OPENDEI

SEPTEMBER 2022

REFERENCE ARCHITECTURES AND INTEROPERABILITY IN DIGITAL PLATFORMS

SEPTEMBER 2022

REFERENCE ARCHITECTURES AND **INTEROPERABILITY** IN **DIGITAL PLATFORMS**

Reference Architectures and Interoperability in Digital Platforms

Work Package	WP7 – Cross-Domain Task Forces
Task	T7.4 - Task Force 3: Digital Platforms, Pilots and Standards
Due Date	30/04/2022
Submission Date	30/04/2022
Version	2.0
Type	White Paper
Dissemination Level	Public
Deliverable Lead	FPM
Authors	Antonio Kung (FPM), Sergio Gusmeroli (FPM), Gabriella Monteleone (FPM), Alberto Dognini (RWTH), Luc Nicolas (EHTEL), Carmen Polcaro (Innovalia), All TF members contributed to the workshops
Reviewers	
Project website address	www.opendei.eu

Keywords

Digital Platforms and Pilots, Reference Architecture, interoperability, Standards, European Digital Platform in four industrial domains: Manufacturing, Agri-food, Energy, Health and Care

TABLE OF CONTENTS

List of Figures

List of Tables

1 INTRODUCTION

The document represents the result of a task force on reference architectures, interoperability frameworks and standardisation in DEI Digital Platforms. The task force is an activity of the project (H2020-872548- OPEN DEI Aligning Reference Architectures, Open Platforms and Large-Scale Pilots in Digitising European Industry[1], a collaborative and support action which contributes to the objective to help make every industry in Europe benefit fully from digital innovations.

The project's focus is on identifying relevant state-of-the-art cross-domain content that can be appropriate to the four sectors of manufacturing, agriculture, energy, and health & care. In each sector, an ambassador is responsible for liaising with its own sector or community. Each sector is represented by a number of large-scale pilots that are funded by the European Commission. Moreover, according to a participative and collaborative Open Innovation approach, Cross-domain Task Forces, which originated by OPEN DEI activities aim to involve not just eminent experts of the four domains projects' ecosystem, but also external experts with the role of Facilitators.

The Task Force, under technical moderation of Antonio Kung (TRIALOG), aims at engaging pool of experts in the domain of Digitising European Industry (DEI). It aims at creating a framework to spur creative thinking, discussing and disseminating reflections and innovative proposals on the definition and the implementation of reference architectures, interoperability frameworks and standards to support the implementation of next generation European digital platforms in the four basic industrial domains: manufacturing, agriculture, energy, health & care.

1.1 Structure of Position Paper

This position paper provides an analysis on how digital platforms can be aligned to enable cross-platforms and cross-domain data exchange:

- Section 2 provides considerations on alignment, at a reference architecture level and at an interoperability framework level. It further covers a number of topics of interest concerning the need for a semiotic approach, trustworthiness, resource management, interoperability approaches and digital twin integration
- Section 3, 4, 5, 6 elaborate on how reference architectures and interoperability frameworks are addressed in four domains: smart manufacturing, agrifood, energy and health and care.

1.2 Acknowledgement

OPEN DEI would like to thank the main editors of this document

- Alberto Dognini (Chapter 5), Antonio Kung (Lead editor, Chapter 2), Carmen Polcaro (Chapter 3), Gabriella Monteleone (TF 3 Lead), Luc Nicolas (Chapter 6), Marianna Faraldi (Chapter 4), Sergio Gusmeroli (TF 3 Lead).

OPEN DEI would like to thank the following standardisation experts

- Anja Hoefer, Antonio Kung, Asbjoern Hovsto, Alexander Samarin, Arne Berre, Bernd Blobel, Jan de Meer, Karim Tobich, Ricardo Villalon Fonseca, Young-Gab Kim.

The following persons have contributed to this position paper:

- Alberto Dognini (RWTH)
- Alessio Graziani (Engineering)
- Alexander Samarin (Consultant)
- Ane Zuluaga (Innovalia)
- Anja Hoefer (DIN)
- Angelo Marguglio (Engineering)

[1] http://www.opendei.eu/

- Antonio Kung (consultant to Fondazione Politecnico Milano)
- Arne Berre (Sintef)
- Asbjoern Hovsto (Hafenstrom)
- Bernd Blobel (University of Regensburg)
- Carmen Polcaro (Innovalia),
- Christos Kloukinas (City University London)
- Chandra Challagonda (Fiware)
- Daoud Urdu (Waveningen University and Research)
- Fabio Coelho (Inesctec)
- Gabriella Monteleone (Fondazione Politecnico Milano)
- Giorgio Micheletti (IDC)
- Golboo Pourabdollahian (IDC)
- Ioannis Basdekis (Sphynx)
- Ingo Martens (Hanse Aerospace)
- Ioannis Soldatos (Intrasoft)
- Jan de Meer (Consultant)
- Jesus Martinez Gadea (Atos)
- John Soldatos (Intrasoft)
- Juanjo Hierro (Fiware)
- Konstantin Pozdniakov (City University London)
- Kristian Helmholt (TNO)
- Karim Tobich (Consultant)
- Luc Nicolas (EHTEL)
- Luis Usatorre (Tecnalia)
- Marianna Faraldi (TechnoAlimenti)
- Myriam Blouin (Streamvision)
- Oscar Lazarl (Innovalia)
- Olga Meyer (Fraunhofer IPA)
- Philippe Sayegh (Verses)
- Ricardo Villalon Fonseca (U. Costa Rica)
- Riccardo Zanetti (Engineering)
- Sergio Gusmeroli (Fondazione Politecnico Milano)
- Silivi Castellvi (IDSA)
- Stefan Rilling (Fraunhofer IAIS)
- Stefano de Panfilis (Fiware)
- Wajid Usman (Information Calalyst)
- Young-Gab Kim (Sejong University)

2 ALIGNING DIGITAL PLATFORMS FOR DEI

2.1 Context

Digital platforms are online infrastructures which facilitate interactions and transactions between users. As stated in the OPEN DEI deliverable D2.1 (Reference Architecture for Cross-domain Digital Transformation)[2], "From an economic viewpoint, Digital Platforms are restructuring the global economy, contributing to the digitalization of organisations, value chains and whole sectors, by resetting entry barriers, changing the logic of value creation and value capture. From a commercial viewpoint, Digital Platforms ease the creation of ecosystems of stakeholders, supporting new forms of innovation and value creation, as well as related business and commercial models, focused on Digital Platforms' underlying vision and value proposition".

The promise of digital transformation requires approaches to allow for cross-platforms and cross-domain data exchange.

In addition to OPEN DEI deliverable D2.1, the following position papers have elaborated on related issues:

- BDVA published in April 2019 a position paper entitled "Towards a European Data Sharing Space: Enabling data exchange and unlocking AI potential"[3]. This was followed by a second version of the position paper November 2020[4].
- AIOTI published in August 2016 a paper entitled "Semantic interoperability for the Web of Things". It was followed in October 2019 by two papers: "Semantic IoT Solutions – A Developer Perspective" and "Towards semantic interoperability standards based on ontologies"[5].
- IDSA published in April 2019 a paper entitled "Reference Architecture Model"[6]
- GAIA-X published in June 2020 a paper entitled "GAIA-X: Technical Architecture"[7], and in June 2021 another paper entitled "GAIA-X Architecture Document"[8].

OPEN DEI consequently published in April 2021 a seminal paper entitled 'Design principles for data spaces"[9]. This was followed by a report from the Finland ministry of transport and communication in October 2021, entitled "State of Data Spaces"[10]

A set of three additional documents provides a consolidated viewpoint:

- AIOTI position paper entitled "Guidance for IoT and Edge integration in data spaces";
- BDVA position paper entitled "Data sharing spaces and interoperability"; and
- this position paper. It completes the two other papers by providing considerations on federated digital platforms, the relation between digital twins and data spaces, and the specialisation of data spaces into four domains (smart manufacturing, agri-food, energy and health and care).

2.2 Reference Architectures

2.2.1 OPEN DEI Reference architecture framework

The Digital Transformation of our EU industries introduces new challenges in terms of data, knowledge, and technology adoption due to critical interoperability challenges for the underlying digital platforms (and the synergies among them) regarding the identification of the most valuable data and information to be exploited.

---

[2] https://www.opendei.eu/case-studies/d2-1-reference-architecture-for-cross-domain-digital-transformation/

[3] http://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpace%20PositionPaper_April2019_V1.pdf

[4] https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpaces%20PositionPaper%20V2_2020_Final.pdf

[5] https://aioti.eu/resources/standardisation-resources/

[6] https://internationaldataspaces.org/wp-content/uploads/IDS-Reference-Architecture-Model-3.0-2019.pdf

[7] https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Publications/gaia-x-technical-architecture.pdf?__blob=publicationFile&v=5

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

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

[10] https://www.opendei.eu/case-studies/state-of-data-spaces-by-finland-ministry-of-transport-and-communications-october-2021/

Reference models represent a common structure and language to describe and specify system architectures; as such, they help a common understanding and promote system interoperability. In the architectural definition of a Digital Platform, the alignment to reference models is positive because they provide a framework for the standardisation of relevant technical systems, considering industry best practices, from development, through integration, as well as to operation.

The current wave of Innovation Actions (falling under the umbrella of OPEN DEI) is tackling the impact on Digital Transformation pathways via adoption of IoT, Big Data, Artificial Intelligence (and some precursory studies/adoption of Data Spaces). The challenge is to make full use of these technologies as well as the most recent ones that enable, in fact digital platforms play a key role in addressing competitive pressures and integrating new technologies, apps and services.

Even though many initiatives (both at R&D&I level, but also commercial ones) have already demonstrated most of the value of such key enabling technologies (KETs), many organisations and critical value chains are struggling into their adoption in concrete business scenarios. The deployment of trusted solutions is emerging as a key prerequisite for the acceptance of these KETs by end-users and other stakeholders (e.g., value chain participants). However, most use cases are still grounded in black-box systems that cannot be easily trusted nor become accepted by end-users. Nevertheless, their use in real word use cases remains very limited and has not been validated at scale, which limits the quality of experience for users and is a setback to human comprehension of the facts.

The activities conducted by OPEN DEI in such a topic are oriented to lower the barriers for all the stakeholders in four targeted domains (i.e., manufacturing, energy, agrifood and healthcare) and their value chains to adopt and fully leverage trusted technologies in the field of Digital Transformation, in ways that will enable end-to-end sustainability toward the innovation market and especially SMEs.

The report D2.1 “Reference Architecture for cross-domain Digital Transformation”[11], provides useful insights to the most relevant work in the field of Reference Architecture for building Digital Platforms to support the Digital Transformation journeys in the four sectors targeted by OPEN DEI (i.e., manufacturing, agriculture, energy, and healthcare). It presents a detailed analysis of the State of the Art on general purpose architectures as well as standard architectures, representing the foundation of the **OPEN DEI Reference Architecture Framework** (RAF) specifications and its six underlying principles (interoperability, openness, reusability, avoid vendor lock-in, security and privacy, support to a data economy).

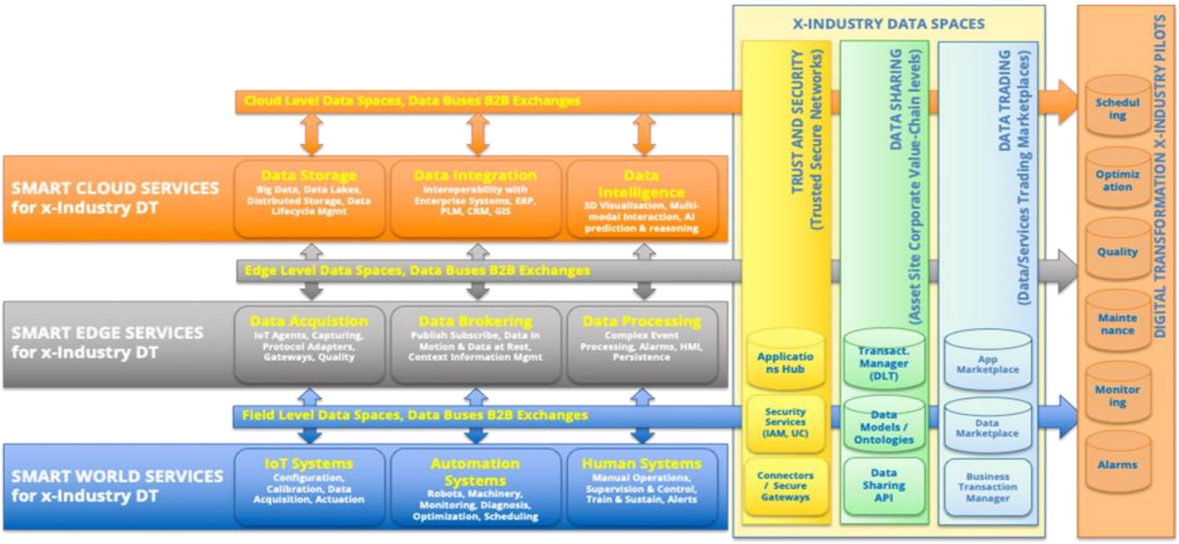


Figure 1 – OPEN DEI reference architecture framework

The OPEN DEI RAF therefore represents a very high-level abstraction of a platform supporting digital transformation of organisations/business/companies, but it does not represent a specific business case (or a set of) or technological approach.

---

[11] https://www.opendei.eu/case-studies/d2-1-reference-architecture-for-cross-domain-digital-transformation/

2.2.2 Digital Platforms Convergence - DSBA Initiative

Digital Technologies and especially AI are progressively migrating from the cloud to the edge, creating a Data Sovereignty computational continuum between IoT and Cloud, supported by initiatives like GAIA-X and the newly borne **Data Space Business Alliance**[12] (a new initiative born in September 2021 to join forces towards a pan-EU trustworthy data space, implementing EU priorities and values. BDVA, IDSA GAIA-X and FIWARE are joining forces for that). The concept of Data Space is gaining momentum, implying a highly distributed architecture following the temporal and spatial life of a product / asset and the interoperability of multi-stakeholders' data repositories. In this context, the alliance, having all the joint capabilities, resources, tools and expertise, supports existing organisations and data spaces to arise, evolve and go to market. IDSA, GAIA-X and DSBA architectural models, specifications and experimental pilots need now to be considered as well as highly distributed applications: FAIR standard Datasets, Sovereignty driven Digital Platforms and configurable and flexible data access and data usage models are at the basis of the EU Data revolution. Moreover, organisations (especially SMEs) could greatly benefit from the on-going shift of data from cloud to edge, as this provides opportunities for applications that reside closer to the field and feature low-latency and real-time performance, while boosting data protection (e.g., protection of intellectual property (IP) and personal and confidential data). One of the first assets to be jointly created in the alliance is a so-called data space radar that will include all data spaces and use cases based on the concept of data space out there providing a unique overview to monitor the data space evolution on a global level.

Finally, the convergence of aims of the associations constituting the DSBA is also among the strongest community efforts to identify, share and assess the availability and maturity of technical solutions in the form of reusable building blocks to build/extend your own digital platform toward embracing the new Data Space opportunities. Among these actors, the FIWARE Community has been really active in the last months, also publishing two relevant positions papers on Data Spaces[13] and Digital Twins[14], providing useful insights on available technology to solve this new challenge.

2.2.3 Purpose of Reference Architectures

A reference architecture is an architecture for making architectures that exhibit known commonalities. One purpose of any reference architecture is to capture commonalities that are the subject of standardisation efforts, for example, in support of *integration-by-design,* i.e., designing so that separate building blocks can be integrated in a solution, and interoperability-by-design, i.e., designing so that required interoperability capabilities are provided. A reference architecture is intentionally incomplete. It can be very detailed with respect to some identified commonalities, rather generic with respect to other identified commonalities and provide no guidance or information on other aspects of a system and its implementation.

Example of ISO/IEC/IEEE 42010 - architecture description and ISO/IEC JTC 1/AG 8 work

ISO/IEC/IEEE 42010[15] provides the basis for describing any architecture via a set of architecture views and architecture models (or view components). This set is governed by architecture viewpoints and model-kinds, respectively. Additionally, ISO/IEC/IEEE 42010 provides some basics for system architecting by introducing such concepts as stakeholders (individual, team, organisation, or classes thereof, having an interest in a system) and concerns (interest in a system relevant to one or more of its stakeholders).

ISO/IEC JTC 1/AG 8 (Meta Reference Architecture and Reference Architecture for Systems Integration) is currently working on the provision of guidelines of developing reference architectures, which will allow for the constructions of solution architectures leveraging individual reference architecture standards.

[12] https://data-spaces-business-alliance.eu/

[13] https://www.fiware.org/marketing-material/fiware-for-data-spaces/

[14] https://www.fiware.org/marketing-material/fiware-for-digital-twins/

[15] The first edition was published in 2011 (https://www.iso.org/standard/50508.html). The new edition will be available shortly (https://www.iso.org/standard/74393.html)

The resulting document entitled *mRA* (for *meta reference architecture*) describes a *canvas* with several viewpoints and model-kinds to provide a standardised way for architecting reference architectures. It also leverages the concept of patterns, or reusable artefacts that can be used in the construction of architectures.

The construction of architectures integrating digital platforms in DEI can be described as follows:

- **Flexible unified construction:** reference architectures are used to define digital platforms for all domains in the scope of DEI. All reference architectures and all domain platforms are methodologically standardised, i.e., they are built on the same principles and logic. The objective is to enable cross-domain and cross-platform interoperability.
- **Specialisation to application domains:** each vertical (or application) domain (i.e., energy, agriculture, manufacturing and healthcare) can have its own vertical domain reference architectures and vertical domain platforms. Some of the architecture artefacts can be conveniently expressed as patterns. Existing reference architectures, such as RAMI and SGAM in the manufacturing and energy domains, can be used as starting points (patterns) for developing standardised vertical reference architectures.
- **Specialisation of digital platforms to large scale pilots needs:** Each technical domain (i.e., Artificial Intelligence, Digital Twin, IoT, Big data, etc.) has its own vertical domain reference architecture and vertical domain platform. Such a platform can be deployed in large scale pilots (LSP) as a separate instance which is interoperable-by-design. Further, each LSP may enrich and adjust its own platform instance for its own needs.

The resulting RA canvas is described in Figure 1. Each vertical domain may have its own extension of any technological domain. An additional platform, called "Unified", can provide a reference IT infrastructure for all LSP.

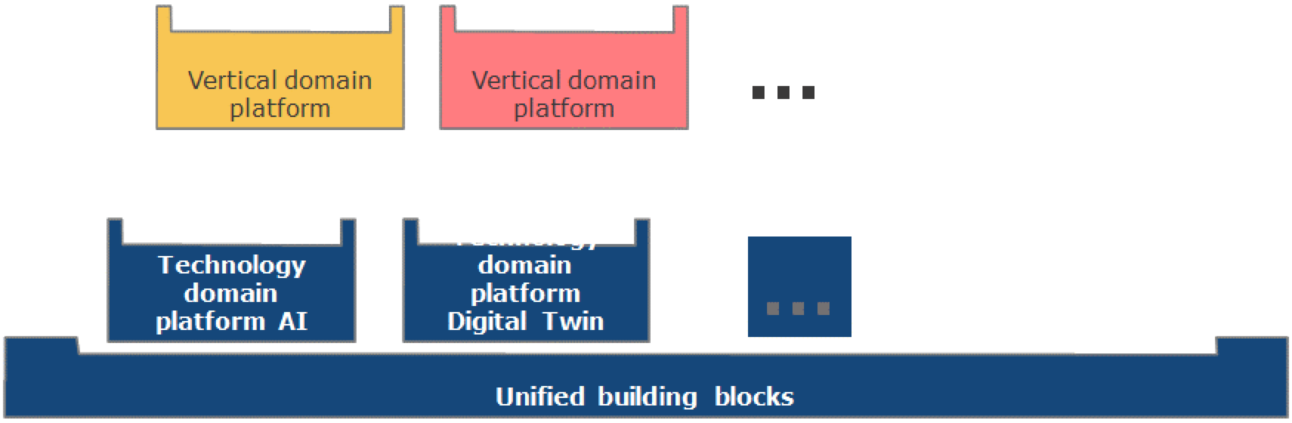


Figure 2 – Reference architecture canvas for digital platforms

Standardisation work related to IoT and Digital twin

The convergence effort carried out at standardisation level on reference architecture as described in the previous section is currently being applied in the area of Internet of Things and Digital Twin. This includes:

- The transition of ISO/IEC 30141 Edition 1 (IoT reference architecture)[16] to ISO/IEC 30141 Edition 2 (IoT reference architecture)[17]. This transition is based on the guidelines provided by ISO/IEC JTC 1/AG 8 (Meta Reference Architecture and Reference Architecture for Systems Integration)
- Guidance on the integration of other architecture related standards such as ISO/IEC 5392 (Reference Architecture of Knowledge Engineering)[18], or PWI JTC1-SC41-5 (Digital Twin - Reference Architecture) [19].

[16] https://webstore.iec.ch/publication/60606
[17] https://www.iec.ch/dyn/www/f?p=103:38:227387890241887::::FSP_ORG_ID,FSP_APEX_PAGE,FSP_PROJECT_ID:20486,23,104064
[18] https://www.iso.org/standard/81228.html
[19] https://www.iec.ch/ords/f?p=103:38:209600985860995::::FSP_ORG_ID,FSP_APEX_PAGE,FSP_PROJECT_ID:20486,23,104896

In most of these solutions, a central role is played by sensors, actuators and digital platforms that collect data and distribute it via the network, where it can be further processed using algorithms at the cloud computing level.

The standard provides a standardised IoT reference architecture based on the vocabulary (ISO/IEC 20924) and best practice for the generic design of smart applications.

It is currently being re-edited in order to follow the ISO/IEC JTC 1/AG 8 meta reference architecture approach.

2.2.4 Building an architecture

The targeted transformation is illustrated by Figure 2 and Figure 3. The first shows the current architecture context consists of siloed domain specific digital platforms. The second shows the transformation into federated cross-domain digital platforms.

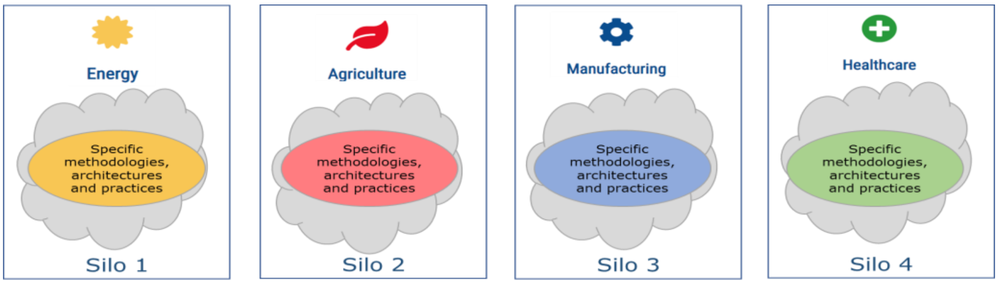


Figure 3 – Siloed Digital Platforms

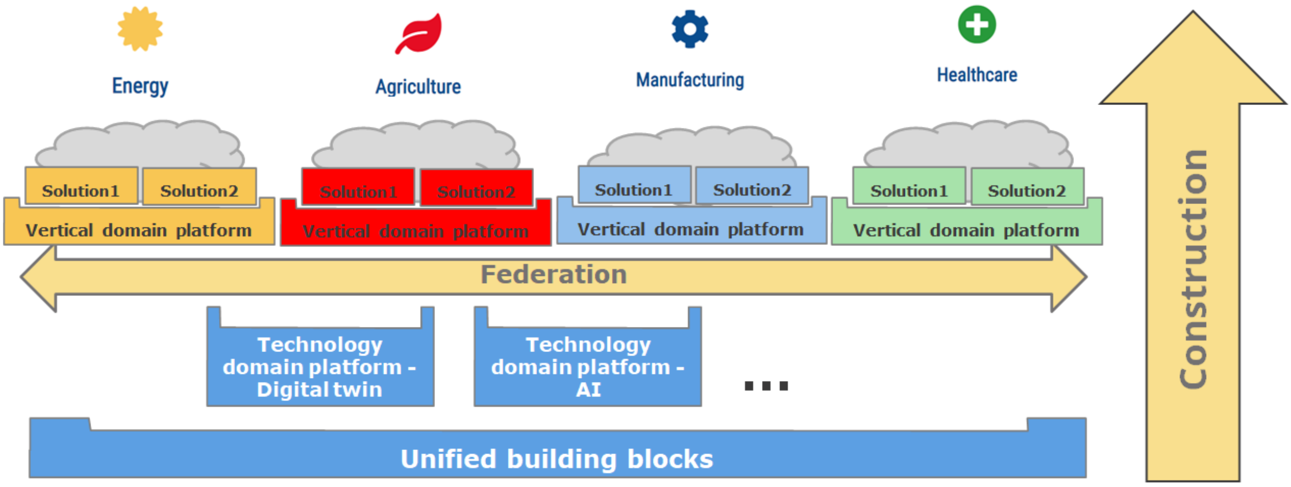


Figure 4 – Constructing an architecture for cross-domain systems

2.3 Interoperability Frameworks

2.3.1 Purpose of interoperability Frameworks

Interoperability is defined as the ability for two or more systems or applications to exchange information and to mutually use the information that has been exchanged (ISO/IEC 21823-1).

There are several definitions of framework, for instance

- a structure of processes and specifications designed to support the accomplishment of a specific task (ISO/IEC 21823-1).
- a particular set of beliefs or ideas referred to in order to describe a scenario or solve a problem (ISO 15638-6)

An interoperability framework can therefore be defined as

- a structure of processes and specifications designed to support the creation of interoperability
- a particular set of beliefs or ideas referred to in order to enable interoperability.

Example of ISO/IEC 21823 6 IoT Interoperability

ISO/IEC 21823 provides an IoT interoperability framework that is described in Figure 5. It focuses on the following interoperability facets:

- transport interoperability which focuses on the compatibility of the communication infrastructure established to exchange data between entities;
- syntactic interoperability which focuses on the ability to exchange information based on formats;
- semantic interoperability which focuses on the ability to understand the meaning of the data model in the context of data exchange;
- behavioural interoperability which focuses on ensuring that the results of the use of the exchanged information match the expected outcome; and
- policy interoperability, which focuses on the ability of entities to interoperate with a legal, organisational and policy framework.

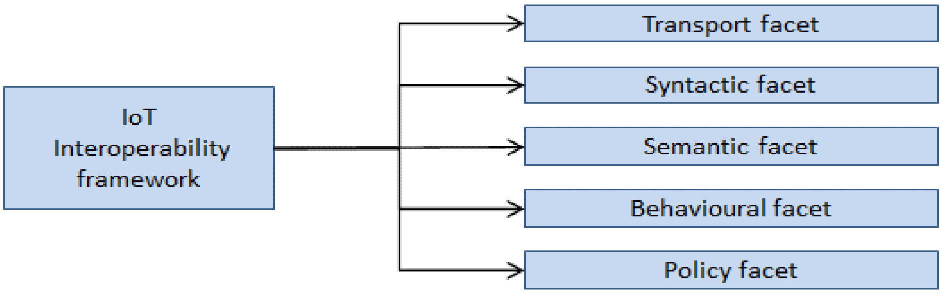


Figure 5 – IoT interoperability framework

Example of ISO 23903 Health informatics — Interoperability and integration reference architecture — Model and framework

The IEEE Standards Glossary defines interoperability as the ability of a system or a product to work with other systems or products without special effort on the part of the customer. Under traditional ICT focus, interoperability is the ability of two or more systems or components to exchange information and to use the information that has been exchanged.

According to ISO 23903:2021, Interoperability has evolved during the last 30 years from structured messaging (e.g. EDI, HL7® messaging) over sharing concepts (e.g. openEHR® or ISO 13606 Archetypes as well as ISO 13940 system of concepts to support continuity of care) – both representing the data/information exchange paradigm – to cooperation at application level. All those solutions focus on information and communication technologies (ICT) systems interoperability using ICT terminologies and ontologies for representing data, information, or even concepts and services, thereby distinguishing the three interoperability levels: a) foundational, b) structural, and c) semantic interoperability. At the next level, we provide knowledge sharing at business concept level to enable agreed service function level cooperation. Solution examples are web services or healthcare specific HL7 FHIR (Fast Healthcare Interoperability Resources). In ecosystems, we have to provide advanced, knowledge-level and business process focused interoperability between all principals acting in those ecosystems such as persons, organisations, devices, applications, components, or objects to achieve the common business objectives (ISO 23903). This results in interoperability levels as described in Table 1.

Table 1 – Interoperability levels[20]

Information Perspective		Organization Perspective
Interoperability Level	Instances	Interoperability Level

[20] after Blobel B (2017) Standardization for Mastering Healthcare Transformation – Challenges and Solutions. European Journal for Biomedical Informatics, 2017; 13,1:9-15.

Technical	Technical plug&play, signal & protocol compatibility	Light-weight interactions
Structural	Simple EDI, envelopes	Data sharing
Syntactic	Messages and clinical documents with agreed vocabulary	Information sharing
Semantic	Advanced messaging with common information models and terminologies	Knowledge sharing at IT concept level in computer-parsable form Coordination
Organisation/ Service	Common business process	Knowledge sharing at business concept level Agreed service function level cooperation
Knowledge based	Multi-domain processes	Knowledge sharing at domain level Cross-domain cooperation
Skills based	Individual engagement in multiple domains	Knowledge sharing in individual context Moderated end-user collaboration

The different interoperability levels deploy different representation styles (languages) with different generative power and restrictions to a special structure ranging from regular through context-free, context-sensitive up to recursively enumerable sets. This results in different means to enable communication and cooperation from simple data definition standards and API for implementing them through data interpretations (semantics), web services platforms up to a systems-oriented, architecture-centric, ontology-based and policy-driven approach.

Example of the European Interoperability Framework

This European Interoperability Framework[21] suggests an interoperability model shown in Figure 6 which is applicable to all digital public services and may also be considered as an integral element of the interoperability-by-design paradigm. It 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'.

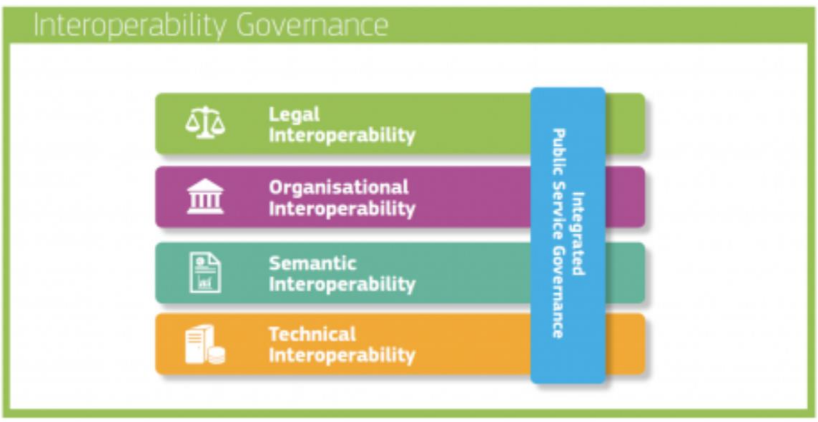


Figure 6 – European Interoperability Framework

[21] The detailed framework is available here : https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/european-interoperability-framework-detail. A brochure is available here : https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf

Extending the interoperability framework

An interoperability framework can be augmented in order to take into account specific needs. Figure 7 describes an extended interoperability framework:

- a **vertical interoperability viewpoint** can be added in order to take into account domain specific needs: smart manufacturing, energy, agriculture or health;
- an **horizontal interoperability viewpoint** can be added in order to take into account technology specific needs: IoT, digital twin, AI, data spaces;
- an **ecosystem interoperability viewpoint** can be added in order to take into account systems of systems specific needs: cross-domain support, cross-geography support (e.g. geo-interoperability), resource management, interoperability points; and
- an **organisational interoperability viewpoint** can be added to take into account governance, human aspects (e.g., human interoperability, empowerment), FAIR practice (Findable, Accessible, Interoperable, and Reusable), and compliance.

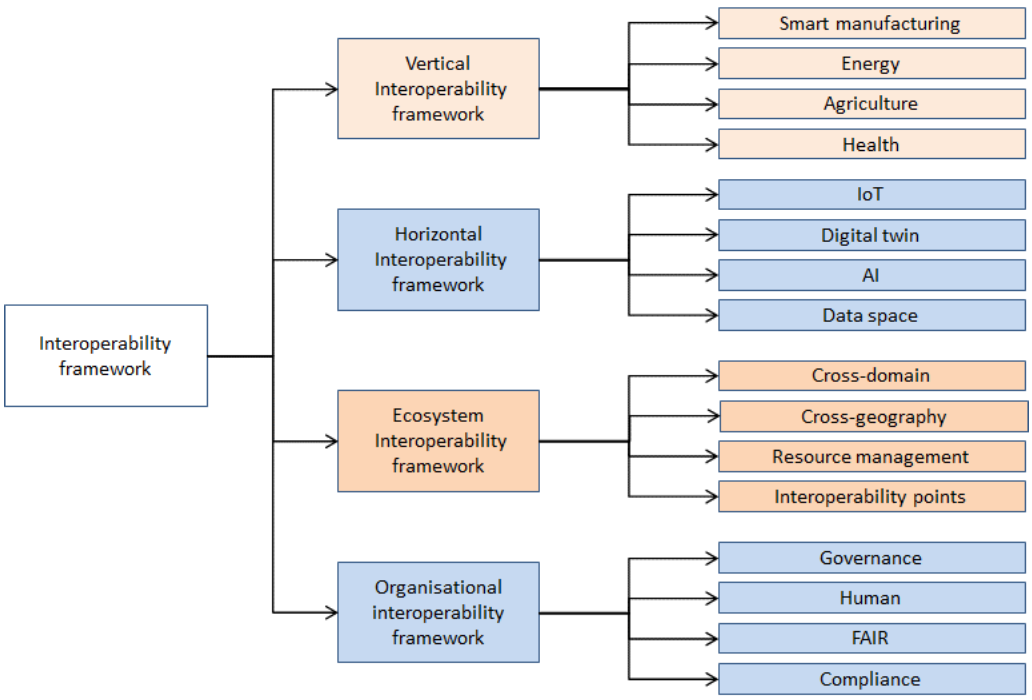


Figure 7 – Extended interoperability framework

2.3.2 Building Interoperability

Interoperability cases

There are many situations where interoperability is needed. When this takes place, two artefacts have been created:

- An **interoperability point**, i.e., a location in the overall system where data is exchanged according to an agreed interoperability specification
- An **interoperability case**, i.e., a documented justification and agreement on an interoperability point.

The availability of these two artefacts leads to the publication of **interoperability profiles**.

Figure 8 explains the relationship between an interoperability case and an interoperability profile: An interoperability case includes both the justification and agreement on an interoperability specification, and an interoperability profile is the main work product resulting from the interoperability case

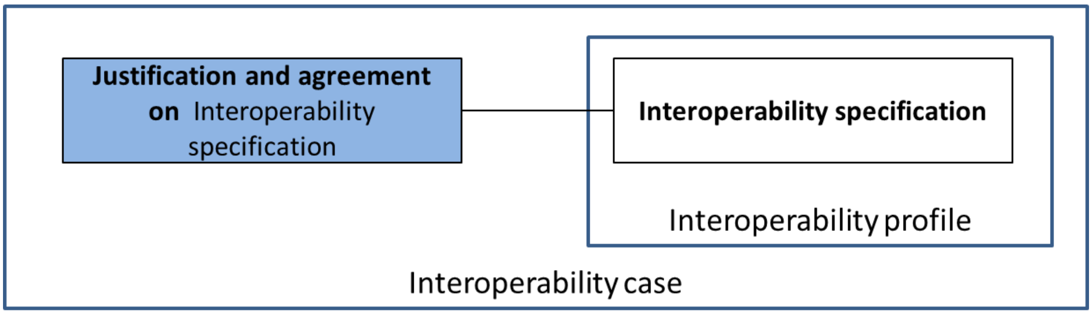


Figure 8 – Interoperability cases

As shown in Figure 9, an interoperability case can include various facets. It can therefore include a justification and agreement on each of the required facets. The interoperability case records all decisions made concerning each facet.

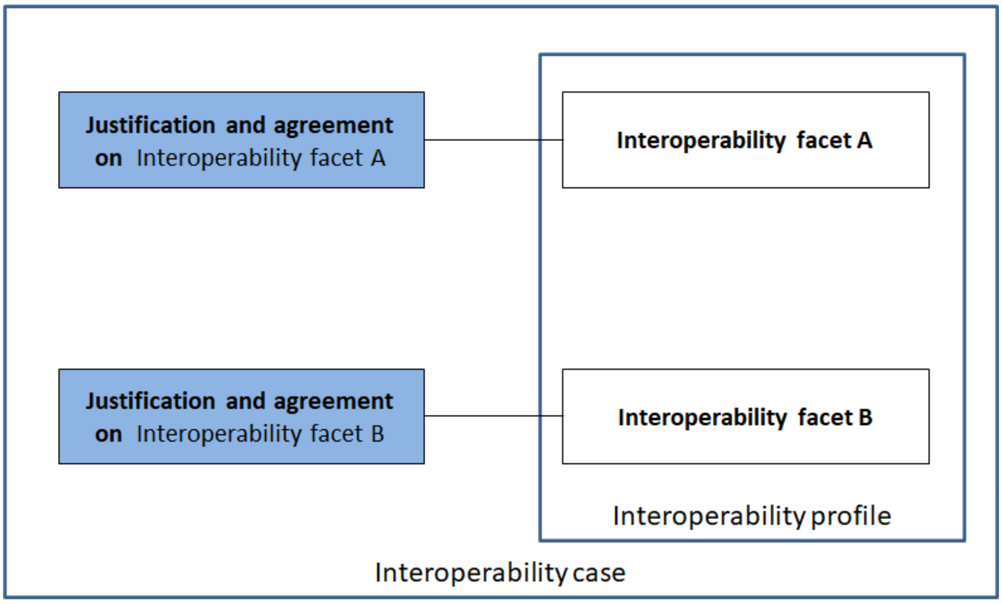


Figure 9 – Interoperability cases integrating multiple facets

The creation of an interoperability case requires detailed and agreed processes. Figure 10 shows the construction of an interoperability case for the semantic facet as carried out in the Interconnect project. The methodology used is based on the SAREF development methodology provided by ETSI[22], which is itself based on the Linked Open Terms (LOT) methodology[23]. The resulting interoperability case includes the following:

- Use case and domain documentation for data exchange
- Ontology purpose and scope
- Ontology requirement specification document (ORSD)
- Ontology implementation
- Ontology maintenance

[22] ETSI TR 103 411 https://www.etsi.org/deliver/etsi_tr/103400_103499/103411/01.01.01_60/tr_103411v010101p.pdf
[23] https://lot.linkeddata.es/

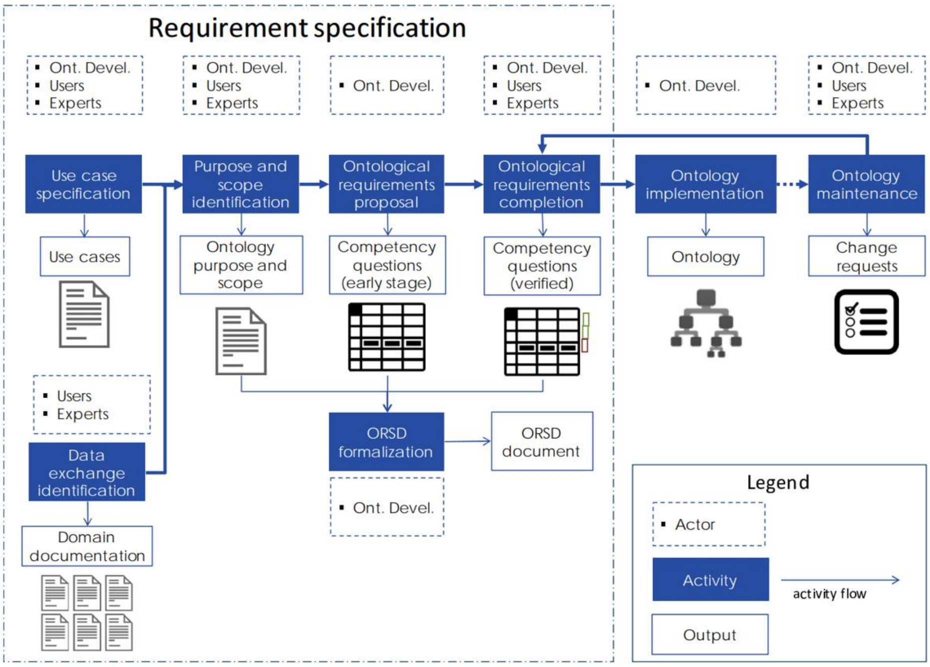


Figure 10 – Example of interoperability case for the semantic facet

Interoperability-by-design

The development of interoperability/integration solutions require a process (interoperability-by-design), which has three outcomes as showed in Figure 11:

- An interoperability point, or a location in an architecture where interoperability is needed. The Figure shows an interoperability point for a digital twin.
- An interoperability case, or a set of reasoned justifications on the interest of a given interoperability profile
- An interoperability profile, which is the information and guidance that can be used to create interoperable systems

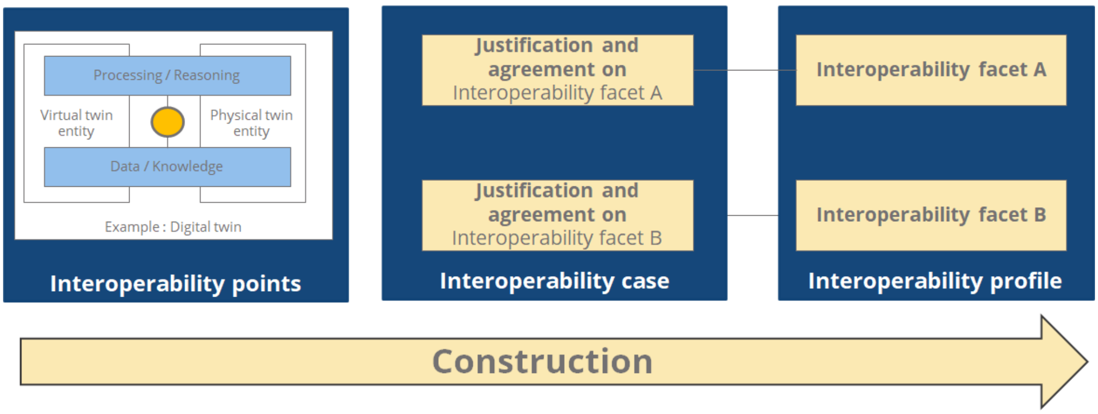


Figure 11 – Constructing interoperability

2.3.3 Hyperdimensional Interoperability

Policy awareness and Context awareness

The vision of a smarter cyber-physical infrastructure requires full awareness of policy and context. AI systems for instance will require context-making to improve performance and explainability: they need to be policy-aware and context-aware. Data needs to be put into context for IoT and cyber physical systems to remain relevant and adaptive to changes in use cases and scenarios.

Context can be defined as follows:

- Context is multi-dimensional. As shown in Figure 12, it encompasses the following dimensions:

- Semantic (meaning and logic)
- Spatial (physical and situational)
- Societal (values and value)
- Systems (networks and ecosystems)

- As shown in Figure 12, Context is represented by meta-data models that describe the activities of people, places and things over time. Context needs to be shared between networks of heterogeneous devices and applications empowering them to proactively offer enriched, situation-aware and usable content, instructions and experiences.
- Context is made up of the elements of relationships between entities, objects, locations and actions—commonly known as the Who, What, When, Where, How and Why of any scenario, situation or circumstance. The answers to these questions are often stored in different data silos and different data spaces. They need to be made interoperable, shareable and addressable by multiple competing AI algorithms that can maintain their coherence at scale.

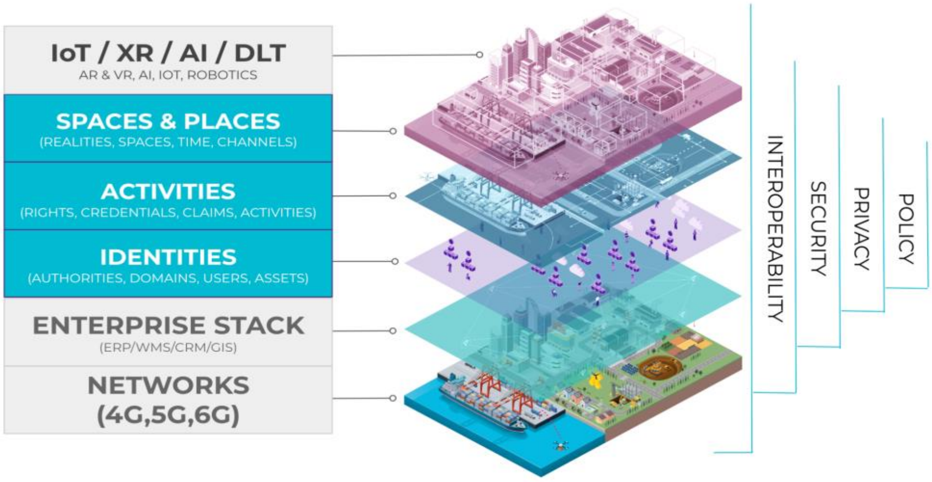


Figure 12 – Multi-dimension Interoperability

Spatial Web initiative and IEEE P2874 standards

The Spatial Web Foundation[24] has defined a contextual model and communication protocol that captures multi-dimension data interoperability. They are being **standardised** within the IEEE WG P2874[25], namely:

- HSML - Hyperspace Modelling Language
  - A common data model that enables adaptive intelligence at scale;
  - A standard that models spatial and hyperdimensional relationships which can exist between any base elements and their purpose.
- HSTP - Hyperspace Transaction Protocol
  - Multi-dimensional range query;
  - A contracting protocol that queries that language and sends and receives the common language's data.
  - Protocols allowing stakeholders to govern identities, activities and spaces and location in an interoperable way and across data domains.
  - Governs interactions between parties to ensure privacy and security.

Note that IEEE has designated HSML and HSTP as "Public Imperative". This designation is typically reserved for critical public infrastructure like nuclear energy, smart grids, and voting machines.

[24] https://spatialwebfoundation.org/

[25] https://sagroups.ieee.org/2874/

Spatial domains

One of the dimensions of the Spatial web inititiative are spatial domains. They are digital titles linked to 3D volumetric locations such as buildings, ports, streets, or larger regions, such as cities, states, continents and trading blocs as shown in Figure 13. Spatial subdomains represent subspaces that have a holonic structure, which allows for the orchestration of hierarchical rights and policies.

Spatial domains enable secure management of digitally mediated rights and permissions for:

- Who/what is authorised to access the domain;
- What content or data is available to view;
- Who can publish and modify content;
- Who can transact or interact with it.

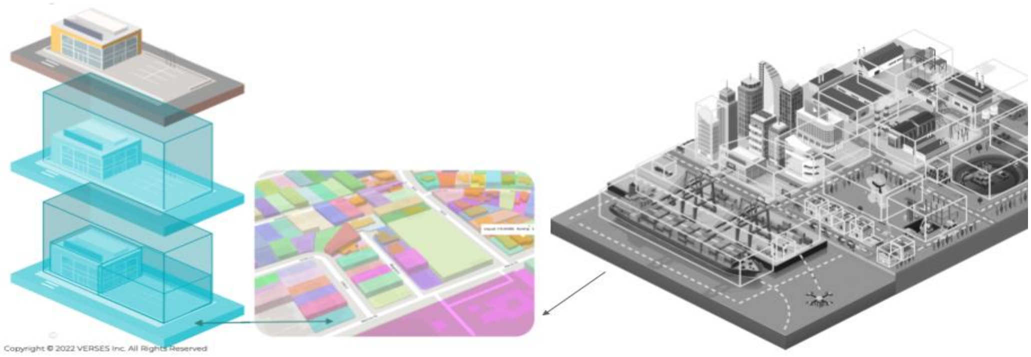


Figure 13 – Examples of Spatial Domains

Spatial domains can be characterized as follows

- They contain policies, rules and regulations: spatial domains contain rules, rights, permissions and fees associated with a geospatial region as defined by their owner or authority. This requires the use of a governance layer of geospatial information written in HSML that can be queried by actors as they approach and move through them.
- They used 3D models, containing geometries, addresses, spatial anchors, location of IoT devices, and information of subdomains. 3D models are dynamic, updated in real time as local conditions change, and represented through a visual digital twin.

Building block for hyperdimensional interoperability

Figure 14 shows a building block that enables hyperdimensional interoperability. This building block is part of the **Metadata and Discovery Services** technology building block[26]. It includes the following capabilities:

- *Internetworking spatial domains and context sharing using machine understandable and executable policies*
- *Operational AI-based assistance, e.g., cross-domain search, AI-based simulation and recommendation, and policy-based AI execution.*

COSM[27] is an example of building block implementation: It can be described as an operating system built upon HSML and HSTP will allow for the deployment of artificial intelligence applications in a way that integrates physical, logical, and operational meta-data to allow for autonomous, efficient, compliant and secure flow of people and things in the real world and/or in a metaverse.

[26] Page 44 of https://design-principles-for-data-spaces.org/ describes 12 building blocsk

[27] https://www.verses.io/cosm-os

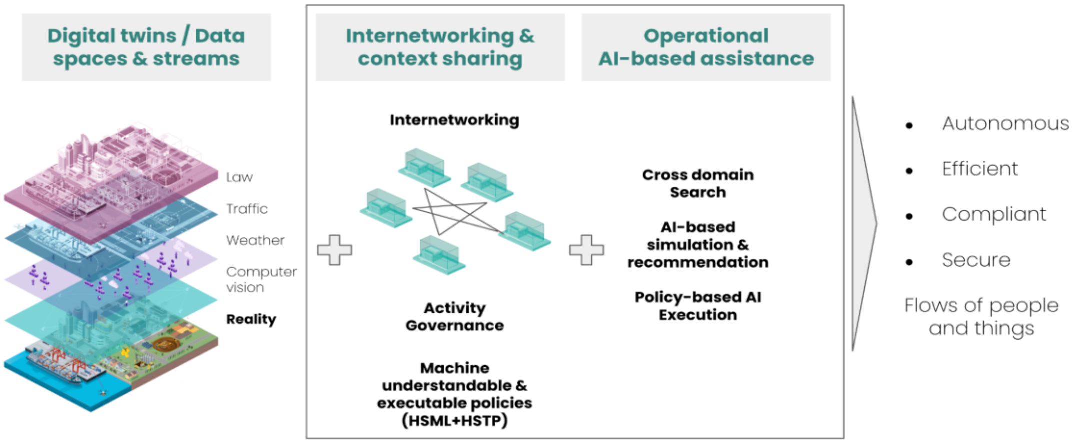


Figure 14 – Multi-dimension Interoperability Building Block

2.4 Topics of Interest for Federated Platforms

2.4.1 Trustworthiness

Trustworthiness is important because humans need trust when working with non-human entities. IEC e-tech platform[28] provides several examples: manufacturing plants of programmed robotic arms and humans working in close proximity side by side; transport systems using automated systems, for instance self-driving vehicles deploying advanced driver assistance systems, or modern airline autopilot and safety systems using manoeuvring augmentation characteristics systems.

In systems of systems, one untrustworthy system can produce all connected systems untrustworthy. As an example, Buchheit M. et al[29] describe how the "*cybersecurity researchers Charlie Miller and Chris Valasek in 2015 and 2016 created a series of technical security exploits that allowed near-total remote control of a consumer vehicle, a 2015 Jeep Cherokee. Their work demonstrates how alignment of assumptions about the operational context is necessary across the different components from the supply chain. The designers of the entertainment unit used different contexts from those engineering the CAN bus*".

Trustworthiness standardisation

Trustworthiness attribute is defined in ISO/IEC TR 24028:2020[30] and ISO/IEC 30145-2:2020[31], for the artificial intelligence and the smart cities' contexts, as the ability to meet stakeholders' expectations in a verifiable way. It may include characteristics such as reliability, security, privacy, safety, resilience, among others. Trustworthiness concept can be applied to products or services, as well as technology or data, for different types of organisations or governments.

There are several trustworthiness related projects under development at ISO / JTC1, a summarized description follows, as described in JTC 1/WG 13 Business plan v1.0[32]:

- Projects from JTC 1 / WG 13 - Trustworthiness
    - JTC 1 Standing document Landscape of trustworthiness: an inventory of trustworthiness and trust terms, as well as a "heat map" of trustworthiness characteristics.
    - ISO/IEC TS 5723, Trustworthiness – Vocabulary: definition of trustworthiness of systems and their associated services along with a selected set of its characteristics.

[28] Establishing trustworthiness is vital in our human-machine world. e-tech magazine. IEC. https://etech.iec.ch/issue/2019-04/establishing-trustworthiness-is-vital-in-our-human-machine-world, July 2019.
[29] Buchheit M., Hirsch F., Martin R.A., Bemmel V., Espinosa A.J., Zarkout B., Hart C.F., Tseng M. The Industrial Internet of Things Trustworthiness Framework Foundations. Version V1.00 – 2021-07-15. https://www.iiconsortium.org/pdf/Trustworthiness_Framework_Foundations.pdf.
[30] ISO/IEC TR 24028:2020 Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence
[31] ISO/IEC 30145-2:2020 Information technology — Smart City ICT reference framework — Part 2: Smart city knowledge management framework.
[32] ISO/IEC JTC 1/WG 13 N 321. JTC 1/WG 13 Business plan v1.0.

  - ISO/IEC PWI 5957, Trustworthiness – Reference architecture: a generic reference architecture for the engineering of trustworthiness in systems and their associated services.
  - ISO/IEC PWI 9814, Trustworthiness – Overview and concepts
- Projects from JTC 1 / SC 27 – Information security, cybersecurity and privacy protection
  - ISO/IEC TS 24462, Ontology for ICT trustworthiness assessment: an inventory of building blocks conceptually associated with different types of assessments, an ontology (i.e., a meta-model) that organises the building blocks, and guidelines for using the inventory of building blocks and the ontology.
- Projects from JTC 1 / SC 38 – Cloud computing and distributed platforms
  - ISO/IEC PWI 11034, Cloud computing – Trustworthiness of cloud services: This document provides an overview, frameworks, and concepts for trustworthiness including the use, provision and management of cloud services.
- Projects from JTC 1 / SC 41 – Internet of things and digital twin
  - ISO/IEC 30147, Internet of things – Integration of IoT trustworthiness activities in ISO/IEC/IEEE 15288 system engineering processes: provides system life cycle processes to implement and maintain trustworthiness in an IoT system or service by applying and supplementing ISO/IEC/IEEE 15288:2015. The system life cycle processes are applicable to IoT systems and services common to a wide range of application areas
  - ISO/IEC 30149, Internet of things – IoT trustworthiness principles: provides principles for IoT trustworthiness based on ISO/IEC 30141 – IoT Reference Architecture
- Projects from JTC 1 / AG 8 - Meta Reference Architecture and Reference Architecture for Systems Integration
  - Meta Reference Architecture Specification – Trustworthiness Annex: the Meta Reference Architecture for Systems Integration (mRA) Specification describes a meta-reference architecture (mRA) suitable for use on creation of a wide range of reference architectures (RAs). The mRA is the root for a family of interoperable-by-design RAs for use within the ISO and IEC JTC 1 standardisation scope. The document is expected to provide an Annex about trustworthiness.

In addition CEN-CENELEC JTC21 (AI) has started an initiative for AI trustworthiness characterization that aims at addressing the need for standards required by the EU proposal for AI regulation[33].

2.4.2 Universal resource management

IoT platforms are an essential factor in providing interoperability because they connect various devices (e.g., sensors, access points, and data networks) and provide services to the user. Therefore, interoperability, such as requesting services and sharing resources among diverse IoT platforms, is important, and it is essential for a real IoT environment.

IoT platforms have many challenges to interoperability, such as support for diverse protocols, device discovery, well-defined semantic management, and processing of data formats in heterogeneous IoT platforms. However, **current diverse IoT platforms and related standards make it difficult to achieve interoperability and collaboration between heterogeneous IoT platforms**. Especially regarding resource interoperability issues, each IoT platform has been developed using a specific and unique resource identifier, including a different type of resource-request format, so it is difficult to identify resources among heterogeneous IoT platforms. Furthermore, the **existing approaches mainly focus on integrating and managing each IoT platform's ontology and a method of duplicating resources for the target IoT platforms**. It makes it a burden for the developer to construct specific ontologies for the diverse IoT platforms

[33] AI act: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

Requirements for resource identifiers

The ISO/IEC 30181 ED1[34] describes the requirements of the resource identifier and an IoT resource name system (RNS), which focuses on mapping and converting the format of a resource identifier among diverse IoT platforms to overcome the heterogeneity of resource identifiers in various IoT platforms.

The requirements for interoperability of resource identifiers in heterogeneous IoT platforms are **uniqueness, equality, unity, persistence, reuse, scalability, and security** which are described below:

- Identifier allocation should be organised in such a way that individual IoT platforms can allocate their own set of identifiers without conflicting with other IoT platforms.
- Various established and emerging identifiers used by the many different IoT platforms should be considered and supported. Mapping identifiers between different IoT platforms should be supported (e.g., mapping related entities' identifiers).
- Because identifiers are requested in many requests and responses during the entity's lifetime, persistence and reuse of identifiers should depend on the application types. These properties also affect scalability, and some identifier types, such as IP addresses, require restrictions on the level of persistence due to their purpose.
- Node management includes node registration and identification and the storage and processing of the massive volume of data generated by physical nodes. In addition, metadata values such as identifiers of nodes used in various IoT platforms require efficient management.
- Issued identifiers should identify the correct entity and remain so during allocation, transfer, and usage. Signing the identifier has been proposed as one method to avoid duplication and identifiers used for other entities, and identifier correctness verification should be achievable online or offline.
- Other requirements include identifier anonymization, non-unique identifiers, identifiers that contain no personal data, disabling identifier tracking, and access control to identifier information.

Collaboration between resource name servers (RNSs)

Figure 12 is the flowchart to map and convert the format of resource identifiers among heterogeneous IoT platforms. Local IoT RNSs are modularized on each server device and the root IoT RNS is connected to the local IoT RNSs to manage the entire resource table.

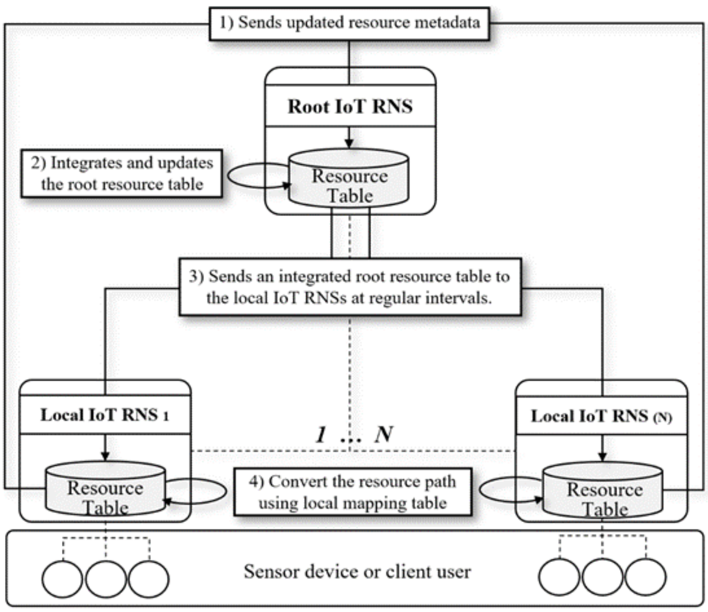


Figure 15 – Flowchart of IoT Resource Name Servers

- The root IoT RNS manages the metadata for connected and disconnected resources. Each local IoT RNS sends related metadata (e.g., device type and device ID) to the root IoT RNS when new

[34] https://www.iec.ch/ords/f?p=103:38:209600985860995::::FSP_ORG_ID,FSP_APEX_PAGE,FSP_PROJECT_ID:20486,23,108552

devices and services are registered and informs the root IoT RNS when devices and services are deleted or disconnected.

- The root IoT RNS updates the resource table with metadata received from each local IoT RNS.
- The root IoT RNS transmits the integrated resource table to each local IoT RNS. Regardless of updates, the integrated resource table is sent to each local IoT RNS at regular intervals.
- The local IoT RNSs convert and store the resource path in the resource table received from the root IoT RNS, depending on the format requested by each platform using the mapping table. Based on the converted path stored on the platform server, the end device for each platform can request the service to the end device of any other platform.

The potential of decentralised verifiable credentials

As mentioned above, ISO/IEC 30181 ED1 is IoT centrally focused but one needs to complete this vision by considering other resources (including humans) and the need overall to support decentralisation when addressing Identity management and distributed agreement.

Today's federated identity management (FIM) systems have some weaknesses, as they are putting Identity providers (IdP) at the centre of the identity ecosystem. This architectural constraint makes them a great target to attack. By making them a central point of storage hackers can regroup skills and resources to target an IdP and gain access to millions of user names and passwords. In addition, some may define them as privacy-invasive as they are able to monitor and track any login event from each user. To avoid such abuse, IdPs have to follow heavy security controls and apply different security mechanisms to fulfil their requirements with privacy laws.

In that context, different security experts came up with an alternative solution, where **decentralised Verifiable Credentials (VC)** would be an alternative and would shift the burden of securing privacy data from those IdPs to its real owner:

- Data owners can control in better ways their data and the granularity of the information or the credentials that they want to release.
- This would replace the IdP sending over all or limited credentials that sometimes are not needed by the service providers (SP) to be known.
- As an example, the combination of Fast IDentity Online (FIDO) protocol with W3C Verifiable Credential could provide a strong authentication and authorization and reduce the risk and cost for those Identity Providers.

In this respect, it is particularly striking to note the proposed short-term evolution of the eiDAS framework in the healthcare domain. The eIDAS regulation has not achieved its full potential. Implementation did not meet expectations. The coverage of the notified eID schemes is limited and final user uptake is highly different from country to country.

Example of the EU Digital Identity Wallet (EDIW)

A new framework called "EU Digital Identity Wallet (EDIW) is in the making: it should also allow users to create and use qualified electronic signatures and seals. The EDIW foresees the adoption of the Self-Sovereign Identity (SSI) paradigm. Digital identities are today managed by de-facto central authorities, namely the identity providers, resulting in users not having full control over their digital identities: in theory, identity data could be shared between service and identity providers without any user involvement. Most recently, new identity management models that focus on decentralised identities emerged and gained traction. One of the notable examples is the model of the SSI. The key concept in such systems is that control over identity data is put back into the users' hand. This gives the holder greater control over how its identity is represented to parties relying on the identity information and, in particular greater control over the personal information that it reveals to other parties. This paradigm change poses new challenges to practically deploy privacy-preserving showings of attributes, which allow the user to select their attributes that should or should not be revealed to the service provider. The EDIW can be used to hold not only an EU Digital Identity but also known attributes and other independently issued credentials of the identified entity, along the lines of the SSI paradigm.

2.4.3 Digital twin and AI integration

Federated architecture is expected to deliver high flexibility and agility among independently cooperating components and at the same time reduces complexity significantly. One way of achieving that would be to use Digital Twin integration. While the current generation of Digital Twins are replicas of existing entities in the physical world, future twins will help design or invent things that have not yet been realised, reducing the gap with virtual reality. Within that context, Working Group 6 of the ISO/IEC JTC1 SC41 is leading the development of international standards on Digital Twins:

- ISO/IEC 30173 ED1 Digital Twin - Concepts and terminology[35], establishes terminology for digital twin (DTw) and describes concepts in the field of digital twin, including the terms and definitions of digital twin, concepts of digital twin (e.g., digital twin ecosystem, lifecycle process for digital twin, and classifications of digital twin), Functional view of digital twin and digital twin stakeholders.
- ISO/IEC TR 30172 ED1 Digital Twin - Use cases[36], provides a collection of representative use cases of Digital Twin applications in a variety of domains, e.g., smart manufacturing, smart cities, etc.
- PWI JTC1-SC41-7 Digital Twin – Maturity model[37], provides a standardised generic digital twin maturity model, definition of assessment indicators, and guidance for a maturity assessment.
- PWI JTC1-SC41-5 ED1 Digital Twin - Reference Architecture[38] describes the rational and solution outlines of digital twin, and specifies a general Digital digital Twin twin (DTw) Reference reference architecture by different views.

The Digital Twin is a powerful concept and provides different levels of integration. The proposed reference architecture has a lot of similarities with FIWARE based Digital Twin Reference Architecture, where the Digital Twin data representation of the world is expected to contain all the information needed by the smart applications, not only measurable data but also other augmented insights and knowledge acquired over time. A Digital Twin approach provides the basis for data integration at the different levels, as illustrated in Figure 13.

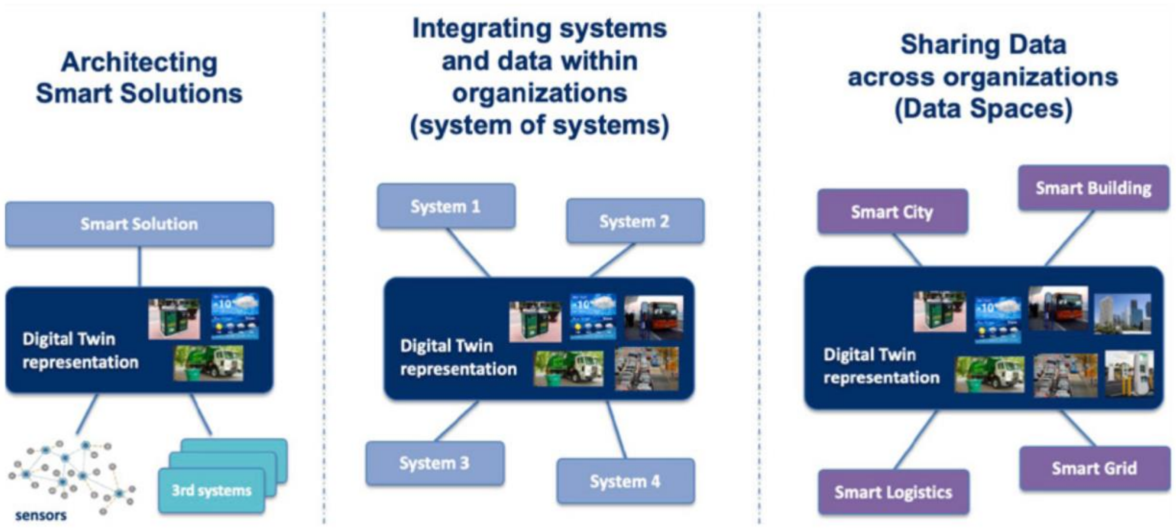


Figure 16 – Three cases of digital twins

Three cases could distinguished:

- Within a vertical Smart Solution, Digital Twin will be solving how main Building Blocks within the architecture can be integrated;

[35] https://www.iec.ch/ords/f?p=103:38:209600985860995::::FSP_ORG_ID,FSP_APEX_PAGE,FSP_PROJECT_ID:20486,23,104883
[36] https://www.iec.ch/ords/f?p=103:38:209600985860995::::FSP_ORG_ID,FSP_APEX_PAGE,FSP_PROJECT_ID:20486,23,104881
[37] https://www.iec.ch/ords/f?p=103:38:209600985860995::::FSP_ORG_ID,FSP_APEX_PAGE,FSP_PROJECT_ID:20486,23,108352
[38] https://www.iec.ch/ords/f?p=103:38:209600985860995::::FSP_ORG_ID,FSP_APEX_PAGE,FSP_PROJECT_ID:20486,23,104896

- Within a Smart Organisation, Digital Twin will be bringing support to the integration of the different systems within a smart organisation following a system of systems approach;
- Within a Smart Data Space, Digital Twin will be establishing the basic “common lingua” that systems linked to the different organisations speak and understand.

Nevertheless, two critical elements need to be standardised in order to support an effective data integration at these three levels:

- The API to get access to Digital Twin data
- The data models describe the attributes and semantics associated with the different types of Digital Twins being considered.

It is important to consider the data quality for analytics and Machine Learning (ML) when it comes to Digital Twins. In that context, the ISO/IEC 5259 series provides a way to understand and tackle such challenges.

- ISO/IEC 5259-1 (Data quality for analytics and ML - Part 1: Overview, terminology, and examples)[39] emphasises the crucial challenges faced by organisations to successfully implement big data and Artificial Intelligence (AI) systems. Therefore, most organisations are working to improve data quality from data collection to data analysis and for data use in model training for machine learning. No matter how good the data analytics and/or AI model performance is, if low quality data is inputted, the result will not be reliable, and if a service using a model trained with low quality data is operated incorrectly, it can be a direct threat to safety.
- ISO/IEC 5259-2 (Data quality for analytics and ML - Part 2: Data quality measures)[40] provides a data quality model, data quality measures, and guidance on reporting data quality in the context of analytics and machine learning (ML).
- ISO/IEC 5259-3 (Data quality for analytics and ML - Part 3: Data Quality Management Requirements and Guidelines)[41] specifies requirements and provides guidance for establishing, implementing, maintaining and continually improving the quality of data used in the areas of analytics and machine learning.
- ISO/IEC 5259-4 (Data quality for analytics and ML - Part 4: Data quality process framework)[42] provides general common organisational approaches, regardless of type, size or nature of the applying organisation, to ensure data quality for training and evaluation in analytics and machine learning. It includes guidelines for supervised machine learning with regard to the labelling of data used for training machine learning systems, including common organisational approaches for training data labelling; unsupervised machine learning; semi-supervised machine learning; and reinforcement machine learning. The proposed process is applicable to training and evaluation data that comes from different sources, including data acquisition and data composition, data pre-processing, data labelling, evaluation, and data use.
- ISO/IEC 5259-5 (Data quality for analytics and ML - Part 5: Data quality governance)[43] provides a data quality governance framework for analytics and machine learning to enable governing bodies of organisations to direct and oversee the implementation and operation of data quality measures, management, and related processes with adequate controls throughout the data life cycle.

In addition, data quality is necessary for ML and Big Data systems to be safe, reliable, and interoperable. ISO/IEC 20547-3 (Big data reference architecture)[44] states that “Data quality management is essential to big data systems, as poor data quality such as incomplete, false or outdated data can disable effective data mining processes, prevent useful findings or lead to wrong output”. Moreover, ISO/IEC TR 24028

[39] https://www.iso.org/standard/81088.html
[40] https://www.iso.org/standard/81860.html
[41] https://www.iso.org/standard/81092.html
[42] https://www.iso.org/standard/81093.html
[43] https://www.iso.org/standard/84150.html
[44] https://www.iso.org/standard/71277.html

(Overview of trustworthiness in artificial intelligence)[45] addresses challenges of data quality in AI systems based on machine learning, including bias in the data used to train the AI system, data poisoning, and adversarial attacks. Finally, ISO/IEC 8183 (Artificial intelligence — Data life cycle framework)[46] provides an overarching data life cycle framework that is instantiable for any AI system from data ideation to decommission. It is applicable to the data processing throughout the AI system life cycle including the acquisition, creation, development, deployment, maintenance and decommissioning.

Obviously, decision-making is a semantical task based on data models and concepts. Since semantics is available in an operational style, semantics can serve for analysis and synthesis purposes performed by a Digital Twin. The analytic results of the Digital Twin can be fed back to the phenomena of the thing analysed. Similar simulation results may lead to a synthetic re-design of a system where the former one is a real-time task the latter one is a disconnected task in the system life cycle.

As almost anything could be represented as a digital twin, the above collection of documents aims to provide guidance to create, develop, maintain, share, and use digital twins and AI. Organisations could work at different levels of their expertise and match the mirroring capability to provide similar visual appearance and renderance up to a federated and autonomous levels where digital twin integration will be used through federated interface among cross-domains without any human intervention. Each of those levels will have a certain level of Artificial Intelligence to help make decisions for small objects to huge ecosystems using semantic tasks based on data models and concepts.

2.4.4 Semiotic approach to support cyber physical systems

Semiotic is the study of signs and symbols. When digital platforms are associated with cyber physical systems, applying a semiotic approach can be used to model knowledge processing.

There are three types of artefacts:

- thing artefacts which are associated with physical phenomena;
- semantic artefacts which are associated with concepts;
- language artefacts which are associated with symbols.

These artefacts are related according to the semiotic triangle relationships:

- physical phenomena are defined by concepts;
- concepts are represented by symbols; and
- symbols are the signs of physical phenomena

Knowledge processing in digital platforms is achieved through iterative assignments called morphisms, that assign a semantic concept to a physical phenomenon to a representing symbol. The relationship is defined as an iteration in which when a symbol represents e.g. a set of requirements of a standard then, the concept behind defines the semantics of these requirements which in turn, can be compiled to a physical device or an entity that fulfils the descriptive requirements of the standard **and** the derived expectations from the anticipated concept. Since adjustments of these three morphisms among the semiotic domains is required for achieving knowledge on a phenomenon of interest the application of the three morphisms is performed in an iterative manner until the gained knowledge is decided to be sufficient for a certain task or proving a phenomenon.

An example of this kind of knowledge iteration can be observed in a self-driving car when it tries to recognize a traffic sign that is shadowed by clouds, rain, buildings etc. Then in real-time trying more than once would help to gain more and better knowledge about the object of interest from its iterative processing.

Note that knowledge is represented in a declarative manner by rules like 'cause-effect' implications which can easily be translated into graph artefacts of ordered pairs of interacting processes representing semantics if a formal manner. Thus since graphs are executable the semantics is executable as well and because of the defined morphisms the symbolic standard is indirectly executable and the considered

[45] https://www.iso.org/standard/77608.html
[46] https://www.iso.org/standard/83002.html

phenomenon respectively the derived knowledge about can be simulated in terms of the declared sets of graph manipulation rules (knowledge implications).

Examples of applying a semiotic approach is showed in table 2.

Table 2 – Applying the semiotic approach

Digital twins	The operation of a digital twin is always based on the data space of an application. Leveraging a semiotic triangle approach, the digital twin can verify whether the intended semantics of a certain concept fits with the current state of the phenomena observed at the system asset. The current state of the system is derived from the data space of an application. The concept may describe a process with its input-output transfer function. Thus, the digital twin simulates the process based on the current state of the according data space.
RAMI 4.0	Existing Reference Architectures like RAMI4.0 and others can be implemented by means of the semiotic triangle. Whereas the means of the semiotic triangle are the three semiotic domains of the semantic concepts, the physical phenomena and the symbols and sentences of the descriptive ontology together with the three morphisms between the three semiotic domains. Similarly, the means of the reference architecture are the three axes, in case of RAMI 4.0 they are the terms of life cycle and value stream, the hierarchical levels of production and devices and, the descriptive layers of interoperation. Life Cycle and value stream are to be represented in semantic terms of anticipated concepts, hierarchies of things and their phenomena are to be represented in terms of engineering artefacts like product, field and control devices, stations, work centres, enterprises and finally the connected world. The requirements of standards and the symbols of observed phenomena are given in a structured way in terms of layers.

2.4.5 Interoperability approaches

Syntactic interoperability refers to the exchange of information between transacting parties based on agreed formats and structure for encoding information. The data model becomes the contract entities must abide to exchange information. An example is the case for the NGSI-LD or CIM data models. Strict compliance with the data model is required to enable transacting parties to exchange data items. When parties do not share the same data model, an intermediary is needed to provide a translation (i.e., a middleware or broker) in case of information exchange., The correct exchange is ensured by who provided the technical implementation. Ultimately it is a human that can disambiguate in case translations are required (i.e., if field 'A' is the same as field 'a').

Semantic Interoperability is the ability that digital systems have to exchange data with unambiguous, shared and agreed meaning, a stepping stone for the implementation of a Digital Single Market. This is often supported by the use of ontologies, such as SAREF, together with ICT knowledge exchange mechanisms. With Semantic Interoperability, the contract is based on the exchange of knowledge rather than exchange of data according to a predefined syntax. This approach provides a new level of freedom to data modelling since as long as the domain is kept (i.e., often a group of ontologies), there is no strict need to comply with the syntax of a data model. Knowledge/information can be encoded in terms of related concepts ('semantic triples'). The use of ontologies makes the relationship between concepts unambiguous, as it is encoded explicitly and machine-readable in the ontology regarding what each piece of exchange data represents (i.e., what it represents in the domain, in what un,it of measure values are provided).

Because the relationship of concepts in the information is encoded explicitly, it is also possible to infer knowledge/information that is implicitly contained. If an information receiving party needs that particular kind of information, it can derive it automatically from the (other) information it has received.

Semantic interoperability allows different providers and companies to offer competitive and cost-effective solutions whilst avoiding lock-ins or closed vertical technology implementations.

Difference between language and semantic representation

Interoperability beyond syntactic interoperability means understanding of a syntactic correctly delivered message in a certain domain. A concept is independent from a chosen language (e.g. natural or formal language) used for the conceptualization of a phenomenon or for providing semantics to standard. Natural languages used in standards have human understandable semantics but formal languages usually have formalised computer interpretable semantics. Thus, both semantic approaches are different. In order to overcome these differences a kind of model semantics is required that can be interpreted by humans and similarly by computers or machines. But what is semantically shared by humans and machines is mathematics. Skilled humans can directly interpret and understand a stated mathematical concept. Machines and computers however need a translation of these mathematical concepts into an algorithm the machine can execute. The translation of formal concepts into a kind of machine executable representation is called **operationalization**. The axiomatization of formal concepts into a kind of human understandable representation is called **declaration**. Hence an operationalized and declarative specification of a phenomenon of interest must be unified by a common style of representation. A model semantics that supports both kinds of representations, i.e. the declaration of axioms of a problem and its operational style allowing simulations of this problem, would be a good choice and would make the approach independent from any kind of descriptive language semantics. In the scope of the semiotic triangle a good choice of appropriate modelling means would be an integrated approach based on graph and abstract data type theories. Research has provided several graph manipulation platforms that are available openly from the shelf. One of the most advanced ones is the KIT platform called GrGen.NET[47].

Example of H2020 Interconnect project

A particular example is provided by the H2020 Interconnect project. It rises above the idea of creating bilateral syntactic interoperability for all smart appliance/device manufacturers and energy/comfort service providers that are connected to a Home/Building Energy Management System (H/BEMS). It offers a consistent set of shared semantic concepts encoded in a set of aligned ontologies. By mapping their (manufacturers, service providers) own data models to the InterConnect set of ontologies, exchange of knowledge/information can take place in terms of shared and interrelated semantic concepts. As the InterConnect set itself is strongly related to widely accepted information / data model standards (EEBUS, ETSI, CENELEC, etc.), the amount of mapping effort across the industry is seriously reduced. To that, the use of ontologies provides for less effort in automatic translation of information across datasets. So, if functionalities change, or if another manufacturer is to be supported, less work is required for implementation. For example, in Interconnect, a H/BEMS can ask any question encodable in semantic triples according to the InterConnect set of ontologies (ICSO2), to any ICSO2 -compliant device or system. There is no need to implement specific adapters for answering specific questions. It enables a border-breaking level of interoperability by providing the ability to combine knowledge from multiple devices and services using automated reasoning.

Interoperability from a semantic viewpoint

The semiotic triangle as explained before in section 2.4.1 states that interoperability among components of things or services in the semiotic domain of observed phenomena shall be achieved by means of semantic concepts. Whereby an interface is an operational thing, interoperability is the semantic concept assigned to the interface by a certain morphism. On the operational level data formats and types must fit on the semantic level concepts must fit. For example, a concept could be energy transport from the place of its generation, via a distribution network and finally to the place of consumption. In this example there are two interfaces between different subsystems which transform formats of energy into other

[47] Edgar Jakumeit et al. The GrGen.NET User Manual Release v6.6 API, Universität Karlsruhe April 2022, http://www.info.uni-karlsruhe.de/software/grgen/

formats of energy. In some cases, system-wide constraints on energy transportation must be proven not being violated like the concept of energy stability in an operating network. Although in case the interfaces locally may operate correctly interoperability between subsystems would not be possible when global axioms would run in danger of violation during system operation.

Semantic interoperability permanently checks the behaviour of interacting processes from various systems where syntactic interoperability is checked once when the designed architecture is deployed. In the first case patterns are generated that can be compiled into a behavioural data model. Data models capturing best practices of interoperability form the basis of decision-taking, reasoning or forecasting.

2.4.6 Executable policies for digital governance

One challenge of interoperability is the support of policy and regulation. There is a need to maintain and enforce policies that can be interpreted and shared by machines and humans. Existing regulations have been drafted as text with little regard for interpretability and executability by machines.

Translating existing regulations into machine-interpretable and machine-executable code can allow for the governance of the behaviour of machines in a policy compliant way and dynamically adapt that behaviour as policies change and evolve over time. The challenge is to discover where and when technology can replace humans when it comes to interpretation, and where and why not.

Example of drones

For example, harmonised liability rules for drones are lacking on the EU level (U-spaces). Consequently, liability law is governed by the traditional rules at the national (Member State) level. An innovation, therefore, lies in (1) the development of liability rules that are (2) machine-interpretable and machine-executable. Finally, regulation is subject to change, and it is often multi-layered, meaning that regulatory bodies at different levels (e.g., EU level, state level, municipality level, company level) jointly contribute to a specific type of regulation. As a result, different pieces of regulation may consist of different languages. To make a regulation machine interpretable and machine-executable, and to make any interpretation model scalable, a semantic representation needs to be available that accommodates regulatory changes and regulation at different levels and in different languages. The technological challenge lies in developing a suitable semantic representation that is applicable to specific geolocations..

Development in the FF2020 Horizon Europen project

The FF2020 Horizon Europe project developed a geospatial infrastructure using the draft IEEE Spatial Web standards and protocols (HSML and HSTP) to govern actions within the domain of Urban Air Mobility to enforce rules and policies systematically. This solution allows current laws to be parsed into machine-readable models and interpreted programmatically, reducing the human interpretation of laws and guidelines into more binary statements. The solution takes the legal text and constructs ontologies and context from automated Natural Language Processing and Machine Learning methodologies. These are then used to construct machine-readable representations of specific legal scenarios. This information is then parsed into HSML, where it is combined with other sources of information (such as IoT devices, user input, cameras, drones, etc.) that enable an AI powering a drone to be able to understand and comply with laws and conditions.

Several laws that have been converted into HSML have been tested in real-world flight environments, including rules about dynamically changing variables. Existing EU laws for Urban Air Mobility were translated into HSML, ensuring that the proposed actions and flight plans automatically conform to the set of parsed rules and regulations at play:

- ensuring that HSML could accommodate the vast array of prohibitive and deontic laws found in the EU literature
- working with legal experts to ensure that the assumptions made during the translation process abide with existing interpretations of the relevant laws
- providing a mechanism for suggesting legal innovations when a machine-readable format cannot straightforwardly accommodate the ambiguity of laws

Capabilities of HSML to express executable policies

The following HSML capabilities were demonstrated

- evaluation of counterfactual actions - what would happen if someone were to perform a certain activity? We can use this to perform (abstract) routing - find (optimal) series of legal activities that get from HSML graph A to HSML graph B
- updates from a range of real-time data sources; allowing autonomous agents to dynamically operate within the confines of the law. If there is a law about maximum wind speed, dynamically update route given changing wind-speed conditions
- extension beyond laws any "rules of the games" can be constructed and abided by using the same schemas and systems
- assistance: the law is inherently ambiguous ("UA operators must fly safely" - how do you define "safely"?). HSML can help automatically identify ambiguous laws and can provide a feedback mechanism for developing future analog laws with machine-readability in mind. This can be the foundation for a virtuous two-way loop between analog law and digital law.

2.5 Aligning with Solutions

Figure 14 shows how reference architectures can be used to build a domain specific solution architecture:

- A digital platform reference architecture includes common blocks and patterns. Common blocks could be IoT related, digital twin related, data space related and so forth.
- Patterns can be domain specific. A SGAM representation is an energy architecture pattern.
- A digital platform solution architecture is constructed as the result of combining the common blocks and specific patterns. This leads to a digital platform implementation.

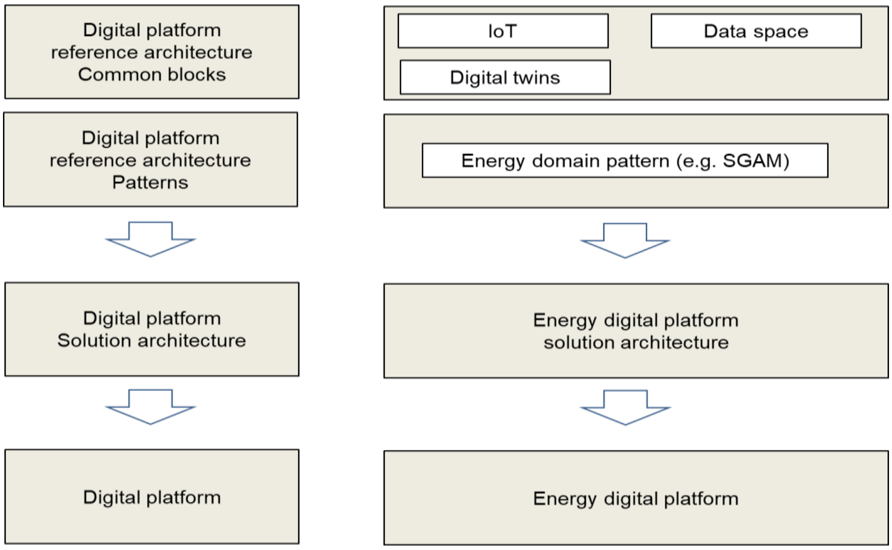


Figure 17 – Constructing solutions architectures

The resulting construction can lead to customised digital platforms that can still provide enough commonalities to allow for cross domain applications and data sharing as shown in Figure 15.

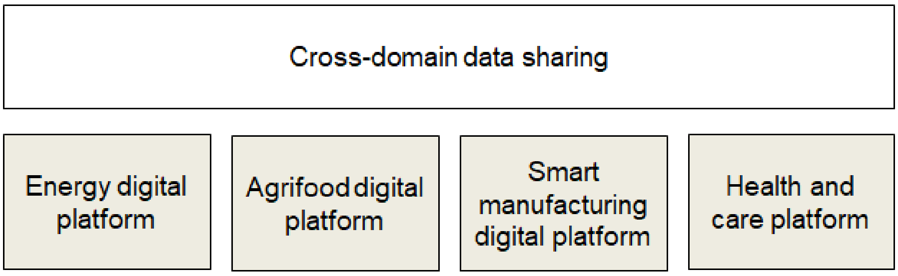


Figure 18 – Constructing federated platforms

This alignment approach is being specified by ISO/IEC JTC 1/AG 8 (Meta Reference Architecture and Reference Architecture for Systems Integration), and applied (as explained above in section 2.2.3 (Reference Architectures) in the development of ISO/IEC 30141 Edition 2 (IoT reference architecture)

3 REFERENCE ARCHITECTURES AND INTEROPERABILITY FOR DIGITAL MANUFACTURING PLATFORMS

3.1 Context for Manufacturing

The digital economy, defined by fundamental changes in the characteristics of information, computing, and communications, is now the preeminent driver of economic growth, social change, and sharing economy. In the digital economy, platform models have strongly proliferated over the past years, providing services on the top of technological building blocks used as a foundation to deal with competitive pressures.

From an economic viewpoint, Digital Platforms are restructuring the global economy, contributing to the digitalization of organisations, value chains and whole sectors, by resetting entry barriers, changing the logic of value creation and value capture. From a commercial viewpoint, Digital Platforms ease the creation of ecosystems of stakeholders, supporting new forms of innovation and value creation, as well as related business and commercial models, focused on Digital Platforms' underlying vision and value proposition.

3.2 Reference Architectures for Manufacturing

This chapter presents an overview of the most known Reference Architectures for Manufacturing, used in several business cases and in most of the cases enabling Digital Transformation pathways for their adopters. Reference models provide a solution-neutral reference architectural model for applications that make use of Internet of Things (IoT), big data analytics, and other technological advancements in manufacturing processes, what is known as smart manufacturing, intelligent manufacturing, or simply Industry 4.0. One of the main objectives once adopted is to be able to communicate the scope and design of the system, to foster collaboration and integration with other relevant initiatives by framing the developed concepts and technologies in a common model.

Probably the most known and used standard in Industry 4.0 is the Reference Architectural Model Industrie 4.0 (**RAMI 4.0**), developed by Platform 4.0 in 2015 and focused on the IoT and Cyber-Physical Systems (CPS) in the industrial manufacturing domain. RAMI4.0 is a three-dimensional model, which describes the Industrie 4.0 space and organises the lifecycle/value streams and the manufacturing hierarchy levels across the six layers of the IT representation of Industry 4.0. It outlines a comprehensive view of manufacturing-related implications to any IoT landscape. The primary topic, the integration of the physical asset and its digital representation, is proposed relying on a common representation called the **Asset Administration Shell** (AAS). The AAS is a standard model aiming to create a bridge between the tangible (real) world and the IoT (digital) world integrating assets into the world of information. An asset is everything that can be connected for implement an Industrie4.0 solution (i.e., machinery, parts, supply material, documents, contracts, etc.). The AAS also defines the data models for the exchange of information between partners in the value chain and a package file format (the Asset Administration Shell Package, AASX), to exchange the full or partial structure of the administration shell. The AAS implements the **Digital Twin** concept meeting requirements of different use cases coming from several domains guaranteeing:

- **Interoperability**, companies can communicate and exchange information
- **Availability**, for every kind of product (non-intelligent and intelligent)
- **Integration** of value chain
- **Covering** of the complete life cycle of products, devices, facilities, etc.
- **Basis** for autonomous systems and AI

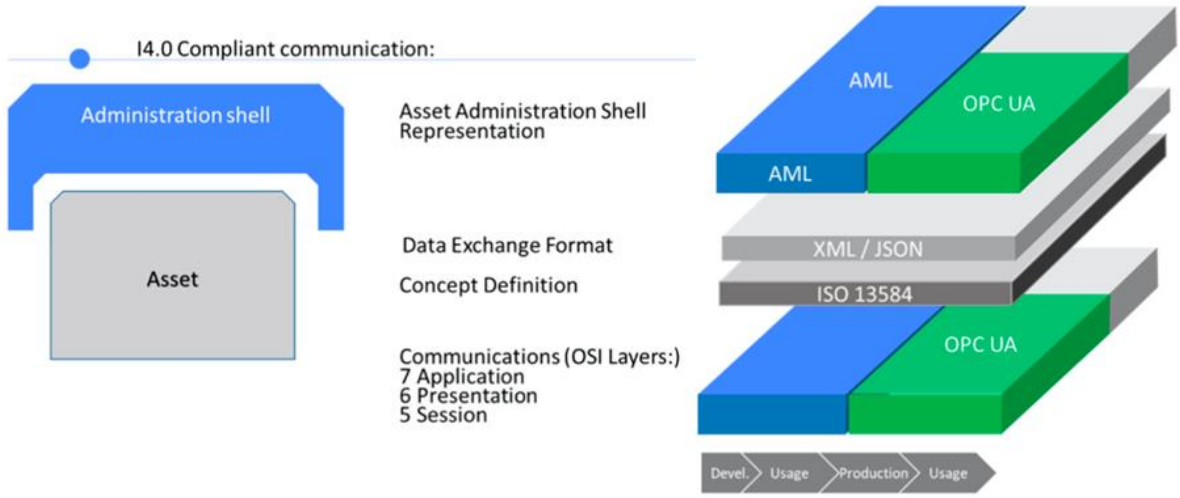


Figure 19 – I4.0 Communication protocol stack

Another prominent Reference Architecture for the manufacturing domain is the Industrial Internet Reference Architecture (**IIRA**), published by the Industrial Internet Consortium (IIC) in the document "The Industrial Internet of Things Volume G1: Reference Architecture"[48] and containing architectural concepts, vocabulary, structures, patterns and a methodology for addressing design concerns. The document identifies the fundamental architecture constructs and specifies design issues, stakeholders, viewpoints, models and conditions of applicability defining a framework by adapting architectural approaches from the ISO/IEC/IEEE 42010-2011 (Systems and software engineering - Architecture description)[49].

This international standard outlines the requirements regarding a system, software, and enterprise level architecture. The ISO/IEC/IEEE 42010 standard recommends identifying the perspectives of the different stakeholders that can be: system users, operators, owners, vendors, developers, and the technicians who maintain and service the systems. The aim is to describe system properties as seen from their viewpoint. Such properties include the intended use and suitability of the concept in terms of its implementation, the implementation process itself, potential risks, and the maintainability of the system over the entire lifecycle.

IIRA addresses concerns about IIoT across industries broadly, while RAMI4.0 focuses mainly on manufacturing in depth. Essentially, the IIRA attempts to identify the most important and common architecture concerns. It then provides an architectural template and methodology that engineers can use to examine and resolve design issues. In addition, the template and methodology suggest ways of addressing the top concerns, allowing designers to glean insights by examining architecture patterns, helping Industrial Internet of Things (IIoT) system designers to avoid missing important architecture considerations and this also helps them to identify design gaps of missing important system functions or components.

48 The Industrial Internet of Things Volume G1: Reference Architecture Version 1.9. (2019, June 19). https://www.iiconsortium.org/pdf/IIRA-v1.9.pdf

49 https://www.iso.org/fr/standard/50508.html

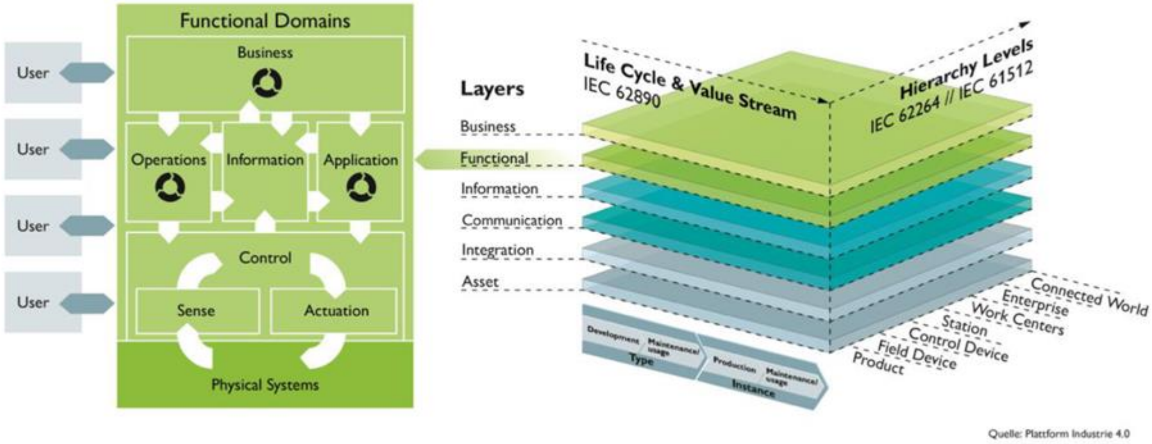


Figure 20 – IIRA viewpoints in RAMI 4.0

Apart from the above-mentioned standards, other communities of experts are developing new approached to enable out European manufacturing industries to exploit the most recent technological trends and benefits of the data economy. Some relevant examples are the community efforts led by FIWARE Foundation on one side, and by International Data Spaces Association (IDSA) on the other side.

FIWARE[50][2] is a curated framework of open source platform components (also referred as Generic Enablers – GE-) which can be assembled together and with other third-party platform components to build "powered by FIWARE" platforms that accelerate the development of interoperable and portable (replicable) Smart Solutions in multiple application domains. FIWARE tries to manage the data within a given smart vertical solution or break the existing information silos within a smart organisation by supporting access to a Context / Digital Twin data representation that manages at large-scale all relevant information. The FIWARE Reference Architecture does not only bring components for an effective exchange and management of digital twin /context data but brings a number of components for publication of data resources and its eventual monetization. Digital Industry is a key pillar in such a vision, aiming at developing an ecosystem of FIWARE-enabled software components, suitable to meet the challenges of Manufacturing Industry business scenarios, as indicated by Industry 4.0 vision.

The **FIWARE Smart Industry RA** defines a real ecosystem compliant with the main Reference Architectures in the manufacturing domain following the success in a domain like Smart Cities. The FIWARE RAs in other domains are also experiencing a growing adoption. The standardisation is a strength point (i.e., support of ETSI) bringing innovation and spreading knowledge. Last but not least, FIWARE brings a comprehensive set of components for data publication and monetization, in the latter case based on TM Forum recommendations, which pave the way for materialising Data Economy concepts.

The International Data Spaces Association (**IDSA**) is the evolution of IDS (Industrial Data Space) which itself was an initiative led by Fraunhofer ISST and promoted by the German Federal Ministry of Education and Research. IDSA is characterised by the focus on information ownership, with the aim of enabling clear and fair exchanges between data providers and consumers. To this end it suggests a reference distributed architecture that accomplishes this goal: the IDS Reference Architecture Model Version 3.0[51].

Broadening the perspective from an individual use case scenario to interoperability and a platform landscape view, the **IDS Reference Architecture Model** positions itself as an architecture that links different cloud platforms through policies and mechanisms for secure data exchange and trusted data sharing (through the principle of **data sovereignty**). Over the so-called IDS Connector (among the core components of the RAM, available as multiple OSS implementations), industrial data clouds, individual enterprise clouds, on-premises applications and individual, connected devices can be connected to the International Data Space ecosystem.

[50] FIWARE developers catalogue. https://www.fiware.org/developers/catalogue/

[51] https://www.internationaldataspaces.org/wp-content/uploads/2019/03/IDS-Reference-Architecture-Model-3.0.pdf

3.3 Interoperability Frameworks for Manufacturing

INTEROPERABILITY is one of the KEY CROSS CUTTING FACTORs defined by CONNECTED FACTORIES 2 project (sister project of OPENDEI, focused on digitalization of manufacturing) and can be defined in many ways and from many perspectives:

- Interoperability is to create a knowledge representation of an agent/asset in modular production.
- Interoperability is the seamless movement of data along different levels of knowledge.

Looking for a new way to define interoperability, it is important to look at the RAMI 4.0 in first place (Figure 18).

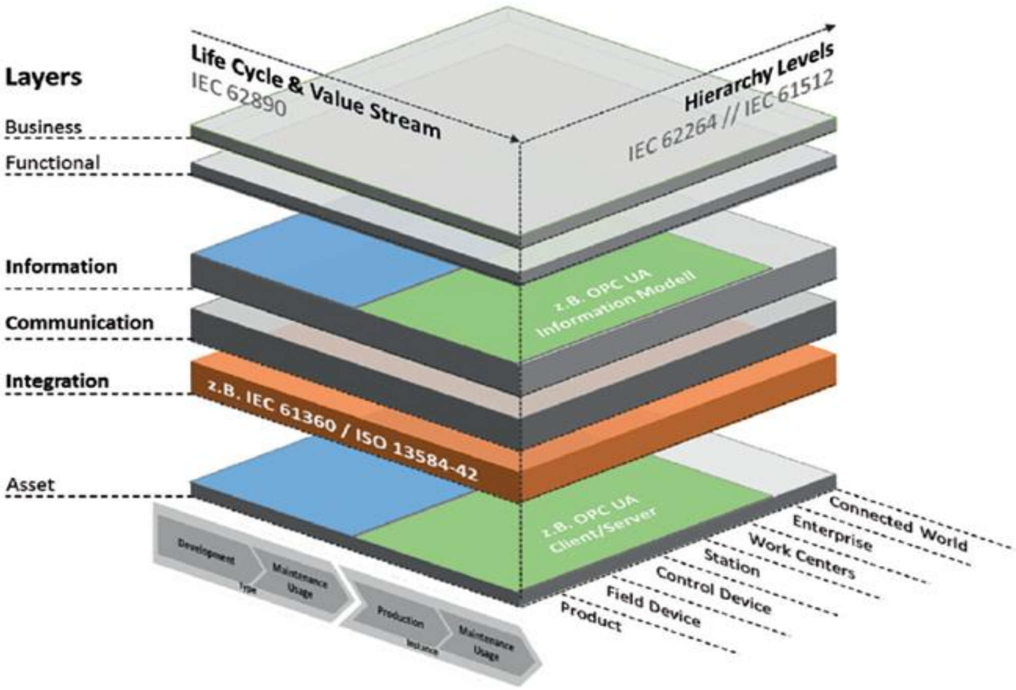


Figure 21 – Interoperability layers in RAMI 4.0

However, there are many ways to look at interoperability, and among them we can start from the different forms of data sources to make interoperable, distinguishing among open data, real time data, and unstructured data (streaming).

But there are other ways to address interoperability, for example, through distinguishing three aspects:

- DATA GENERATION: Data is generated on an asset, and integrated filtering the data (usually on the edge).
- DATA MANAGEMENT: then it is communicated through the information layers, transforming into structured data. This is usually the fog at platform level.
- DATA USAGE: And then is analysed either for improving functional aspects (performance, quality, productivity….) or business aspects. This is the cloud.

At last, another way to look at interoperability is:

- Technical interoperability: ability to exchange data (ISO/OSI).
- Syntactic interoperability: Description of the data with all its type attributes in uniform formats (OPC UA).
- Semantic interoperability: ability to interpret the exchanged data in such a way that intended actions can be recognized and triggered.
- Organisational interoperability: ability to act in non-technical processes.

As a preliminary conclusion of this chapter, we could say that RAMI4.0 permits the mapping of the different interoperability approaches of different projects and initiatives.

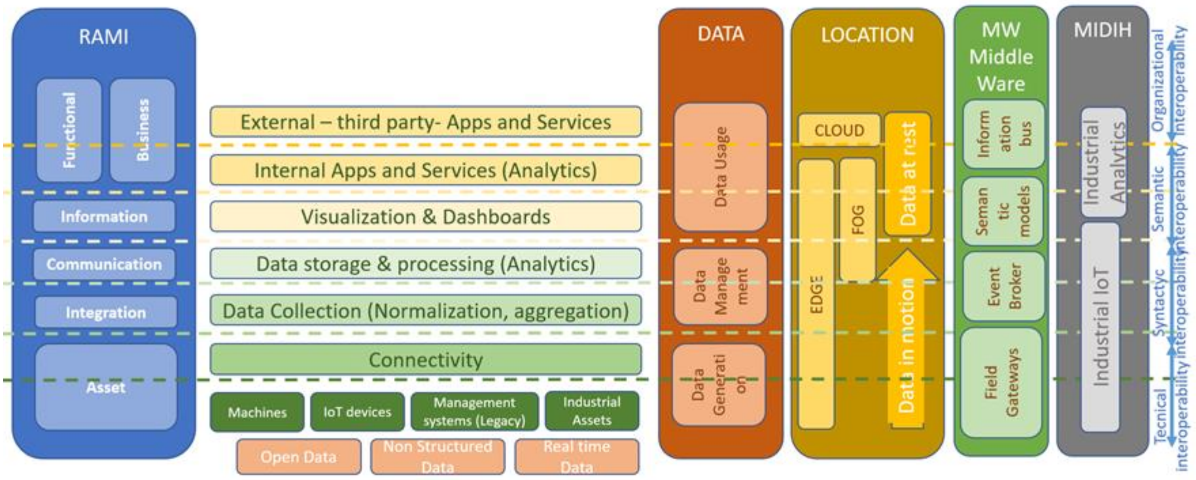


Figure 22 – Mapping RAMI 4.0 Interoperability layers

There are several approaches to the interoperability topic that need to be addressed when implementing interoperability:

- Proprietary approach: Decision of Interoperation of different commercial tools and platforms (against data protectionism). From a more technical point of view, interoperability is extremely relevant. This is not only since systems and tools will very likely come from different vendors, but also since many legacy systems have to be integrated.
- International approach: Data exchange amongst borders.
- Sector approach: Cross-sectorial reference architectures.
- Data resiliency approach: to guarantee the integrity of common repositories. (Blockchain technology).
- Integration approach: Amongst levels on the plant: Edge, fog, cloud.
- Semantic/information approach: Modelling correlation between information sources. Like Open ontologies. Or AAS Asset administration shell.
- Data sovereignty approach: to guarantee the trust amongst participants on data transactions (Industrial Data Space Association).
- Digital Twin & Digital Factory: exchanging data between real and virtual assets?

At the same time, interoperability can be implemented considering different company goals (Table 6):

- Quality goals: improvements on quality control.
- Productivity goals: improvements on productivity control.
- Maintenance goals: improvements on maintenance control.
- Traceability goals for material, resources and products.
- Production visibility goal: M2M & HMI in factory environment: evaluation of human machine interfacing HMI.
- Cooperation approach new forms of cooperation amongst companies and sectors either at supply chain or value chain.
- Logistic approach: improvements on logistics control.

At last, several requirements need to be taken into account when addressing interoperability, as e.g.:

- Safety when addressing the interoperability amongst human and machines/ robots.
- Privacy when addressing interoperability amongst different companies.
- Reliability when data are required for very accurate decisions (i.e. at machine level).

3.4 Aligning the Manufacturing domain for DEI

Following the alignment approach elaborated in section 2.5 (Aligning with Solutions), ISO/IEC 30141 Edition 2 (IoT reference architecture) is currently following the guidance of ISO/IEC

JTC 1/AG 8 (Meta Reference Architecture and Reference Architecture for Systems Integration) to show how smart manufacturing architecture based on RAMI can be aligned with ISO/IEC 30141 Edition 2 in order to enable cross-domain interoperability.

4 REFERENCE ARCHITECTURES AND INTEROPERABILITY FOR DIGITAL AGRIFOOD PLATFORMS

4.1 Context for Agrifood

The Agrifood domain has been the target of many recent EU projects and initiatives such as IoF2020[52], DEMETER[53] and ATLAS[54]. It is also related to the wider area of BioEconomies with projects like DataBio[55], that besides Agriculture also includes Fishery, Aquaculture and Forestry. Common for these initiatives is that there has been a focus on platform interoperability among the diverse set of proprietary platforms that is typical for these sectors/domains.

Further the Agrifood domain is often related to geospatial and spatiotemporal models and is has thus also been a focus on Geospatial reference architectures such as the standards from ISO/TC211, CEN/TC287 and OGC.

4.2 Reference Architectures for Agrifood

Within the Agri-food domain, there are currently multiple different proposals for reference architectures evolving from multiple (European) research projects and industry-driven initiatives like GAIA-X. Looking at the different architectures, two principal concepts can be identified: *data-model and ontologies oriented* architectures, and *service oriented* architectures.

One example for a system based on a service oriented architecture is the *ATLAS Interoperability Network*. The ATLAS Interoperability Network implements the technical solutions and the infrastructure to enable *trusted interoperability* between digital tools in agriculture. The technology developed in ATLAS provides the means to establish a trusted interoperability between the different systems and to establish the flow of data between these systems. The basic building blocks to realize this flow of data are the services offered by the participating systems. These standardized services enable the exchange of data in a formally defined and documented way. The exchange of data between services is designed to be peer to peer without any network centralistic component for specific distribution or steering of data flows, and with no centralized data storage. With such a decentralized approach, the system is highly scalable and resilient, and offers increased security compared to a central HUB approach. The target audience for the ATLAS Interoperability Network are companies offering digital software tools in the agricultural domain. ATLAS provides vendor and technology agnostic formal specifications of the APIs and data formats to which associated ATLAS Services must conform, the so-called *ATLAS Service Templates*. With these ATLAS Service Templates, existing systems can be leveraged to connect to other systems, or to be connected to by other systems from other providers with an comparably low effort.

The DEMETER project is using multiple reference architecture, including the BDVA and ADRA reference models as well as an AI and Big Data pipeline which also have a mapping to Digital Twin architectures.

The IoF2020 project have defined a layered architecture with a set of Interoperability Points.

The following are books/articles with chapter of the BDVA Reference Model - and further extensions and usages for Digital Twins and Bioeconomies and Agrifood

- The Elements of Big Data Value. Foundations of the Research and Innovation Ecosystem (2021)[56]

[52] https://www.iof2020.eu/
[53] https://h2020-demeter.eu/
[54] https://www.atlas-h2020.eu/
[55] https://www.databio.eu/
[56] Open access: https://link.springer.com/book/10.1007/978-3-030-68176-0

- Technologies and Applications for Big Data Value (2022)[57]
- Big Data in Bioeconomy (2021)[58]

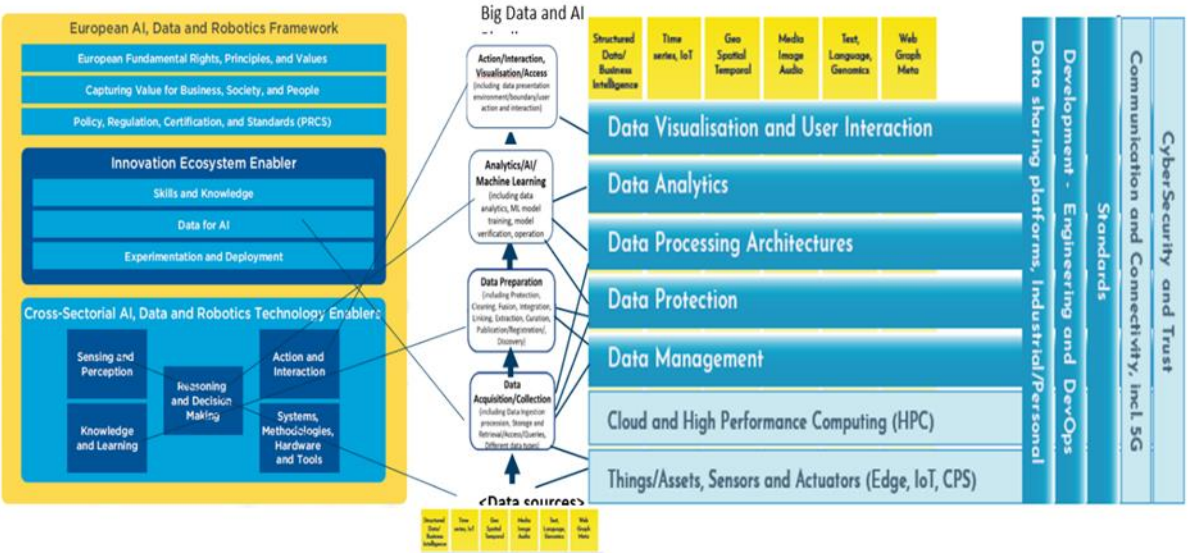


Figure 23 – BDVA and ADRA reference models

Figure 20 shows the relationship between the Big Data and AI Pipeline Framework and the BDVA and ADRA Reference Models, which is explained in more detail below.

The Big Data and AI Pipeline Framework

The Big Data and AI Pipeline Framework is based on the elements of the Big Data Value (BDV) Big Data Value Reference Model, developed by the Big Data Value Association (BDVA) [1]. In order to have an overall usage perspective on Big Data and AI systems a top level generic pipeline has been introduced in order to understand the connections between the different parts of a Big Data and AI system in the context of an application flow. Figure 21 depicts this pipeline, following the Big Data and AI Value chain.

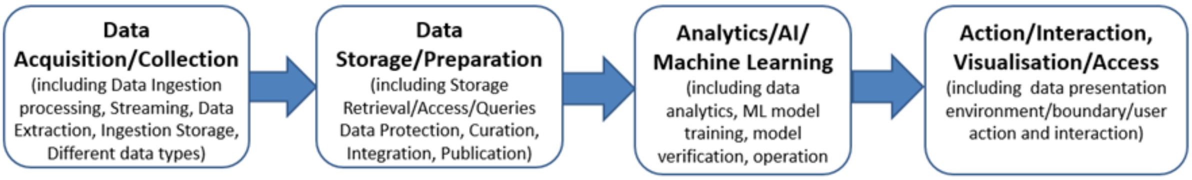


Figure 24 – Top level Generic Big Data and AI Pipeline pattern

The steps of the Big Data and AI Pipeline Framework are also harmonised with the ISO SC42 AI Committee standards [2]. It is in particular harmonised with the steps of *Collection, Preparation, Analytics and Visualization/Access* steps within the Big Data Application Layer of the recent international standard ISO 20547-3 Big data reference architecture within the functional components of the Big Data Reference Architecture [2], [6]. Figure 22 shows how the Big Data and AI Pipeline also can be related to the BDV Reference Model and the AI PPP Ecosystem and Enablers (from SRIDA AI). Benchmarks often focus on specialized areas within a total system typically identified by the BDV reference model. This is in particular useful for the benchmarking of particular technical components. Benchmarks can also be directly or indirectly linked to the steps of a Big Data and AI pipeline, which is useful when benchmarks are being considered from a Big Data and AI application perspective, where practical project experiences has shown that these steps easily can be recognized in most application contexts.

Benchmarks are useful both for the evaluation of alternative technical solutions within a Big Data and AI project and also for comparing new technology developments/products with alternative offerings. This

[57] Open access: https://link.springer.com/book/10.1007/978-3-030-78307-5
[58] https://link.springer.com/book/10.1007/978-3-030-71069-9

can be done both from a technical area perspective for selected components and from a pipeline step perspective when seen from the steps of a Big Data and AI application.

As it can be seen in Figure 21, this pipeline is quite high level. Therefore, it can be easily specialised in order to describe more specific pipelines, depending on the type of data and the type of processing (e.g. IoT data and real-time processing). The 3D cube in Figure 22 depicts the steps of this pipeline in relationship with the type of data processing and the type of data being processed. As we can see in this figure, the type of data processing, which has been identified as a separate topic area in the BDV Reference model, is orthogonal to the pipeline steps and the data types. This is due to the fact that different processing types, like Batch/data-at-rest and Real-time/data-in-motion and interactive, can span across different pipeline steps and, can handle different data types, as the ones identified in the BDV Reference Model, within each of the pipeline steps. Thus, there can be different data types like structured data, times series data, geospatial data, media, Image, Video and audio data, text data, including natural language data, and graph data, network/web data and metadata, which can all imply differences in terms of storage and analytics techniques.

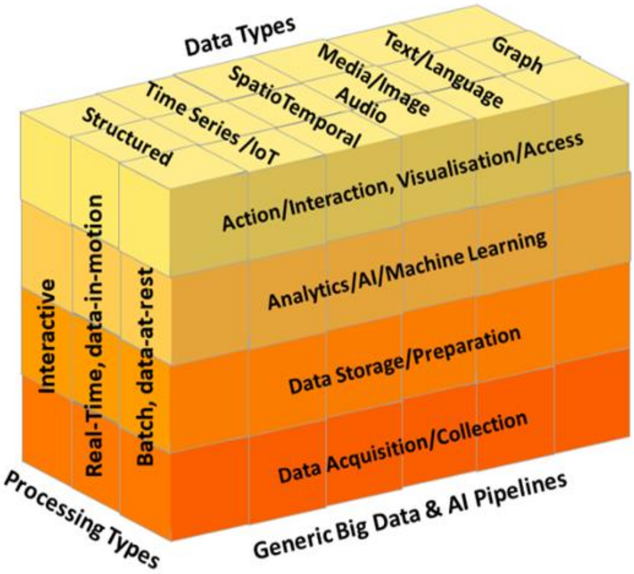


Figure 25 – Top level Generic Big Data and AI Pipeline cube

Other dimensions can similarly be added for a multi-dimensional cube, e.g. for Application domains, and for the different horizontal and vertical technology areas of the BDV Reference model, and for the technology locations of the Computing Continuum/Trans Continuum – from Edge, through Fog to Cloud and HPC – for the actual location of execution of the four steps, which can happen on all these levels. The same orthogonality can also be considered for the area of Data Protection, with Privacy and anonymisation mechanisms to facilitate data protection. It also has links to trust mechanisms like Blockchain technologies, smart contracts and various forms for encryption. This area is also associated with the area of CyberSecurity, Risk and Trust.

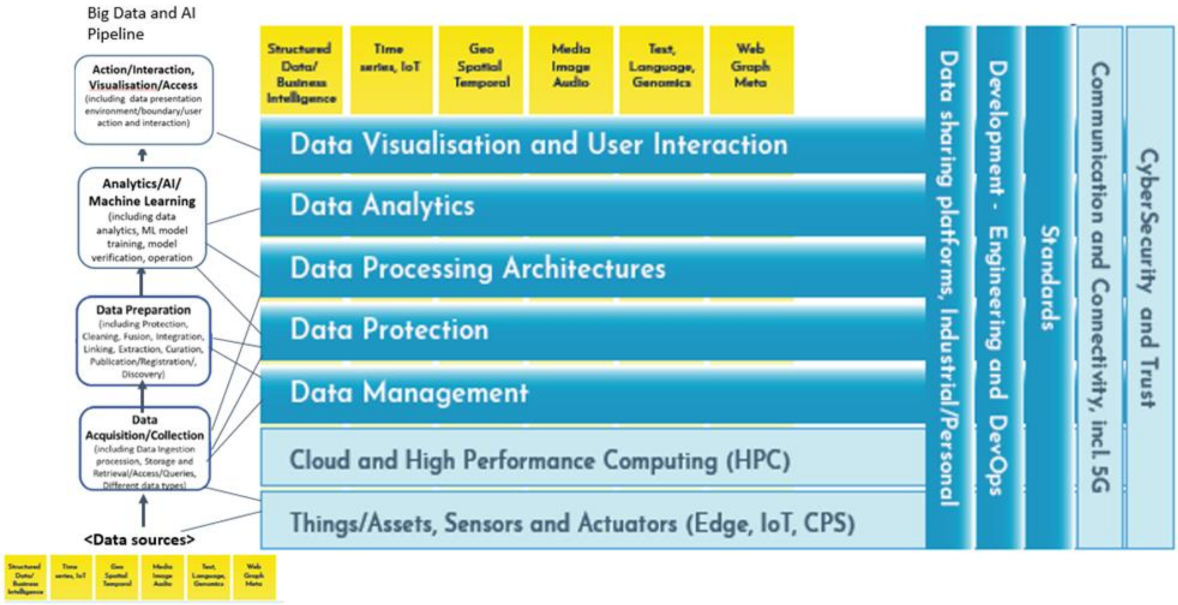


Figure 26 – Big Data and AI Pipeline using technologies from the BDV reference model

The BDV Reference Model shown in Figure 23 has been developed by the BDVA [1], taking into account input from technical experts and stakeholders along the whole Big Data Value chain as well as interactions with other related Public-Private Partnerships (PPPs). An explicit aim of the BDV Reference Model in the SRIA 4.0 document is to also include logical relationships to other areas of a digital platform such as Cloud, High Performance Computing (HPC), IoT, Networks/5G, CyberSecurity etc.

The following describes the steps of the Big Data and AI Pipeline showed to the left of the BDV Reference Model in Figure 23, with lines related to the typical usage of some of the main technical areas.

Data Acquisition/Collection: This step includes acquisition and collection from various sources, including both streaming data and data extraction from relevant external data sources and data spaces. It includes support for handling all relevant data types and also relevant data protection handling for this step. This step is often associated with the use of both real-time and batch data collection, and associated streaming and messaging systems. It uses enabling technologies in the area using data from things/assets, sensors and actuators to collect streaming data-in-motion as well as connecting to existing data sources with data-at-rest. Often, this step also includes the use of relevant communication and messaging technologies.

Data Storage/Preparation: This step includes the use of appropriate storage systems and data preparation and curation for further data use and processing. Data storage includes the use of data storage and retrieval in different databases systems – both SQL and NoSQL, like key-value, column-based storage, document storage and graph storage and also storage structures such as file systems. This is an area where there historically exists many benchmarks to test and compare various data storage alternatives. Tasks performed in this step also include further data preparation and curation as well as data annotation, publication and presentation of the data in order to be available for discovery, reuse and preservation. Further in this step, there is also interaction with various data platforms and data spaces for broader data management and governance. This step is also linked to handling associated aspects of data protection,

Analytics/AI/Machine Learning: This step handles data analytics with relevant methods, including descriptive, predictive, and prescriptive analytics and use of AI/Machine Learning methods and algorithms to support decision making and transfer of knowledge. For Machine learning, this step also includes the subtasks for necessary model training and model verification/validation and testing, before actual operation with input data. In this context, the previous step of data storage and preparation will provide data input both for training and validation and test data, as well as operational input data.

Action/Interaction, Visualisation and Access: This step (including data presentation environment/boundary/user action and interaction) identifies the boundary towards the environment for action/interaction, typically through a visual interface with various data visualisation techniques for human users and through an API or an interaction interface for system boundaries. This is a boundary where interactions occur between machines and objects, between machines, between people and machines and between environments and machines. The action/interaction with the system boundaries can typically also impact the environment to be connected back to the data acquisition/collection step, collecting input from the system boundaries.

The above steps can be specialised based on the different data types used in the various applications, and are set up differently based on different processing architectures, such as batch, real-time/streaming or interactive. Also, with Machine learning there will be a cycle starting from training data and later using operational data.

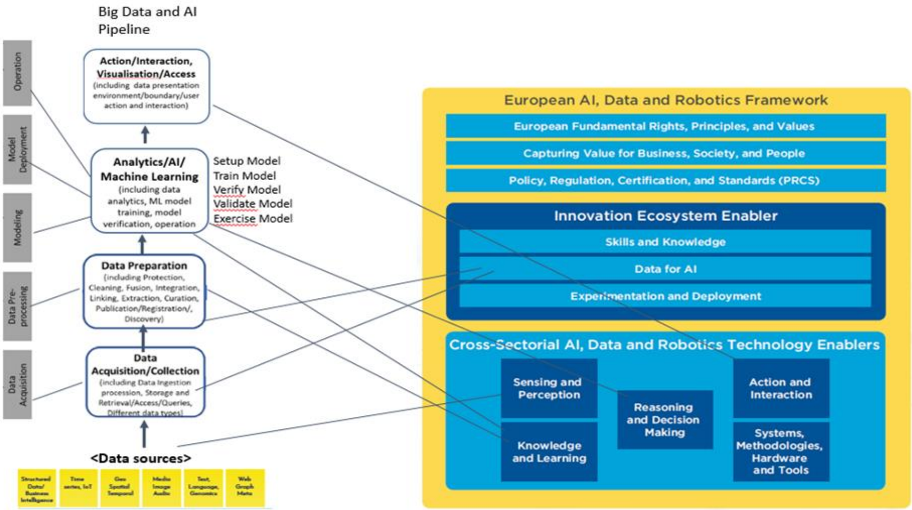


Figure 27 – Mapping Generic Big Data and AI Pipeline - the European AI and Robotics Framework

The steps of the Big Data and AI Pipeline can relate to the AI enablers as follows:

Data Acquisition/Collection: using enablers from *Sensing and Perception* technologies, which includes methods to access, assess, convert and aggregate signals that represent real-world parameters into processable and communicable data assets that embody perception.

Data Storage/Preparation: using enablers from *Knowledge and learning technologies*, including data processing technologies, which cover the transformation, cleaning, storage, sharing, modelling, simulation, synthesising and extracting of insights of all types of data both that gathered through sensing and perception as well as data acquired by other means. This will handle both training data and operational data. It will further use enablers for *Data for AI* which handles the availability of the data through data storage through data spaces, platforms and data marketplaces in order to support data driven AI.

Analytics/AI/Machine Learning: using enablers from *Reasoning and Decision* making which is at the heart of Artificial Intelligence. This technology area also provides enablers to address optimisation, search, planning, diagnosis and relies on methods to ensure robustness and trustworthiness.

Action/Interaction, Visualisation and Access: using enablers from *Action and Interaction* – where Interactions occur between machines and objects, between machines, between people and machines and between environments and machines. This interaction can take place both through human user interfaces as well as through various APIs and system access and interaction mechanisms. The action/interaction with the system boundaries can typically also be connected back to the data acquisition/collection step, collecting input from the system boundaries.

These steps are also harmonised with the emerging pipeline steps in ISO SC42 AI standard of "Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML), ISO/IEC 23053, with Machine learning pipeline with the related steps of *Data Acquisition, Data Pre-processing, Modeling, Model Deployment and Operation*.

Benchmarks can be identified related both to technical areas within the BDV Reference Model and the AI Frameworks and to the various steps in the DataBench Toolbox that supports both perspectives.

Big Data and AI Pipeline Examples for IoT, Graph and SpatioTemporal data (DataBio project)

In the following, we present example pipelines which handle different data types. Specifically, they handle IoT data, Graph data and Earth Observation/Geospatial data. Each pipeline is mapped to the four phases of the top level Generic Big Data and AI Pipeline pattern, presented in Section 2. All these pipelines have been developed in the DataBio project [5] which was funded by the European Union's Horizon 2020 research and innovation programme. DataBio focused on utilizing Big Data to contribute to the production of the best possible raw materials from agriculture, forestry, and fishery/aquaculture for the bioeconomy industry in order to produce food, energy and biomaterials, also taking into account responsibility and sustainability issues. The pipelines that are presented below are the result of aggregating Big Data from the three aforementioned sectors (agriculture, forestry, and fishery) and intelligently process, analyse and visualize them.

Pipeline for IoT data real-time processing and decision making

The "Pipeline for IoT data real-time processing and decision making" has been applied to three pilots in the DataBio project from the agriculture and fishery domain, and, since it is quite generic, it can also be applied to other domains. The main characteristic of this pipeline is the collection of real-time data coming from IoT devices to generate insights for operational decision making by applying real-time data analytics on the collected data. Streaming data (a.k.a. events) from IoT sensors (e.g. are collected in real-time, for example: agricultural sensors, machinery sensors, fishing vessels monitoring equipment. These streaming data can then be pre-processed in order to lower the amount of data to be further analysed. Pre-processing can include filtering of the data (filtering out irrelevant data and filtering in only relevant events), performing simple aggregation of the data, and storing the data (e.g. on cloud or other storage model, or even simply as a computer's file system) such that conditional notification on data updates to subscribers can be done. After being pre-processed, data enters the complex event processing (CEP) [26] component for further analysis, which generally means finding patterns in time windows (temporal reasoning) over the incoming data to form new more complex events (a.k.a. situations or alerts/warnings). These complex events are emitted to assist in decision-making processes either carried out by humans ("human in the loop" [27]) or automatically by actuators, e.g., sensors that start irrigation in a greenhouse as a result of a certain alert. The situations can also be displayed using visualization tools to assist humans in the decision-making process (as for example in [27]). The idea is that the detected situations can provide useful real-time insights for operational management (e.g. preventing a possible crop pest or machinery failure).

Figure 25 shows the steps of the pipeline for real-time IoT data processing and decision making that we have just described and their mapping to the steps of top level Generic Big Data and AI Pipeline pattern that we have analysed in Section 2.

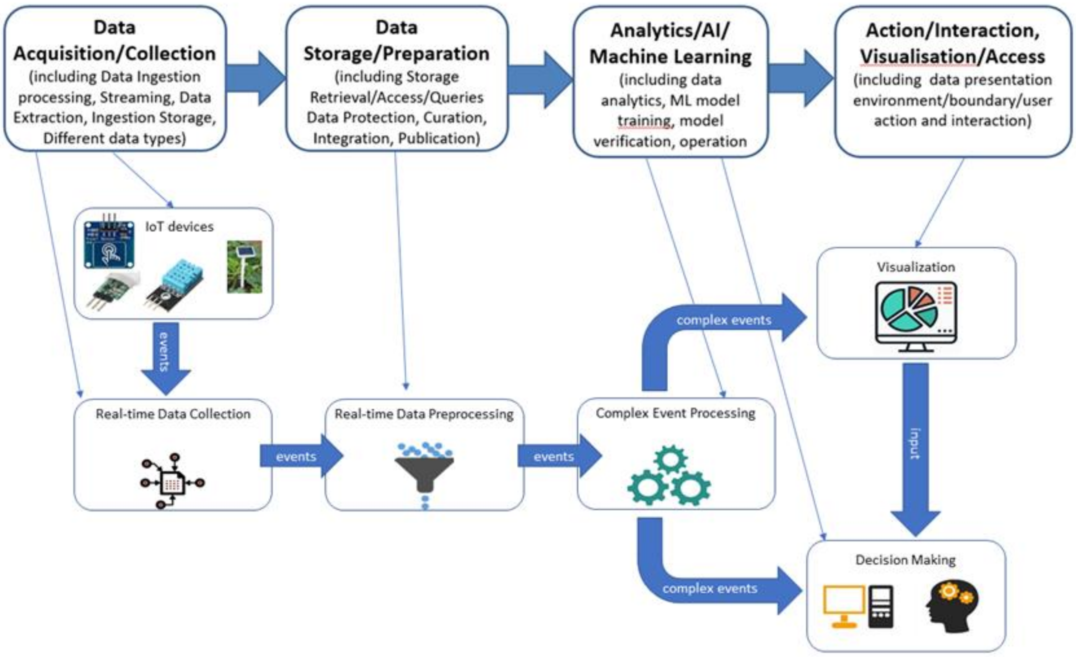


Figure 28 – Mapping of steps of the "Pipeline for IoT data real-time Processing and decision making" to the "Generic Big Data and AI Pipeline" steps

Pipeline for Linked Data Integration and Publication

In DataBio project and some other agri-food projects, Linked Data has been extensively used as a federated layer to support large scale harmonization and integration of a large variety of data collected from various heterogeneous sources and to provide an integrated view on them. The triplestore populated with Linked Data during the course of DataBio project (and few other related projects) resulted in creating a repository of over 1 billion triples, being one of the largest semantic repositories related to agriculture, as recognized by the EC innovation radar naming it the "Arable Farming Data Integrator for Smart Farming". Additionally, projects like DataBio have also helped in deploying different endpoints providing access to the dynamic data sources in their native format as Linked Data by providing a virtual semantic layer on top of them. This action has been realised in DataBio project through the implementation of the instantiations of a 'Pipeline for the Publication and Integration of Linked Data", which has been applied in different uses cases related to the bioeconomy sectors. The main goal of these pipelines instances is to define and deploy *(semi-) automatic processes* to carry out the necessary steps to transform and publish different input datasets for various heterogeneous sources as Linked Data. Hence, they connect different data processing components to carry out the transformation of data into RDF format [7] or the translation of queries to/from SPARQL [8] and the native data access interface, plus their linking, and including also the mapping specifications to process the input datasets. Each pipeline instance used in DataBio is configured to support specific input dataset types (same format, model and delivery form)

A high-level view of the end-to-end flow of the generic pipeline and its mapping to the steps of the Generic Big Data and AI Pipeline is depicted in Figure 26. In general, following the best practices and guidelines of Linked Data Publication [9], [10], the pipeline takes as input selected datasets that are collected from heterogeneous sources (shapefiles, GeoJSON, CSV, relational databases, RESTful APIs), curates and/or pre-process the datasets when needed, selects and/or creates/extends the vocabularies (e.g., ontologies) for the representation of data in semantic format, processes and transforms the datasets into RDF triples according to underlying ontologies, performs any necessary post-processing operations on the RDF data, vi) identify links with other datasets, and publishes the generated datasets as Linked Data and applying required access control mechanisms.

The transformation process depends on different aspects of the data like the format of the available input data, the purpose (target use case) of the transformation and the volatility of the data (how dynamic is the data). Accordingly, the tools and the methods used to carry out the transformation were determined firstly by the format of the input data. Tools like D2RQ [11] were normally used in case of data coming from relational databases, tools like GeoTriples [12] was chosen mainly for geospatial data in the form of shapefiles, tools like RML Processor [13] for CSV, JSON, XML data formats, services like Ephedra [14] (within Metaphactory platform) for Restful APIs.

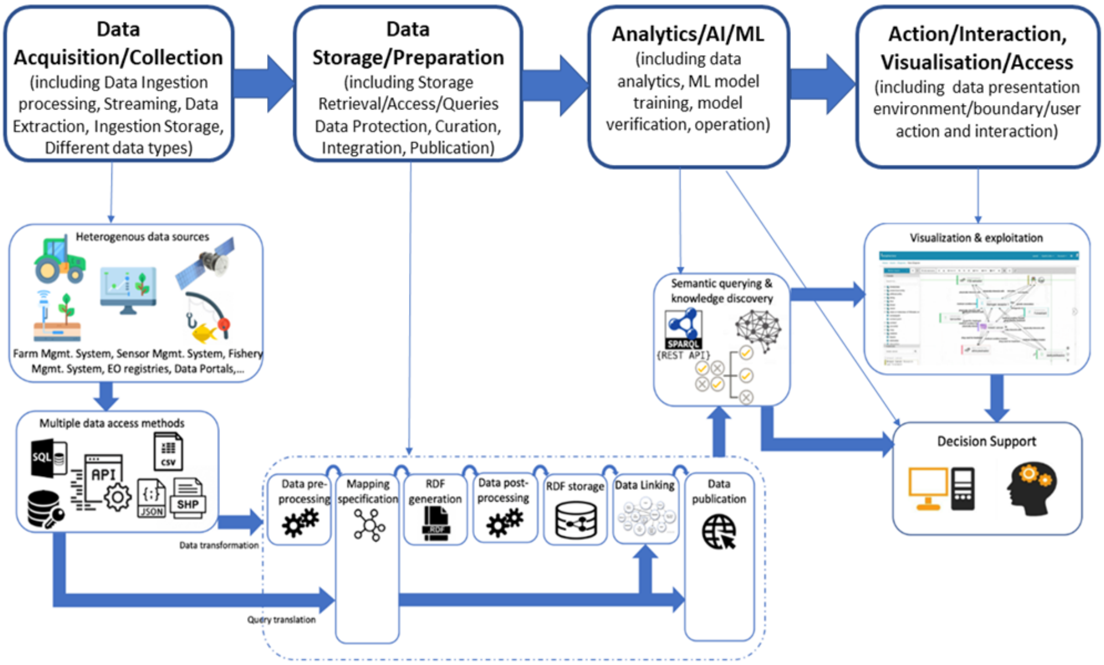


Figure 29 – Mapping of steps of the "Pipeline for Linked Data Integration and Publication" to the "Generic Big Data and AI Pipeline" step

Pipeline for Earth Observation and Geospatial Data Processing

The pipeline for Earth Observation and Geospatial data processing [28], developed in the DataBio project, depicts the common data flow among six project pilots, four of which are from the agricultural domain and two from the fishery domain. To be more specific,, from the agricultural domain there are two smart farming pilots, one agricultural insurance pilot and one pilot that provides support to the farmers related to their obligations introduced by the current Common Agriculture Policy [29]. The two pilots from the fishery domain were in the areas of oceanic tuna fisheries immediate operational choice and oceanic tuna fisheries planning.

Some of the characteristics of this pipeline include the following:

- Its initial data input is georeferenced data [30], which might come from a variety of sources such as satellites, drones or even from manual measurements. In general, this will be represented as either in the form of vector or raster data [31]. Vector data usually describes some spatial features in the form of points, lines or polygones. Raster data, on the other hand, is usually generated from imaging-producing sources such as Landsat or Copernicus satellites.
- Information exchanged among the different participants in the pipeline can be either in raster or vector form. Actually, it is possible and even common that the form of the data will change from one step to another. For example, this can result from feature extraction based on image data or pre-rendering of spatial features.
- For visualisation or other types of user interaction options, information can be provided in other forms like: images, maps, spatial features, time series or events.

Therefore, this pipeline can be considered as a specialization of the top level Generic Big Data and AI Pipeline pattern, presented in Section 2, as it concerns the data processing for Earth Observation and Geospatial data. The mapping between the steps of these two pipelines can be seen in Figure 27.

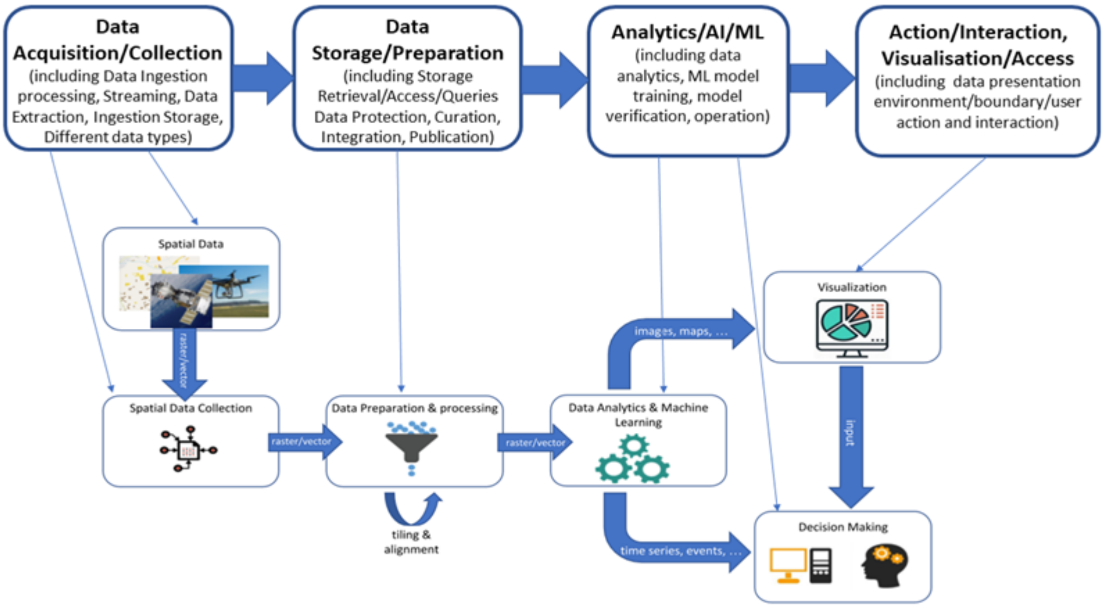
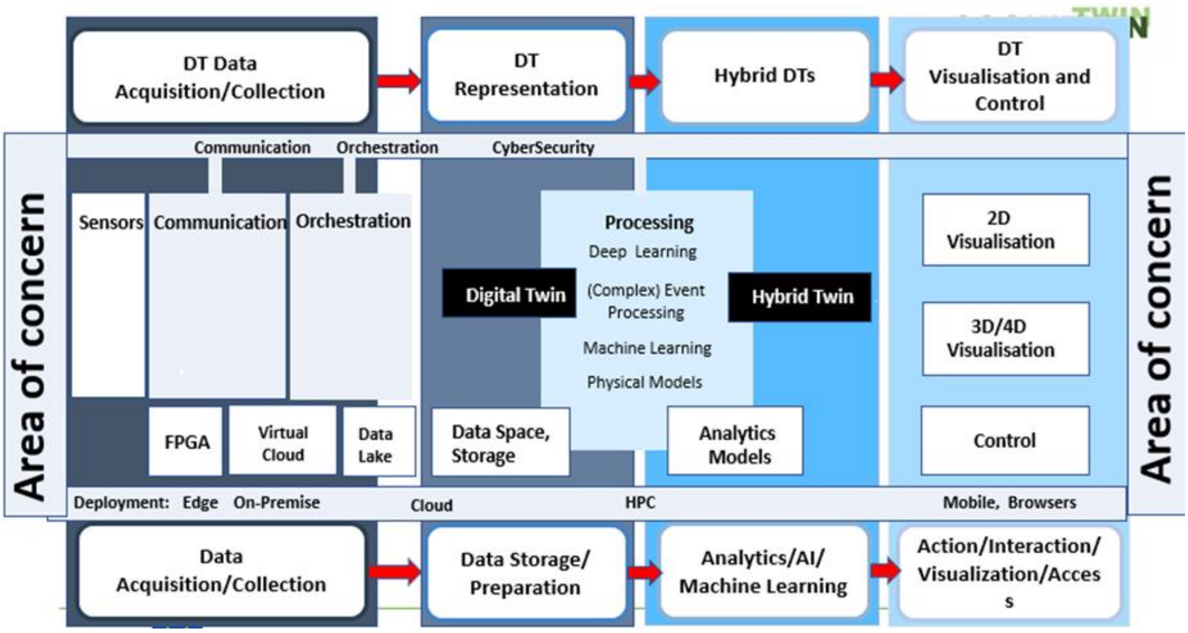


Figure 30 – Mapping of steps of the "Pipeline for Earth Observation and Geospatial Data Processing" to the "Generic Big Data and AI Pipeline" steps



COGNITWIN Digital Twin Reference Model (cognitwin.eu)

Figure 31 –Digital Twin Reference Model of COGNITWIN project

Digital Twin Pipeline

Many of the advancements and requirements related to Industry 4.0 are being fulfilled by use of Digital Twins (DT). We have in earlier papers introduced our definitions for Digital, Hybrid and Cognitive Twins (Abburu et.al, 2020 (2), Albayrak and Unal, 2020, Jacoby and Uslander, 2020) – which also aligns with definitions of others (Eriniakis et. al. 2020, Fuller et.al. 2020): "A DT is a digital replica of a physical system that captures the attributes and behaviour of that system", (Abburu et. al, 2020). The purpose of a DT is to enable measurements, simulations, and experimentations with the digital replica in order to gain understanding about its physical counterpart. A DT is typically materialized as a set of multiple isolated models that are either empirical or first-principles based. Recent developments in artificial intelligence (AI) and DT bring more abilities to the DT applications for smart manufacturing.

A hybrid twin (HT) is a set of interconnected DT, and being an extension of a HT, a cognitive twin (CT) is a self-learning and proactive system (Abburu et. al., 2020). The concepts of HT and CT are introduced as elements of the next level of process control and automation in the process and manufacturing industry. We define a HT as a DT that integrates data from various sources (e.g., sensors, databases, simulations, etc.) with the DT models, and applies AI analytics techniques to achieve higher predictive capabilities, while optimizing, monitoring and controlling the behaviour of the physical system. A Cognitive Twin (CT) is defined as an extension of HT incorporating cognitive features to enable sensing complex and unpredicted behaviour and to reason about dynamic strategies for process optimization. A CT will combine expert knowledge with the power of HT.

The Big Data and AI Framework was introduced by Berre et.al. 2021 (Figure 1). Figure 2 displays the top level big data and AI pipeline pattern. We introduce Digital Twin pipeline architecture which can be used for generation of HT and CT used in various industries. Inspired by the Big Data and AI Pipeline (Figure 1), Digital Twin – COGNITWIN Pipeline Architecture, Cognitive Twin Toolbox conceptual architecture with Digital Twin Pipeline architecture is proposed (Figure 3).

The proposed pipeline architecture starts with data acquisition and collection to be used by the DT. This step includes acquisition and collection of data from various sources, including streaming data from the sensors, and data at rest.

Following the data acquisition and collection, the next step is the DT data representation. This step is used for storing and preprocessing the acquired data used by the DT. DT (Hybrid) Cognitive Analytics Models step of the pipeline enables integration of multiple models and addition of cognitive elements to the DT by means of analytics conducted. Finally, the DT Visualisation and Control step of the pipeline provides a visual interface for the DT, and it provides interaction between the twin and the system.

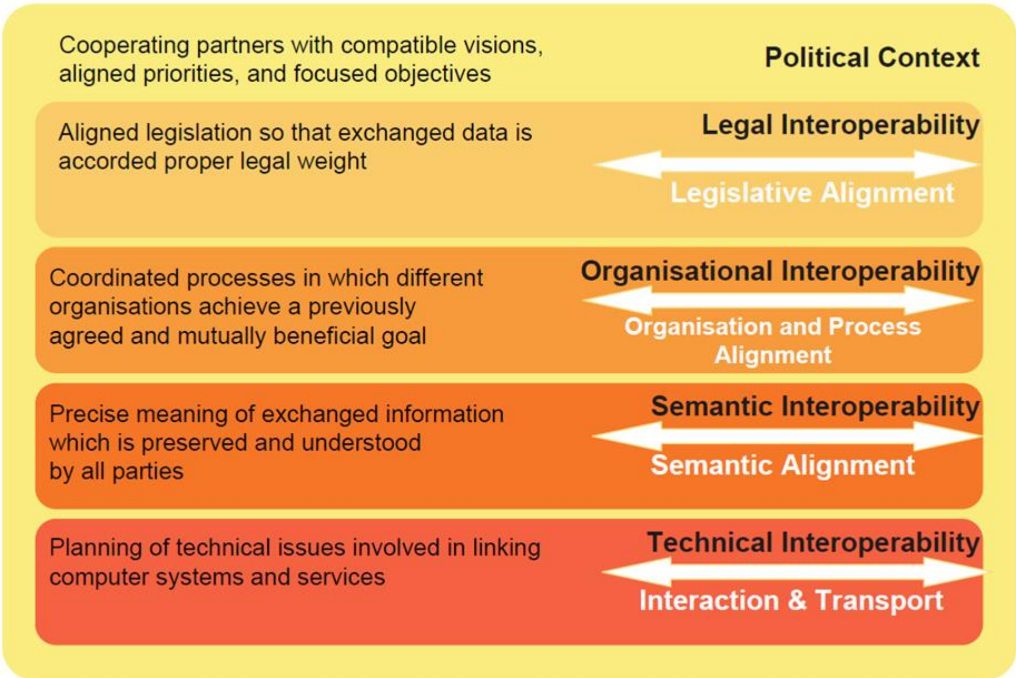


Figure 32 – EIF - European Interoperability Framework

The Interoperability dimensions introduced in the EIF - European Interoperability Framework - is domain independent and can thus be applied to all domains.

4.3 Interoperability Frameworks for Agrifood

Within the Agrifood domain effort is put from different initiatives, communities and projects on interoperability in the last decades. For example, main findings of the IoF2020 project developments with regard to semantic interoperability should take place on a global level, it is important to consider the Model Driven Approach (MDA) and Platform Independent Modelling (PIM) as guiding principles for core processes and procedures, and, to assess the FAIR ecosystem within the Agrifood sector, the FAIR implementation profiles could provide supportive ground. Finally, the Open Platforms idea could ensure an architecture that meets compliance and realises the System of Systems approach.

More recently, the IoF2020, ATLAS and Demeter projects have put effort to compare and analyse different interoperability approaches. Main reusable components of the IoF2020 projects are based on the SmartDataModels, which exploits the NGSI-LD tools for information management purposes. The initiative is led by the FIware foundation and defines typical data 'entities' in an objective manner. Main findings of the ongoing ATLAS project focus rather more on interoperability on services than semantics and syntax. For instance, the Service Template specifications enable users to define various characteristics of a service, such as identifiers, versioning, prerequisite knowledge, etc. Further information, such as usage scenarios, could be provided to clarify motivation behind a service design.

Interestingly, the ongoing Demeter project has put more emphasis on semantics and interoperability compared with the other projects. The main relevant output of the project is the Agricultural Information Model (AIM). The NGSI-LD metamodel is reused with central concepts of Entity, Property, Relationship and value to develop the AIM ontology. Moreover, the AIM is aligned with other agricultural models, such as SAREF4AGRI, ADAPT, FOODIE, AgroVoc, INSPIRE and Earth Observation standards. Finally, the model employs SHACL based validation expressions that increases data quality. A non-exhaustive list of important ontologies and data models are presented in Table 3.

Table 3 – List of ontologies and data models

Model	Maintainer/Source
ADAPT	AgGateway https://adaptframework.org/
AIM	DEMETER http://defs-dev.opengis.net/profiles/object?uri=https://w3id.org/demeter/agri
AgroVoc	FAO http://aims.fao.org/vest-registry/vocabularies/agrovoc
AgroXML	KTBL http://www.agroxml.de
DAPLOS	UN/EDIFACT http://www.unece.org/trade/untdid/d08a/trmd/daplos_c.htm
E-Crop	UNCEFACT https://www.unece.org/fileadmin/DAM/cefact/brs/BRS_eCROP_v1.pdf
GSMA / FIWARE Agri Models	GSMA / FIWARE Foundation https://github.com/GSMADeveloper/NGSI-LD-Entities
ISO 11783-11 / ISOBUS	ISO / VDMA https://www.isobus.net/isobus/

rmAgro	Wageningen University & Research *https://rmagro.org/*
SAREF-4-AGRI	ETSI / TNO *https://mariapoveda.github.io/saref-ext/OnToology/SAREF4AGRI/ontology/saref4agri.ttl/documentation/index-en.html*

4.4 Aligning the Agrifood Domain for DEI

An analysis of the relevant dimensions in interoperability architectures for Agrifood shows that there are many commonalities with other domains. We envisage a possibility to adapt to other reference architectures with the addition of the Agrifood life cycle for production and consumption (as an alternative to the SGAM energy life cycle from production to consumption) - and also with a selection of other relevant dimensions - for a 3D, 4D, 5D + reference architecture - as a set of connected cubes,

5 REFERENCE ARCHITECTURES AND INTEROPERABILITY FOR DIGITAL ENERGY PLATFORMS

5.1 Context for Energy

The ambitious decarbonisation objectives of the energy system strictly depend on the "Digital and Green Transition", involving the interconnection of multiple sectors. The infrastructure of the digital energy system is going to be complex: numerous devices and components (smart meters, electric vehicles (EV), heat pumps, renewable energy sources, home appliances, etc.) have to be smoothly interrelated and synchronised for balanced operations of electrical grids. In particular, the main transformation corresponds to the increasing share of renewables and, consequently, the decentralisation of the energy system; to deal with this challenge, the actors in the energy market need for easy access to data and innovative solutions as energy data services. Accordingly, a key action consists of developing an European energy data exchange approach to integrate the three strategic sectors: electricity, mobility (EV) and buildings.

Interoperability is a key building block in achieving the open and scalable data exchange: a single platform or a central data hub cannot achieve the objectives; on the contrary, the infrastructure must be built incrementally with applications and systems capable of interoperating and exchanging data across different deployments. New use cases would drive the design, development and implementations of data exchange solutions, addressing the interoperability barriers and diversity of standards. The deployment of semantic interoperability and ontology based approaches have fostered the interconnection of vendor-agnostic devices and enhanced the platform's adoptions. Moreover, the energy architectures have to account for effective governance mechanisms, driven by the increasing numbers of market players and decentralised solutions. In this regard, governance system shall be extended to local data exchange, where data is kept at the source; in general, a performant platforms that accounts for the interconnection of multiple energy sectors shall allow federated data exchange mechanisms that dynamically vary from central to decentralised functions (from central markets to decentralised physical assets), enabling the achievement of flexibility services.

In the following paragraphs, the most relevant approaches for data exchange in the energy system are presented, highlighting their innovative features and technical specifications.

5.2 Reference Architectures for Energy

In the following paragraphs, the most relevant approaches for data exchange in the energy system are presented, highlighting their innovative features and technical specifications.

In the recent decades the need for data exchange in the energy system has grown. At first there was the demand for more efficiency that required an increase of operational situational awareness and analysis of historical developments. For example, studying information on the actual usage of energy infrastructures enables better planning and design. More insight into cause-and-effect relationships of material and construction degradation allows for better design, production, and installation of (physical) parts of the energy system. Secondly, the need to communicate about demand and supply has increased, especially on the electrical grid where there is a strong need to balance all demand and supply on a continuous basis. In the past more demand or less 'only' required more output or less output of power plants as they were largely adjustable. Relatively unpredictable renewable power sources (e.g., wind, solar) are less adjustable and thus attaining balance requires more agreement on flexible usage and thus more exchange of information in the energy system between parties as the production and consumption of electricity can be highly dynamic in time.

As energy systems are geographically distributed and involve many different stakeholders (e.g., equipment manufacturers, installation companies, network operators, energy markets, energy carrier producers and consumers, etc.) the useful exchange of data for achieving higher efficiency in energy systems requires:

- The **availability** of **data**, which in turn requires sensors (and/or human data entry operators) and storage and retrieval systems of the data.
- The **technical ability** to **exchange**, which requires telecommunication between parties
- Efficiency in processing data into information and/or knowledge.

These in turn require the parties involved to agree on why, with whom, where, when and how they exchange data. Several organisations have been working on this. For example, there is the Common Information Model (CIM) in the 61970 series of standards from the International Electrotechnical Commission (IEC) which focuses on the information model for components of an electrical system. Then there is the Smart Grid Architectural Model (SGAM) from a collaboration of Comité Européen de Normalisation (CEN), the Comité Européen de Normalisation Electrotechnique (CENELEC) and the European Telecommunications Standards Institute (ETSI). This is a multidimensional model that allows parties to agree on the definition of (semantical) concepts in an electrical grid and how they are interconnected.

Next to more (traditional) grid-oriented reference architectures and information model standards, there is the emergence of standardisation of data exchange in energy management in smart homes and buildings. The information contained in that data can be of great value for aggregating parties that want to trade in flexibility of energy production and consumption, or for Distribution and/or Transmission System Operators (D/TSO) who want to have more information from the endpoints of their grids. An example is the EN 50491 series from CENELEC that deals with the requirements for Home and Building Electronic Systems (HBES) and Building Automation and Control Systems (BACS). Another example is EEBUS, an internet protocol-oriented set of agreements between parties for data exchange. The different reference architectures, information models and protocols have overlap in concepts, but are not always completely aligned and/or harmonised and this currently remains a challenge for projects like H2020 InterConnect that currently also work on interoperability between (Internet-of-Things) devices and services in the domain of energy management. Note that the SGAM-based European energy Data Exchange Reference Architecture (DERA) from the Bridge does provide an overview on how different concepts are related at different layers of abstraction, which will help in further harmonisation and alignment, also at the more technologically detailed level.

One of the greatest hurdles in the convergence of reference architectures is dealing with the high demands regarding safety, security, and reliability of the energy system. Distribution of energy carriers like electricity comes with high risk. Not only because of the direct physical dangers involved, but also because of the consequences of non-delivery: without a flow of energy many societal sectors breakdown. Organisations responsible for production and delivery therefore have very strict requirements for allowing parties to be part of control of the energy system. This means parties are reluctant to add new parties in data exchange and/or change/give up carefully crafted standards in data exchange.

5.3 Interoperability Frameworks for Energy

The energy sector is linked with multiple domains like IoT (smart buildings, homes and appliances), smart cities, electric mobility or industry 4.0. This effectively makes all interoperability challenges cross-domain. The multi-domain complexity is what interoperability (semantic, syntactic and higher levels) solutions for the energy systems and markets must address. In the multi-domain ecosystem of the energy sector there are many proven digital solutions and platforms capable of providing high performance and reliable management of operations and resources. If these existing digital systems are to be made interoperable in a cost-effective manner, the challenge needs to be approached from a framework perspective.

Interoperability is usually facilitated by intermediary platforms responsible for translating different communication protocols and remapping data models. Challenges to this are: 1. Technical and business dependability; 2. 3rd party on path of potentially privacy and business sensitive data; 3. Limited control over performance bottlenecks of a facilitating platform; 4. Scalability and innovation depends on the facilitating platform operator; 5. The facilitating platform can become the weakest link in the cybersecurity chain; 6. Pilling new computational and storage resources increases the energy footprint of the achieved interoperability ecosystem. Based on these limitations, a more **decentralised approach for enabling interoperability for the energy sector is needed.**

Since digitalization of the energy sector and its closely interrelated domains (IoT, industry, mobility etc.) is largely underway, the interoperability solutions must consider legacy software systems. Moreover, the operation in such a regulated environment where EU directives and national regulations establish the rules for interaction with market actors and consumers, needs to be flexible and to incorporate such specificities. The interoperability approach to these software systems comes in the form of software adapters capable of directly interfacing with existing technology stacks, on one side, while exposing unified/interoperable interfaces and data models on the other side. Adapters, with centralised or distributed orchestrators, semantically powered discovery/reasoning mechanisms and integration methodology build today's interoperability frameworks. They allow existing systems to adapt data interfaces to be interoperable while managing the business logic with full control over the data flows and access control policies. Frameworks also provide guidance and boundaries as well as proven best practices and plug-n-play software modules for building new digital systems which will be interoperable with the sector standards (e.g., IEC 61850, IEC CIM, SAREF), ensuring a high level of replicability of the use cases and achieved results between different countries and in different energy system setups.

The Energy sector and European Commission are moving towards the software framework approach over dedicated digital platforms as the main enablers for syntactic and semantic interoperability. This is evident in the data spaces initiatives like Gaia-X where the guiding principles include: Data sovereignty, Federated deployments, or Data Spaces. This is also in-line with the European Commission's Action Plan on the Digitalization of the Energy System listing the following pillars for stable and sustainable energy sector: a Common European Energy Dataspace and a data-drive and citizen-centric energy market while strengthening security and data protection and playing ICT as the frontrunner for energy efficiency.

The European Commission initiated work on interoperability enablers and different approaches (centralised platforms, distributed enablers, and frameworks) with its European IoT Platform Initiative which resulted in 7 collaborative projects and 7 different solutions for tackling interoperability between IoT systems. These projects together with FIWARE and Web Of Things using NGSI-LD data modelling, provide a solid basis for building interoperability frameworks for various cross domain challenges including the ones from the energy sector.

The H2020 BRIDGE initiative together with EDSO brings together H2020 projects as Interconnect (Semantic interoperability Framework based on SAREF), PLATOON (IDS syntactic based interoperability layer), INTERFACE (semantic data models and secure plug-n-play interfaces), OneNet (energy platform federation), SYNERGY (framework for interconnecting data sources and consumers) or BG4NRG (big data governance).

5.4 Aligning the Energy Domain for DEI

Energy is a cross-sector domain from which all sectors from DEI will require an interface. This departs from the energy supply *per se*, but most importantly because of the increasing engagement stakeholders have with the grid to support the grid planning and operations directly or indirectly. Particularly important for the alignment of the energy domain is the SGAM (Smart Grid Architecture Model), shown in Fig. 1, an architecture concept that joins stakeholders to allow them to establish a positioning. The SGAM focuses on supporting a neutral positioning towards the creation of smart grid use-cases, allowing a representation of interoperability viewpoints in a technologically neutral approach.

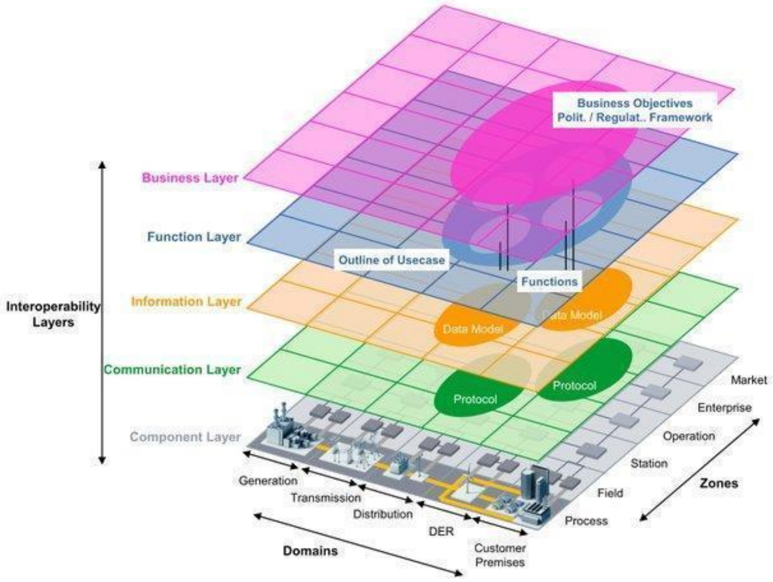


Figure 33 – The Smart Grid Architecture Model (SGAM) Framework

The key alignment of the Energy domain towards DEI stands on **the provision of interoperability** (at different levels) between systems and platforms, enabling the **exchange and creation of data driven use-cases** to monitor, forecast and provide control over the domain; that is, Energy informatics to **enable cross-sector solutions**. Navigating the SGAM's interoperability levels and domains, key standards arise such as the IEC 61850 providing guiding principles for communication among substations components in the Distribution domain; additionally, the IEC Common Information Model (CIM) is one of the key data models for data exchange at the Information layer, providing a common vocabulary and a basic ontology for the systems in electric power industry. In fact, this layer should remain the focus from R&D with syntactic canonical data modelling as NGSI-LD for advanced interoperability between different platforms and legacy systems becoming widely used. Moreover, of vital importance is the wide adoption of shared vocabularies and ontologies, such as SAREF and its multiple extensions to remove domain uncertainty from data modelling or to guide data exchange between systems. The InterConnect project is working on bringing together different semantic understandings of the energy sector, for example by determining how to relate and/or map Home/Building Energy Management System (H/BEMS) interface concepts in SAREF4ENER from ETSI to those in EN50491 from CENELEC.

Semantics and particularly semantic interoperability should become a main driver for cross-industry/sector digitising, as it allows to describe distinct domains and explore how they are related. These strategies free adopters from becoming restricted to data representations and to freely approach data by *asking* instead of *requesting,* which is crucial for interoperability across domains. Initiatives such as GAIA-X, build atop this vision, by establishing several data domains (data spaces) where semantics and interoperability become staple features. GAIA-X aims to create a federated digital infrastructure for data exchange, specifying the standards, services and technical requirements for a decentralised architecture aligned with the International Data Spaces (IDS) building blocks.

The "Data Management" working group of H2020 BRIDGE initiative has proposed the "European Energy Data Exchange Reference Architecture (DERA)". This high-level architecture starts from the SGAM approach, particularly from its interoperability layers, and aims at the applicability across different domains. The architecture proposes the set of components, processes, protocols and data models for the deployment of data spaces and interoperability to extend the energy applications. Among other projects, the H2020 Interconnect project is deploying a semantic interoperability framework that enables the exchange of semantically encoded information, in an ontology agnostic way. making it applicable to other domains as well, not just IoT and Energy, which is particularly interesting for cross-domain approaches.

Organisations like the Association of Issuing Bodies and EnergyTag work on digitising the guarantees of the origin of (European) energy. This will become more important as the demand to provide proof for the reduction of CO2 emissions will increase. There is an industry wide challenge to determine where ICT for operations has to interface with the ICT for providing such proof. From a technological, organisational and also a legal point of view. An important aspect is to avoid using more energy for bookkeeping than what was actually used.

6 REFERENCE ARCHITECTURES AND INTEROPERABILITY FOR DIGITAL HEALTH & CARE PLATFORMS

6.1 Context for Health and Care

Clinical data denote patients, their complaints, signs, diseases, operations, drugs, lab values, etc. Recorded in information systems of different genres (electronic health records, disease registries, clinical trial documentations, mortality databases) they are heterogeneous, context-dependent, often incomplete and sometimes incorrect. Clinical data are shaped according to the specific needs for which they are collected, such as reporting, communicating, and billing. Wherever statistical analyses or case-based reimbursement are needed, data has to be in a structured form, with a trade-off regarding scope and granularity. Where communication between health professionals is paramount, poorly structured narratives tend to prevail over structured and coded data, because text is richer in detail and faster to create. The transformation of textual sources into structured output is a main driver for human language technologies. The application of such techniques, alone, does not, however, guarantee interoperability and standardisation[59].

Interoperability in the healthcare domain has been on the political agenda of both national and European levels for more than 15 years. Implementations at national levels usually relied on existing interoperability initiatives, such as HL7 (**Health Level Seven)** but to due to the insufficient maturity of the standards, early adopters have usually adapted or completed those standards to support their priority use cases and developed an incentive/compliance strategy to have those standards adopted by the major industrial players operating in a given country or region. Interoperability in the Healthcare domain is thus mainly **publicly** driven and had as first objective to support continuity of care.

The question of the alignment with other high-level use cases such as epidemiology, research, knowledge management, big data analytics and AI has only come on the political agenda in recent years. Each of those domains has developed the resources it thought would best serve its goals. In the private domain, IHE (Integrating the Healthcare Initiative) has played a very important role in driving interoperability in Europe and beyond: with a pragmatic approach, it has developed a wide community of experts who have created "profiles" for the most important use cases.

Those "profiles" rely as much as possible on existing resources but are completed by implementation guidelines and an appropriate testing and validation environment. They are also meant to support cross-cutting functionalities which are used for many different use cases. At European level, the creation of the eHealth Digital Service Infrastructure (eHDSI)[60] has provided a first backbone to support cross-border interoperability with a first focus on two use cases: the Patient Summary and ePrescription. The EU Commission has however recently taken new important initiative with the proposed Data Act which should provide a major contribution to have data extracted from the silos they are currently entangled in and the Health Data Space Act with a specific focus on personalised medicine. With as a consequence the identification of a set of requirements, including interoperability and security to be respected by Electronic Health Record systems. In May 2022, the EU Commission also made public its regulation proposal for a "European Health Data Space" which will significatively accelerate the sharing of data both for primary and secondary use, addressing all aspects of interoperability and referring in particular directly to interoperability profiles and compulsory certification. Since a few years, the Fast Health

[59] Schulz, S., Stegwee, R., Chronaki, C. (2019). Standards in Healthcare Data. In: Kubben, P., Dumontier, M., Dekker, A. (eds) Fundamentals of Clinical Data Science. Springer, Cham.

[60] https://health.ec.europa.eu/ehealth-digital-health-and-care/electronic-cross-border-health-services_en

Interoperability Initiative (FHIR)[61] is increasingly becoming the reference standard and is now there to stay. The need to put the citizen/patient at the centre of the system and consequently to include the services targeted at him rather than healthcare providers is a key driver in this respect. Most European countries have initiated a transition journey to FHIR.

When referring to Interoperability in the healthcare domain, people usually just consider the communication challenge. Therefore, they do not use reference architectures, but address the four interoperability layers: legal, organisational, semantic and technical. However, when moving to more complex, interdisciplinary systems, an architectural approach is inevitable[62,63].

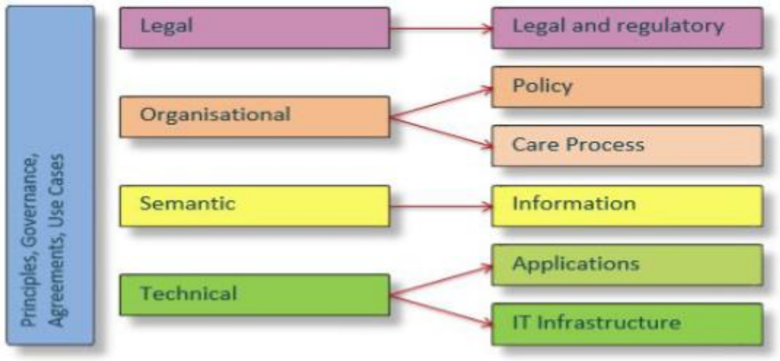


Figure 34 – eHealth Interoperability layers

Those layers are also mapped to all the different stakeholders who intervene in the value chain at strategic, tactical and operational levels.

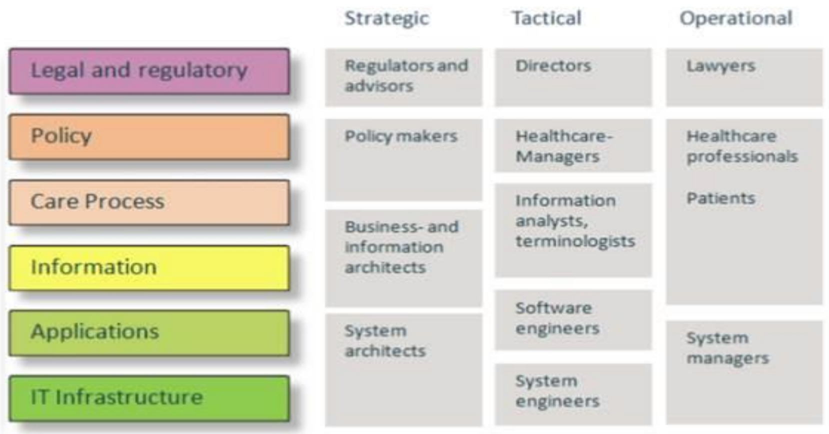


Figure 35 – eHealth Strategic, Tactical and Operational viewpoint

This viewpoint is elaborated in the European Interoperability Framework[64] and has been refined for the health domain[65,66].

[61] https://fhir.org/

[62] Blobel, B., Ruotsalainen, P., Oemig, F. (2020) Why Interoperability at Data Level Is Not Sufficient for Enabling pHealth? Stud Health Technol Inform. 2020; 273: 3-19

[63] Blobel, B., Oemig, F., Ruotsalainen, P. Lopez, D.M. (2022) Transformation of Health and Social Care Systems—An Interdisciplinary Approach Toward a Foundational Architecture. Front. Med. 2022;9:802487. doi: 10.3389/fmed.2022.802487

[64] The detailed framework is available here : https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/european-interoperability-framework-detail. A brochure is available here : https://ec.europa.eu/isa2/sites/isa/files/eif_brochure_final.pdf

[65] https://eufordigital.eu/wp-content/uploads/2021/03/Common-Guidelines-for-eHealth-Harmonisation-and-Interoperability.pdf

[66] Angelina Kouroubalia, Dimitrios G.Katehakis. The new European interoperability framework as a facilitator of digital transformation for citizen empowerment : https://www.sciencedirect.com/science/article/pii/S153204641930084X

Finally, the implementation of FAIR principles for clinical data imply that clinical data must follow shared standards. Such standards should describe[67]:

- Data provenance, i.e. their originators, creation times and related processes;
- Information templates in which data are embedded;
- Vocabularies / terminologies / ontologies used to attach meaning to data;
- The semantic descriptors or representational units (codes, labels) in these vocabularies;
- Formal or textual definitions of these representational units;
- The formal languages used for the above

6.2 Reference Architectures for Health and Care

Experts from the large-scale pilots' projects of the OPEN DEI healthcare cluster have come to a number of conclusions[68] when comparing their Reference Architectures.

Reference architecture should be technology agnostic. Architecture descriptions supporting platforms selection/development should at least identify:

- system stakeholders (including users, operators, owners, developers, maintainers);
- fundamental concerns (including the purpose of the system, suitability of the architecture to fulfil the set objective, feasibility, risks, maintainability, evolution);
- architecture views (representing a related set of concerns as seen from a perspective a view is taken, a viewpoint); and
- the rationale for each important architecture decision.

The create-IoT 3D RAM was considered as sufficiently broad to describe each project's architecture comprehensively. For this reason, the experts have proposed the CREATE-IoT 3D RAM model as a standard RAM for future digital healthcare and Ambient Assisted Living (AAL) projects.

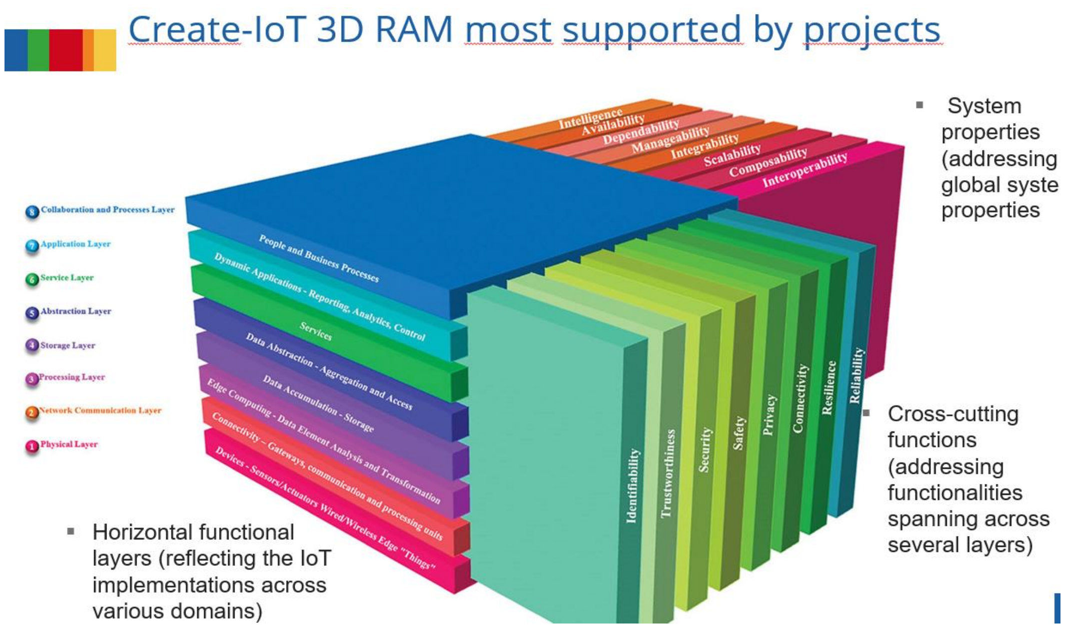


Figure 36 – Create-IoT 3D model

The other key findings were:

- Architectural choices, and resulting architectures, are most often made considering functional and non-functional requirements, while technical and business constraints are in most cases only implicit.

[67] Schulz, S., Stegwee, R., Chronaki, C. (2019). Standards in Healthcare Data. In: Kubben, P., Dumontier, M., Dekker, A. (eds) Fundamentals of Clinical Data Science. Springer, Cham.

[68] https://www.mdpi.com/2079-9292/10/14/1616

- Quality attributes such as performance, interoperability, reliability, maintainability, usability, and security are often vaguely described
- Trade-offs, e.g., between maximum cybersecurity and usability, are also necessary to balance the system
- Clinician-facing functions and systems should be included as an extension to current AAL taxonomies
- There is a specific need to apply privacy-enhancing techniques in smart and healthy living solutions.
- Performance reports, especially ones that observe more extended running platforms and services, are missing, since projects end before collecting them.
- There is a real need to compare attributes linked to Performance (in terms of latency and throughput), Usability (in terms of learnability and user interaction design), and security (in terms of confidentiality, integrity, availability)

6.3 Interoperability Frameworks for Health and Care

In Health and care, interoperability relies on different kinds of standards like terminologies, ontologies and information models.

Thesauri provide the terminology and some simple semantic relations between terminology items like synonymy, whereas ontologies aim at giving precise mathematical formulations of the properties and relations of entities i.e. they provide formal semantics together with syntactic rules for composition. However, the use of a code from a terminology standard is not sufficient, as long as pragmatic or contextual aspects are missing. Provision of such contextual and provenance information is the domain of (clinical) information models. Several standards for clinical models and their specifications have been proposed, in order to prevent data silos which, even if they are well structured, are buried in proprietary and non-interoperable formats. However, the adoption of such standards (e.g. detailed clinical models (DCMs, ISO/TS 13972:2015) by manufacturers and the embedding of standardised terminologies within them has been low until now. The overlap between standards of either kind poses major challenges to prevent so-called iso-semantic models, which tend to arise e.g. when using terminologies and information models

There are many known and widely used standards in the health and care sector from organisations such as FCAT, HL7, IHE, ISO, NEMA, OpenEHR, Regenstrief Institute, PCHAlliance, SNOMED, WHO, WONCA[69].

As mentioned in section 6.1, Interoperability frameworks in the health care domain are heavily influenced by the activities carried on by the Health Level Seven International (HL7) organisation. Founded in 1987, and having as a primary objective the standardisation in the health domain, HL7 has produced several standards that have become very popular during the years. HL7 first defined a standard focused on how data should be requested to a health care service, this standard is known as V2. V2 is a messaging standard, probably one of the most adopted in the IT world, it defines a messaging protocol for the exchange between health systems independently of the technology each system would use. As a fact, V2 specifies the structure and the semantics of (request and response) messages only and it doesn't determine how data will be transported or how it should be secured. HL7v3 [9] is based on a Reference Information Model[70] (the RIM) and is designed primarily for the messaging environment. The Clinical Document Architecture (CDA®) has been the second popular standard, adopted by the majority of the hospitals all over the world, based on XML, it defines the structure and the semantics of clinical documents for the purpose of exchange between healthcare providers and patients. It defines a clinical document as having the following six characteristics: 1) Persistence, 2) Stewardship, 3) Potential for authentication, 4) Context, 5) Wholeness and 6) Human readability; examples of documents are:

[69] Schulz, S., Stegwee, R., Chronaki, C. (2019). Standards in Healthcare Data. In: Kubben, P., Dumontier, M., Dekker, A. (eds) Fundamentals of Clinical Data Science. Springer, Cham.

[70] https://www.w3.org/wiki/HCLS/ClinicalObservationsInteroperability/HL7_RIM

Discharge Summary, Imaging Report, Pathology Report, Laboratory Report. As CDA is very focused on data structure (mainly it defines a data model).

Since 2000, another initiative has emerged: openEHR[71] is an open standard specification in health informatics that describes the management and storage, retrieval and exchange of health data in electronic health records (EHRs); openEHR is based on two levels modelling approach, a stable reference information model constitutes the first level of modelling, while formal definitions of clinical content in the form of archetypes and templates constitute the second level. Both HL7 CDA and openEHR are based on the RIM and can thus be aligned although OpenEHR has a broader scope.

More recently, HL7 produced a completely new standard called FHIR (Fast Healthcare Interoperability Resources). This standard defines both how data should be represented and how data should be requested to a health service. Differently from V2, FHIR is more technology related, because it is based on RESTful services accessible over the Internet and in this sense FHIR adopts a simple yet powerful transport technology that is a de-facto standard in contemporary distributed systems. Apart from defining what is the transport protocol, adopters of FHIR are to use whatever technology that suits their IT systems. Data in FHIR are represented with XML or with JSON, the conceptual definitions of the data model can be represented in both ways, clients requesting data must be aware of what representation the server is able to provide, and must use the proper parser to convert it into an object oriented representation. The reference implementation of FHIR, is named Hapi FHIR and provides both server and client side libraries and utilities to ease the adoption of this standard.

6.4 Aligning the Health and Care domain for DEI

One possible alignment process lies in the use of the "system development process" - RUP (Rational Unified Process)- standardised in ISO 10746 (Figure 34). The development process is characterised by specific views appropriately represented through specific ICT languages/ontologies from both programming languages for the **Engineering View** through UML for the Information and Computational Views up to BPML for the **Enterprise View**. The elements/components of the systems must consider design aspects regarding their composition/decomposition and their relationships within each viewpoint. Consequently, our system architecture model has two dimensions: **System Viewpoint and System Component Composition**. For developing ecosystem solutions meeting the requirements and objectives of the users, we have to describe the real-world business system, using the different users' domain-specific perspectives, methodologies, and representation styles (languages/ontologies).

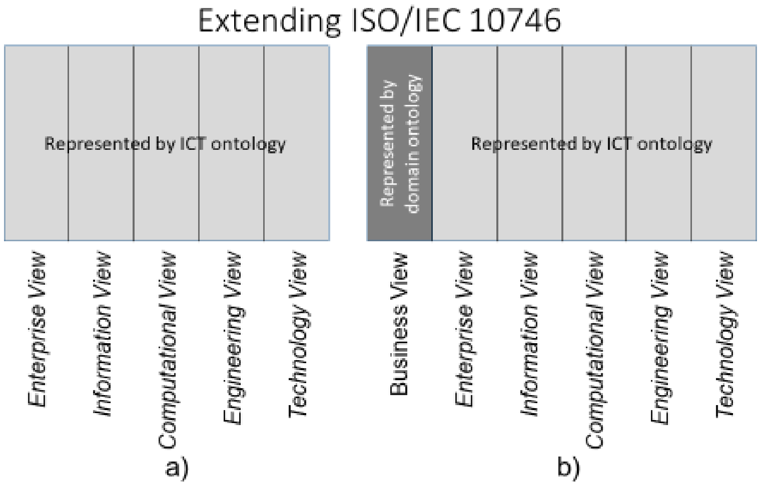


Figure 37 – ISO/IEC 10746 Extension

To represent domain-specific ecosystems, ISO 10746 has been extended by adding a Business View (Figure 34 b). The resulting system architecture model, first introduced in the late nineties as Generic Component Model (GCM) and meanwhile standardised in ISO 23903, has three dimensions (Figure 35). Beside the aforementioned dimensions System Viewpoint and System Component Composition, we have

[71] https://www.openehr.org/

to introduce **System Domain dimension**. For representing the Business View, we have to deploy the domain-specific languages/ontologies managed using top level ontologies according to ISO 21383.

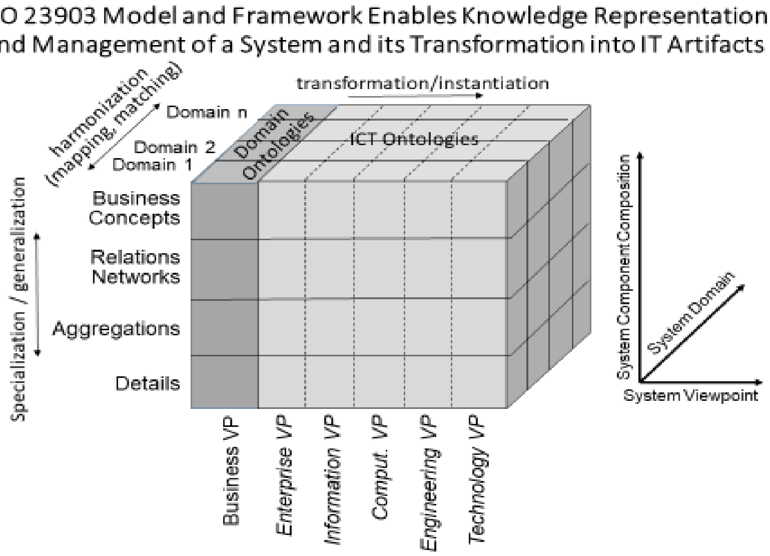


Figure 38 – ISO 23903 Model and Framework

Interoperability, i.e. communication and cooperation amongst the elements/components is enabled by transforming proprietary element representations into standardised one using the view specific representation standard, as shown in Figure 36 for the Information View and implementing the solution through Application Programming Interfaces (APIs). A concrete representation reference model/standard is, e.g., ISO/HL7 21731 HL7 version 3 – Reference information model.

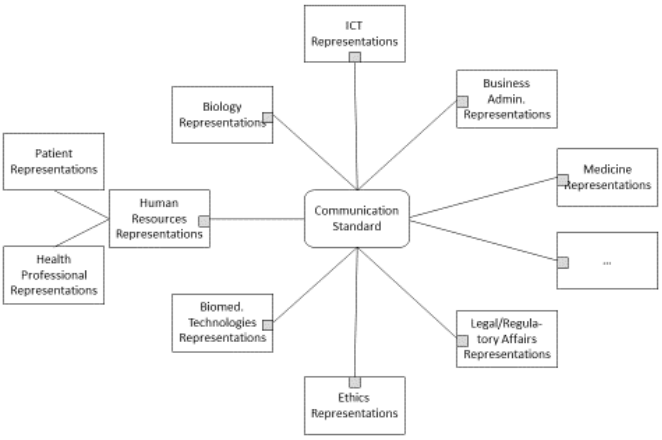


Figure 39 – ISO 23903 Information View

For the Business View, the ISO 23903 Interoperability and integration reference architecture – Model and framework shall be used, architecturally and ontologically re-engineering all elements/components (Figure 37). This approach is applicable for all principals (persons, organisations, devices, components, objects) and can therefore be deployed not only in the healthcare domain but in all domains addressed in the OPEN DEI project

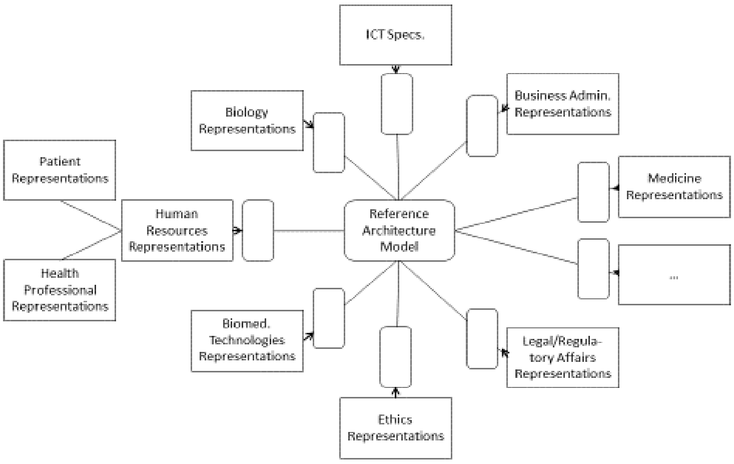


Figure 40 – ISO 23903 Reference architecture model

A newer information view representation with the inclusion of the computational view and the engineering view is the aforementioned HL7 FHIR standard.

An example for re-engineering a model to correctly integrate/interrelate it with other models and specification is demonstrated in Figure 7 with re-engineering the incorrectly layered European Interoperability Framework (EIF) already mentioned in section 6.1 (Figure 31)

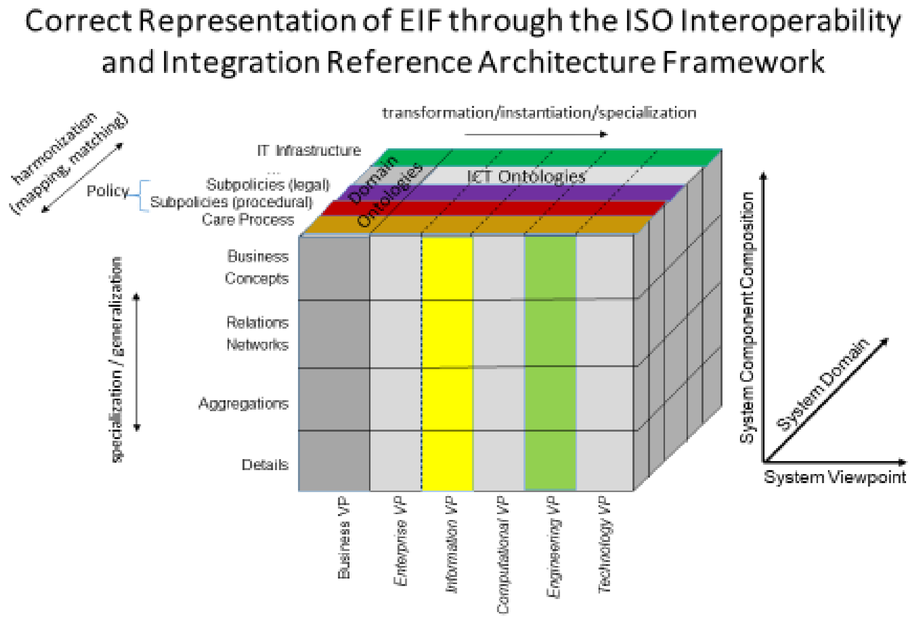


Figure 41 – Expressing EIF with ISO 23903

The different views deploy different representation languages/ontologies with growing expressivity and formalisation, using more restrictive grammars according to the Chomsky classification of grammars. This results in less generative power and more restrictions to special structure, but simpler execution. At that level, the correctness of representations and relations regarding real-world components from a domain-specific perspective and in specific context is not decidable. Therefore, the modelling process has to follow the Good Modeling Best Practices, starting with the use-case-specific definition of elements/components and their naming through domain experts before transforming the resulting model into the other views of the development process. The same holds for any integration challenges.

6.5 Example of InteropEHRate Research Project

The InteropEHRate project (www.interopehrate.eu) is an European research project funded from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826106 and is also part of the OPEN DEI healthcare cluster.

The main breakthrough behind this project is to enable patients to be in full control of the use of their health data. The central instrument, being laid in "patients' hands" is the Smart EHR (S-EHR), leveraging a set of new protocols (defined by the project itself) for secure and cross border exchange of health data. S-EHRs App can communicate with applications of health care providers and researchers of different countries via highly secure channels of two kinds: local (based on a Device-to-Device protocol that does not use the Internet), remote (using Internet-based protocols). Using applications based on the InteropEHRate protocols, health data of the citizen flows from producer organisations (private or public) to consumer organisations (European hospitals or European Research Centres), under the control of the citizen's mobile acting as a central hub of the overall communication.

The InteropEHRate project defines a standard architecture, that is a high-level view of the InteropEHRate open specification. It enables citizen-centred and decentralised health data sharing, through the secure storage of health data on Citizen's personal mobile device avoiding the sharing of health data with app vendors or other third parties. The specification(s) defines a family of open-source communication protocols and a set of constraints for mobile applications and optional cloud services that support the secure cross any border exchange of health data with or without Internet, with or without cloud storage, in a GDPR-compliant way.

The InteropEHRate open specification is open in the sense that each one of the specified protocols and applications may have different implementations, possibly provided by different competing vendors. Conformance to the open specifications assure the interoperability among implementations of different vendors.

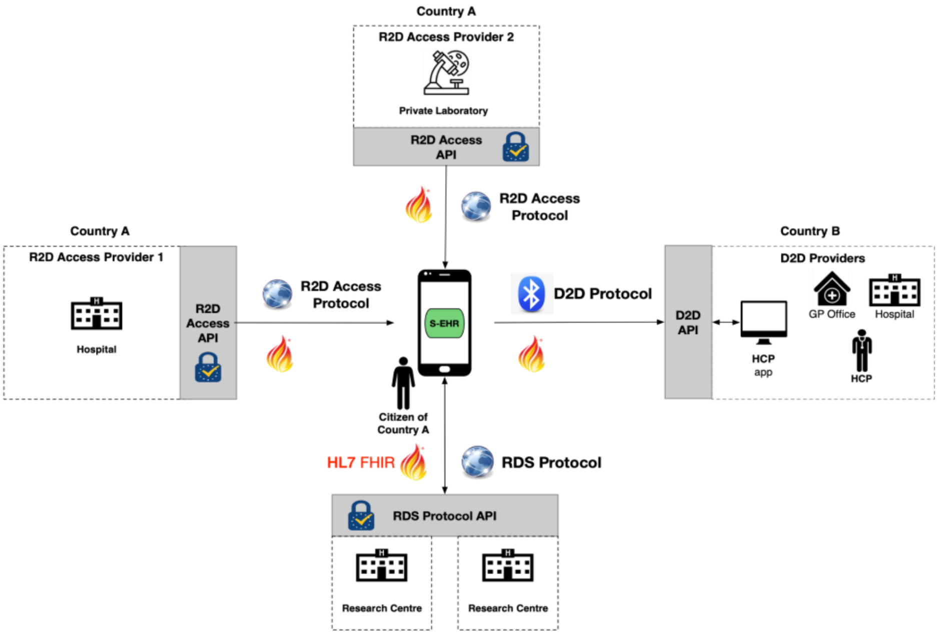


Figure 42 – InteropEHRate architecture

The InteropEHRate architecture assumes that in the near future the EU citizens will own standard kinds of mobile applications called Smart EHRs (S-EHRs). Note that a S-EHR is not a specific software, but a standard kind of software. Citizens will be able to choose among different S-EHRs, conformant with InteropEHRate, offered by different vendors.

More specifically, the S-EHR uses the so called Remote-to-Device (R2D) protocols to exchange health data at distance (on the Internet) with healthcare organisations while the Device-to-Device (D2D) protocol allows exchanging health data with healthcare organisations during face-to-face encounters (without the usage of the Internet, but adopting short range communication technologies like Bluetooth). The portion of the information system of the healthcare organisation used by HCPs to interact with the S-EHR is called the HCP App.

The InteropEHRate project makes large use of interoperability frameworks specifically defined for the eHealth domain. It makes large use of the HL7-FHIR (simply FHIR from here on) standard, both for what concerns the RESTful API (defines how data can be requested or sent to a FHIR server) and the data model. Regarding the latter, the communication protocols of the project allows only to exchange with FHIR data model. Moreover, leveraging the customisation capabilities provided by FHIR itself, the InteropEHRate project has defined two Implementation Guides (set of profiles), defining every detail of the data exchanged. FHIR specification allows adopters to customize the FHIR reference data model, adding more specific constraints (attributes mandatory, attributes binding to specific value set, definition of additional attributes, removal of defined attributes). This activity produces a set of FHIR profiles that can be used by FHIR software validators to validate the data against the profile in order to measure the level of conformance of exchanged data, such profiles must be publicly available.

But interoperability does not only concerns a common representation of data, it also regards how data are requested / sent to an eHealth server. From this point of view the InteropEHRate project adopted part of the FHIR RESTFul API for the R2D family protocols (internet based protocol). FHIR specification allows adopters to also customize its RESTFul API (the list of service exposed by a health server), specifying what services are provided by a server, and what parameters each service accepts. Such information are provided by the server itself to clients, by mean of a special meta data (defined by FHIR) called CapabilityStatement. Clients can automatically detect what operations are provided by a FHIR server by requesting the *CapabilityStatement* to a server and by inspecting it. The *CapabilityStatement* does not provide any information regarding the semantic of a specific service, because the semantic has already been specified by the FHIR specification and cannot be altered. The *CapabilityStatement* only provides information regarding what subset of the set of services defined by the specification is provided by a specific server instance. Having defined the R2D protocol as a subset of the FHIR RESTful API implies that if a hospital already has a server exposing FHIR interface, it is very likely that this server will also be compliant to one or more of the R2D family protocols.

The project also addresses the need to support organisations with data standardisation[72]: the InteropEHRate system takes local datasets as input and produces representations that fulfil FAIR (findable, accessible, interoperable, reusable) principles. The process consists of the in-depth understanding of the data, followed by the in-depth adaptation of the data to the required formats (e.g. FHIR), terminologies and codings. Elements of technologies are already present in existing systems. Examples include formal domain knowledge (e.g. SNOMED CT) and terminology servers that help automate mappings towards standards; information extraction that helps formalise unstructured data; and graphical extract, transform, and load (ETL) tools that provide a more agile method for defining transformations. InteropEHRate innovates by incorporating all these components into an end-to-end data integration methodology, supported by a suite of graphical tools, while simultaneously pushing the boundaries of precision and automation

Another important interoperability framework adopted by the InteropEHRate project regards security and more in detail the citizen digital identity. In this case the standard adopted is eIDAS (Electronic Identification and Trust Services Regulation - eIDAS N. 910/2014). eIDAS is part of the Connecting European Facility programme, it is an EU regulation that defines digital identification for citizens and services (provided by public and private trusted organizations) for electronic transactions in the European market. Within the context of the InteropEHRate (cross border health data exchange) it makes perfect sense to use a (federated) unique identity provider acting as a Single Sign On and allowing European Citizens to access all the InteropEHRate services using one single identity, without requiring citizens to have several digital identities one per health service.

[72] https://www.interopehrate.eu/blog/2022/05/17/white-paper-towards-interoperable-health-data/

7 CONCLUSION AND RECOMMENDATIONS

This document has studied the creation of cross domain capability focusing on four domains, manufacturing, agrifood, energy and health/care.

While ISO 23903 (Interoperability and Integration Reference Architecture) was published in the health domain, it can be used as a general cross-domain reference. It provides an abstract and generic model and framework to represent any system of systems using a system-theoretical, architecture-centric, ontology-based, policy-driven approach:

- Each system, component, specification can be represented in its specific language and representation style without requiring revision.
- The model is based on universal type theory and universal logics, good modelling practices as well as systems and language theories. The model is presented by three dimensions:
  - systems domains;
  - systems composition/decomposition using four generic granularity levels; and
  - systems development process or evolution.
- To represent the components, related ontologies such as the domain ontologies establishing the business view and IT ontologies for the other views are deployed, which can be mapped for interoperability, integration and transformation. The approach is technology independent and allows for the integration of any domain, any reference architecture or any model/specification including those presented in this document. The feasibility of this approach for integration and interoperability has been demonstrated in the multidisciplinary health and social care domain to mapping existing specification, to manage security, privacy and trust, and to manage languages and ontologies.

Two recommendations are made. The **first recommendation** is to agree and standardise a cross-domain convergence framework. As showed in Figure 40, it involves construction processes:

- An architecture construction process which is based on the metaRA guidance specified by ISO/IEC JTC 1/AG 8 to specify reference architectures and construct solutions architectures. The resulting construction platform solution architectures can integrate the OPEN DEI reference architecture framework as well as the architecture orientation of the BSDA initiative.
- An interoperability construction process which is based on the use of the interoperability framework developed by ISO/IEC JTC 1/SC41 to enable the identification of interoperability points (where interoperability takes place in an architecture), interoperability cases (why interoperability is needed), and interoperability profiles (specification leveraging transport, syntactic, semantic, behavioural and policy interoperability).

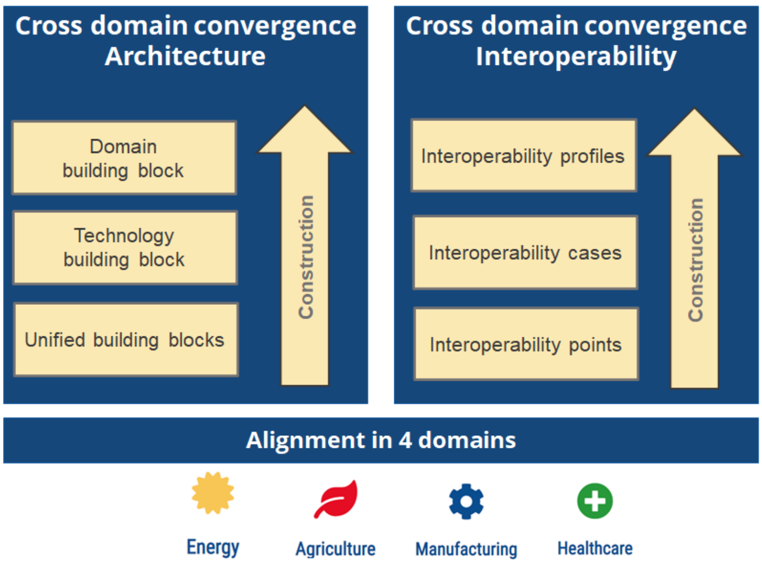


Figure 43 – Cross-domain architecture and interoperability convergence

The **second recommendation** is to agree and standardise associated building blocks in the design of data spaces. The OPEN DEI **design principles for data spaces** document identifies four categories of building blocks:

- the infrastructure category which includes (1) the data models and formats, (2) the data exchange APIs, and (3) the provenance and traceability building blocks;
- the trust category which includes (1) the identity management, (2) the access & usage control / policies, and (3) the trusted exchange building blocks;
- the data value category which includes (1) the metadata & discovery protocol, (2) the data usage accounting, and (3) the publication & market place services building block; and
- the governance category which includes (1) the overarching cooperation agreement, (2) the operational, and (3) the continuity model building blocks.

As showed in Figure 41, this paper has contributed the following:

- semiotic considerations for the data models and formats building block;
- semantic considerations for the data models and formats building block;
- trustworthiness considerations for the trusted exchange building block;
- resource management considerations for the identity management building block;
- policy and context awareness for the metadata and discovery services building block;
- interoperability cases considerations for the governance operational building block; and
- digital twin and AI integration considerations.

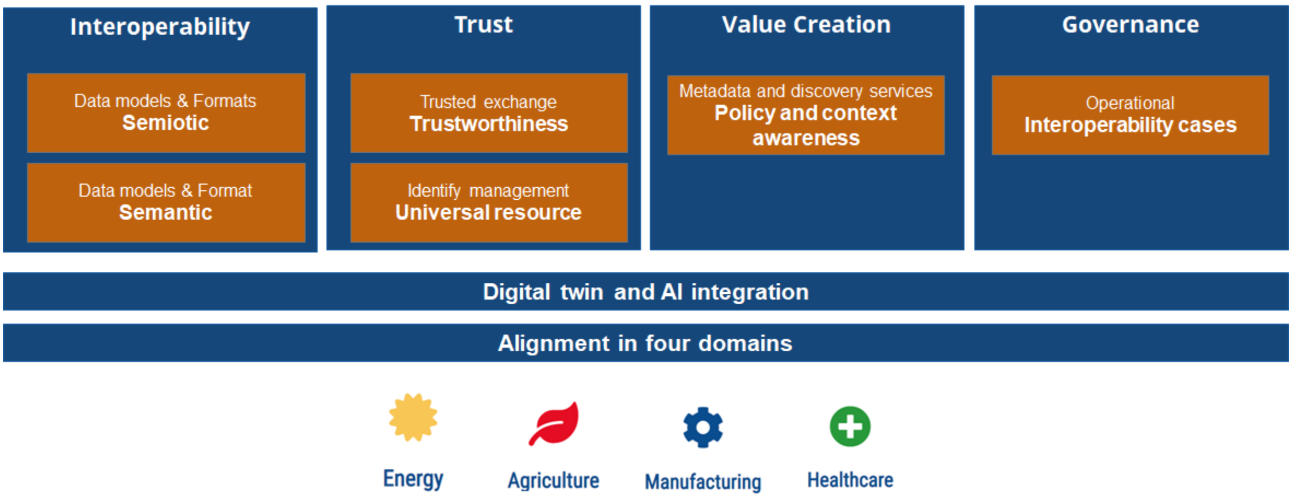


Figure 44 – Contributions to data space building blocks

OPENDEI

www.opendei.eu

@opendei_eu

opendei-eu