Trust Components for the input data are still sufficiently catered for. The discussion below proposes the following three-phases plan for sharing the data.

Phase 1: Onboarding

Any type of data transfer, in reality, implies a handover of accountability. By becoming a data holder, you are becoming a custodian of values and regulatory obligations, but you will also be assigned with moral duties towards other entities (e.g. other departments within a specific organisation). During the 'onboarding' phase, it is of paramount importance to take measures towards data validation; in particular, such measures may, mainly, serve the following:

a. Validation that your organization *must* receive the data by law
b. Validation that your institution *wishes* to receive the data

c. Validation that your organization does receive the data

Phase 2: Strategic/tactics

After data validation, it is important to be able to provide answers to the following key questions relating to Phase 2:

a. What does the organizations have to do with the data?
b. What does the organization wish to do with the data?
c. What is the organization allowe*d* to do with the data?

Phase 3: Operational

After defining the strategy, the operational phase starts. Relevant queries to raise at this phase could be summarized as follows:

a. Storage: where and how will the data be stored?
b. Disclosure: where and how will the data be disclosed?
c. APIs: what APIs will be used, where and for what purposes?
d. Security: are the right security measure in place to prevent data leakage or unauthorised access?
e. Enrichment: how can I enrich datasets? What interoperability requirements are needed?

Finally, organizations acting in their capacities as data holders and data recipients may agree to use the following policy instruments, possibly, further enabling trusted data sharing: a Code of Conduct/Code of Engagement, meaning, a set of *guidelines or principles* that governing the relationship between data holders and Terms and Conditions, meaning, a set of *rules and agreements* that outline the terms under which a service is provided, establishing a contractual relationship between the data holder and the recipient.

5.4 STEP 4: Identification of Data Output

Data output is about 'making available' the requested data[23]. Public bodies need to comply with the legal requirements established by the Open Data Directive, national law and the High-Value Datasets. Therefore, when preparing to making data available, an organization needs to consider the following:

- Does the data meet the ODD requirements?
- Does the dataset meet the HVD requirements?
- How is the organization going to disclose the data?

Furthermore, if AI systems are going to be used, the AI Act could become applicable at this stage. It is important to assess the risk classification of any AI system used for the processing of the datasets. A high risk level could lead to further requirements for the data processing, such as transparency obligations.

Finally, it is important to check if the Trust Components for the input data are still duly considered. For instance, if an organization receives the data, the acquired custodianship comes with accountability obligations. As shown in Table 1, though, in the course of this step and in addition to the trust components, it is essential to

[23] Commission Implementing Regulation (EU) on list of specific high-value datasets and the arrangements for their publication and re-use.

have sufficient clarity on the starting date of the data collection, the moment of the data deletion, key contacts of data holder and data recipient, technical contacts, and, if applicable, other notable details (Table 1).

A.	Data availability and accessibility	
B.	Data classification	
C.	Data quality and reliability	
D.	Data control	
E.	Interoperability	
F.	Licensing and reuse	
G.	Transparency and accountability	
H.	Starting Date	
I.	Moment for Data Deletion	
J.	Key Contact	Data Holder: Data Recipient:
K.	Technical Contact	Data Holder: Data Recipient:
L.	Other Notable Details	

Table 2: Criteria for Identification of Data Output

5.5 Step 5: Continuous Validation

Nevertheless, especially, when AI technology is used, it is intended that the steps above are repeated periodically. In the context of continuous validation, various (good case, bad case and worse case) scenarios plotting could take place, related impact assessments could be conducted, while the effectiveness, for instance, of any technical and/or organisational measures considered to serve BeOpen Trust Components should be re-examined periodically, as captured in the Figure 3 below.

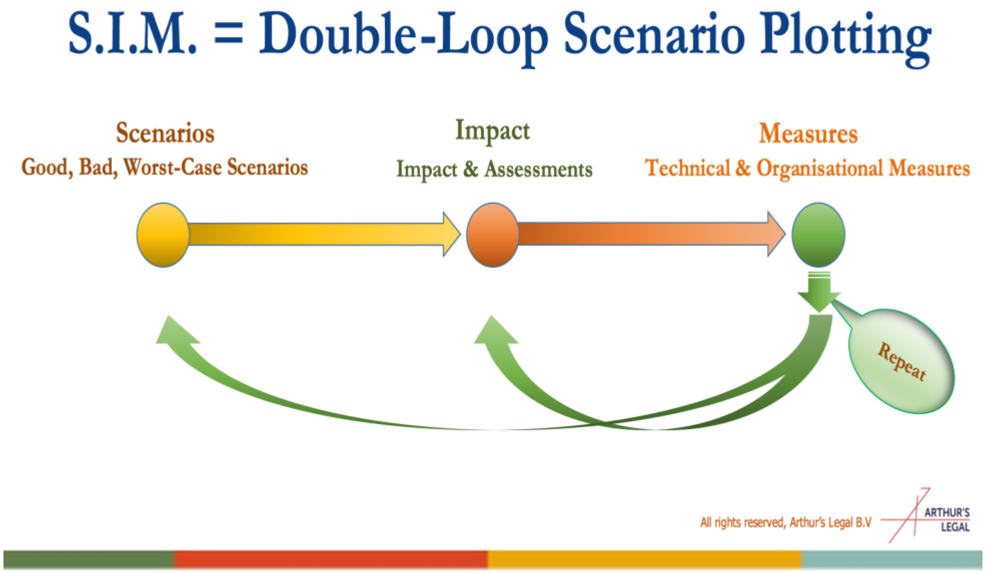
S.I.M. = Double-Loop Scenario Plotting
Scenarios
Good, Bad, Worst-Case Scenarios
Impact
Impact & Assessments
Measures
Technical & Organisational Measures
Repeat
All rights reserved, Arthur's Legal B.V
ARTHUR'S LEGAL

Figure 3: Continuous Validation

5.6 Guidelines Based on Learnings from Use Case Leaders

Managing open data involves multiple stakeholders with varying levels of control. Successful data sharing requires agreements and platforms that facilitate collaboration among different stakeholder groups. The data architecture needs to accommodate diverse interests and responsibilities.

1. Data sharing agreements, open data portals and other platforms can be used for data sharing among different stakeholder groups.
2. Investing in training and skills development is crucial for effective data management. Skills in data anonymisation, processing, and the ability to work in interdisciplinary teams are highlighted. Bridging gaps between policy, business, and legal departments is essential for streamlined data sharing processes.
3. Organisations should be proactive in understanding the requirements for open data and high-value datasets. This includes compliance with regulations and ethical considerations.

4. The presence of personal data should not lead to discarding valuable datasets or as a motivation to withdraw from Open Data obligations. Rather, it should encourage creativity in proper anonymization and inferencing based on the data.
5. For data quality and reliability, there are a range of concrete indicators to be used. These include accuracy, precision, F1 and completeness. The right indicators are dependent on the data type and the context of data collection.
6. Data quality and reliability should also involve the reduction of biases in the datasets. It is recommended to consider whether your dataset might include biases and if so, how this could affect certain groups and how this can be avoided. For instance, does the mobility dataset include more granular information of certain areas in cities, and could this lead to positive or negative discrimination?
7. Trust in data providers is not a substitute for verifying the quality and reliability of datasets. Organisations must take responsibility for ensuring the integrity of the data they use.
8. Dashboards and data management tools can facilitate the restructuring of datasets without compromising full access. This allows for efficient updates while maintaining data security.
9. It is recommended to put effort in considering how transparency can be improved in data supply chains, as well as data processing. Transparency is a crucial element for accountability and trust.
10. Cybersecurity is identified as a foundational element of trust. Safeguarding software, networks, and data is essential for achieving Trust Components such as data access, control, interoperability, and accountability. The absence of consideration for cybersecurity by pilot leaders is noted as a potential gap.

6 Concluding Remarks

This deliverable provided valuable insights as to how fundamental components for trust are reflected across the practices of BeOpen Use Case Leaders during the first year of the project. The main observations in this respect can be summarized as follows:

1. In line with the project objectives, ensuring availability and accessibility of data has been clearly a priority across all project Use Cases.
2. Raising awareness, also, with respect to compliance with applicable regulations is a prerequisite both for the proper handling of open data and high data value datasets, as well as for the maximization of the resulting benefits.
3. Investing in skills' development is crucial for the proper handling of open data and High Value Datasets. For instance, skills relating, among other, to data anonymization techniques and interdisciplinary collaboration appear to be particularly relevant not only for public sector, but also for the private sector where departments in charge of policy affairs, legal affairs and business development collaborate to establish and streamline data sharing processes.
4. Quality of data presumes putting sufficient emphasis, also, on accuracy, precision and completeness. Quality of data plays a key role in terms of reducing the risk for bias in the compiled datasets that may significantly negatively impact specific groups of people.
5. Consideration of transparency mechanisms is important before any data disclosed. Transparency is a key, also, in data supply chains. Overall, enhancing accountability, contributes to transparency and, possibly, strengthens trust among the stakeholders involved in data sharing.

The concept of "trust" is notoriously hard to define, and it may acquire different meanings across disciplines and cultures.. In any event,it is a requirement for economical flourishing and societal development. Taking this into account and building on the findings captured in this document providing for Year 1 of the project, Task 5.4 will explore the Value Models, the Feasibility and the Sustainability of BeOpen outcomes. Notably, the resulting deliverable D5.6 due in M36 will report on the existing gaps through the analysis of the state of play and state of the art in the Use Cases, in order to provide assessments on potential exploitable value models ensuring feasibility and sustainability.

7 Appendices

This section provides for the request for input addressed all Use Case Leaders in November 2023.

7.1 Appendix I

The questionnaire below aimed to gain an understanding of the degree of implementation or adherence to the relevant BeOpen Trust Components identified. The respective input provided by all Use Case Leaders is available on the project's repository.

Essential trust components	Did you consider this trust component in your piloting activities?* *(yes/no/not applicable)*	Please specify the **aspects or activities of the Use Case** where the implementation of this trust component may be relevant.	If any, please provide a short description of **any practices, approaches or efforts** employed to implement this trust component.	If any, please provide a short description of any **challenges** you experienced or foresee in the consideration & implementation of this trust component	Other remarks (e.g. if you gained any learnings or takeaways of your relevant efforts so far).
Data protection and privacy *(e.g. limiting pers. data collection, anonymisation, psyeudonymiation, respecting data protection principles)*					
Data sharing and accessibility *(e.g. data sharing agreements, processes in place for trusted data sharing, use of open data portals)*					
Data ownership and control *(e.g. access rights and data control mechanisms)*					
Cybersecurity *(e.g. measures against cyberthreats and data breaches, enhancing system integrity, encryption)*					

Transparency and accountability *(e.g. making data management processes transparent, attributing responsibilities to stakeholders, liability for harms)*					
Licensing and reuse *(e.g. clear licensing terms for free and open transmission, use, and resuse of data)*					
Quality and reliability *(e.g. mechanisms for maintaining high data quality standards, incl accuracy, timeliness and completeness)*					
Standards and certification *(e.g. adherance to certain standards and certification)*					
Education and awareness *(e.g. educating stakeholders within the organisation about digital risks and best practices, such as raising awarenesss about cyberthreats and data/AI ethics)*					
Compliance and enforcement *(e.g. compliance monitoring and enforcement mechanisms such as reporting requirements and oversight)*					

Data ethics *(e.g. addressing biases in databases, acknowledging power imbalances, respecting human rights)*					
Human oversight *(e.g. mechanisms that ensure human oversight judgement play a role in automated processes such as decision making)*					

*If selected 'not applicable', please, explain under the 'other remarks' field why this particular trust component is not applicable in your Use Case.