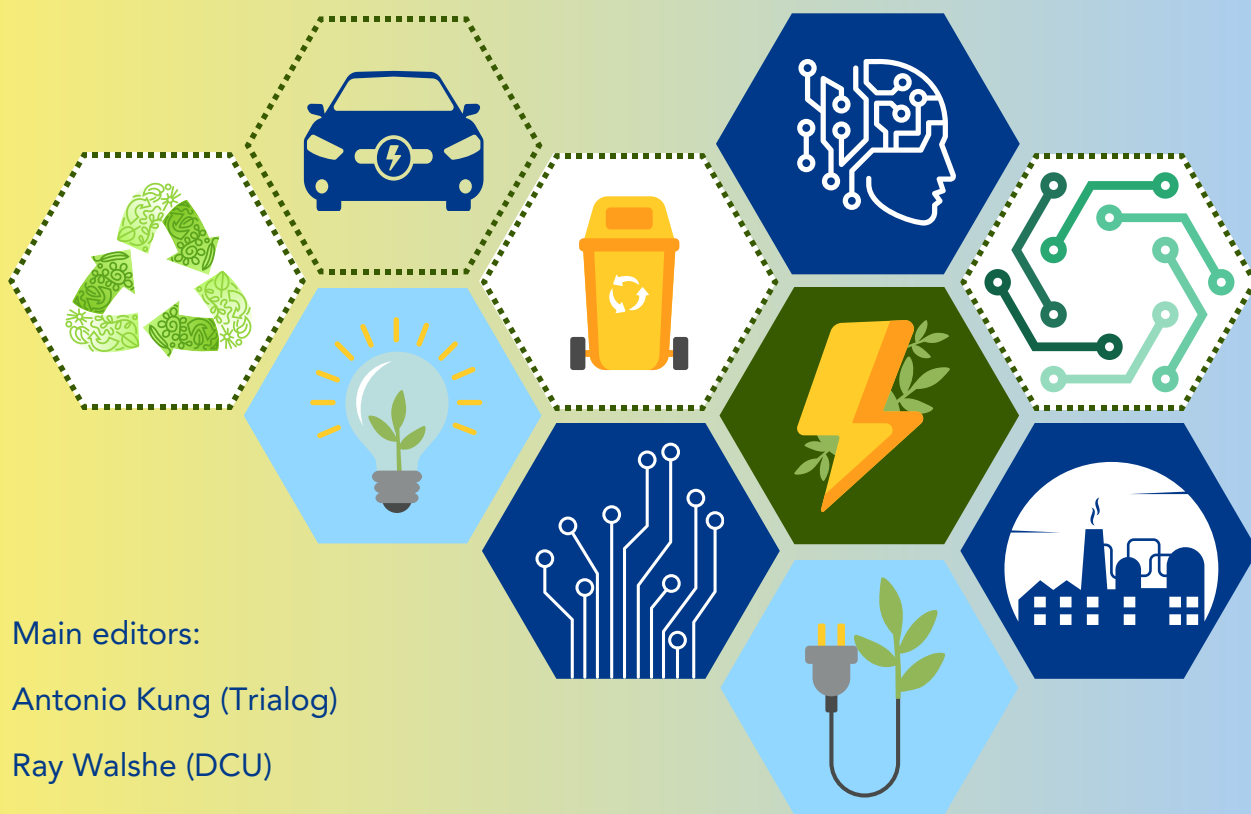


DATA SHARING SPACES AND INTEROPERABILITY



Main editors:

Antonio Kung (Trialog)

Ray Walshe (DCU)

Rigo Wenning (ERCIM)

This paper is the result of a cooperative work by a group of Big Data Value Association members, part of the Task Force on Data Spaces and Task Force on Standards and representatives from collaboration partners. Published by BDVA in December 2023.



I	Executive Summary	8
I.1	Executive Summary	9
I.2	Authors	10
II	Introduction	11
III	Context	13
III.1	Definitions	14
III.1.1	Data and metadata	14
III.1.2	Interoperability and portability	15
III.1.3	Data spaces	16
III.2	Interoperability as a key concern in data spaces	21
III.2.1	Ecosystems and interoperability	21
III.2.2	Data lifecycle	26
IV	Value of Metadata Interoperability	29
IV.1	Enabling the Data Sharing Value Wheel	30
IV.2	Enabling FAIR Data	31
IV.3	Enabling Governance	32
IV.4	Empowering People	33
IV.5	Improving Organisation Operations	33
IV.6	Technology Enablers	34
V	The Mechanics of Metadata Interoperability	35
V.1	Creating knowledge on data	36
V.2	Representing Knowledge on the Web	37
V.3	Organising Interoperable Exchange	39
V.4	Acquiring Metadata Automatically	39
VI	Metadata interoperability for data sharing spaces	40
VI.1	Overview of viewpoints	41
VI.1.1	Interoperability Framework Viewpoint	41
VI.1.2	The Daunting Challenge of Smart Cities	42
VI.1.3	Cyber Physical Systems Viewpoint	43
VI.1.4	AI viewpoint	44
VI.2	Recommendation for an inventory of metadata information	45

VII	Examples of Initiatives and Projects	46
VII.1	IDS Reference Architecture	47
VII.1.1	Overall Characteristics	47
VII.1.2	Interoperability	50
VII.1.3	Metadata Interoperability	51
VII.2	GAIA-X Reference Architecture	52
VII.2.1	Overall characteristics	52
VII.2.2	Interoperability	55
VII.2.3	Metadata interoperability	57
VII.3	FIWARE Reference Architecture	58
VII.3.1	Overall Characteristics	58
VII.3.2	Interoperability	59
VII.3.3	Metadata interoperability	59
VII.4	PLATOON Research Project	60
VII.4.1	Overall characteristics	60
VII.4.2	Interoperability	61
VII.4.3	Metadata interoperability	61
VII.5	InterConnect Research Project	62
VII.5.1	Overall characteristics	62
VII.5.2	Interoperability	64
VII.5.3	Metadata interoperability	65
VII.6	SmartBear Research Project	66
VII.6.1	Overall characteristics	66
VII.6.2	Interoperability	68
VII.6.3	Metadata interoperability	68
VII.6.4	Use case	68
VIII	Examples of Data Space Use Cases	69
VIII.1	Agriculture	70
VIII.2	Renewable energy	71
VIII.3	Health	72
VIII.4	Industry 4.0	73
VIII.5	Smart living	74
IX	Conclusions	75
X	About BDVA	77



Figure 1 – Abstract Data Space Model	16
Figure 2 – System of systems (source ISO/IEC/IEEE 21839)	21
Figure 3 – Emergent capability (left) and emergent risks (right)	22
Figure 4 – Point of interoperability (PI) conceptual model	23
Figure 5 – Notation used to describe a point of interoperability (PI) and example	23
Figure 6 – Notation used to describe a PI involving metadata and example	24
Figure 7 – Metadata for privacy management in smart cities	25
Figure 8 – Metadata for privacy preferences management	25
Figure 9 – Metadata for digital twins	26
Figure 10 – Data Lifecycle “evolution”	27
Figure 11 – Example of data lifecycle	27
Figure 12 – Metadata and interoperability issues in the data lifecycle	28
Figure 13 – Data sharing wheel	30
Figure 14 – Creation of knowledge based on existing interoperable knowledge	36
Figure 15 – Knowledge as web resource	37
Figure 16 – European Interoperability framework	42
Figure 17 – ISO/IEC 30145 Smart city framework	42
Figure 18 – AI viewpoint of interoperability	44
Figure 19 – Main interaction patterns between IDSA components	48
Figure 20 – IDSA on top of abstract data space model	48
Figure 21 – IDSA certification levels for component certification	50
Figure 22 – IDSA information model dimensions	51
Figure 23 – Gaia-X ecosystem	52
Figure 24 – GAIA-X on top of abstract data space model	53
Figure 25 – Gaia-X conceptual model	54
Figure 26 – Gaia-X federated identity model	55
Figure 27 – Gaia-X federated trust model	56
Figure 28 – INTERCONNECT architecture	62
Figure 29 – INTERCONNECT integration with GAIA-X	64
Figure 30 – SMARTBear architecture	67

Table 1 – Definitions on data	14
Table 2 – Definitions on interoperability and portability	15
Table 3 – Existing Definitions of data spaces	18
Table 4 – Existing descriptions of data spaces	20
Table 5 – Interoperability facets	41
Table 6 – Cyber physical system concerns	43
Table 7 – IDS-RAM characteristics	47
Table 8 – IDSA interoperability	50
Table 9 – IDSA metadata interoperability	51
Table 10 – GAIA-X Interoperability	52
Table 11 – GAIA-X interoperability	55
Table 12 – Components of the Federation Model	55
Table 13 – GAIA-X metadata Interoperability	57
Table 14 – FIWARE interoperability	58
Table 15 – FIWARE interoperability	59
Table 16 – FIWARE Metadata interoperability	59
Table 17 – PLATOON characteristics	60
Table 18 – PLATOON interoperability	61
Table 19 – PLATOON meta interoperability	61
Table 20 – InterConnect characteristics	62
Table 21 – InterConnect interoperability	64
Table 22 – InterConnect metadata interoperability	65
Table 23 – SmartBear interoperability	66
Table 24 – SmartBear interoperability	68
Table 25 – SmartBear metadata interoperability	68
Table 26 – SmartBear use cases	68
Table 27 – Agriculture use case	70
Table 28 – Renewable energy use case	71
Table 29 – Health use case	72
Table 30 – Industry 4.0 use case	73
Table 31 – Smart living use case	74

I Executive Summary

Executive Summary

This discussion paper focuses on the problem of how to achieve interoperability in and between Data Spaces. It collects inputs and provides insights that can be useful to future standardisation activities in the area.

In the past 20 years data interoperability improved significantly allowing one to showcase the economic and societal benefits of the data economy. But in the data economy, data sharing and monetisation is subject to constraints that shape the framework within which economic exploitation can happen and business models can be effective. On the one hand, data spaces have generic mechanisms for exchanging, ingesting, processing, sharing and understanding of data. On the other hand, sophisticated mechanisms are needed to express the constraints coming with the shared data. Properties like data quality, data protection constraints and data rights are called metadata. Hence it is important to not only allow for seamless inclusion of data into a data space, but also for the automatic inclusion and understanding of metadata that express the constraints and permissions associated with the data.

This document provides information on the state of the art, achievements, as well as gaps concerning metadata interoperability. After describing the importance of metadata and the principles of meta-data interoperability, a contribution of this discussion paper is to analyse available standards, clarifying what their added value for the ecosystem is. Furthermore, it provides guidance on achieving interoperability solutions for the realisation of a functioning and frictionless European-governance data sharing space.

In more detail, this discussion paper:

- describes a value perspective of interoperability using the wheel (data, governance, people, organisations and technology) described in the BDVA discussion paper on data sharing spaces [1];
- provides a technical perspective of interoperability addressing the facets (transport, syntactic, semantic, behavioural, policy) described in ISO/IEC 19941 [2];
- elaborates on ecosystem and lifecycle perspectives addressing the various agreement models to reach interoperability;
- provides examples of interoperability solutions such as the web of things;
- explains the integration of existing solutions, such as IDSA, FIWARE and Gaia-X; and
- provides practical guidelines to achieve meta-data interoperability and recommendations on a possible roadmap.

[1] Towards a European-Governed Data Sharing Space. Enabling data exchange and unlocking AI potential. November 2020 https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpaces%20PositionPaper%20V2_2020_Final.pdf

[2] ISO/IEC 19941:2017 Information technology – cloud computing – Interoperability and portability. Freely available standard https://standards.iso.org/ittf/PubliclyAvailableStandards/c066639_ISO_IEC_19941_2017.zip

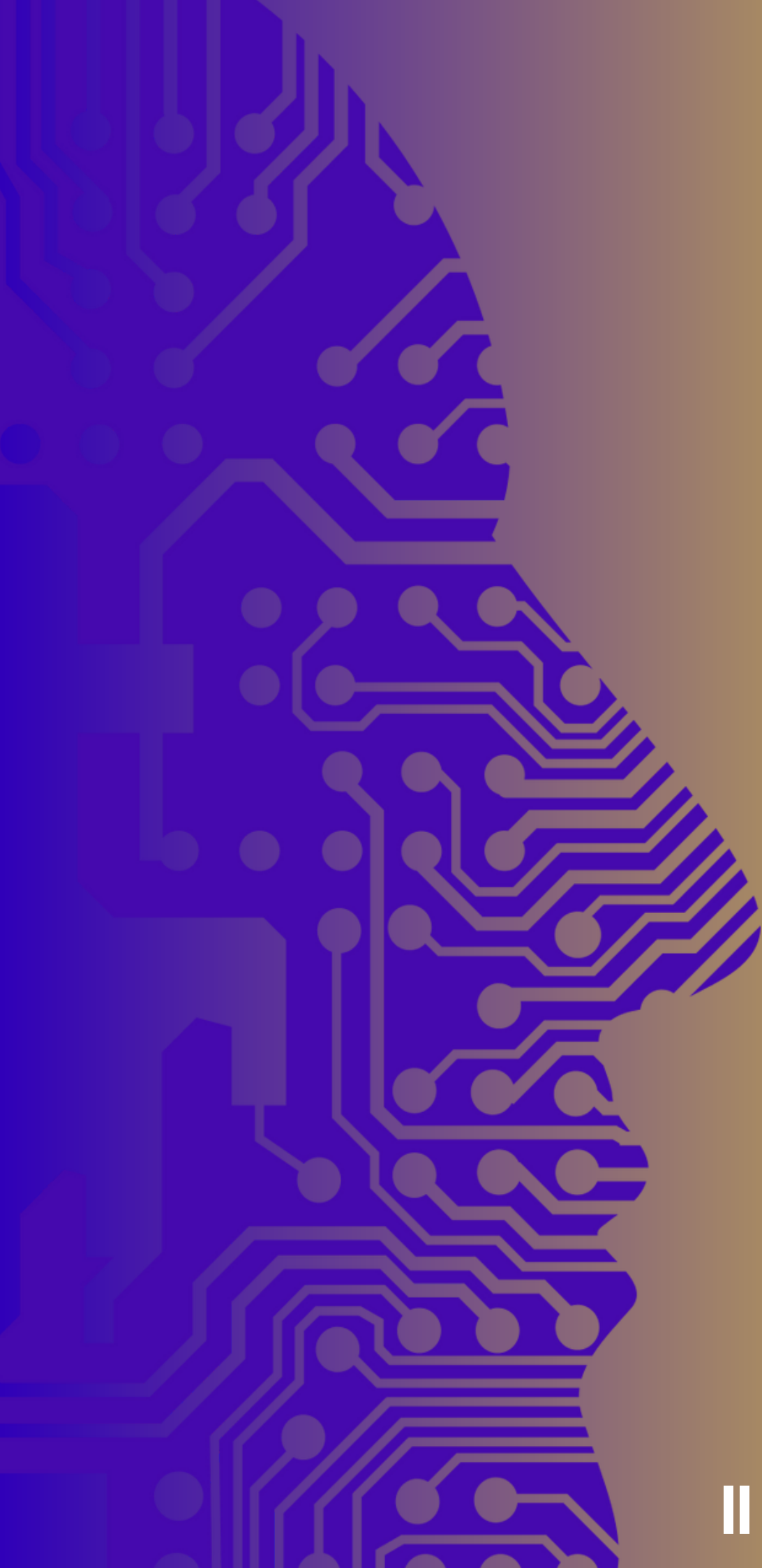
Authors

This paper is the result of a cooperative work by a group of Big Data Value Association (BDVA) [3] members part of the Task Force on Data Spaces and Task Force on Standards and representatives from collaboration partners.

Main contributors (in alphabetical order):

Vladimir Alexiev (Ontotext)
Daniel Alonso Román (BDVA)
Chandra Challagonda (Fiware)
Edward Curry (Insight centre)
Davide Dalle Carbonare (Engineering)
Diego Collarana Vargas (Fraunhofer IAIS)
Oscar Corcho (UPM)
Mark Dietrich (EGI)
Valerio Frascolla (Intel)
Gonzalo Gil (Tekniker)
Robert Ginhör (Know-centre)
Marta Gutierrez (EGI)
Ernoe Kovacs (Nec lab)
Antonio Kung (Trialog)
Luca Malinverno (Porini)
Andrea Manzi (EGI)
Gregoris Mentzas (NTUA)
Irena Pavlova (Gate-ai)
Christian Racca (TOP-IX)
Daniel Saez Domingo (ITI)
Arjan Stoter (TNO)
Sarah Stryek (TU Graz)
Rizkallah Touma (i2CAT Foundation)
Tuomo Tuikka (VTT)
Daniel Vander Vorsl (Vicomtech)
Ray Walshe (DCU)
Rigo Wenning (ERCIM)

The core of this publication has been developed in 2022, with a revision and update in 2023 (Q3). Final public publishing date: December 2023.



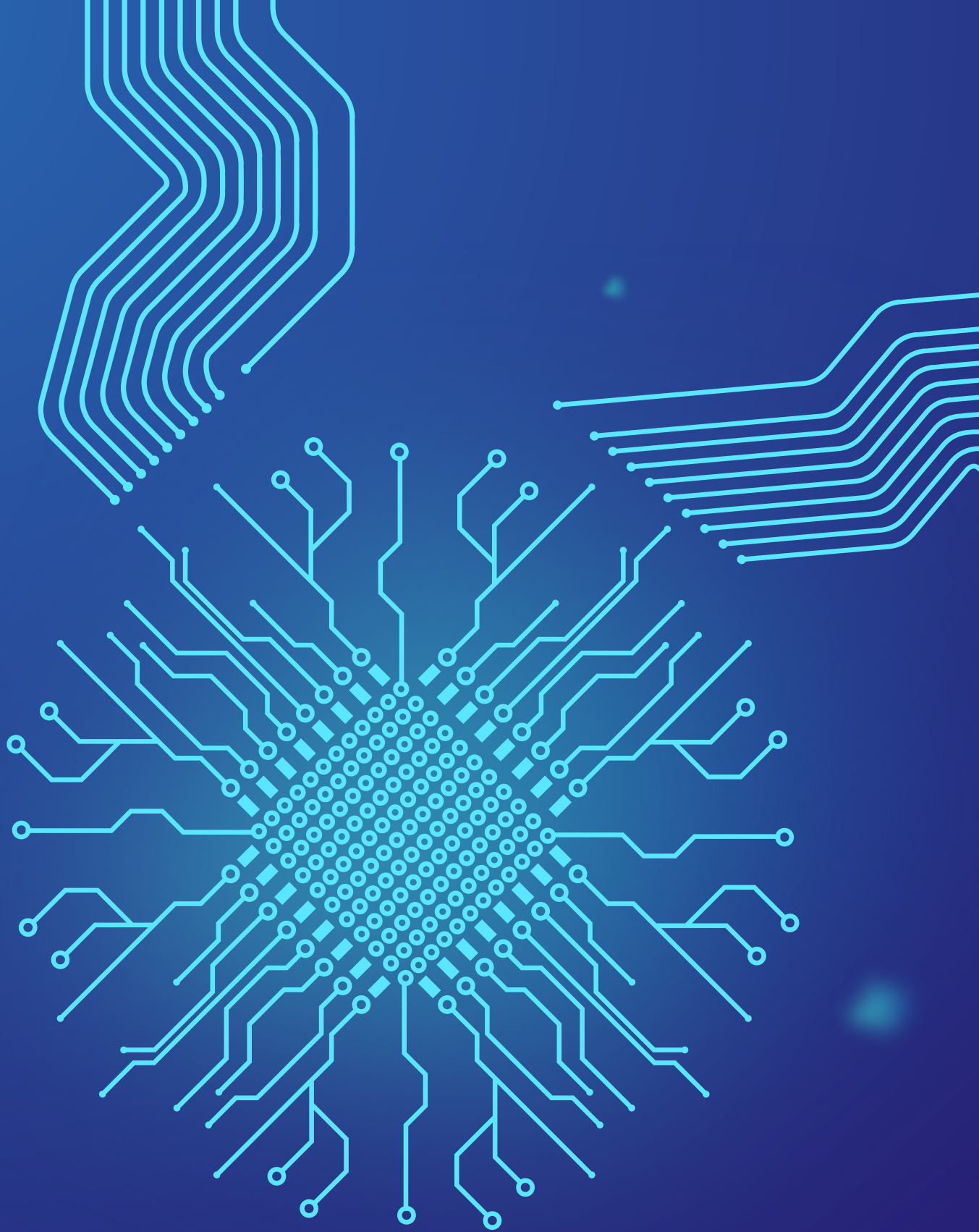
II Introduction

Data Spaces

Data Spaces are the technical-regulatory manifestation of the European grand vision that the benefits of data sharing must be done within the guidelines of European values, e.g., human rights, data sovereignty and fair use of data. While many organisations are working on technical standards for data spaces, it must be recognised that meta-data is the base for data space interoperability. Understanding this principle has led to the creation of this discussion paper and its emphasis on meta-data interoperability.

Structure of the publication

In the following chapters, we first summarise the most important definitions, standards and concepts (Chapter III “Context”). Then we dive deeper into the meta-data interoperability problem (Chapter IV) and define meta-data interoperability mechanisms (Chapter V). Next, we elaborate on the Meta-Data Interoperability Architecture (Chapter VI). Finally, we show how metadata interoperability facilitates ecosystems and how it can be used (“Outside Integration”, Chapter VII), how it is embedded into the current emerging Data Space architectures (“Inside Integration”, Chapter VIII) and describe important use cases (Chapter IX). At the end, we offer recommendations on a roadmap.



Definitions

It is important to align the definition of key terms, such as data, metadata, interoperability or data spaces, with definitions proposed by existing standards [4] to ensure common understanding. We add some specific comments to clarify distinctions.

Data and metadata

Table 1 lists definitions on the concept of data as they have been provided by ISO/IEC 20546:2019 (Big data – Overview and vocabulary).

Table 1 – Definitions on data

Data	Re-interpretable representation of information in a formalised manner suitable for communication, interpretation, or processing Note 1 to entry: Data can be processed by humans or by automatic means.
Data Analytics	Composite concept consisting of data acquisition, data collection, data validation, data processing, including data quantification, data visualisation and data interpretation
Data Processing	Systematic performance of operations upon data Note 1 to entry: Example: Arithmetic or logic operations upon data, merging or sorting of data, or operations on text, such as editing, sorting, merging, storing, retrieving, displaying, or printing.
Dataset	Identifiable collection of data available for access or download in one or more formats
Information	Data that are processed, organised and correlated to produce meaning. Note 1 to entry: Information concerns facts, concepts, objects, events, ideas, processes, etc.
Metadata	Data about data or data elements, possibly including their data descriptions and data about data ownership, access paths, access rights and data volatility

These definitions imply the following:

- Information includes data and metadata.
- Processing operations involve
 - accepting various input information,
 - applying calculations and transformations and
 - producing one or more output information.
- Processing operations involve metadata, which are parts of input information and which are created as part of output information and related to the processing of input information. Examples of metadata are purpose, provenance, or ownership.
- Other operations such as storing, retrieval, display, or printing are distinct operations that are no processing operations. They can also involve specific metadata. Examples of metadata are access rights, visibility, retention, replicability or cache-ability policies.

Interoperability and portability

Table 2 provides definitions on interoperability that have been proposed by ISO/IEC 19941:2017 (Cloud computing – interoperability and portability).

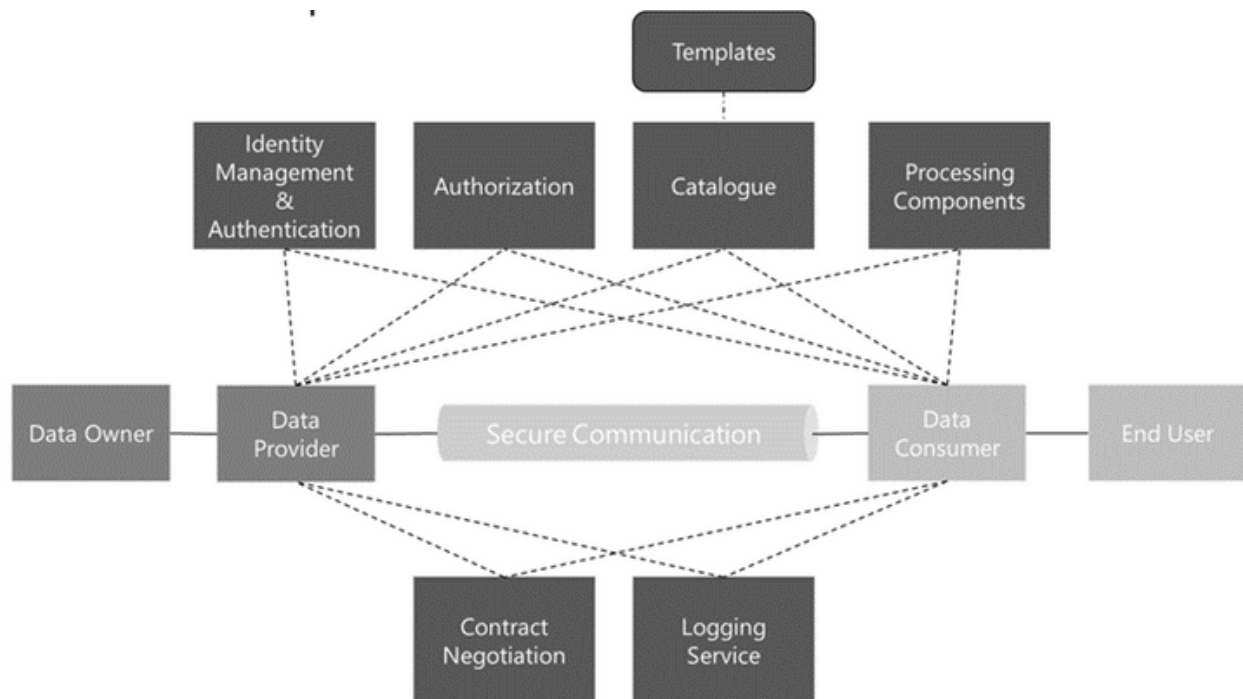
Table 2 – Definitions on interoperability and portability

Interoperability	Ability of two or more systems or applications to exchange information and to mutually use the information that has been exchanged.
Data Portability	Ability to easily transfer data from one system to another without being required to re-enter data

Data spaces

Data spaces are currently the subject of intensive discussions in standardisation and industry alliances. In order to ease the understanding of data spaces, we have developed an *“Abstract Data Space Architecture”*, described in Figure 1.

Figure 1 - Abstract Data Space Model



The abstract architecture shown in Figure 1 identifies the most commonly used building blocks for data spaces. We will use it in later sections to map existing approaches to the abstract architecture, allowing us to identify which modules of a data space are covered by a given standard and which are not.

The abstract architecture puts the “data processing chain” in the middle:

- Data Owners make data available to data providers.
- Data Providers make the data available to other participants of the data space using a secure communication channel.
- Data Consumers receive the data and make the data available to the end user of the data.
- As part of the data exchange, there will be a negotiated contract that defines who can access data (“Access Control”) and what a data consumer is allowed to do with the data (“Data Usage Control”). This contract may take the form of a bilateral agreement between the Data Holder/Provider and Data Consumer/End user, a multilateral agreement among all the parties to the Data Space combined with an exchange-specific record of the agreed conditions of access and use, or potentially the use of smart contracts as contemplated in the Data Governance Act [5].
- Logging services support the secure exchange of data by recording and potentially auditing the exchanged data helping the enforcement of the established contract and its policies.
- Identity Management and Authentication support secure communication in synergy with a module for authorisation.
- The catalogue component allows consumers to understand the available data, for example, which ontology is used or where the data is stored.
- Further components can be used for enabling (potentially restricted and secured) processing of the data.

The term data space has been used in a variety of contexts, e.g., as a term to describe the concept of allowing disparate data sources to be integrated without requiring upfront semantic alignment of the data schema [6] of each source, or in the Fraunhofer Industrial Data Space Initiative in 2014 and subsequently incorporated into a scheme for inter-organisational data sharing that has since been championed by the International Data Spaces Association [7] and first instantiated in 2017 [8]. The concept of data space has since then been defined in several ways. Table 3 provides excerpts found in various references.

[5] <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52020PC0767>

[6] Halevy, A., Franklin, M. & Maier, D. (2006). Dataspaces: A New Abstraction for Information Management.. Sigmod Record. 34. 1-2.

[7] International Data Spaces Association, <https://internationaldataspaces.org/>

[8] Otto, B., Lohmann, S. Auer, S. (2017). Reference Architecture Model for the Industrial Data Space. <https://doi:10.13140/RG.2.2.17352.11529>.

Table 3 - Existing definitions of data spaces
DEFINITIONS OF DATA SPACES

OpenDEI design paper [9]	<p>→ From a technical perspective, a data space can be seen as a data integration concept which does not require common database schemas and physical data integration, but is rather based on distributed data stores and integration on an “as needed” basis on a semantic level. Abstracted from this technical definition, a data space can be defined as a federated data ecosystem within a certain application domain and based on shared policies and rules</p>
BDVA position paper v1 [10] BDVA position paper v2 [11]	<p>→ Data Spaces (v1) Ecosystem of data models, datasets, ontologies, data sharing contracts and specialised management services (i.e., as often provided by data centres, stores, repositories, individually or within ‘data lakes’), together with soft competencies around it (i.e., governance, social interactions, business processes)</p> <p>European governed data space (v2) Singular but federated virtual space connecting several other interoperable spaces</p>
IDSA[12]	<p>→ Data Space Architecture model for data integration; characterised by distributed management of data from multiple data sources and by not using a common semantic model</p> <p>International Data Spaces Distributed network of Data Endpoints (i.e., instantiations of the International Data Spaces Connector), allowing secure exchange of data and guaranteeing Data Sovereignty</p>
GAIA-X architecture document [13]	<p>→ A Data Space is a virtual data integration concept defined as a set of participants and a set of relationships among them, where participants provide their data resources and computing services and data are made available in a decentralised manner, for example, to combine and share data of stored in different cloud storage backends.</p> <p>Data Spaces have the following design principles:</p> <ul style="list-style-type: none"> • data resides in its sources; • only semantic integration of data and no common data schema; • nesting and overlaps are possible; • spontaneous networking of data, data visiting and coexistence of data are enabled. <p>Within one Data Ecosystem, several Data Spaces can emerge.</p>

[9] <https://design-principles-for-data-spaces.org/>

[10] Towards a European-Governed Data Sharing Space. Enabling data exchange and unlocking AI potential. April 2019 https://bdva.eu/sites/default/files/BdVA%20DataSharingSpace%20PositionPaper_April2019_V1.pdf

[11] Towards a European-Governed Data Sharing Space. Enabling data exchange and unlocking AI potential. November 2020 https://www.bdva.eu/sites/default/files/BdVA%20DataSharingSpaces%20PositionPaper%20V2_2020_Final.pdf

[12] <https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/>

[13] <https://gaia-x.eu/wp-content/uploads/2022/06/Gaia-x-Architecture-Document-22.04-Release.pdf>

The following definitions have been suggested while preparing this report:

- A **Data Ecosystem** is a “purposeful collaboration or partnership consuming, producing and providing interoperable data and related resources”. This definition aligns with a number of articles in the literature on data ecosystems [14] [15] [16] [17]. The “purposeful” characteristic aligns with SITRA [18] and the Data Sharing Coalition [19], as does the implication that the participants are identified a priori. A trust and governance framework is implied but may not be formal. Many effective data ecosystems (e.g., in research) work well based only on a strong community of like-minded researchers.
- A **Data Space** is a “public collection of findable, accessible, interoperable and reusable (FAIR), quality data and related resources consumed, produced and provided by identified participants, each respecting societal values and operating within an explicit framework of trust and governance”. This definition builds on the data ecosystem definition but is extended to include concepts introduced by the EC [20] when it discusses data spaces. Note that both the purpose and the idea of a defined (possibly restricted) community are removed from the definition, although data providers and consumers are still identified. The definition of data ecosystems implies that participants are identified a priori. It can involve a specific trust and governance framework. This framework can be an informal framework, as for instance in research data ecosystems which are based only on a strong community of like-minded researchers.

Distinct trust and governance frameworks are defined for each data ecosystem and data space. Consequently, it may not be possible to transfer data that might exist in one data space or data ecosystem to another ecosystem/space without agreement at the governance level.

These definitions are aligned with ISO/IEC 27570 (Privacy guidelines for smart cities) [21] which further defines five collaboration processes: governance, data management, risk management, engineering and citizen engagement.

[14] Otto, B., Lis, D., Jürjens, J. et al. (2019). Data Ecosystems. Conceptual Foundations, Constituents and Recommendations for Action.

[15] Lury S. Oliveira, M. & Barros Lima, G. & Lóscio, B. (2019). Investigations into Data Ecosystems: a systematic mapping study. Knowledge and Information Systems. 61. <https://doi.org/10.1007/s10115-018-1323-6>.

[16] Lury S. Oliveira, M. and Lóscio, B. (2018). What is a data ecosystem? In Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age (dg.o '18). Association for Computing Machinery, New York, NY, USA, Article 74, 1–9. DOI: <https://doi.org/10.1145/3209281.3209335>. This article also references earlier definitions.

[17] Lis, D. and Otto, B., (2020) "Data Governance in Data Ecosystems – Insights from Organizations". AMCIS 2020 Proceedings. 12. https://aisel.aisnet.org/amcis2020/strategic_uses_it/strategic_uses_it/12

[18] SITRA. <https://www.sitra.fi/en/publications/rulebook-for-a-fair-data-economy/#download-the-rulebook>

[19] Data Sharing Coalition. <https://datasharingcoalition.eu/app/uploads/2021/04/data-sharing-canvas-30-04-2021.pdf>

[20] “A European strategy for data” COM(2020) 66 final. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066&qid=1619802547376>

[21] ISO/IEC TS 27570:2021 Privacy protection — Privacy guidelines for smart cities. <https://www.iso.org/obp/ui/#iso:std:iso-iec:ts:27570:ed-1:v1:en>

Alternatively, other contributions focus on describing the architecture or the features of a data space, as shown in Table 4.

Table 4 - Existing descriptions of data spaces

DESCRIPTIONS OF DATA SPACES	
GAIA-X technical architecture [22]	<p style="text-align: center;">→</p> <p>This paper provides an architecture description: GAIA-X is set to be an Infrastructure and Data Ecosystem according to European values and standards. This overall mission drives its architecture. The architecture employs digital processes and information technology to facilitate the interconnection between all participants in the European digital economy. By leveraging existing standards, open technology and concepts, it enables open, consistent, quality-assured and easy-to-use innovative data exchange and services. Additionally, GAIA-X will become a facilitator for interoperability and interconnection between its Participants, for data as well as services</p>
European Commission working document [23]	<p style="text-align: center;">→</p> <p>Instead of defining a data space, this paper provides a description of the features that are present in a common data space:</p> <ul style="list-style-type: none"> • A secure and privacy-preserving infrastructure to pool, access, share, process and use data. • A clear and practical structure for access to and use of data in a fair, transparent, proportionate and/non-discriminatory manner and clear and trustworthy data governance mechanisms. • European rules and values, in particular personal data protection, consumer protection legislation and competition law, are fully respected • Data holders will have the possibility, in the data space, to grant access to or to share certain personal or non-personal data under their control. • Data that is made available can be reused against compensation, including remuneration, or for free. • Participation of an open number of organisations / individuals

Interoperability as a key concern in data spaces

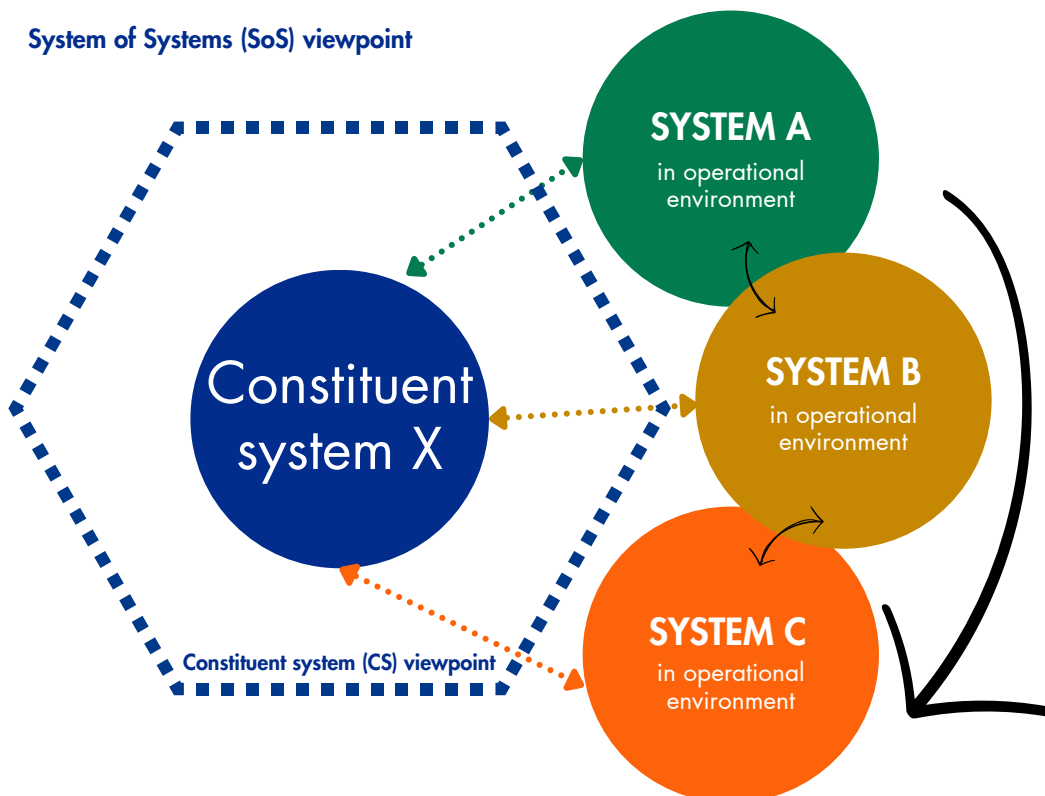
Ecosystems and interoperability

Ecosystems

An ecosystem is defined by ISO/IEC 27570 [24] as an infrastructure and services based on a network of organisations and stakeholders. This term is used to address complex Systems of system (SoS).

Figure 2 shows a system, which interacts with other systems in an operational environment: each system in an SoS is operated and managed independently and it can be distributed geographically; the collective operation of each individual system can cause a more complex behavior, which is called emergent behaviour.

Figure 2 – System of systems (source ISO/IEC/IEEE 21839) [25]

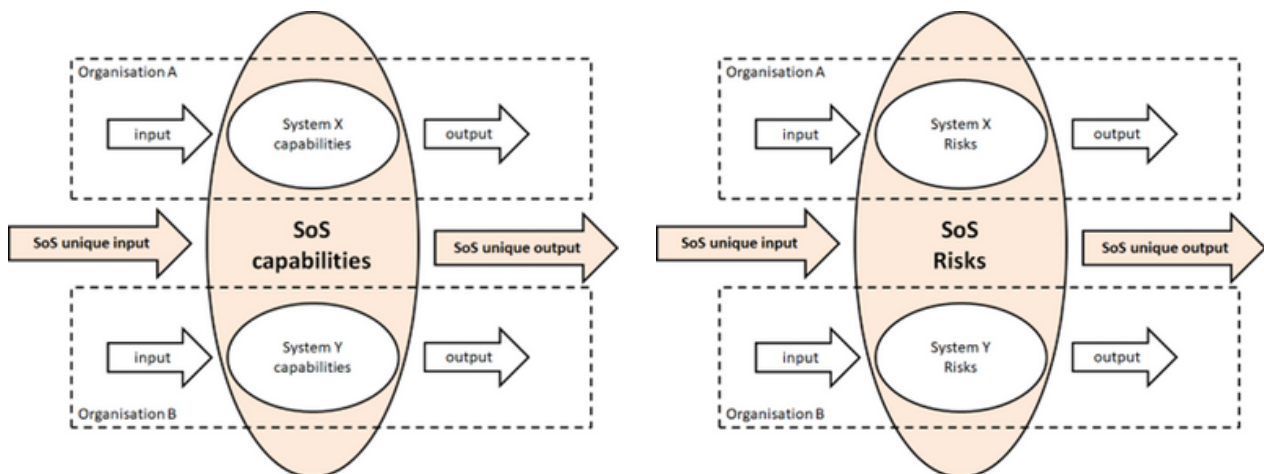


[24] ISO/IEC TS 27570:2021 Privacy protection — Privacy guidelines for smart cities. <https://www.iso.org/standard/71678.html>
 [25] ISO/IEC/IEEE 21839:2019 Systems and software engineering — System of systems (SoS) considerations in life cycle stages of a system. <https://www.iso.org/obp/ui/#iso:std:iso-iec-ieee:21839:ed-1:v1:en>

The impact of SoS capabilities in terms of trustworthiness has to be evaluated carefully: the combination of system X with system Y creates an emerging effect, at the level of the resulting SoS capability, but also at the level of the resulting SoS risks as shown in Figure 3:

- The picture on the left shows the capability of the SoS, consisting of system X capability, system Y capability and the resulting SoS emergent capability.
- The picture on the right shows the risks of the SoS, consisting of system X risks, system Y risks and the resulting SoS emerging risks

Figure 3 – Emergent capability (left) and emergent risks (right)



Points of interoperability in ecosystems

Interoperability is the ability for two or more systems or applications to exchange information and to mutually use the information that has been exchanged. A point of interoperability (PI) is the artefact in a system architecture that focuses on this exchange. Figure 4 shows a conceptual model of a PI:

- the owner/organisation operates a system-of-interest,
- the system-of-interest support a point of interoperability and
- the owner, system-of-interest and point of interoperability belong to an ecosystem.

Figure 4 – Point of interoperability (PI) conceptual model

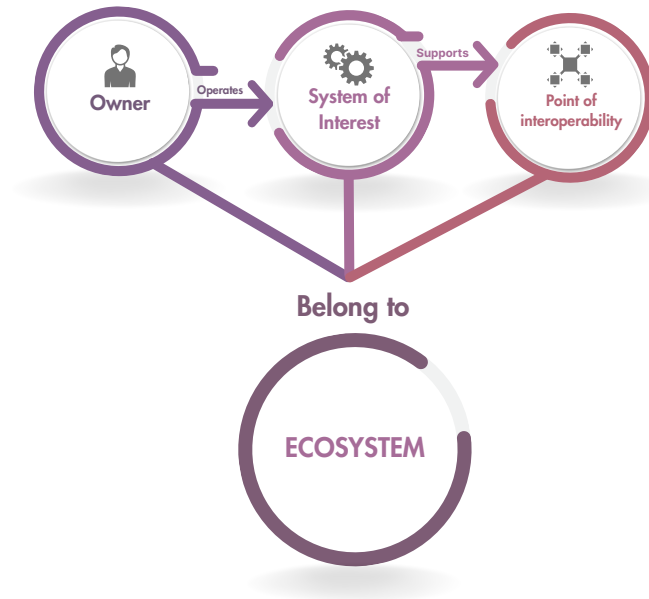


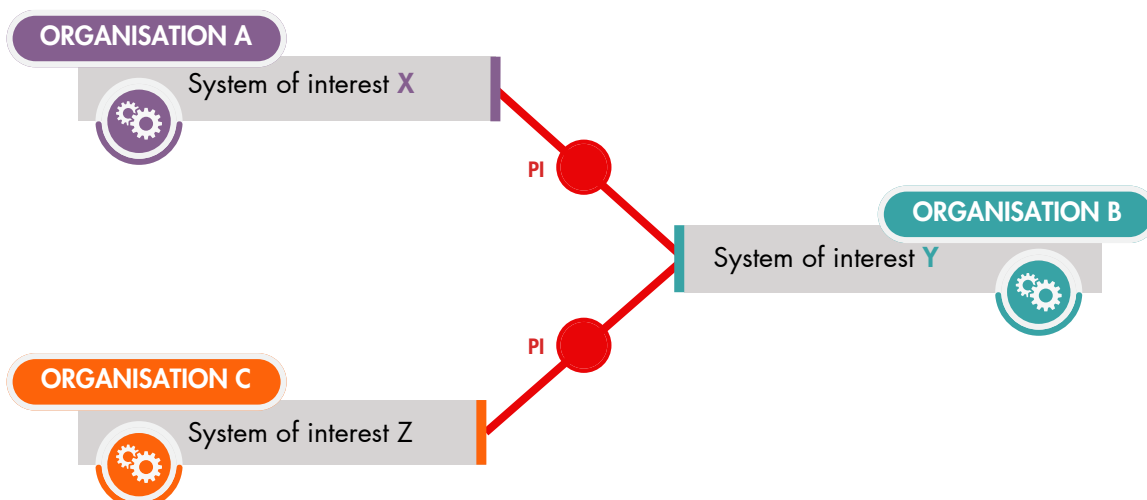
Figure 5 shows the conventions used to describe PIs:

- coloured boxes are owners/organisations,
- grey boxes are systems and
- red circles are PIs

The example on the right shows the following:

- system X, Y, Z are operated by A, B, C;
- system X and system Y interoperate using one PI; and
- system Y and system Z interoperate using another PI.

Figure 5 – Notation used to describe a point of interoperability (PI) and example



Metadata in points of interoperability

A PI can be used to exchange metadata, or information on data. Examples of metadata could be

- policy and governance information (e.g., plans, roadmaps, rules, behavior);
- business protection requirements information (e.g., IPR obligations, contracts);
- trustworthiness requirements information (e.g., security, privacy, safety, resilience, reliability); or
- usage information (e.g., consent, privacy preference)
- assurance information (e.g., evidence of enforcement)

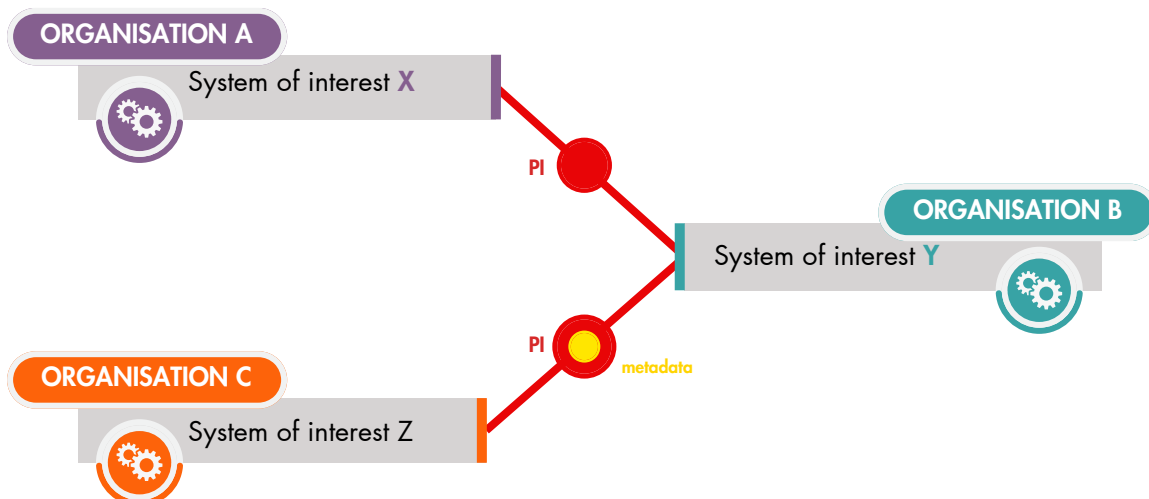
Figure 6 shows the conventions used to describe PIs involving metadata:

- coloured boxes are owner/organisations,
- grey boxes are systems,
- yellow circles are metadata
- red circles are PIs.

The example on the right shows the following:

- system X, Y, Z are operated by A, B, C;
- system X and system Y interoperate using one PI
- system Y and system Z interoperate using one PI involving metadata exchange.

Figure 6 – Notation used to describe a PI involving metadata and example



Example of privacy

Figure 7 shows an example of metadata exchange in a smart city from a privacy management point of view. It involves exchange of metadata of the following types:

- privacy policies,
- privacy risk information,
- privacy protection capabilities and
- transparency capabilities.

Figure 7 – Metadata for privacy management in smart cities

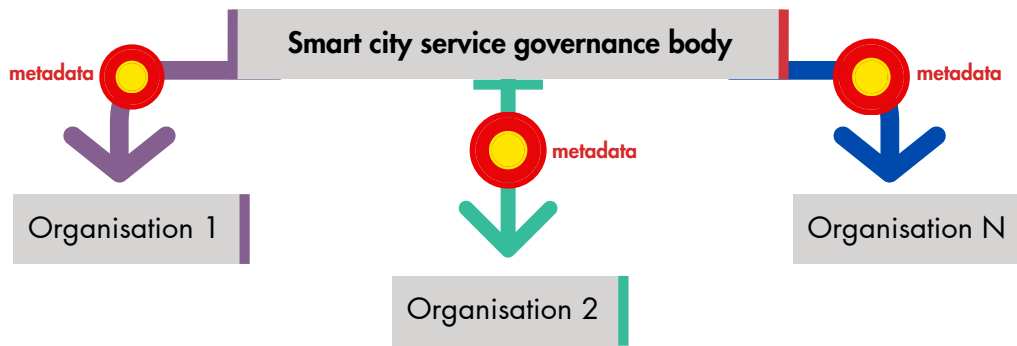
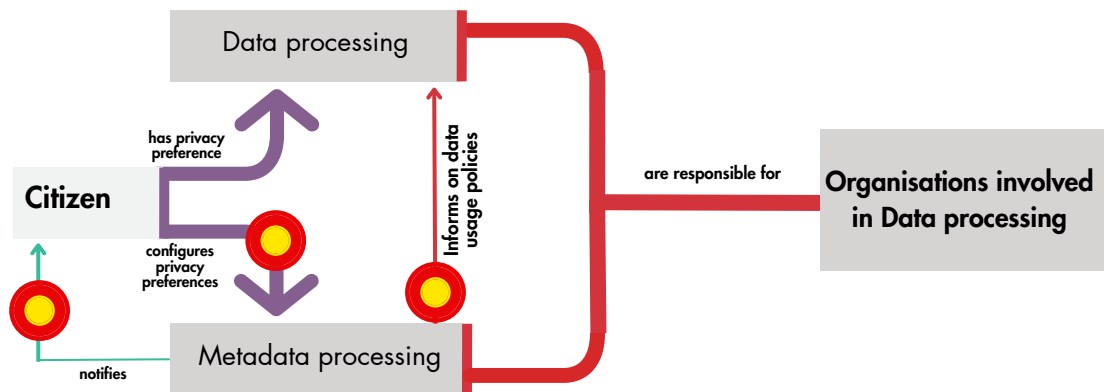


Figure 8 shows metadata exchanges concerning privacy preference management [26]. They involve exchange between citizens (and a system):

- citizens can configure privacy preferences,
- organisations can notify citizens about events (such as a privacy breach), and
- the metadata processing component can inform the data processing component on data usage policies.

Figure 8 – Metadata for privacy preferences management

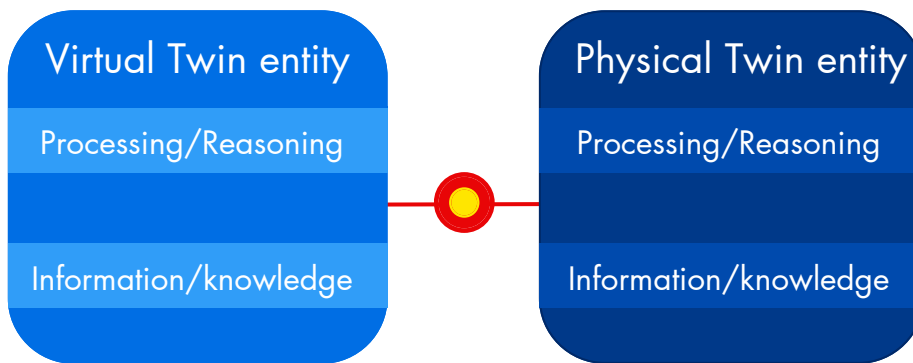


[26] ISO/IEC 27556 Information security, cybersecurity and privacy protection – User-centric privacy preferences management framework. <https://www.iso.org/standard/71674.html>

Example of digital twins

Figure 9 shows an example of metadata exchange in digital twins. The digital twin system includes a virtual twin entity and a physical twin entity. Each twin accesses and updates information and knowledge, using processing and reasoning capabilities. Metadata exchange can involve new/modified policies, knowledge, configurations.

Figure 9 – Metadata for digital twins



Data lifecycle

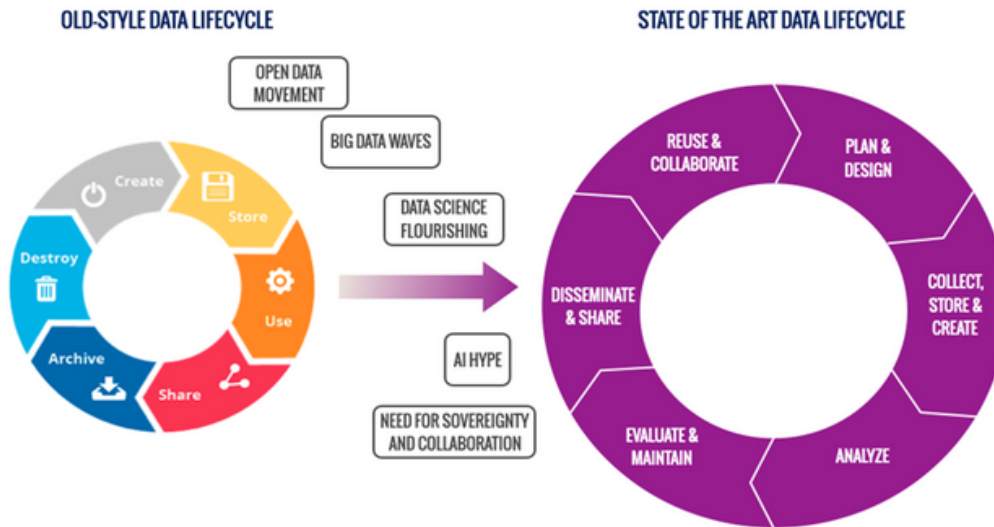
Data today can be augmented with external data sources and manual or automatic annotation. Data can also be obtained, through totally external datasets, acquired or purchased through specific channels (marketplace, data portals, data brokerage services).

Once ingested, the data are processed, analysed and enhanced with the most recent Data Science techniques. From the data and through algorithms we (or sometimes the machines) generate predictions or classifications which can be the basis for making more informed and aware decisions.

Therefore, the data exploitation is not typically for a single use, the data can be manipulated, reused by the same organisation for different purposes or made available to third parties through open access channels or specific collaboration agreements. The reuse of data can take place both outside and within the same organisation that generated them, for example by different departments or business units.

Finally, there can be multiple data lifecycles based on different scenarios: internal and external, profit and non-profit with different stakeholders.

Figure 10 compares an “old-style” data lifecycle against a more recent approach that takes into account current market needs (reuse, sharing, collaboration). Note that ISO/IEC JTC1/SC42 is developing a standard on the topic [27].

Figure 10 – Data Lifecycle “evolution”


Guidance on data lifecycle can be provided by several vendor-agnostic, not-for-profit associations, such as the Data Governance Institute (DGI), the Data Management Association (DAMA), the Data Governance Professionals Organisation (DGPO) and the Enterprise Data Management Council, all of which provide great representations of Data Lifecycle and Data Governance process.

It is important to remark that Data Lifecycle management is strongly dependent on the field of usage. A nice example in this sense is represented by the following research study in the field of biomedical and clinical data management. The Biomedical Data Lifecycle is therefore a representation of stages that occur in research in regards to how data is collected, used and stored.

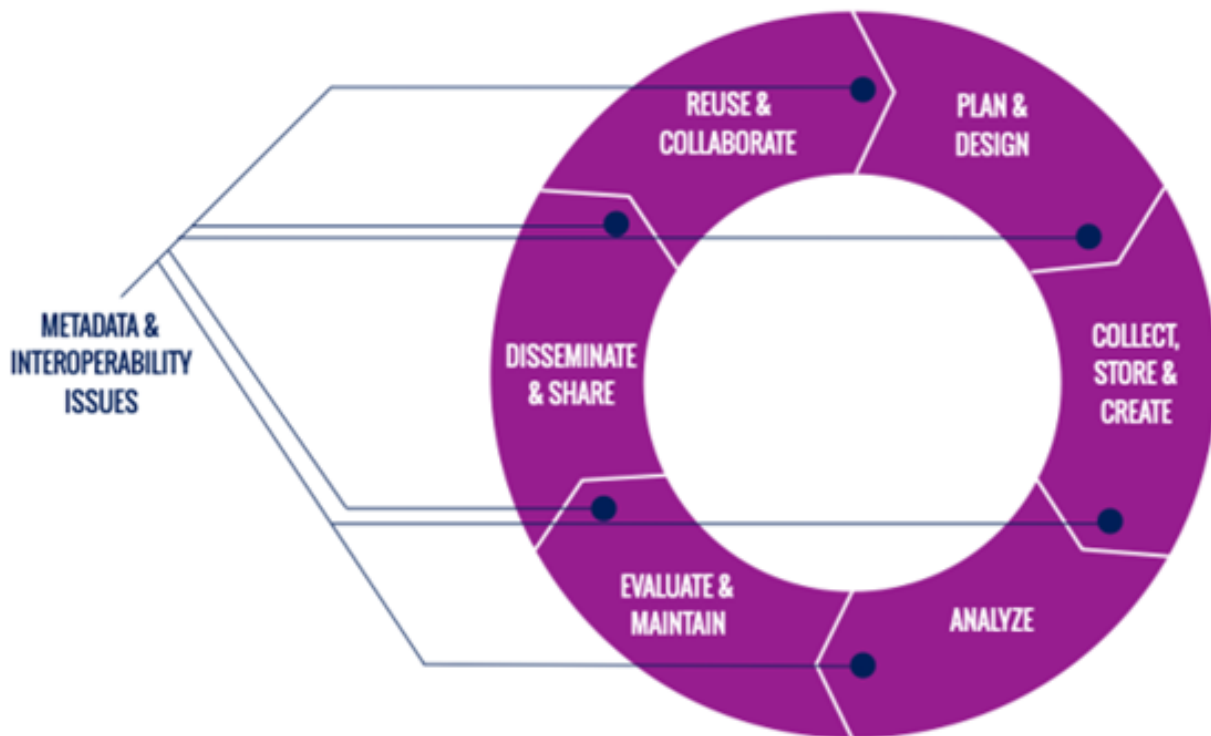
Figure 11 – Example of data lifecycle [28]


[28] This picture on Research Data Lifecycle from LMA Research Data Management Working Group is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Metadata and Interoperability play an important role in all the stages of Data Lifecycle as represented in Figure 12:

- Plan and design stage: metadata can describe the meaning of data and purpose of processing such as environmental data and pollution measurement
- Collect, store and created stage: metadata can describe collecting and storage policy such consent for personal data
- Analyse stage: metadata can describe analysis methods, rationales for results (e.g., transparency)
- Evaluate and maintain stage: metadata can describe evaluation approaches, auditability features
- Disseminate and share stage: metadata can describe access and usage rights
- Reuse and collaborate stage: metadata can describe governance policies

Figure 12 – Metadata and interoperability issues in the data lifecycle



The data life cycle is a true cycle -- data can be used and potentially re-used and data can be used (i.e., processed or analysed) to create new data -- continuing the cycle. Metadata is critical to make the “connections” from one item of data to the next, to link both data and relevant processing capabilities (services, software) and to support discovery of any item of data. Metadata is the glue that makes the data life cycle possible.



IV Value of Metadata Interoperability

This section explains how metadata can help create value in data sharing spaces.

Enabling the Data Sharing Value Wheel

BDVA published in November 2020 a position paper [29] promoting a vision based on the data sharing value wheel showed in Figure 13. The wheel visualises five items:

- data based on FAIR,
- governance,
- people,
- organisations,
- and technology.

Figure 13 – Data sharing wheel



The wheel is a representation that data life cycle is a true cycle:

- data can be used and reused
- data can be used (processed/ analysed) to create new data, thus continuing the cycle.

Metadata is critical to make the connections from one item of data to the next, to link both data and relevant processing capabilities (services, software) and to support discovery of any item of data.

Metadata sharing and interoperability is the instrument to enable this wheel.

Enabling FAIR Data

The FAIR Data Guiding Principles [30] were developed with the specific intent to “... act as a guideline for those wishing to enhance the reusability of their data holdings”. The “I” of FAIR stands for “interoperable” -- and both data and metadata describing data must be interoperable. Interoperability for data and metadata spans all the “layers” of the interoperability framework (see section 5), from the bottom (transport, syntactic, semantic) to the top (behavioural, policy). Semantic interoperability allows the meaning of any data to be properly understood, making use and re-use possible and enabling the creation of value. At another layer, policy interoperability gives users confidence that they are allowed to use the data in the desired way and defines how any derived data products can be used in the future.

Following the FAIR principles requires the exchange of related metadata:

- Findable: information on how data is identified and searchable
- Accessible: information on how and whether data is accessible
- Interoperable: information on format
- Reusable: information on provenance, license, format

Enabling Governance

As noted above, it is possible to have multiple data ecosystems, data spaces and other structures for data sharing. A given organisation or a given private individual might choose to participate in multiple data sharing initiatives (potentially acting as a data provider and/or a data consumer in each). Similarly, a given item of data might be exposed by its holder to the same variety of data sharing initiatives. The governance of each initiative is a critical factor helping data holders decide what initiatives to join and which initiatives they will expose data to.

Governance metadata allows potential participants to share policies and to evaluate whether they wish to join certain data ecosystems and whether they wish to provide data to various data sharing ecosystems (both closed and open). The governance structures of each data sharing initiative should be described in a consistent way using agreed metadata structures. This also allows two data sharing initiatives to evaluate their compatibility with one another and decide whether to engage in harmonisation efforts (such as described in more detail by the Data Sharing Coalition [31]).

Empowering People

Metadata plays an essential role in enabling and protecting individuals and their rights in a jurisdiction that want to be fit for the digital age. Metadata plays a strong role in effective identity management systems, extending the ability of individuals to “live” digitally by correctly identifying data over which individuals have rights (e.g., personal data), facilitating access to a growing number of valuable services, and augmenting digital personas with attributes and certificates, such as verified credentials [32], that give individuals more power to thrive digitally.

Improving Organisation Operations

Metadata exchange can enable the convergence towards common practices and values for organisations as are mentioned above for individuals. In addition, metadata “maturity” is critical for organisations that hope to use data they generate themselves or re-use data provided by others. In both cases, organisations need to keep careful track of the rights associated with each item of data (e.g., data subjects (term use in the Data Act)).

[32] <https://www.w3.org/2017/vc/WG/> and <https://www.w3.org/TR/vc-data-model/>

Technology Enablers

In early 2000, XML [33], itself derived from ISO SGML [34], brought a boost to interoperability for data exchanges. Data was extracted from the silos and transformed into XML in order to be shared. On the receiving side, the system had and still has a transformation module that allows it to import the XML data into the receiving system. Over the years, many other formats were born and the use of those formats is now a matter of zeitgeist and trends. Currently, the JSON format [35] is en vogue. And nowadays, applications even natively use interoperable data formats without needing an additional transformation step.

But with the interoperability induced increase in data sharing, with the commercialisation of data sharing, with the creation of data value chains, social constraints and properties of the shared data became more important. The notion of data space with its load of constraints and all the needed additional information mentioned in the previous chapters can be technically seen as properties of data. Those properties can determine the permissions for processing and downstream sharing. Properties on data quality can be important to determine whether data is exploitable for a certain field of use, e.g., in the medical sector. Limitations on collection and sharing, like GDPR, but also commercial secrets and other use limitations are now subject to complex legal contracts that distribute the liability in case data hasn't the properties previously agreed to in complex and costly negotiations.



V The Mechanics of Metadata Interoperability

Creating knowledge on data

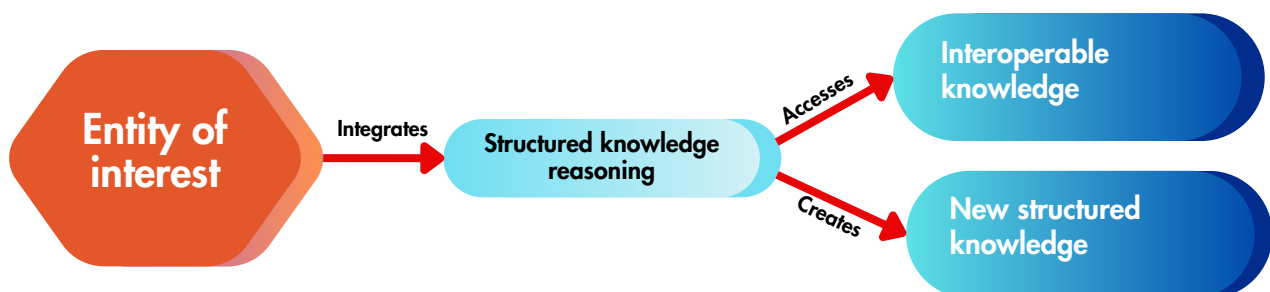
Section *Ecosystems and interoperability* showed that metadata can be considered as knowledge on data. The process creation is as follows:

- data is collected;
- Information on the context is gained;
- knowledge on a given topic is gained through conventional learning (experience and lessons learned, study, ...);
- a representation of this knowledge is created in order to make available structure knowledge on topic, e.g., using natural language processing (NLP);
- this representation must be annotated/adapted so that it can be exchanged.

Figure 14 shows the resulting effect of allowing for metadata exchange:

- An entity of interest (e.g., a digital twin, a connected vehicle application, an fintech application) integrates reasoning capability
- The entity of interest accesses structured knowledge, i.e., knowledge which has a format that it can use)
- The entity of interest creates new knowledge.

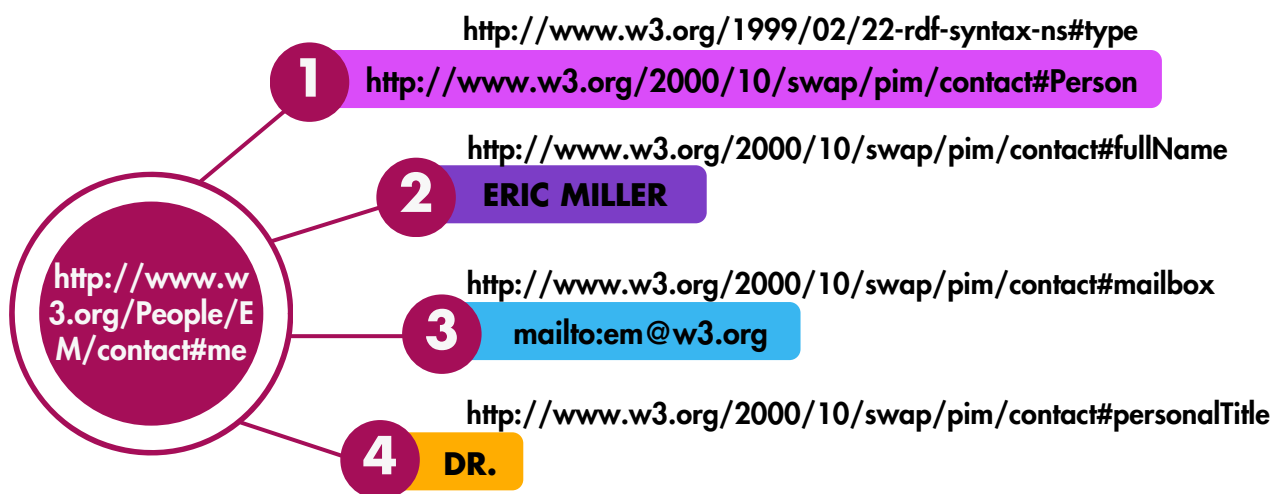
Figure 14 – Creation of knowledge based on existing interoperable knowledge



Representing Knowledge on the Web

In order to add properties to data and attach them persistently, knowledge representation can be used. Knowledge representation is a field of AI dedicated to representing information about the world in a form that a computer can use to solve complex tasks [36]. The EU has a long history of very successful research in this area. With the advent of the web as a global information system, the web of data as a global knowledge system received a considerable boost. People realised very quickly that links between web pages were not only pointers, but often expressed more, e.g., the fact that someone considered a web page to be important. This phenomenon was used by early versions of Google to create their page rank algorithm. Ramanathan V. Guha used knowledge representation in combination with the emerging XML language to create a precursor of RDF, the Resource Description Framework [37]. The initial goal of RDF was to represent metadata about Web resources, such as the title, author, and modification date of a Web page, copyright and licensing information about a Web document, or the availability schedule for some shared resource. However, by generalising the concept of a “Web resource”, RDF can also be used to represent information about things that can be identified digitally, even when they cannot be directly accessed digitally [38].

Figure 15 – Knowledge as web resource



[36] Wikipedia https://en.wikipedia.org/wiki/Knowledge_representation_and_reasoning visited 2022-03-11

[37] Wikipedia https://en.wikipedia.org/wiki/Ramanathan_V._Guha visited 2022-03-11

[38] RDF Primer 1.0 <https://www.w3.org/TR/rdf-primer/>

Annotation and identification capability

If things are identified on the web, RDF can make statements about them. And if the thing is a data object, RDF can make statements about that data, it can annotate data. The challenge turns into identifying objects in order to be able to talk about them in machine readable ways using knowledge representation. And as for the web of documents, the unique global resource identifier (URI) is a necessary condition for the system to function. Identifying an object via its URI is not a necessary condition for closed systems, even if using knowledge representation. But as soon as information leaves a system and is shared beyond its own boundaries, local identifiers have to be disambiguated and turned into URIs. This disambiguation is often called semantic lifting.

For the receiving side to understand the delivered metadata, syntactic interoperability is not enough. It is not sufficient to just encode things the same way, say in JSON or XML using UTF-8. The semantic meaning of the metadata must be understood by the receiving machine. In order to understand the received data, *the receiving system must gain knowledge about the meaning of the received elements. And this meaning can be encoded in taxonomies, vocabularies and ontologies. Those represent an agreement between actors to encode a certain meaning in a certain way. Hence the enormous need to standardize those agreements and allow them to be reusable to the highest extent possible.* But this does not argue for unification. Because understanding a term does not prescribe a specific reaction to that term by the system. And this allows to create knowledge about the environment the system is operating in, while allowing for an infinite variety on how to react to the environment. And this is key in systems facing society and having social significance.

Organising Interoperable Exchange

Taking into account what has been said so far, semantic interoperability is a challenging task. Data and properties of data have to be gathered and put in relation to each other. Categorisation of data and metadata needs to be organised. Having data categorised by hand does not scale. Categories and their meaning need to be modeled. Data and metadata have to be transported around the system without losing their relation. And a reaction to the system should create new metadata that is added to the system and trigger appropriate reactions. Consequently, metadata interoperability needs consideration at data ingestion time, data processing time and upon further interactions with the environment.

Acquiring Metadata Automatically

Scaling up to the data economy, metadata has to be acquired automatically: There is therefore a need for semantisation during data acquisition.

As the data economy goes hand in hand with the emergence of the Internet of Things (IoT), the things that generate data must be enabled to also provide semantic information about the data they provide. This way, a system can automatically acquire the metadata needed. This can be either done by having a sensor itself providing semantic information in the payload it provides, or it can be provided by some middleware that watches a sensor network and adds semantic information before giving data to the next step.



**VI Metadata interoperability
for data sharing spaces**

Overview of viewpoints

Interoperability Framework Viewpoint

Metadata information could be defined as a refinement of interoperability frameworks. Table 5 shows definitions that have been proposed concerning interoperability frameworks at standardisation level [39].

Table 5 – Interoperability facets

Transport interoperability	interoperability where information exchange uses an established communication infrastructure between the participating systems
Syntactic interoperability	interoperability such that the formats of the exchanged information can be understood by the participating systems
Semantic interoperability	interoperability so that the meaning of the data model within the context of a subject area is understood by the participating systems
Behavioural interoperability	interoperability so that the actual result of the exchange achieves the expected outcome
Policy interoperability	interoperability while complying with the legal, organisational and policy frameworks applicable to the participating systems

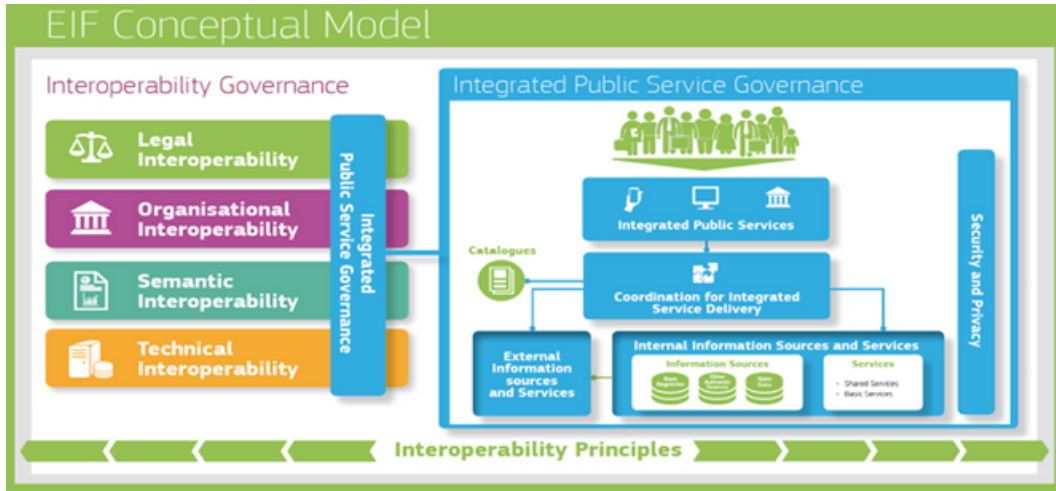
The European Interoperability Framework [40] describes an interoperability model shown in Figure 16 which includes:

- four layers of interoperability: legal, organisational, semantic and technical;
- a cross-cutting component of the four layers, ‘integrated public service governance’; and
- a background layer, ‘interoperability governance’.

[39] ISO/IEC 19941:2017 (Cloud computing – interoperability and portability) and ISO/IEC 21823-1:2019 (Interoperability for IoT systems – Part 1: Framework)

[40] <https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/european-interoperability-framework-detail>

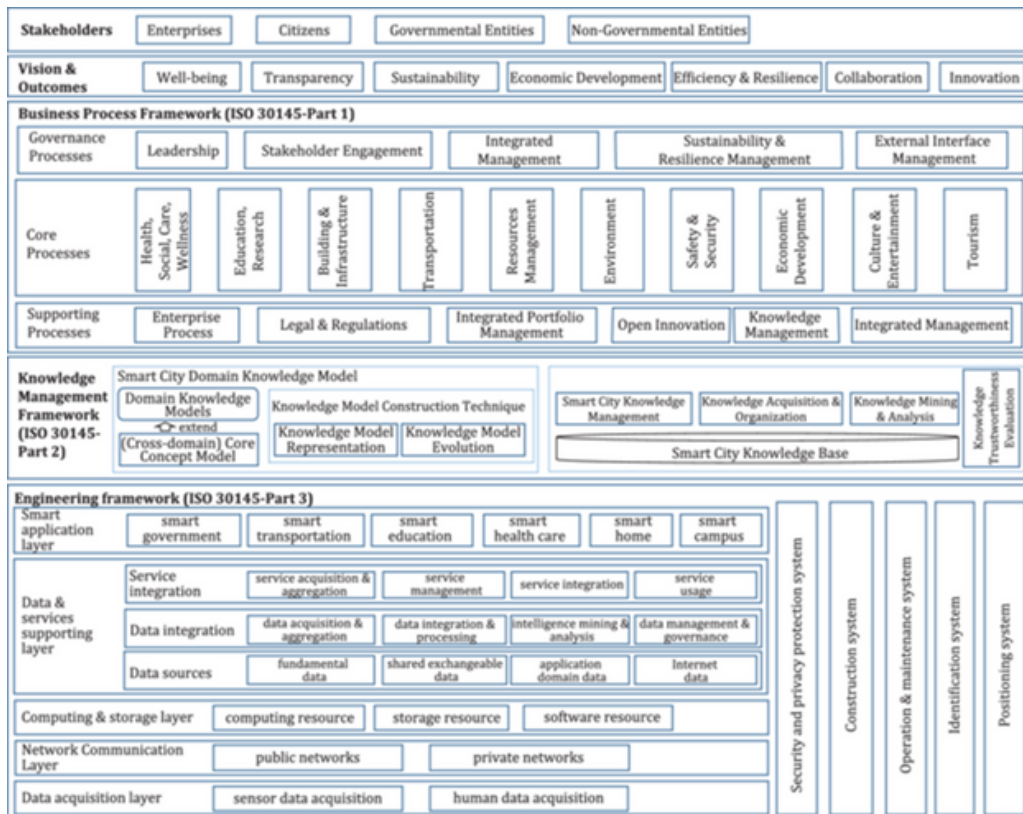
Figure 16 – European Interoperability framework



The Daunting Challenge of Smart Cities

Figure 17 shows the overall smart city framework as proposed by the ISO/IEC 30145 Smart city reference framework [41], highlighting the next challenge, exchanging metadata at the cross-domain level.

Figure 17 – ISO/IEC 30145 Smart city framework



Cyber Physical Systems Viewpoint

The exchange of data in a cyber physical system may require the exchange of specific behavioural information. The information can be related to concerns as listed in the NIST framework for cyber-physical systems [42]. This is showed in Table 6.

Table 6 – Cyber physical system concerns

CATEGORY OF CONCERN	DESCRIPTION
Functional	→ Concerns about function including sensing, actuation, control, communications, physicality, etc.
Business	→ Concerns about enterprise, time to market, environment, regulation, cost, etc.
Human	→ Concerns about human interaction with and as part of a CPS
Trustworthiness	→ Concerns about trustworthiness of CPS including security, privacy, safety, reliability, and resilience.
Timing	→ Concerns about time and frequency in CPS, including the generation and transport of time and frequency signals, timestamping, managing latency, timing composability, etc.
Data	→ Concerns about data interoperability including fusion, metadata, type, identity, etc.
Boundaries	→ Concerns related to demarcations of topological, functional, organisational, or other forms of interactions.
Composition	→ Concerns related to the ability to compute selected properties of a component assembly from the properties of its components. Compositionality requires components that are composable: they do not change their properties in an assembly. Timing composability is particularly difficult.
Lifecycle	→ Concerns about the lifecycle of CPS including its components.

[42] NIST special publication 1500-201, June 2017 <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-201.pdf>

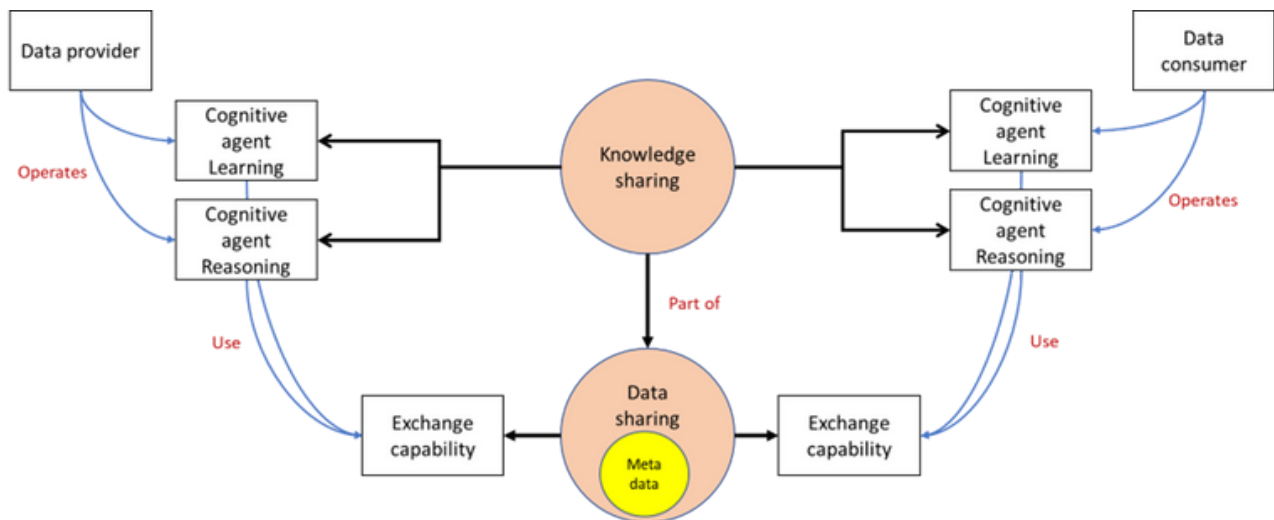
AI viewpoint

Figure 18 shows how data sharing takes place when knowledge is involved:

- Data sharing involved knowledge sharing between a data provider and a data consumer,
- Data providers and data consumers operate two agents: a learning agent and a reasoning agent.
- The agents are involved in knowledge sharing, either to learn together (federated learning), or to reason together (federated reasoning),
- Agents requires the use of exchange capability,
- The exchange capability provides data sharing, including metadata.

From an AI point of view, data will consist of core knowledge (e.g., knowledge about a patient) while metadata will contain information on the modalities of knowledge sharing (e.g., provenance of the knowledge about a patient).

Figure 18 – AI viewpoint of interoperability



Recommendation for an inventory of metadata information

The building of a metadata interoperability model requires an understanding of the type of information that is considered metadata, versus data. We have covered a number of viewpoints and conclude on the need to collect an inventory of metadata information. Such repository:

- requires the use of one or several taxonomies, e.g., descriptive information, structural information, administrative information, reference information, statistic information, or legal information;
- integrates consideration on the environment (i.e., the ecosystem);
- supports the various lifecycle processes
- allows for the management of indicators (e.g., FAIR, compliance)
- supports access to other metadata building blocks (e.g., DPVCG privacy ontology [43]).



VII Examples of Initiatives and Projects




This section focuses on examples of initiatives and projects where data space capabilities have been developed, providing insight on their implementations. The initiatives focus on architecture (IDS, GAIA-X, FIWARE), while the projects focus on pilots (PLATOON, Interconnect, Smartbear).

IDS Reference Architecture

Overall Characteristics

The IDS reference architecture model (IDS-RAM) sets the standard for building data-driven ecosystems, products and services.

Table 7 – IDS-RAM characteristics

	REFERENCE	https://internationaldataspaces.org/use/reference-architecture/
	DESCRIPTION	<p>The IDS-RAM sets the standard for building data-driven ecosystems, products and services that enable standardised, trustworthy and self-determined data exchange based on European values. The IDS-RAM upholds the data sovereignty of the creator of the data, guarantees trust among participants and ensures data privacy and security throughout the data exchange. The most recent version of the reference architecture at the time of writing is IDS-RAM 4.2 [44].</p>
	STAKEHOLDERS AND CONCERNS	<p>The continuous development of the IDS-RAM is governed by the International Data Spaces Association (IDSA). IDSA also promotes and supports the creation of Data Spaces in sectors such as Industry 4.0, Smart Mobility, Smart Energy and Healthcare, along with a wide variety of stakeholders including large enterprises, SMEs, NGOs, technology providers and universities and research institutions.</p>

[44] https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0/tree/v4.2.0



ARCHITECTURE PRINCIPLES

The figure below illustrates the interactions between the core IDS-RAM components. The main component that makes the data exchange possible is the IDS Connector, which is responsible for forwarding the data from the Data Provider to the Data Consumer. The participants in a Data Space have their identities verified before any data exchange by a certification authority and a DAPS service. The connector is aided by the Broker, which is an intermediary that receives, stores and provides metadata about the available data products within the Data Space. The App Store is another intermediary which manages and provides reusable data apps that can be deployed within the Connector to facilitate data processing workflows. Furthermore, the Clearing House provides logging and settlement services for all financial and data exchange transactions. Finally, the Vocabulary Hub offers domain-specific vocabularies to participants of the data space [45].

Figure 19 – Main interaction patterns between IDSA components [46]

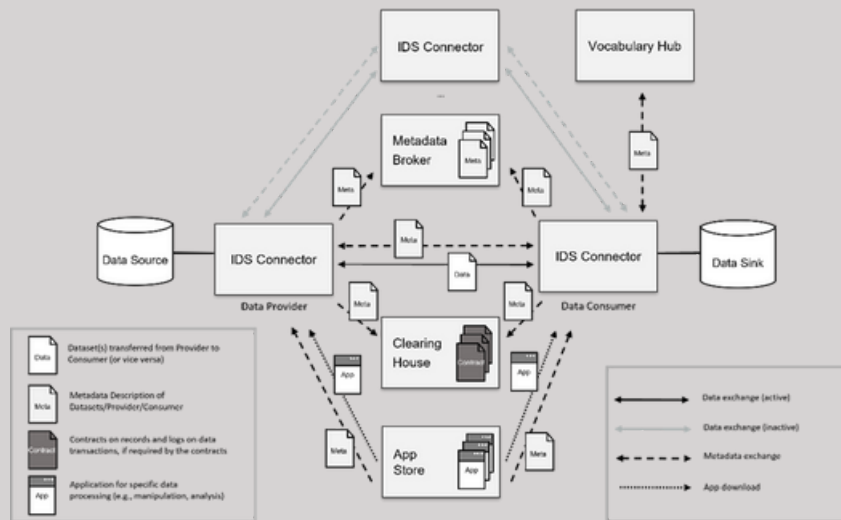
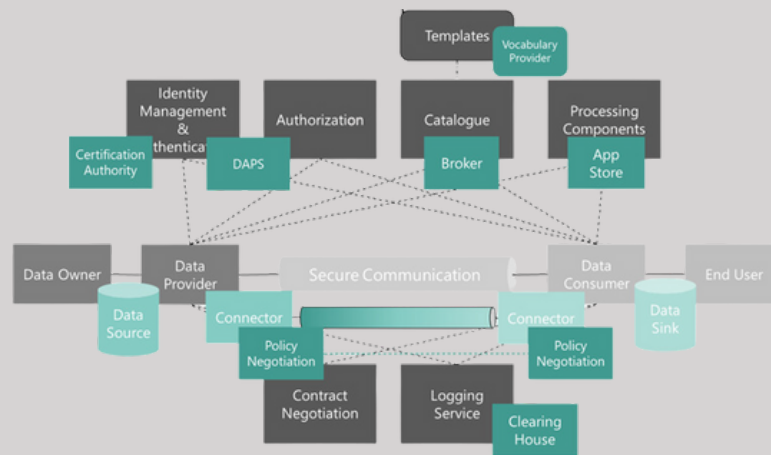


Figure 20 identifies how the IDSA architecture and its components are mapped onto the abstract Data Space architecture introduced in figure 1.

Figure 20 – IDSA on top of abstract data space model





DATA MANAGEMENT CAPABILITIES

The IDS-RAM contemplates various data management capabilities across its different layers:

- **Data Encryption:** through the IDS Communication Protocol (IDSCP) which supports point-to-point data encryption and end-to-end authorisation to establish confidential and authenticated data exchange between the Data Provider and the Data Consumer.
- **Identity Management and Authentication:** through an Identity Provider which issues digital X.509 certificates to all participants. These certificates are used for identification, authentication and encryption before establishing any (meta)data exchange.
- **Trust Management:** through the use of cryptographic methods, in particular the Public Key Infrastructure (PKI), which allows each entity to authenticate against other participants across the entire Data Space.
- **Access Control:** the IDS-RAM gives Data Providers the possibility to define Attribute-Based Access Control (ABAC) which can be implemented within the Connector using technologies such as XACML or JAAS to regulate access requests from other participants to resources.
- **Usage Control Definition:** data usage control in IDS-RAM works by attaching data usage policy information, expressed in a machine-readable format such as LUCON, to the data being exchanged between the Provider and the Consumer.



ROADMAP

The IDS-RAM provides the possibility to implement processes regarding the following data management aspects, but they are not included as an integral part of the architecture:

- **Usage Control Enforcement:** IDS-RAM cannot, and does not intend to, replace legal contracts or licensing agreements. Instead, the IDS provides a framework to technically enforce usage controls in addition to existing, legally binding contracts. Such enforcement would require the Connectors data flows to be modified to add monitoring and interception points (i.e., Policy Enforcement Points, PEPs) that request permission or denial of an action from a centralised decision engine (i.e., a Policy Decision Point, PDP).
- **Data Quality:** can be assessed by extending the functionality of the Connector with self-implemented Data Apps that perform data quality checks before a data exchange is carried out.
- **Data Provenance:** can be controlled through local tracking components integrated into the Connectors and a centralised provenance component that analyse the logs from all data exchanges.

Interoperability

Table 8 – IDSA interoperability



INTEROPERABILITY

Standardised data exchange between participants is the fundamental aspect of the IDS-RAM. The **IDS Connector** is the main technical component for this purpose, given that it acts as a standardised and interoperable interface for data exchange between Data Providers and Data Consumers, as well as for metadata exchange with other components (Brokers and App Stores). The Connector is implemented in different variants and can be acquired from different vendors, but each implementation is interoperable with any other Connector and other technical components, in the IDS ecosystem. On the other hand, transport interoperability of the (meta)data is supported by the use of the **IDSCP** in all exchanges and communication between IDS-RAM components and metadata interoperability is governed by the **IDS Information Model** (see Table 9 for details).



ASSURANCE OF INTEROPERABILITY

Given that the IDS-RAM is envisioned to be adopted by a wide variety of industries and use cases, different implementations of the IDS Connectors and other components, most importantly Brokers and App Stores, may vary greatly. Thus, the IDSA has defined a Core Component Certification process in order to ensure the interoperability and compatibility of the different implementations of IDS components. The process includes interoperability testing and also considers security aspects of the implemented component. All Connector and other component implementations must go through this process before they can be used within the IDS ecosystem [47].

Figure 21 - IDSA certification levels for component certification [48]

Requirements to be fulfilled	Evaluation effort and assurance		
	ASSURANCE LEVEL 1 Checklists self-assessment and automated interoperability testing	ASSURANCE LEVEL 2 External concept review including functional and security testing	ASSURANCE LEVEL 3 External evaluation including concept review, testing and source code audit
TRUST LEVEL 1 Data space interoperability	✓	✓	
TRUST LEVEL 2 Feature complete for data usage control		✓	✓
TRUST LEVEL 3 Additional protection from internal attacks		✓	✓

[47] Menz, Nadja, Resetko, Aleksei, & Winkel, Jonas. (2019). IDS Certification explained (1.0). Zenodo. <https://doi.org/10.5281/zenodo.5269021>

[48] [https://github.com/International-Data-Spaces-Association/IDS-RAM-4-0/tree/v4.2.0/documentation/4 Perspectives of the Reference Architecture Model/4 2 Certification Perspective](https://github.com/International-Data-Spaces-Association/IDS-RAM-4-0/tree/v4.2.0/documentation/4_Perspectives_of_the_Reference_Architecture_Model/4_2_Certification_Perspective)

Metadata Interoperability

Table 9 – IDSA metadata interoperability



METADATA INTEROPERABILITY

Metadata interoperability between IDS components is governed by the IDS Information Model [49], a formal model that defines a minimal, domain agnostic set of metadata concepts to support the description, interchange and identification of data products and Data Apps, known in the model as Digital Resources. The model is used by all IDS components (Connectors, Brokers, App Stores, etc.) to describe a Digital Resource, thus guaranteeing metadata interoperability across the IDS ecosystem. Moreover, the Information Model is expressed as an RDF ontology and includes links to widely-used concepts of upper-level ontologies (DCAT, ODRL, etc.), thus fostering acceptance and metadata interoperability beyond the IDS ecosystem.

For the sake of clarity, the Information Model splits the metadata concepts used to describe a Digital Resource into 6 different concern dimensions, each considering only one aspect of the Resource. The figure below shows these dimensions, along with the main concepts included in each of them. Concerns beyond the scope of modeling Digital Resources and their interchange are out of scope. Furthermore, domain-specific modeling is delegated to other shared vocabularies and data schemata.

Figure 22 – IDSA information model dimensions



Metadata interoperability is further supported in IDS-RAM by the **Vocabulary Provider**, which manages and offers domain-specific vocabularies (i.e., ontologies, reference data models, or metadata elements) that can be used to annotate and describe datasets. One example of such a vocabulary is the Quality Information Framework (QIF) ontology in the Industry 4.0 domain, an ISO standard used for software systems to communicate manufacturing data structures [50]. This type of vocabularies can complement the IDS Info Model and improve the metadata interoperability between data space participants in a particular domain.

[49] International Data Spaces Information Model (<https://w3id.org/idsa/core>)

[50] Mertens, C. et al. (2022). A Framework for Big Data Sovereignty: The European Industrial Data Space (EIDS). In: Curry, E., Scerri, S., Tuikka, T. (eds) Data Spaces . Springer, Cham. https://doi.org/10.1007/978-3-030-98636-0_10

GAIA-X Reference Architecture

Overall characteristics

Table 10 – GAIA-X Interoperability

DESCRIPTION

<https://www.gaia-x.eu>

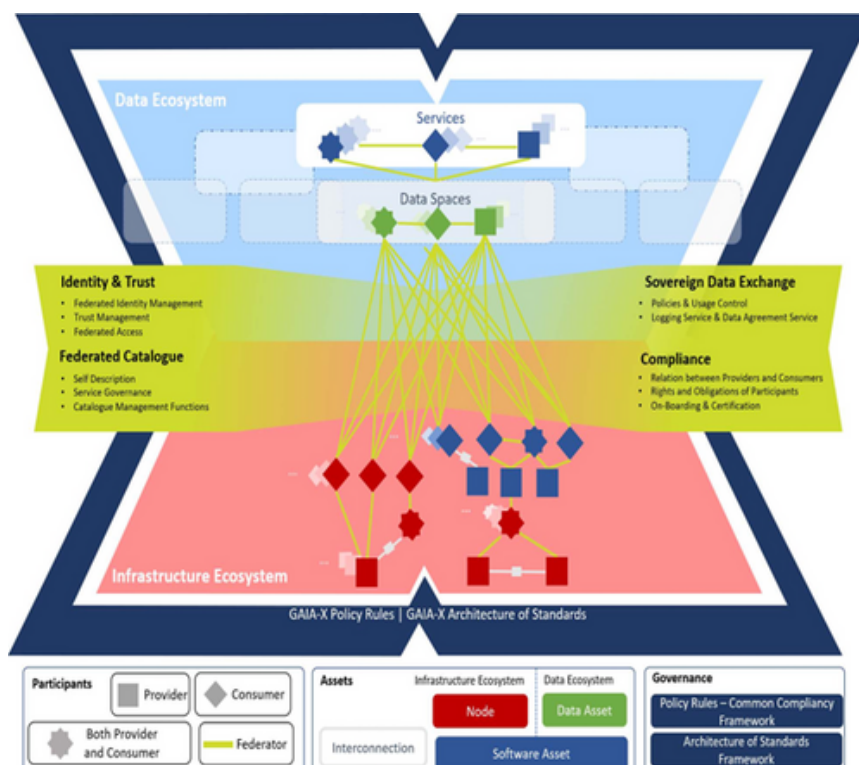
The Gaia-X Architecture enables Ecosystems and data spaces using the elements explained in the Gaia-X Conceptual Model in general and the Federation Services in particular.

An Ecosystem is an organising principle describing the interaction of different actors and their environment as an integrated whole, like in a biological Ecosystem. In a technical context, it refers to a set of loosely coupled actors who jointly create an economic community and its associated benefits.

Gaia-X proposes to structure a Data Ecosystem and an Infrastructure Ecosystem, each with a different focus on exchanged goods and services. Despite each of them having a separate focus, they cannot be viewed separately as they build upon each other, i.e., they are complementary.

The Gaia-X Ecosystem consists of the entirety of all individual Ecosystems that use the Architecture and conform to Gaia-X requirements. Several individual Ecosystems may exist (e.g., Catena-X) that orchestrate themselves, use the Architecture and may or may not use the Federation Services open source software.

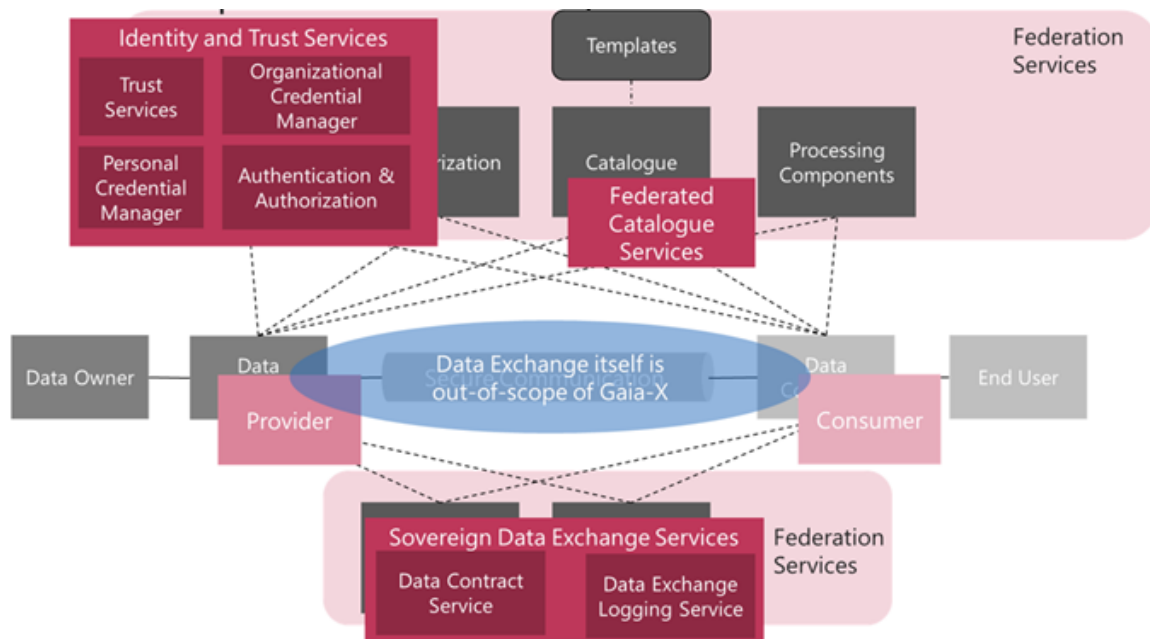
Figure 23 – Gaia-X ecosystem



DESCRIPTION

Figure 24 identifies how the GAIA-X architecture is mapped on the abstract Data Space architecture in Figure 1.

Figure 24 – GAIA-X on top of abstract data space model



ARCHITECTURE PRINCIPLES

The architecture [51] is used to address the following requirements:

- Interoperability of data and services: The ability of several systems or services to exchange information and to use the exchanged information in mutually beneficial ways.
- Portability of data and services: Data is described in a standardised protocol that enables transfer and processing to increase its usefulness as a strategic resource. Services can be migrated without significant changes and adaptations and have a similar quality of service (QoS) as well as the same Compliance level.
- Sovereignty over data: Participants can retain absolute control and transparency over what happens to their data. This document follows the EU's data protection provisions and stresses a general 'compliance-by-design' and 'continuous-auditability' approach.
- Security and trust: Gaia-X puts security technology at its core to protect every Participant and system of the Gaia-X Ecosystem (security-by-design). An Identity management system with mutual authentication, selective disclosure and revocation of trust is needed to foster a secure digital Ecosystem without building upon the authority of a single corporation or government.

[51] https://www.gaia-x.eu/sites/default/files/2021-06/Gaia-X_Architecture_Document_2106.pdf

ARCHITECTURE PRINCIPLES

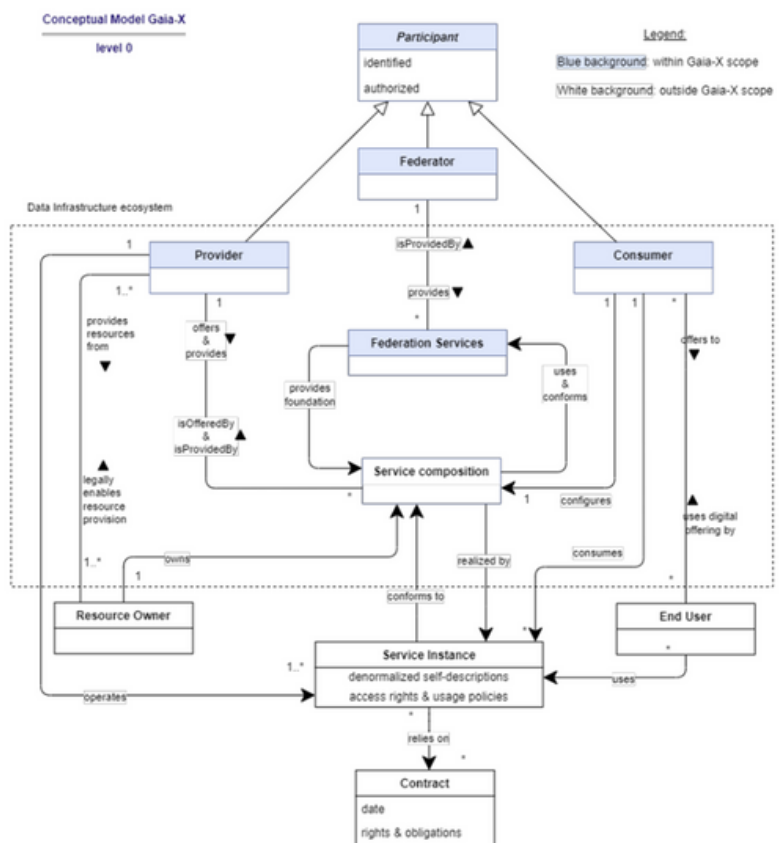
Design Principles:

- **Federation:** Federated systems describe autonomous entities, tied together by a specified set of standards, frameworks and legal rules. The principle balances the need for a minimal set of requirements to enable interoperability and information sharing between and among the different entities while giving them maximum autonomy. The principle defines the orchestrating role of Gaia-X governance elements and implies interoperability within and across Gaia-X Ecosystems.
- **Decentralisation:** Decentralisation describes how lower-level entities operate locally without centralised control in a self-organised manner. (The federation principle enables this self-organisation by providing capabilities for connectivity within a network of autonomously acting Gaia-X Participants.) The principle of decentralisation implies individual responsibility for contributions and no control over the components, which fosters scalability.
- **Openness:** The open architecture makes adding, updating and changing of components easy and allows insights into all parts of the architecture without any proprietary claims. In this way, Gaia-X is open to future innovation and standards and aware of evolving technologies. The documentation and specifications of Gaia-X architectures and technologies are openly available and provide transparency as technology choices will be made to encourage the distribution of collaboratively created artefacts under OSD compliant open source licenses.

Conceptual Model:

The Gaia-X conceptual model, shown in the figure below, describes all concepts in the scope of Gaia-X and relations among them. The Gaia-X core concepts are represented in classes. An entity highlighted in blue shows that an element is part of Gaia-X and therefore described by a Gaia-X Self-Description. The upper part of the model shows different actors of Gaia-X, while the lower part shows elements of commercial trade and the relationship to actors outside Gaia-X.

Figure 25 – Gaia-X conceptual model



Interoperability

Table 11 – GAIA-X interoperability

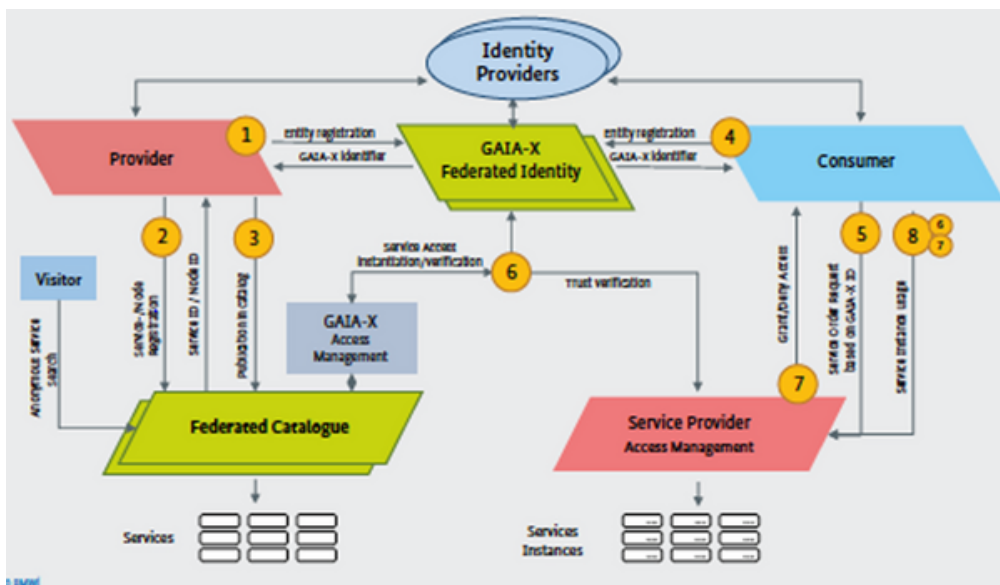
INTEROPERABILITY

Interoperability is enabled by a federated identity model as showed in Figure 26 and Table 12.

Table 12 – Components of the Federation Model

1 Components	2 Description
<ul style="list-style-type: none"> ● GAIA-X Federated Identity 	<ul style="list-style-type: none"> ● This component guarantees identity proofing of the involved Participants to make sure that GAIA-X Participants are who they claim to be.
<ul style="list-style-type: none"> ● GAIA-X Federated Catalogue 	<ul style="list-style-type: none"> ● The Federated Catalogue is a logical combination of a Self-Description repository and search algorithms so that Self-Description-based attribute searches can be processed.
<ul style="list-style-type: none"> ● Service Provider AM 	<ul style="list-style-type: none"> ● The Service Ordering Process will involve the Consumer and the Service Provider. The Service Provider will create the Service Instance and will grant access for the Consumer.

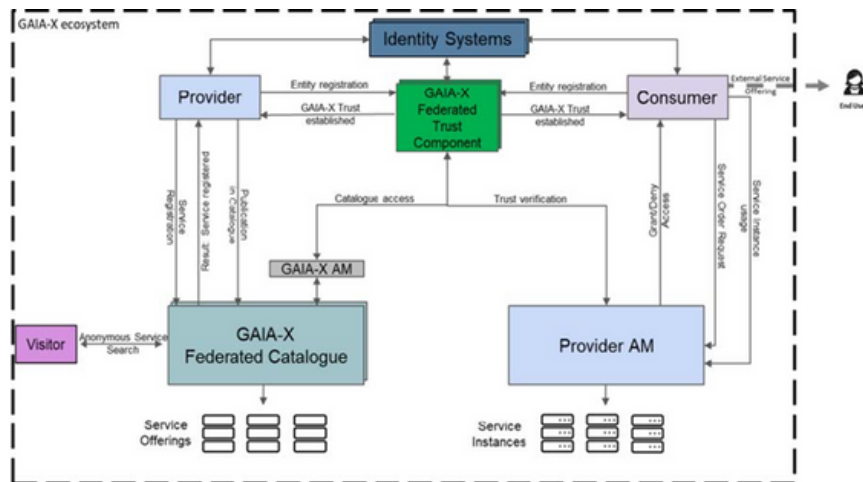
Figure 26 – Gaia-X federated identity model



ASSURANCE OF INTEROPERABILITY

Assurance of interoperability is based on a federated trust model.

Figure 27 – Gaia-X federated trust model



The Federated Trust Model achieves Trust between Consumers and Providers. This is realised with the components shown in Figure Above. While the Federated Trust Component and the Federated Catalogue have been defined before, the Federated Trust Model further involves the Gaia-X AM, which is an internal Gaia-X access management component responsible for authorising Principals' interactions within the Gaia-X Portals and the Provider Access Management (Provider AM), which the Provider will use to grant access for the Consumer to Service Instances. Within the federated approach, Identities are built up of verifiable Claims and shared on a need-to-know basis.

The Gaia-X Portals

The Gaia-X Portals support Participants to interact with Federation Services functions via a user interface, which provides mechanisms to interact with core capabilities using API calls. The goal is consistent user experience for all tasks that can be performed with a specific focus on security and Compliance. The Portals provide information on Assets, Resources and Service Offerings and interaction mechanisms for tasks related to their maintenance. Each Ecosystem can deploy its own Portals to support interaction with Federation Services. The functions of the Portals are further described below.

A Portal supports the registration of organisations as new Participants. This process provides the steps to identify and authorise becoming a Participant. Additionally, organisations are assisted in signing up as members of Gaia-X association AISBL. Participants are supported in managing Self-Descriptions and organising Credentials. This includes Self-Description editing and administration. A Portal further offers search and filtering of Service Offerings and Participants, based on Federated Catalogues. Additionally, solution packaging refers to a composition mechanism for the selection and combination of Service Offerings into solution packages to address specific use cases possible with a Portal. To orchestrate the various APIs, an API framework to create a consistent user and developer experience for API access and lifecycle is introduced. An API gateway will ensure security for all integrated services. An API portal will provide a single point of information about available API services and version management.

Metadata interoperability

Table 13 – GAIA-X metadata Interoperability

METADATA INTEROPERABILITY

The Gaia-X Ecosystem includes a huge variety of Participants and Service Offerings. Therefore, interoperability needs to be ensured on different levels (Infrastructure as a Service [IaaS], Platform as a Service [PaaS], Software as a Service [SaaS], data assets and others). Regarding interoperability of data, core elements to be identified in this endeavour are API specifications and best practices for semantic data descriptions. The use of semantic data interoperability is seen as a foundation to eventually create a clear mapping between domain-specific approaches based on a community process and open-source efforts.

FIWARE Reference Architecture

Overall Characteristics

Table 14 – FIWARE interoperability

<p>FIWARE fiware.org</p>	<p>FIWARE is a curated system of open-source components enabling the creation of interoperable data spaces. The core of the system is a sophisticated context broker that is exchanging semantic annotated data typically described by ontologies. Further elements of the data ecosystem realise specific software features. Examples are persistent storage of data, big data processing of the context information in existing big data systems, complex event processing, spatial-temporal processing of data, edge programming using a intent-based programming model, security components, data marketplaces, open linked data publishing, visualisation and knowledge management. FIWARE follows the NGSI-LD standards. Its context broker is an approved building block of the Connecting Europe Facility (CEF). States like Slovenia are recommending the Context Broker together with the OASC MIMs as the standard elements for their nationwide smart city ecosystems. The same standard (NGSI-LD) is also an official standard in India and used for the “Indian Urban Data Exchange” platform that will connect all Indian cities. Phase 1 of 14 cities have already concluded the implementation of IUDX. NGSI-LD and context brokering is also a core element of the Japanese CityOS, a reference architecture defined in a SIP (Strategic Innovation Project) project by the Japanese cabinet. The latest FIWARE brochure lists about 250 cites on all continents of the world actively utilising the context broker technology and FIWARE.</p>
<p>Architecture principles</p>	<p>Formal Standard: FIWARE is based on the ETSI specification NGSI-LD (“Next Generation Service Interface - Linked Data”) as defined in the ETSI ISG “Context Information Management (CIM)” [52]. NGSI-LD defines a system model, a set of APIs and a data model. The core syntactical model of NGSI-LD is JSON-LD, a data standard very popular in the Internet and Web world. NGSI-LD defines a meta-model that prescribes certain mandatory elements of a NGSI-LD entity, among them an ID of the entity, a type, as well as how to define properties such as attributes or relationships between entities. Metadata can be defined on the level of entities, but also for attributes and relationships. Further specialisation are defined for geographic information and for temporal information. The NGSI-LD meta-model can be mapped to an RDF representation triples giving its concepts and properties also a formal specification. Furthermore, NGSI-LD can be automatically mapped into knowledge graphs and processed by respective semantic and knowledge-based systems. Utilising the @context clause of JSON-LD, NGSI-LD type information for entities, attributes and relationships can be automatically linked to ontologies those having a semantic grounding of the NGSI-LD data structures.</p>
<p>Data management capabilities</p>	<p>Supporting Standards: The abstract ETSI NGSI-LD standard is complemented by separate specification for data models. This is done by the Smart Data initiative driven by TMForum, GSMA, OASC, FIWARE; as well as many other initiatives. This has resulted in around 750 formally specified data models. These data models cover the needs of Smart Cities, Smart Industries as well as many other domains.</p>

Interoperability

Table 15 – FIWARE interoperability

Interoperability

To summarise, syntactical interoperability in FIWARE is achieved through the JSON-LD based data structure as well as the standardised API. Semantic interoperability is achieved through the NGSI-LD meta-model as well as through the semantic grounding of NGSI-LD data models using the @context clause to reference ontologies. Behavior interoperability is achieved through the communication patterns supported by the NGSI-LD API (Data Publishing, sophisticated query operations for synchronous communication, as well as a subscribe-notify communication mechanism for asynchronous communication).

Metadata interoperability

Table 16 – FIWARE Metadata interoperability

Metadata interoperability

Towards Interoperable Data Spaces: FIWARE is currently working intensively with BDVA, IDSA and GAIA-X to achieve a common data space model supported by the strong FIWARE ecosystem of interoperable open-source software component. This includes especially metadata needed for data space contracts, as well as access control and data usage control.

PLATOON Research Project

Overall characteristics

Table 17 – PLATOON characteristics

Reference

PLATOON (Digital PLAtform and analytic TOOlS for eNergy), <https://platoon-project.eu/>
Information kindly provided by Erik Maqueda (Technical manager, PLATOON project).

Description

PLATOON has developed a federated platform for the energy sector focusing on the following pillars: interoperability, trust and data analytic services. The project has developed, implemented and validated into seven large scale pilots scalable and replicable solutions that accelerate energy transition.

Stakeholders and concerns

PLATOON has developed a federated platform for the energy sector focusing on the following pillars: interoperability, trust and data analytic services. The project has developed, implemented and validated into seven large scale pilots scalable and replicable solutions that accelerate energy transition.

Architecture principles

PLATOON has developed an open-source reference architecture based on widespread open reference architectures such as FIWARE, IDSA, COSMAG [53] and SGAM.

Data management capabilities

The following features are supported:

- Market place with IDS capabilities (Broker + App Store + Clearing House)
- Access control and Authentication, using IDS data access protocols
- Federation: defined reference architecture (see above), defined common semantic data models based on standards (SAREF, CIM, SEAS, OntoWind)
- Trustworthiness management with the development of an IDS open source connector (TRUE connector)
- Usage control: Data Usage Data App compatible with IDS connector.
- Privacy compliance: Data Privacy Data App (CAPE) compatible with IDS connector.

Roadmap

- Open source IDS Connector with data usage and privacy features [54]
- Open source Broker
- Features validated in large scale pilots.

Interoperability

Table 18 – PLATOON interoperability

Interoperability

Interoperability has been one of the main pillars of the project. PLATOON has defined an open common semantic data models based on standards (SAREF, CIM, SEAS, OntoWind).

Assurance of interoperability

Assurance of interoperability has been one of the main pillars of the project. It is based on mapping rules and interoperability constraints associated with the Unified knowledge base and the Federated Query Processing Engine developed by PLATOON.

Metadata interoperability

Table 19 – PLATOON meta interoperability

Metadata interoperability

Metadata interoperability has been one of the main pillars of the project. PLATOON has defined common data sources and data applications metadata. PLATOON has also developed an open-source IDS vocabulary provider for managing different vocabularies and linking with IDS.

The open-source IDS vocabulary provider is available in PLATOON GitHub repository.

Processes for data and metadata specification and consensus reaching

PLATOON has defined and implemented a methodology for defining the common semantic data models, data sources and application metadata.

InterConnect Research Project

Overall characteristics

Table 20 – InterConnect characteristics

✓ **INTERCONNECTPROJECT.EU**

Description

Interoperable solutions/services connecting (devices in) Smart Homes, Buildings and Grids for the democratisation of efficient energy management, through a flexible and interoperable ecosystem where demand-side flexibility can be soundly integrated with effective benefits to end-users.

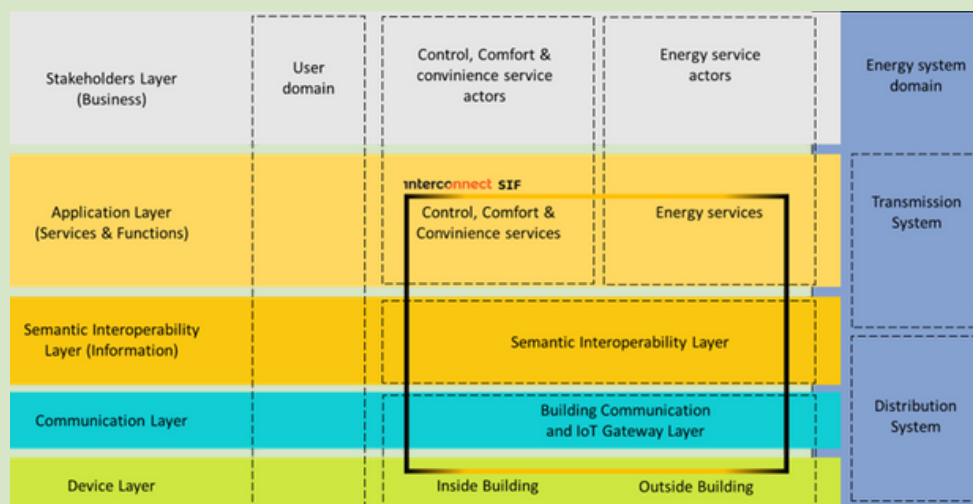
Stakeholders and concerns

An ecosystem of stakeholders with a need to interconnect devices and services: owners, facility managers and inhabitants of buildings, device manufacturers, IoT platform providers, energy service providers, energy (i.e., electricity) providers/retailers, Distribution System Operators (DSO), etc.

Architecture principles

Multiple architectural viewpoints (i.e., Energy, IoT, Interoperability Framework and Semantic) provide bridges for collaboration by domain experts versed in existing reference architectures (SGAM, RAMI, AIOTI, etc.). All viewpoints are based on separation of concerns and abstraction of functionality.

Figure 28 – INTERCONNECT architecture



Data management capabilities

Each InterConnect devices and/or service has a so-called Knowledge Base (KBs) associated with it. Knowledge Graphs (KG) are used for the exchange of Knowledge between these KBs. KGs are encoded using semantic web technology. Specific ontologies have been defined by the InterConnect project to have a shared understanding of the many concepts in the InterConnect ecosystem. The basic principle underlying the exchange of knowledge is to 'share on a need-to-know basis'. Each KB determines if it wants to share knowledge based on the content of received knowledge graphs. There is a Service Store where services can register (as a KB) and where they can be discovered by other KBs for the purpose of the exchange of knowledge. As InterConnect is a layer 'on top of' web/internet technology layers for the semantic exchange of information, it can make use of underlying functionality regarding access control, authentication, etc. This layer is also known as the Semantic Interoperability Layer (SIL).

Roadmap

The specific InterConnect technology that enables device manufactures, IoT platform providers, (energy) service providers and DSOs to interconnect using the Semantic Interoperability Layer is being finished to be used in 7 large scale pilots across Europe right now. It is proposed to investigate if it is possible to include finer-grained access control at the semantic level into Knowledge Graphs.

Interoperability

Table 21 – InterConnect interoperability

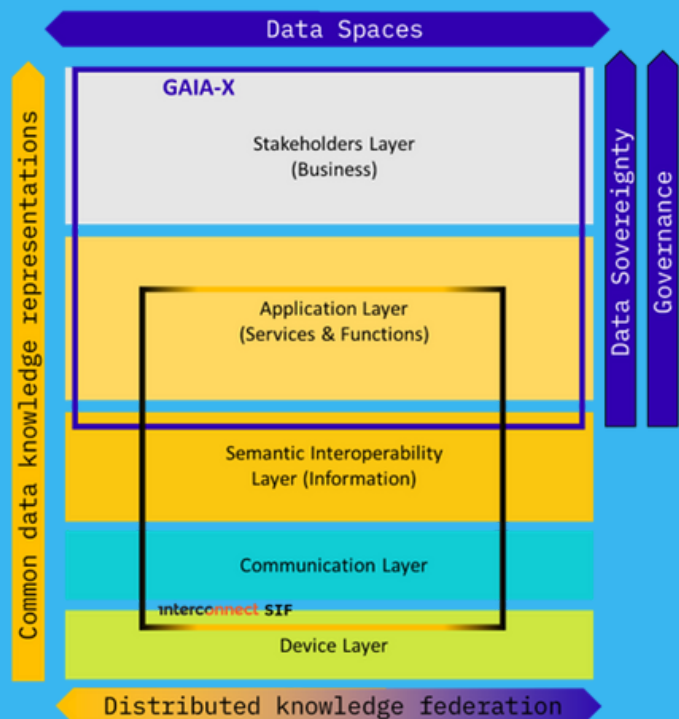
Interoperability

Use cases in InterConnect are about interoperability of devices and services, interconnected by the (electrical) grid and the cloud.

Technical/communication interoperability is achieved by making use of standardised semantic web technology. Semantic interoperability is achieved by having a shared understanding of (semantic) concepts as defined in the InterConnect set of ontologies. A quintessential part of this set is the Smart Applications REFerence (SAREF) ontology suite as defined by ETSI. It is investigated if adaptations of SAREF are needed for better support of certain use cases. Also, standardised information models from EEBUS and CENELEC are used to create interoperability in energy flexibility services.

The picture below shows how Interconnect plans to be integrated in future data space initiatives.

Figure 29 – INTERCONNECT integration with GAIA-X



Assurance of interoperability

Adoption of the InterConnect set of ontologies requires assurance of interoperability at the semantic and communication technology level.

Currently the project develops standardised 'adapters' for the communication of knowledge graphs, which creates a uniform and interoperable way for connecting. The ontologies are developed using input from all pilots involved as well as from domain experts who bring in expertise on existing information models.

InterConnect has a dedicated work package that investigates how and in which (standardisation) organisations the ontologies can be made available. A long-term view is taken beyond the current view.

Metadata interoperability

Table 22 – InterConnect metadata interoperability

Metadata interoperability

All data exchange between devices and services in pilots is seen as exchange of knowledge, which is encoded in knowledge graphs that contain meta-data in terms of using and/or referring to the InterConnect ontologies.

InterConnect has defined a set of ontologies which enable interconnected parties to encode knowledge using semantic web technology. These ontologies have been created based on requirements from pilots and alignment with existing standardised information models in the domain of energy and IoT.

Processes for data and metadata specification and consensus reaching

Without reaching consensus on the metadata specification, it is not possible to create pilots with interoperable interfaces for device and service components.

InterConnect has had interviews with pilot and domain experts and through means of iterative workshops (existing) ontologies have been extended and/or created, resulting in a set of InterConnect ontologies. An important aspect of this process is alignment with existing information models (e.g., from EEBUS, CENELEC, ETSI, increasing the chance of industry support).

SmartBear Research Project

Overall characteristics

Table 23 – SmartBear interoperability

Reference

<https://www.smart-bear.eu/>

Description

The primary goal of SMARTBEAR project is to develop an integrated platform gathering numerous health related data flows, to analyse the day-by-day patients' activities and their health status. These analyses are then used to provide evidence-based, personalised interventions towards improving the degree of healthy and independent living of the patients.

Stakeholders and concerns

Consortium partners form a synergy of the health sector and IT technologies: large hospitals owning large patients' datasets, IT companies providing secure collection, storage and analysis of the data, medical devices vendors.

Concerns: lack of centralised big data platforms to analyse and process the medical data in a privacy-preserving manner, to prevent the development/deterioration of various conditions of the elderly and reduce medical costs.

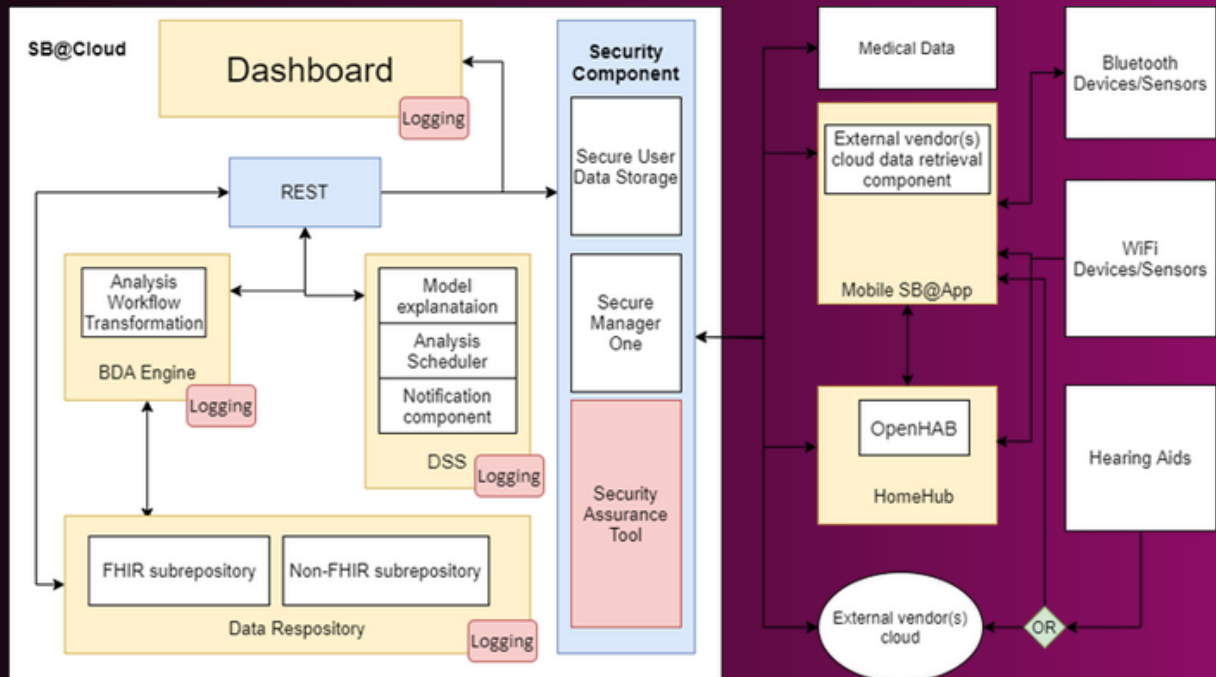
Architecture principles

The overall system is divided into three main parts – the Mobile SB@App (top-right), the HomeHub (bottom-right) and the SB@Cloud (left) subsystems. The first two are used to coordinate the collection of personal data from different devices and transmit them to the latter subsystem, where the data are anonymised, stored, analysed and then suggested personalised interventions are produced by the Decision-Support Subsystem (DSS) and transmitted back to the Mobile SB@App, which runs on a smartphone operated by each patient, to inform the patients about actions they could take to improve their health.

Medical data in the SB@Cloud are held in a FHIR-compliant repository. There are also medical data that come from external sources (e.g., hospital EHR systems) and systems developed by other EU research projects (Smart4Health [55], Holobalance [56]).

Architecture principles

Figure 30 – SMARTBear architecture



Data management capabilities

Following features are supported:

- Access control and Authentication: security component consisting of secure assurance tool, secure manager one and secure user data storage
- Federation (partially): through data models that follow the FHIR standard and in one direction only (data digestion – no data exporting).
- Privacy-by-design: all data are pseudo-anonymised before storage by replacing their original person ID with another one that has been allocated to that person from the project upon their registration. In this manner the original data provider (e.g., hospital EHR) cannot link back a person's data to that person.

Roadmap

Full data federation (exporting data to other providers/projects as well) is being considered at the moment. The validation on the large-scale pilots will be finished by 2024.

Interoperability

Table 24 – SmartBear interoperability

Interoperability

The SmartBear platform is compatible with other EU projects collecting medical data (Smart4Health & Holobalance).

SmartBear platform stores the data received from the other projects using FHIR-compliant repository for interoperability purpose. At the moment, data exchange is in one direction only (digestion – there's no data exporting).

Assurance of interoperability

No assurance of interoperability is present in SmartBear project.

Metadata interoperability

Table 25 – SmartBear metadata interoperability

Metadata interoperability

The SmartBear platform is compatible with other EU projects collecting medical data (Smart4Health & Holobalance).

SmartBear platform stores the data received from the other projects using FHIR-compliant repository for interoperability purpose. At the moment, data exchange is in one direction only (digestion – there's no data exporting).

Processes for data and metadata specification and consensus reaching

The SmartBear data/metadata is specified at the design stage.

The specification is FHIR-compliant and is developed by engaging actors from the different organisations/projects.

Use case

Table 26 – SmartBear use cases

Interoperability

<https://www.smart-bear.eu/about/#usecases>

The project includes 6 large scale pilots (Portugal, Spain, France, Italy, Romania and Greece). The SmartBear platform will provide the mechanisms for secure collection, storage and analysis of the medical data to cover 5 comorbidities:

- Hearing Loss
- Cardiovascular Diseases
- Cognitive Impairments
- Mental Health Issues
- Balance Disorders
- Frailty

Data interoperability

The SmartBear platform can digest data from external sources (e.g., hospital EHRs) and other EU projects collecting medical data. This is achieved through the usage of pre-agreed FHIR-compliant representations.



VIII Examples of Data Space Use Cases

This section shows some examples of data space use cases in different domains. For each case, the data space solution is described, the benefits of the use of data are explained and considerations on architecture, metadata and standards are provided.

Agriculture

Table 27 – Agriculture use case

SOLUTION

DJUSTCONNECT AGRICULTURAL DATASHARING PLATFORM [57] (BELGIUM)

1

- Data sharing platform for farmers and horticulturists, data providers and data consumers
- Integration of diverse data sources from different data providers, suppliers, buyers, producer organisations, agricultural coops
- One contract solution with clear rules and in accordance with legislation and Code of Conduct on agricultural data sharing ensuring privacy, safety and consent approved data exchanges

BENEFITS

2

- Sustainable business models for data sharing
- Farmers keep sovereignty of the data
- Data portability and interoperability

ARCHITECTURE/METADATA/STANDARDS

3

- IDSA based cloud architecture
- Metadata to provide company, resource information, API specification

Renewable energy

Table 28 – Renewable energy use case

SOLUTION

COMMON TAXONOMY DEFINITION – IEC STANDARDS [58] (FRANCE)

1

- Alignment (automatic) of systems between different information models
- Development of a matrix of information models associating energy verticals (Generation, Transmission, Distribution, DER, Customer) and sectors (Process, Field, Station, Operation, Enterprise, Market) via means of ontologies
- Transverse interoperability for electricity, gas and heat

BENEFITS

2

- Alignment of standards IEC, Data Models with ontologies or linked data formats for advanced cross-sector interoperability

ARCHITECTURE/METADATA/STANDARDS

3

- Standard data models: CIM, COSEM, IEC 61850, CGMES
- Cross sector models and ontologies: SAREF, OneM2M, CIM+, NGIS-LD (FIWARE)
- Smart energy Grid Architecture model SGAM IEC 63200

[58] https://www.gaia-x.eu/sites/default/files/2021-06/Gaia-X_Data-Space-Energy_Position-Paper.pdf

Health

Table 29 – Health use case

SOLUTION

RESEARCH PLATFORM GENOMICS [59] (GERMANY)

1

- Cloud based genomics platform compliant with the Global Alliance for Genomics & Health (GA4GH)
- Secure and GDPR compliant storage of sensitive large datasets and availability in the context of cancer research.
- Analysis of data at petabyte scale

BENEFITS

2

- Complex integrative analysis with cross domain data (imaging, clinical) allowing personalised medicine and better prediction of cancer
- Standardised pipelines aligned with GA4GH that can be used in beyond 1 Million Genomes initiative.
- Integration with other omic platforms German Human Genome-Phenome Archive (GHGA)

ARCHITECTURE/METADATA/STANDARDS

3

- Standard semantic models for genetic data: WGS, BAM, FASTQ
- Organisational and Policy metadata specified by GA4GH. For example the Data Use Ontology (DUO) model of GA4GH allows usage conditions for specific data items to be described with consistent metadata.

Industry 4.0

Table 30 – Industry 4.0 use case

SOLUTION

SMART CONNECTED SUPPLIER NETWORK [60] (NETHERLANDS)

1

- Provision of an open communication standard for exchanging data within the supply chain covering all complex sectors such as the machine building industry
- Establishment of a foundation that agrees on domain specific semantics, legal aspects and processes
- Horizontal standards provided by widespread Data Spaces architectures
- Registration of ~ 100 manufacturers and ~ 12 service providers

BENEFITS

2

- 20% higher productivity of the supply chain through fast, secure and interoperable exchange of information between companies
- Connect once, communicate with the entire supply chain
- Integration of OEM (Original Equipment Manufacturers), 1st 2nd and 3rd tier suppliers, wholesalers and most available ERP (Enterprise Resource Planning) software

ARCHITECTURE/METADATA/STANDARDS

3

- IDSA architecture
- Messaging format: UBL (Universal Business Language) OASIS, ISO/IEC 19854:2015
- Connected organisations uniquely identified by GLN (Global Location Number)

[60] <https://internationaldataspaces.org/the-smart-connected-supplier-network-by-tno/>

Smart living

Table 31 – Smart living use case

SOLUTION

ENERGY EFFICIENCY OF SMART BUILDINGS [61] (GERMANY)

1

- Integration of data-based solutions from energy consumers, producers and storage systems across buildings
- Service discovery registry
- Intelligent controls systems/services for balancing energy networks (e.g. charging infrastructures for e-cars, heat pumps, PV feeds) switching on and off the supply based on learned demand based on IoT smart metres and switching devices
- Integration of storage solutions for smart home IoT sensors, IoT devices appliances and power-to-heat (PtH), power-to-gas (PtG), power-to-mobility (PtM)

BENEFITS

2

- Increased energy efficiency via automated energy management and monitoring services that make use of house residents consumption
- Balance of generation, storage and consumption of energy

ARCHITECTURE/METADATA/STANDARDS

3

- Semantic descriptions of data points (smart controllers) and resources with W3C SENSE Web of Things and iotschema.org (IoT ontology)
- Service metadata for catalogue of services and individual service description using WoT and Smart Readiness Indicator for Buildings from EC.
- Federated catalogue based on RDF-triple-stores and SPARQL



IX Conclusions

This white paper shows the value of metadata interoperability in data sharing spaces. As conclusions from this work, the authors of the paper have identified a number of further works of interest for the future:

- **Integration of interoperability “in” and “between” data space architectures:** the building of interoperability requires a mapping on the data space architecture in order to provide a rationale on the interoperability needs. Furthermore, it is recommended that different data space initiatives being simultaneously developed, each with its own standards and interoperability requirements, collaborate on the convergence of the proposed solutions, guaranteeing the inter-solution interoperability of future data space implementations.
- **Interoperability construction tools:** the mechanics of metadata interoperability should be promoted and supported via interoperability construction tools and related standards.
- **Inventory of metadata information:** an inventory of metadata information should be made available and maintained, facilitating metadata interoperability through different aspects:
 - o transversal concerns e.g., integration of trustworthiness, cyber physical considerations, etc .
 - o domain concerns e.g., agriculture, health, energy



X About BDVA

About the Big Data Value Association

BDVA is an industry-driven international not-for-profit organisation with 250 members all over Europe and a well-balanced composition of large, small and medium-sized industries as well as research and user organisations. Our mission is to develop an innovation ecosystem that enables the data-driven digital transformation of the economy and society in Europe, delivering maximum benefit. To reach this goal, we focus on advancing areas such as big data technologies and services, data platforms and data spaces, industrial AI, datadriven value creation, standardisation and skills.

BDVA enables existing regional multi-partner cooperation, to collaborate at the European level through the provision of tools and know-how to support the cocreation, development and experimentation of pan-European data-driven and AI applications and services and know-how exchange.

Through BDVA, our members contribute to the European data and AI R&I agenda and develop guidelines and strategic roadmaps for industry and policymakers in BDVA Task Forces and our events give opportunities to build new collaborations and co-create new projects. Being part of the BDVA community, the members gain higher visibility on the European level and our services are designed to give timely updates on all the latest developments in the fields of data and AI.

BDVA believes in collaborations! BDVA has been the private side of the H2020 partnership Big Data Value PPP, it is a private member of the EuroHPC JU and it is a founder member of the AI, Data and Robotics Partnership. BDVA has developed a strong and growing cooperation with Gaia-X, IDSA and FIWARE through the Data Spaces Business Alliance (DSBA), it is a partner of the Transcontinuum Initiative (TCI) and collaborates with many industry-driven AI national initiatives and other European communities.

BDVA is open to new members!

Visit [BDVA.EU](https://bdva.eu) to learn more about members and activities. You can contact us anytime at info@bdva.eu.

Note

This document should be referenced as follows: Kung A., Walshe R.,
Wenning R. - Data sharing spaces and interoperability. BDVA. 2023.



BDV BIG DATA VALUE
ASSOCIATION

BDVA Office
Data, AI and Robotics (DAIRO) aisbl
Avenue des Arts, 56
1000 Bruxelles
Belgium

BDVA.eu
info@bdva.eu