Deliverable 2.5: BD4NRG Reference Architecture First Version

WP2 – System Requirements and Specifications

August 2021

www.bd4nrg.eu

Disclaimer

Copyright Message

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613

Grant Agreement Number: 872613 Acronym: BD4NRG

Full Title	Big Data for Next Generation Energy		
Topic	DT-ICT-11-2019 Big Data solutions for energy		
Funding scheme	H2020- IA: Innovation Action		
Start Date	January 2021	Duration	36
Project URL	http//www.bd4nrg.eu		
Project Coordinator	ENG		
Deliverable	D2.5 - BD4NRG Reference Architecture First Version		
Work Package	WP2 - System Requirements and Specifications		
Delivery Month (DoA)	M8	Version	1.0
Actual Delivery Date	31/08/2021		
Nature	Report	Dissemination Level	Public
Lead Beneficiary	RWTH		
Authors	Katharina Wehrmeister (RWTH), Alexander Pastor (RWTH), Leonardo Carreras (RWTH), Stefan Dähling (RWTH), Marzia Mammina (ENG), Alessandro Rossi (ENG), Davide Profeta (ENG), Efthimios Bothos (NTUA), Babis Magoutas (NTUA), Vagelis Karakolis (NTUA), Haris Doukas (NTUA), Ruben Bäumer (Centrica), Natalie Samovich (ENERC), Tiago Alves (ENERC), Joao Campos (ENERC), Sonia Jiménez (IDSA), Giovanni Di Orio (NOVA), Pedro Maló (NOVA), Elisa Herrmann (ATOS), Apostolos Kapetanios (ED), Alberto Abella (FIWARE)		
Quality Reviewer(s):	Massimo Bertoncini (ENG), Olivier Aine (ENTSOE)		
Keywords	Reference architecture, Data Spaces, IDSA, GAIA-X, FIWARE, BRDIGE RA, analytics, governance, marketplace		

Preface

BD4NRG focuses on addressing emerging challenges in Big Data management for energy sector with an innovative open holistic solution for smart grid-tailored, near real time, energy-specific and AI-based open Big Data Analytics modular framework. The vision is to deliver holistic services for techno-economic optimal management of Electric Power and Energy Systems value chain. Services range from optimal risk assessment for energy efficiency investments planning, to optimized management of grid and non-grid owned assets, improved efficiency, and reliability of electricity networks operation, while at the same time contributing to achieve fair energy prices to the consumers and laying the foundations for an EU-level energy-tailored data sharing economy. The BD4NRG Toolbox will be implemented and validated in real-life pilots in 12 large-scale demo sites across 10 countries for:

- Increasing the efficiency and reliability of the electricity network BD-4-NET
- Optimising the management of assets connected to the grid BD-4-DER

- De-risking investments in energy efficiency and increasing the efficiency and comfort of buildings BD-4-ENEF

Who We Are

	Participant Name	Short Name	Country Code	Logo
1	ENGINEERING – INGEGNERIA INFORMATICA SPA	ENG	IT	
2	NATIONAL TECHNICAL UNIVERSITY OF ATHENS	NTUA	GR	
3	RHEINISCH-WESTFAELISCHE TECHNISCHE HOCHSCHULE AACHEN	RWTH	DE	
4	EUROPEAN DYNAMICS LUXEMBOURG SA	ED	LU	
5	INTERNATIONAL DATA SPACES EV	IDSA	DE	
6	EUROPEAN NETWORK OF TRANSMISSION SYSTEM OPERATORS FOR ELECTRICITY AISBL	ENTSO-E	BE	
7	PANEPISTIMIO DYTIKIS ATTIKIS	UNIWA	GR	
8	ATOS SPAIN SA	ATOS	ES	
9	FUNDACION CARTIF	CARTIF	ES	
10	UNIVERZA V LJUBLJANI	UNILJ	SL	
11	ENEL X SRL	ENELX	IT	
12	REN - REDE ELECTRICA NACIONAL SA	REN	PT	
13	CENTRO DE INVESTIGACAO EM ENERGIA REN - STATE GRID SA	RDN	PT	
14	UNINOVA-INSTITUTO DE DESENVOLVIMENTO DE NOVAS TECNOLOGIASASSOCIACAO	UNINOVA	PT	
15	ENERCOUTIM - ASSOCIACAO EMPRESARIALDE ENERGIA SOLAR DE ALCOUTIM	ENERC	PT	
16	FIWARE FOUNDATION EV	FIWARE	DE	
17	CENTRICA BUSINESS SOLUTIONS BELGIUM	CENTRICA	BE	
18	NEDERLANDSE ORGANISATIE VOOR TOEGEPAST NATUURWETENSCHAPPELIJK ONDERZOEK TNO	TNO	NL	

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

	Participant Name	Short Name	Country Code	Logo
19	ASM TERNI SPA	**ASM**	IT	
20	VIDES INVESTICIJU FONDS SIA	**LEIF**	LV	
21	COMSENSUS, KOMUNIKACIJE IN SENZORIKA, DOO	**COMSENSUS**	SL	
22	HOLISTIC IKE	**HOLISTIC**	GR	
23	INTERUNIVERSITAIR MICRO-ELECTRONICA CENTRUM	**IMEC**	BE	
24	TERRASIGNA SRL	**TS**	RO	
25	UBIMET GMBH	**UBIMET**	AT	
26	ELEKTRO LJUBLJANA PODJETJE ZADISTRIBUCIJO ELEKTRICNE ENERGIJE D.D.	**EKL**	SL	
27	BORZEN, OPERATER TRGA Z ELEKTRIKO, D.O.O.	**BORZEN**	SL	
28	AJUNTAMIENTO DE SANT CUGAT DEL VALLES	**AJSCV**	ES	
29	ELES, D.O.O., SISTEMSKI OPERATER PRENOSNEGA ELEKTROENERGETSKEGA OMREŽJA	**ELES**	SL	
30	E-LEX - STUDIO LEGALE	**ELEX**	IT	
31	OSMANGAZI ELEKTRIK DAGITIM ANONIM SIRKETI	**OEDAS**	TR	
32	VEOLIA SERVICIOS LECAM SOCIEDAD ANONIMA UNIPERSONAL	**VEOLIA**	ES	
33	STICHTING EG	**EGI**	NL	
34	CINTECH SOLUTIONS LTD	**CN**	CY	
35	EMOTION SRL	**EMOT**	IT	

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Contents

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Figures

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Tables

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Executive Summary

The BD4NRG project encompasses a large number of stakeholders in the electricity sector as well as cross-sectorial areas such as mobility and buildings. The Big Data analytic services, which are to be implemented over the course of the project, will be demonstrated in 12 Large Scale Pilots. This results in a heterogeneous landscape of requirements and goals, which WP2 collects, analyses and integrates.

This document provides the first version of the BD4NRG reference architecture. The reference architecture (RA) provides the overall view of all the components, e.g., hardware, services and actors, which compose the various use cases in the Large Scale Pilots, and their relation. Hence, it integrates the aforementioned requirements and goals of the different stakeholders into one picture. In turn in can also be used to derive the concrete architectures for the individual use cases.

To express the reference architecture, this deliverable analyses different architecture models. The BRIDGE RA is chosen to express the BD4NRG reference architecture, since it builds upon an existing model in the electricity sector, i.e., SGAM, and extends it to other domains. This document provides a mapping of the use case descriptions from the Large Scale Pilots to the BRDIGE RA model. This mapping reveals that for some use cases more information is required, especially regarding used data models.

Since many use cases comprise the need for data sharing among different stakeholders, the BD4NRG project also envisions to incorporate existing efforts in the area of Data Spaces for this purpose. Existing project coalitions and efforts are analysed to this end. The most interesting initiatives in this area are IDSA, GAIA-X and FIWARE, which define their own reference architectures for the realisation of data spaces. However, some integration effort among these initiatives is currently ongoing we expect that the result of these efforts will be available within the project duration and can hence be used.

The individual mappings of use cases to the BRIDGE RA and the Data Space efforts are merged into the first version of the overall BD4NRG reference architecture. The BRIDGE RA depicts a starting point for this. It is complemented by a vertical pillar, which is concerned with the integration of Data Space enablers. The Grant Agreement introduces three layers for the architecture: Analytics Services, Data Governance and Marketplace. For the Analytics Services layer ALIDA, a Big Data analytics platform made available by ENG, will be used. It provides a catalogue of Big Data Analytics services and functionality for data processing. WP 4 will adopt and extend ALIDA during the project. The Data Governance layer will build upon the integration efforts of IDSA, GAIA-X and FIWARE mentioned before. WP3 will coordinate the corresponding tasks. The Marketplace layer will refactor and adapt some complementary developments from ENG and ATOS. ENG initially designed its tokenized DLT/blockchain/smart contract energy-centered marketplace within the H2020 projects eDREAM and COORDINATE. A further maturation followed in the H2020 PlatONE project, with the goal to coordinate and optimally manage flexibility and/or energy trading in a P2P way. ATOS developed its marketplace in the CROSS-CPP project and connects the different actors by managing the data flow among them. WP5 will integrate both efforts.

The last contribution of this document is to give concrete recommendations for the implementations within the project. For this purpose, the deliverable reviews existing Linux Foundation Energy projects

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

and derives general recommendations. The main recommendations are: the use of a microservice architecture to achieve a modular, scalable and extensible system; the use of popular design patterns for APIs, such as REST; the use of a common data representation, such as NGSI-LD; the consideration of security issues from the beginning of the implementation phase; and a proper documentation to allow the reusability of implementations.

While this document depicts the first version of the BD4NRG reference architecture, deliverable D2.6 of the BD4NRG project will describe the final version. The main efforts towards deliverable D2.6 are the integration of other WP2 activities into the reference architecture and the further collection of information from the Large Scale Pilots, which were not yet available at this stage of the project.

Moreover, some project partners (ENG, TNO, FIWARE, IDSA) actively participate in the emerging EU-wide GAIA-X effort to design, develop and validate technology building blocks to support the EU Federated infrastructure of the future, aimed at supporting the rising data economy. Therefore, this deliverable is also devoted to representing a first-of-a-kind workbench which leverages on the proposed use cases down to the development of data-space compliant interoperable building blocks able to support the energy domain in its digitization pathway. This is becoming one of the more and more strategic actions of the European Commission to facilitate, enable and accelerate the transition towards a consumer-centric smartened renewable and more efficient energy value chain.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

1 Introduction

1.1 Scope of the Document

Deliverable D2.5 *"BD4NRG Reference Architecture First Version"* is produced within the WP2 *"System Requirements and Specifications"* as a part of Task 2.6 *"Big Data, AI and IoT Reference Architecture and alignment with Extended COSMAG and GAIA-X Data Space Architectures"*. The purpose of this document is to define a first version of the BD4NRG Reference Architecture (RA). The RA integrates all use cases into one architectural model. Moreover, it incorporates existing Data Space RAs, which are concerned with data governance and data sharing across different stakeholders.

To this end, the document leverages insights from other WP 2 Tasks as well as information already available from BD4NRG Large Scale Pilots (LSPs). Moreover, several existing efforts in the area of architectural models as well as Data Space RAs are reviewed. Taking a step towards a more concrete specification, it gives implementation recommendations based on relevant Linux Foundation Energy (LFE) projects.

The requirements will evolve over time as the project goes on, and D2.5 serves as a starting point for further efforts in the WP, working towards the final RA in deliverable D2.6.

1.2 Relation to Other Documents

This document ties together the efforts made in WP 2 in the first 8 months of the project. It relies on deliverable D2.1 to provide necessary information about the project's LSPs as well as existing data-driven reference architectures. Considerations of data models and privacy concerns from deliverables D2.3 and D2.4 respectively are also relevant for this content. Finally, while deliverable D2.2 approaches individual components of the BD4NRG architecture, D2.5 considers the overarching structure among them.

This deliverable is relevant to WPs 3, 4 and 5, named *"Data Governance and Management"*, *"Data Modelling & Open Modular AI-based edge-level Analytics Toolbox"* and *"Marketplaces for realizing EU-level Energy Data Economy"*, respectively. It defines the corresponding layers of the BD4NRG architecture and puts them into context of existing work along with the BD4NRG LSPs, as well as providing implementation recommendations for these technical work packages.

1.3 Abbreviations

Abbreviation	Expansion
AAS	Asset Administration Shell
AI	Artificial Intelligence
AIOTI	Alliance for Internet of Things Innovation
ALM	Application Lifecycle Management

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Abbreviation	Expansion
API	Application Programming Interface
B2B	Business to Business
BD	Big Data
BDA	Big Data Analytics
BDD	Behaviour Driven Development
BDVA/DAIRO	Big Data Value Association/Data, AI and Robotics
CEF	Connecting Europe Facilities
CIDM	Common Industrial Data Model
CNC	Computerized Numerical Control
CoMPAS	Configuration Modules for Power Industry Automation Systems
COSMAG	Comprehensive Architecture for Smart Grid
DaaS	Data as a Service
DAPS	Dynamic Attribute Provisioning Service
DCS	Distributed Control System
DDD	Domain Driven Development
DER	Distributed Energy Resources
DIP	Dependency Inversion Principle
DLT	Distributed Ledger Technology
DMS	Data Management System
DSO	Distribution System Operator
DT	Digital Transformation
DTM	Dynamic Trust Monitoring
EBSI	European Blockchain Service Infrastructure
eIDAS	Electronic Identification, Authentication and Trust Services

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Abbreviation	Expansion
ENEF	Energy Efficiency
ER	Enterprise Reporting
FTP	File Transfer Protocol
GDPR	General Data Protection Regulation
GUI	Graphical User Interface
GXF	Grid Exchange Fabric
HLA	High-Level Architecture
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
ICT	Information and Communication Technology
IDSA	International Data Space Organization
IDS-RAM	International Data Spaces Reference Architecture Model
IIC	Industrial Internet Consortium
IIoT	Industrial Internet of Things
IIRA	Industrial Internet Consortium Reference Architecture
IoT	Internet of Things
JDBC	Java Database Connectivity
JSON	JavaScript Object Notation
LFE	Linux Foundation Energy
LI	Location Intelligence
LSP	Large Scale Pilot
MLaaS	Machine Learning as a Service
MQTT	Message Queueing Telemetry Transport
OEM	Original Equipment Manufacturer

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Abbreviation	Expansion
OLAP	On-Line Analytical Processing
OO	Object-Oriented
OPEN DEI	OPEN Digitising European Industry
OSGP	Open Smart Grid Platform
PA	Predictive Analysis
PACS	Protection, Automation & Control Systems
PLC	Programmable Logic Controller
PLM	Product Life Cycle Management
PM	Performance Management
PPP	Private Public Partnership
PSD2	2^{nd} Payment Services Directive
QoS	Quality of Service
RA	Reference Architecture
RAF	Reference Architecture Framework
RAMI	Referenzarchitekturmodell Industrie 4.0 (German: Reference Architecture Model Industry 4.0)
REST	Representation State Transfer
RIAPS	Resilient Information Architecture Platform for Smart Grid
SCADA	Supervisory Control and Data Acquisition
SCL	System Level Configuration
SDO	Standard Development Organization
SGAM	Smart Grid Architecture Model
SI	Smart Intelligence
SIMaaS	Simulation as a Service

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Abbreviation	Expansion
SOGNO	Service-Based Open-Source Grid Automation Platform for Network Operation of the Future
SoS	System of Systems
TDD	Test Driven Development
TSO	Transmission System Operator
WP	Work Package
XML	Extensible Markup Language

Table 1 – Abbreviations

1.4 Document Structure

This document is organized as follows. Chapter 2 reports the methodology for collecting, analysing and leveraging existing architecture efforts. Chapter 3 presents relevant architecture models and puts them into the context of BD4NRG LSPs and WP2 efforts, and Chapter 4 discusses established data space RAs and their possible integration with each other. Subsequently, Chapter 5 builds upon these chapters to converge the discussed efforts into a first version of the BD4NRG RA. Chapter 6 then presents relevant LFE projects and gives implementation recommendations accordingly, before Chapter 7 finally concludes the document.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

2 Methodology

2.1 Overview of Reference Architecture Efforts

This section enumerates existing reference architectures to consider toward the specification of the BD4NRG reference architecture. Additionally, it gives a perspective on which aspects to consider for BD4NRG reference architecture alignment with the LSPs.

The reference architecture efforts can be divided coarsely into four groups: architecture models, project coalitions, data space reference architectures, and microservice architectures. These groups are explained in more detail in the next subsection (Section 2.2).

The architecture models are BRIDGE, COSMAG, and RAMI. For project coalitions, OPEN DEI, BDVA/DAIRO, and AIOTI are considered. Furthermore, the data space reference architectures IDS/GAIA-X, FIWARE, and the AIOTI Marketplace are examined. SOGNO and other Linux Foundation Energy (LFE) projects serve as examples of concrete microservice architectures.

The objective in aligning these efforts is not only to capture the data exchange requirements such as interoperability, sovereignty, trust and security of BD4NRG use cases but also to provide a flexible and general, yet sufficiently detailed reference architecture as a guideline for subsequent implementation. At the same time BD4NRG is aiming at representing a first-of-this-kind validation of IDSA conceptual architecture along the energy domain, hence providing significant contribution to the parallel work currently being carried out within the Energy Working Group of GAIA-X. To these ends and to promote acceptance, emphasis is to be put on building it following eminent architectural models, mainly BRIDGE which in turn extends SGAM. Furthermore, evaluating maturity and identifying reusable building blocks from the more concrete data space reference architectures, particularly GAIA-X/IDS and FIWARE is paramount for the most efficient use of resources possible.

2.2 Classification

As mentioned in the previous section, this document classifies existing efforts into four classes in descending levels of abstraction: architecture models, project coalitions, data space reference architectures and microservice architectures. In the following, the nature of these classes is more explicitly characterized.

Architecture models are rather abstract meta-models that define and specify architectures. They are less concerned with direct platform-building facilitating data exchange but strive for the holistic modelling of industrial domains by establishing industry-wide common and structured vocabulary.

Project coalitions provide discussion, dissemination, and networking opportunities for industrial and academic partners. In broad terms, they serve the advancement of European economic and scientific interests by providing a platform for knowledge sharing and coordination of agendas such that efforts develop in the same direction and final solutions are tacitly compatible.

Data space reference architectures provide the necessary specifications and, to a certain degree, implementations of the building blocks that recognize the specific data space requirements. They explicitly define class APIs and are designed to be compatible with and extending standards. The IDS reference architecture is built with GAIA-X and FIWARE with the Semantic Web in mind, respectively.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

On the other hand, their modularity allows for developing third-party extensions and/or implementations of their architectures.

A microservice architecture is a concrete design pattern for applications and can be used for the implementation of a reference architecture. In contrast to monolithic architectures, it allows for separating concerns into multiple independent components, so-called services that can more easily be developed, maintained, tested, and deployed by independent parties. Furthermore, such an approach is more in spirit with data federation and sovereignty as decentralized, independent components can be comfortably integrated into such an architecture. Lastly, implementations with orchestration in mind support frictionless dynamic scaling and robustness against single points of failure.

The following chapters describe the architecture models, project coalitions, and data space reference architectures, respectively. Subsequently, their possible convergence into the BD4NRG RA is explained.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

3 Architecture Models

The first class of reviewed efforts are architecture models. These can be used to express the BD4NRG RA. As most relevant efforts in this area, the chapter presents the BRIDGE RA, COSMAG and RAMI. Subsequently, the BRDIGE RA is chosen to express the individual use cases of the project as a starting point for the BD4NRG RA, which is introduced later in Chapter 5.

3.1 COSMAG

COSMAG (Comprehensive Architecture for Smart Grid) is an effort to define specifications for any possible data-exchange interaction in the Smart Energy sector, with the aim of helping the transition of the energy system to a Data Economy, addressing the key issues related to the needs for a Digital Single Market within Europe.

COSMAG does not aim to define any new standard and is instead based on existing efforts and architectures. Among these are the Smart Grid Architecture Model (SGAM)[1] and the work by EC Smart Grids Task Force – Expert Group 3[2].

SGAM is a three-dimensional model, represented by Figure 1, that comprises five interoperability layers (Business, Function, Information, Communication and Component), as well as a Smart Grid plane with two dimensions: Zones, representing the hierarchical level of the power system (Process, Field, Station, Operation, Enterprise, Market) and Domains, depicting the electrical energy conversion chain (Generation, Transmission, Distribution, DER, Customer Premises).

[1] https://ec.europa.eu/energy/sites/ener/files/documents/xpert_group1_reference_architecture.pdf

[2] Regulatory Recommendations for the Deployment of Flexibility (https://ec.europa.eu/energy/sites/ener/files/documents/EG3%20Final%20-%20January%202015.pdf)

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

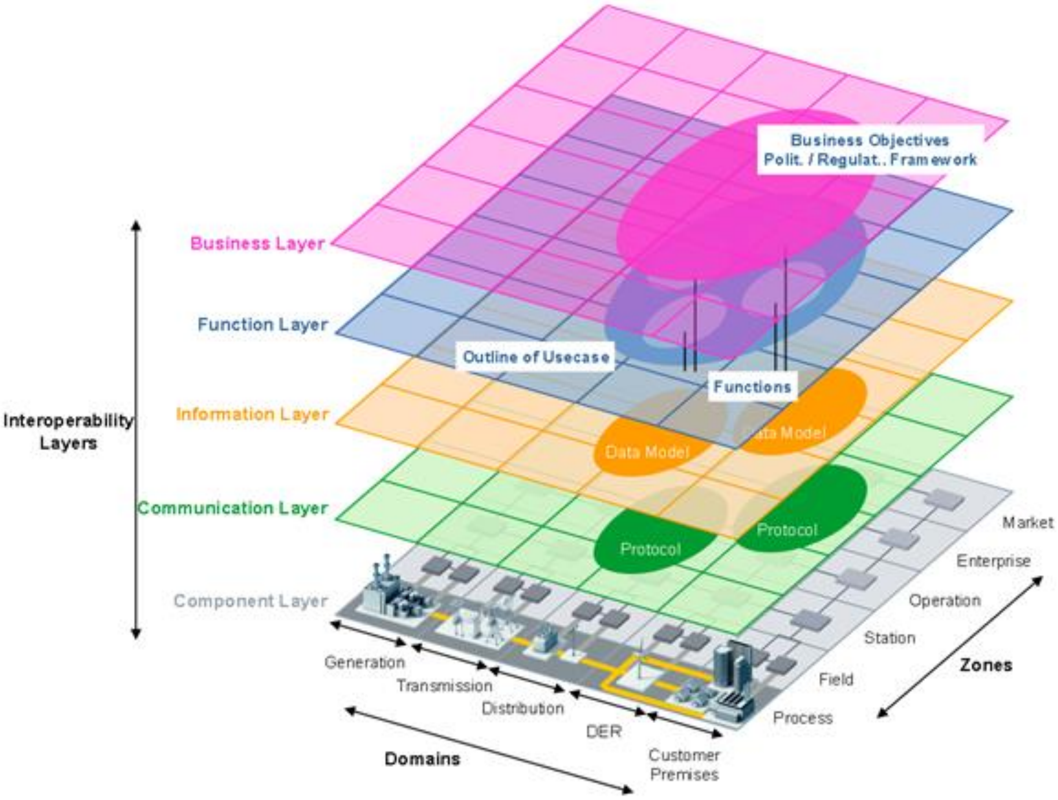


Figure 1 – The Smart Grid Architecture Model.

Figure 2 depicts the possible roles, market interactions and actors that are envisioned as part of a modern electricity market in the SGTF – EG3 Report.

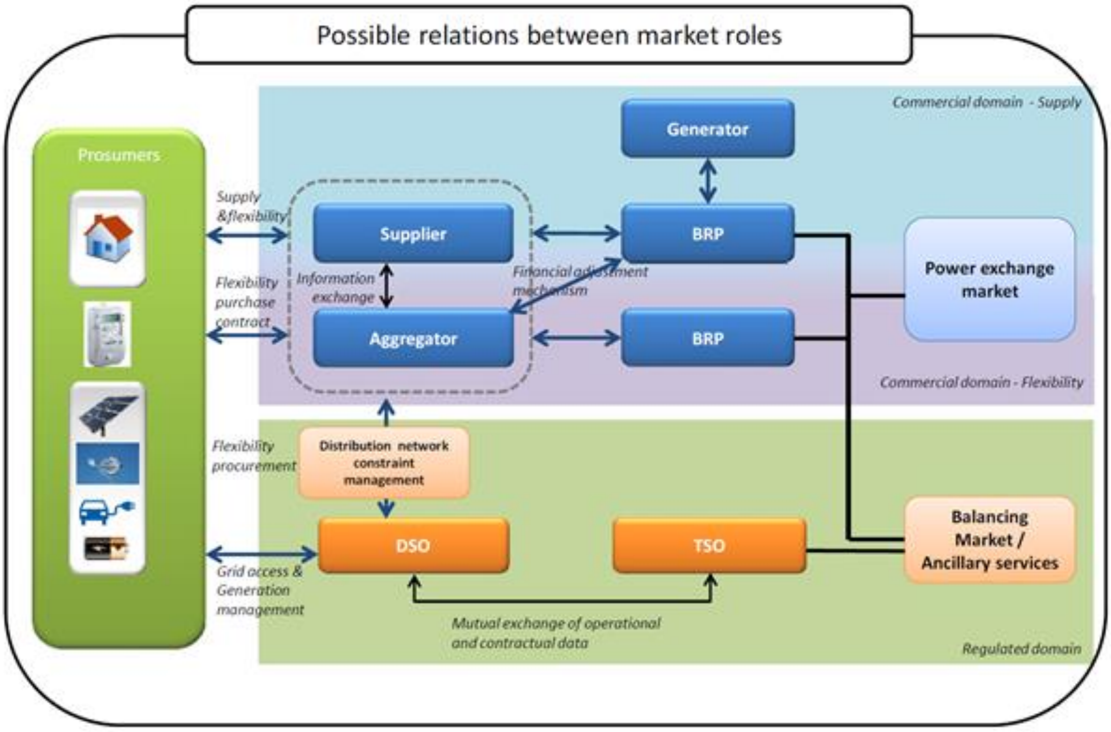


Figure 2 – Possible roles, market interactions and actors envisioned in the SGTF – EG3 Report.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

This representation allows an analysis of the most important communication flows and their maturity, considering both the status of the protocol definitions and data models.

The list of actors derived from the Figure 2 is not exhaustive. Other actors could be identified, some of which are listed below. Interfaces with local markets have not been defined as there is no existing local market up to date, but COSMAG allows to consider its future presence.

- TSO
  - TSO – Market (balancing)
  - TSO – DSO
- DSO
  - DSO – Local Market (no local markets yet)
  - DSO – Prosumer
- Aggregator
  - Aggregator – Local Market (no local markets yet)
  - Aggregator – DSO (integration of network constraints)
  - Aggregator – Prosumer (e.g. OpenADR)
- Prosumer
  - Prosumer – Retail (mainly metering, e.g. DLMS/COSEM)
  - Prosumer – Prosumer (local energy communities)

A defining goal of COSMAG is sector coupling. More specifically, achieving cross-sector and cross-stakeholder data exchange through common standards regarding data models and ontologies, comprehensive documentation and Open APIs. Additionally, the use of DLTs for secure and trusted data share among stakeholders and exploitation of Big Data by avoiding the creation of data silos will allow links to other platforms, like those related to smart cities, heating, transportation, building technologies, industries, etc.

More information on COSMAG can be found in Chapter 3.5 of BD4NRG deliverable D2.1.

3.2 RAMI

3.2.1 Reference Architectural Model for Industrie 4.0 (RAMI 4.0)

Within the Plattform Industrie 4.0 project, the working group "Reference architectures, standards and norms" has developed the Reference Architectural Model Industrie 4.0 (RAMI 4.0), to establish a set of guidelines and specifications on how to design and implement Industry 4.0 solutions in practice. It describes fundamental aspects of an Industry 4.0 manufacturing enterprise and how to approach them in a structured and organized manner. In particular, RAMI4.0 is a three-dimensional map aiming at fully describing and representing the Industry 4.0 space. It highlights the most important aspects of Industry 4.0 while establishing a common understanding between all the relevant actors and participants involved in Industry 40 discussions (see Figure 3). As explained in [3], RAMI 4.0 is an

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

adaptation and extension of the Smart Grid Architecture Model (SGAM). SGAM deals with the electricity grid, from generation passing through transmission and distribution to the consumer. Industry 4.0 focuses on production scenarios and product development. Hence, it deals with development processes, manufacturing production lines, machinery, field devices, and final products.

The three-dimensional model enables the strong connection between the world of Information Technology (IT), the manufacturing company and the product lifecycle. Each dimension refers to one of these worlds. The vertical axis refers to the IT perspective which – in turn – relies on several layers with distinct abstraction levels, i.e., spacing from the physical asset to the functional and business layers (e.g. services). This axis deals – concretely – with the digitization of the manufacturing enterprise and thus with IT/OT software architectures. The left-hand horizontal axis refers to the product life cycle management (PLM) and is grounded on the IEC-62890 standard. It relies on two main concepts namely: type and instance. The type refers to the initial stages of the development of a product (design and prototyping) while the instance refers to the instantiation of the product itself, i.e., when the product is being manufactured. Finally, the right-hand horizontal axis refers to the hierarchical levels and organization of a manufacturing company and it is grounded on the IEC-62264 and IEC-61512 standards while extending them by including the Product and the Connected World layers at the bottom and top respectively. Therefore, the right-hand axis represents the generic model of a new/modern manufacturing company.

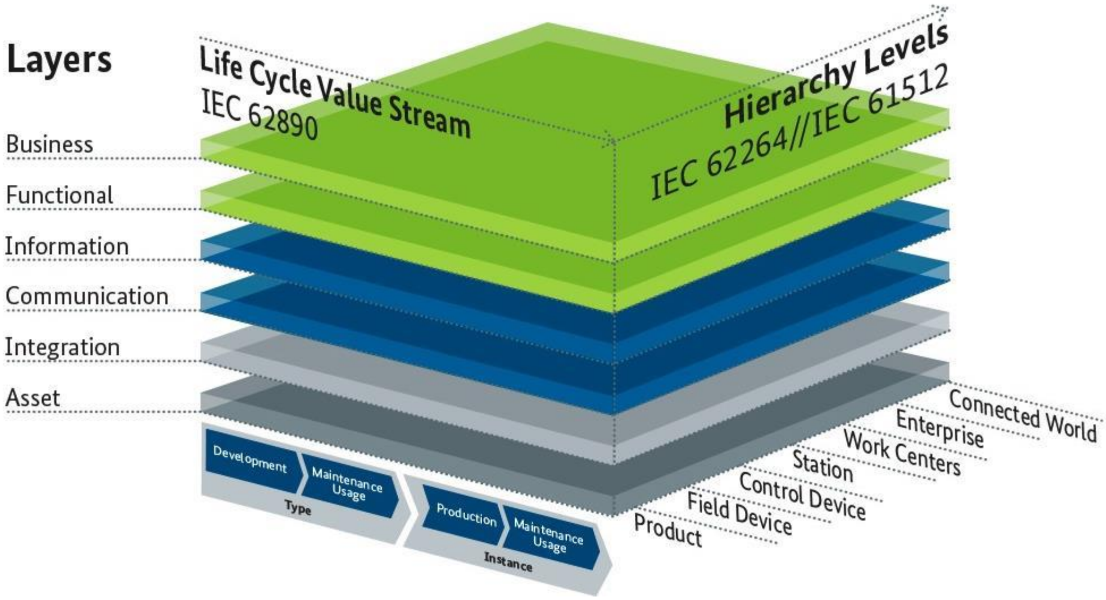


Figure 3 – Reference Architecture Model for Industrie 4.0 (RAMI 4.0)[3]

3.2.1.1 Vertical Axis: Layers

The following layers are part of the vertical axis, starting from the bottom to the top:

[3] ZVEI, "The Reference Architectural Model RAMI 4.0 and the Industrie 4.0 Component," Industrie 4.0, 2015. http://www.zvei.org/en/subjects/Industry-40/Pages/The-Reference-Architectural-Model-RAMI-40-and-the-Industrie-40-Component.aspx (accessed Nov. 11, 2015).

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

- **The Asset Layer** contains the representation of the real/physical elements of a system, e.g., linear axes, metal parts, documents, circuit diagrams, CNC machines, sensors, actuators, etc. Humans are also part of the Asset Layer and are connected to the virtual world via the Integration Layer. The asset layer contains all the elements of a manufacturing production system that could be a data source.
- The **Integration Layer** is responsible for delivering all the necessary mechanisms to allow the integration of components from the Asset Layer in the digital world. Practically, this layer deals with the design and implementation of software connectors for the physical world.
- The **Communication Layer** deals with the standardization of the communication protocols, data format and Application Programming Interfaces (APIs) to be provisioned to the Information Layer.
- The **Information Layer** deals with the pre-processing of the events from virtualized physical components (the ones in the Asset Layer), the execution of event-related rules and complex-event processing.
- The **Functional Layer** deals with application specific rules and decision-making logic as well as services that are executed on the top of the data provided by the Information layer. Depending on the application scenario the models and logic generated within this layer can also be executed in the lower layers.
- The **Business Layer** deals with business rules, processes and models typically combining the services in the functional layer.

3.2.1.2 Left-hand Horizontal Axis: Life Cycle Value Stream

The left-hand horizontal axis considers the lifecycle of the products as well as the associated value streams. The standard IEC 62890:2020[4] establishes the basic principles for Life-Cycle-Management of industrial physical assets. It describes the necessary elements and mechanism for developing and implementing lifecycle strategies and analysis for suppliers, system integrators and product users. For that purpose, IEC 62890 defines essential elements:

- **Type** is an instantiable component with a clearly defined set of properties that are passed on when it is instantiated. An engineered product in digital form is an example of a type. A type is equivalent to what is called a ‘class’ in object-oriented (OO) terminology.
- **Instance** is defined as a specific, uniquely identifiable component of a particular type that, during instantiation, becomes an individual entity imparted through the special set of features of a type. A produced unit is an example of an instance. An instance is equivalent to an ‘object’ of a class (or a type) in OO terminology.
- **Lifecycle** refers to the time from when a product is created (as a type) and to when the maintenance phase has ended (end-of-life).
- **Service life** refers to the time from when a particular product is produced (as an instance) to when it is disposed of.

[4] https://webstore.iec.ch/publication/30583

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

The RAMI 4.0 model depends on these four fundamental concepts.

3.2.1.3 Right-hand Horizontal Axis: Hierarchy Levels

The right-hand horizontal asset includes the functional hierarchy and roles within the manufacturing context based on the standard models described in the IEC 62264 and IEC 61512 standards. RAMI 4.0 extends these standards by adding the functional elements “Field Device”, “Product” and “Connected World”. The “Field Device” layer considers smart and intelligent devices such as Internet-of-Things (IoT) devices, single-board computers, etc. Furthermore, “Product” itself is a fundamental source of data, information and insights that can be used by manufacturing companies to identify and create new value streams connected to both final consumers and machines/components providers. Finally, the “Connected World” functional level highlights the intra-enterprises interactions, i.e., it takes into account the data flow outside the enterprise boundaries while enforcing the collaboration between enterprises, suppliers and consumers.

3.2.2 Asset Administration Shell: Implementing Digital Twins for use in Industry 4.0

The successful implementation of Industry 4.0 solutions in practice strictly depends on the presence of standards. Standards are essential for establishing interoperability, whenever individual solutions need to be integrated. The use of Industrie 4.0 solutions requires interoperability so that components, devices, and applications can communicate seamlessly across companies, industries and countries. The concept of Industry 4.0 components, in general, and of the Asset Administration Shell (AAS) in particular take a fundamental role in this (see Figure 4). Industry 4.0 components are globally and uniquely identifiable elements of an Industry 4.0 system capable to communicate and offer services in accordance with Quality-of-Service (QoS) properties. The basic idea behind Industry 4.0 components is to wrap each asset with an “administration shell” or AAS that is responsible to deliver a minimal and standardized description of a physical asset in the digital world. Practically, the AAS provides the basis for the development and use of unified and open Industrie 4.0 standards.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

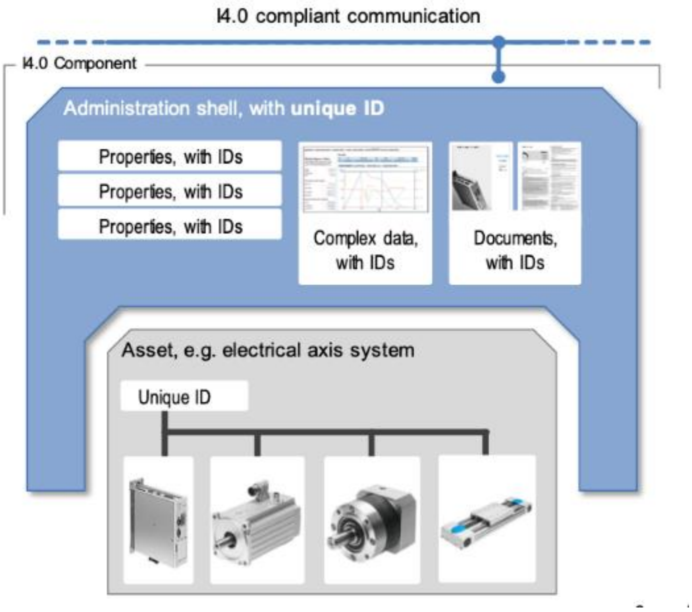


Figure 4 – Asset Administration Shell Representation[5]

The RAMI4.0 model is building the whole digital transformation strategy on the top of the AAS concept, representing a standardized digital representation of the asset. This depicts the corner stone of the interoperability between the applications managing the manufacturing systems. It identifies the Administration Shell and the assets represented by it, holds digital models of various aspects and describes technical functionality exposed by the Administration Shell or respective assets. The key features of the AAS are:

1. Allowing the integration of both non-intelligent and intelligent products in and Industry 4.0 solutions.
2. Supporting the complete lifecycle of products, devices, machines and facilities.
3. Allowing for integrated value chains.
4. Providing the basic infrastructure for the development of autonomous and AI systems.

[5] Plattform Industrie 4.0, "Details of the Asset Administration Shell - Part 1 The exchange of information between partners in the value chain of Industrie 4.0." Federal Ministry for Economic Affairs and Energy, Nov. 2020. [Online]. Available: https://idtwin.org/wp-content/uploads/2021/02/Details_of_the_Asset_Administration_Shell_Part1_V3.pdf

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

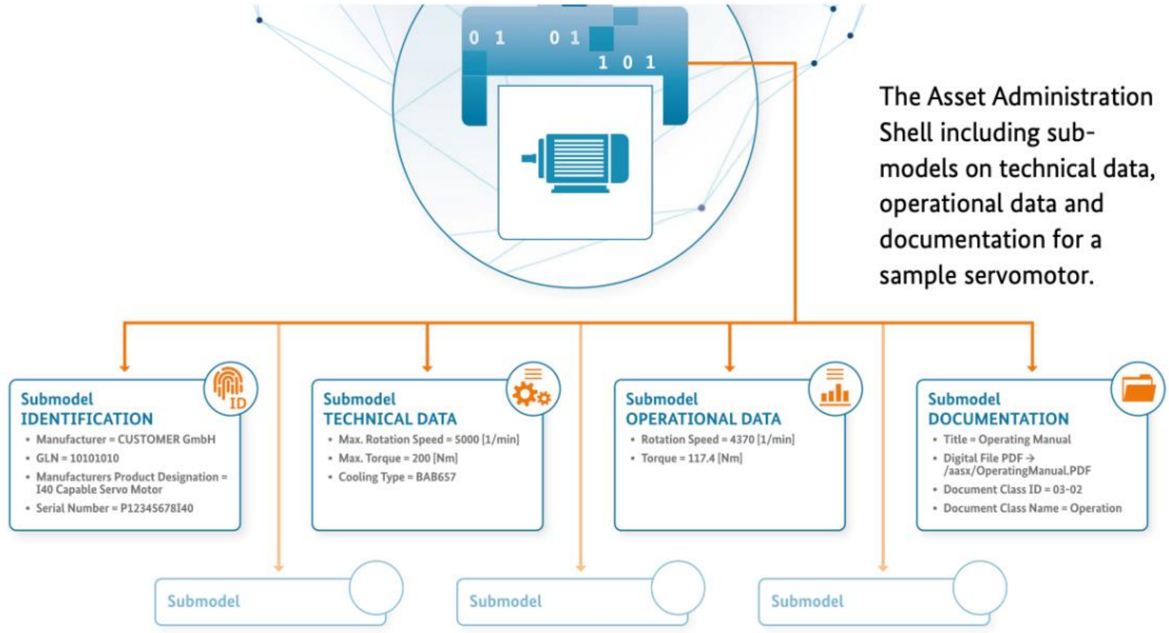


Figure 5 – Asset Administration Shell Structure and submodels composition[6]

In the Industry 4.0 world, each asset is represented by an AAS – its own "Internet presence" or "I4.0 digital twin" describing all the information and functionalities in submodels (see Figure 5).

3.2.3 Industrial Internet Consortium Reference Architecture (IIRA)

The IIRA is a standards-based open architecture for Industrial Internet-of-Things (IIoT) systems[7]. The IIRA maximizes its value by having broad industry applicability to drive interoperability, to map applicable technologies, and to guide technology and standard development. The architecture description and representation are generic and at a high level of abstraction to support the requisite broad industry applicability. The IIRA is the result of an abstraction and distillation process where the Industrial Internet Consortium (IIC) considered defined use cases for extracting common characteristics, features and patterns. It is a technical document that will be refined and revised continually as feedback is gathered from its application in the testbeds developed in IIC as well as real-world deployment of IIoT systems. The IIRA design is also intended to transcend today's available technologies and so can identify technology gaps based on the architectural requirements. This will in turn drive new technology development efforts by the industrial internet community. The IIC outlined the IIRA with a focus on its four key viewpoints (business, usage, functional and implementation), key system characteristics and their assurance (especially safety, security, and resilience) and several cross-cutting concerns (see Figure 6).

[6] Plattform Industrie 4.0, "Details of the Asset Administration Shell - From Idea to Implementation." [Online]. Available: https://www.plattform-i40.de/PI40/Redaktion/EN/Downloads/Publikation/vws-in-detail-presentation.pdf?__blob=publicationFile&v=12

[7] Industrial Internet Consortium, "The Industrial Internet Reference Architecture," Technical Report, 2015. [Online]. Available: www.iiconsortium.org/IIRA.htm

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

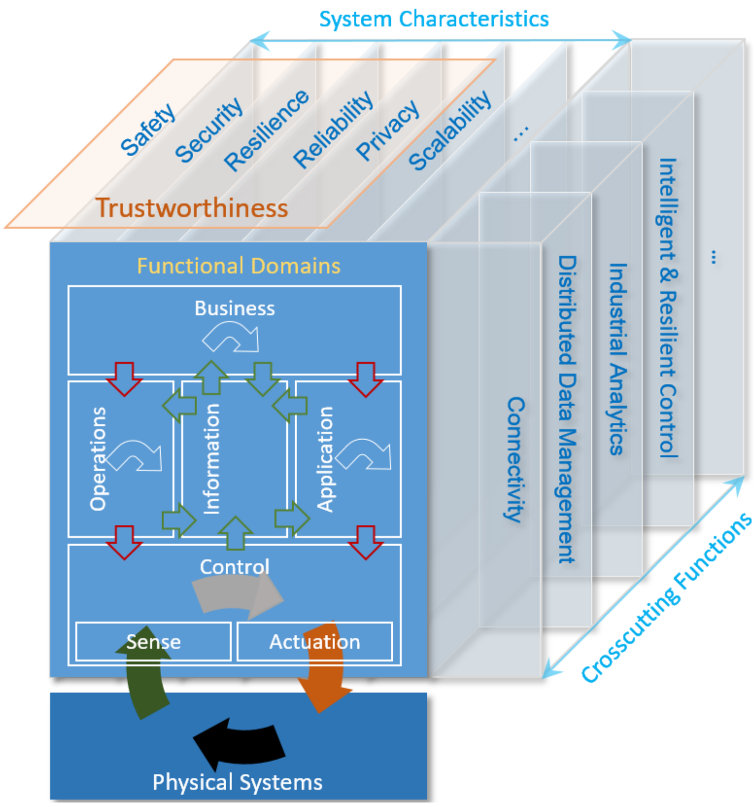


Figure 6 – IIRA Overview: Functional domains, Crosscutting functions and System characteristics

In contrast to RAMI4.0, IIRA crosses multiple application domains.

3.3 BRIDGE Reference Architecture

BRIDGE is an initiative of the European Commission where 64 Horizon 2020 projects, in the context of Smart Grid, Energy Storage, Islands, and Digitalisation, are involved to pool field experience, feedback, and lessons learned, to create a structured view of cross-cutting issues that are encountered in the demonstration projects to help overcome the barriers to effective innovation. In that sense, BRIDGE aims at providing coordinated and coherent recommendations to strengthen the messages and maximize their impacts towards policy makers.

The BRIDGE initiative fosters continuous knowledge sharing amongst projects with the aim to deliver, in a single voice, conclusions and recommendations about the exploitation of results from the different projects through four different Working Groups representing the main areas of interest:

- Data management
- Business models
- Regulations
- Consumer and Citizen engagement.

Within BRIDGE Working Groups, the Working Group on Data Management is the most relevant for the work related to the BD4NRG RA. It is working on:

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

- Communication Infrastructure, embracing the technical and non-technical aspects of the communication infrastructure needed to exchange data and the related requirements, including issues faced by TSO and DSO.
- Cybersecurity and Data Privacy, entailing data integrity, customer privacy and protection.
- Data Handling, including the framework for data exchange and related roles and responsibilities, together with the technical issues supporting the exchange of data in a secure and interoperable manner, and the data analytics techniques for data processing.

After the 2019 BRIDGE General Assembly, Task Forces have been launched to work on specific topics, such as Energy Communities, Replicability / Scalability Analysis, Joint Communication and R&I Priorities.

Horizon2020 smart grid projects under the umbrella of BRIDGE Initiative have investigated the usage of Data Exchange platforms and interoperability of those platforms to define recommendations for a business process agnostic data exchange reference architecture in the energy domain. The results of this activity expect to facilitate cross-sector and cross-border data exchange at European level, ensuring free flow of data among stakeholders, while ensuring GDPR compliance and data owner's control over their data. In view of this, the BRDIGE initiative has analysed eleven other R&D projects by gathering general information on the projects through surveys and requesting specific information on organisational, informational, and technical aspects of the projects. A subsequent mapping of results to the smart grid architecture model (SGAM)[8] interoperability layers results in the BRDIGE RA. These layers are:

- Business Layer, which represents the business view on the information exchange related to smart grids.
- Function Layer, which describes functions and services including their relationships from an architectural viewpoint.
- Information Layer, which describes the information that is being used and exchanged between functions, services and components. It contains information objects and the underlying canonical data models. that represent the common semantics for functions and services to allow an interoperable information exchange via communication means.
- Communication Layer, which describe protocols and mechanisms for the interoperable exchange of information between components in the context of the underlying use case, function or service and related information objects or data models.
- Component Layer, which includes system actors, applications, power system equipment, protection and tele-control devices, network infrastructure (wired / wireless communication connections, routers, switches, servers), and any kind of computers.

BRIDGE is looking only at one of the three dimensions of SGAM, the dimension of the five layers above described, sometimes called interoperability layers, as shown in the BRIDGE Reference Architecture in

[8] https://ec.europa.eu/energy/sites/ener/files/documents/xpert_group1_reference_architecture.pdf

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Figure 7[9]. The top part of the architecture shows the other scales: electricity domain on the left, other domains (gas, water, transport, health, etc) on the right side, and cross-sector in the middle.

The idea is to bring together the input from individual domains (mainly electricity but also other domains). In the cross-sector domain section, the commonalities of the different domains, essentially related to the data management, are placed. The energy sector is comprised of multiple vectors, mainly electricity, natural gas and district heating utilities. The possible cross-vector integration within the energy sector is being considered by the BRIDGE initiative. In view of this, dependencies among the vectors in the energy sector, due to seasonality and variability effects, need to be taken into consideration.

As depicted in Figure 7, each of the interoperability layers has been divided into sub-layers and for each sub-layer some examples are provided.

The Business Layer is split into:

- Regulation
- Associations
- Business Roles
- Business Processes.

The Function Layer includes functional processes, such as grid modelling, flexibility management.

The Information Layer is split into:

- Information Models, ontologies
- Profiles, data models.

The Communication Layer is split into:

- Data Formats
- Protocols

The Component Layer is split into:

- Data exchange platforms
- Applications
- Hardware.

On the right side of the architecture, there are recommendations provided for each layer and sub-layer. Data source integration is the key feature of the utilities; thus, it is essential to identify interdependencies and data exchange flows among these utilities. Moreover, the architecture integrates other sectors (gas, heating, water, etc.).

9 https://ec.europa.eu/energy/sites/default/files/documents/bridge_wg_data_management_eu_reference_architcture_report_2020-2021.pdf

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

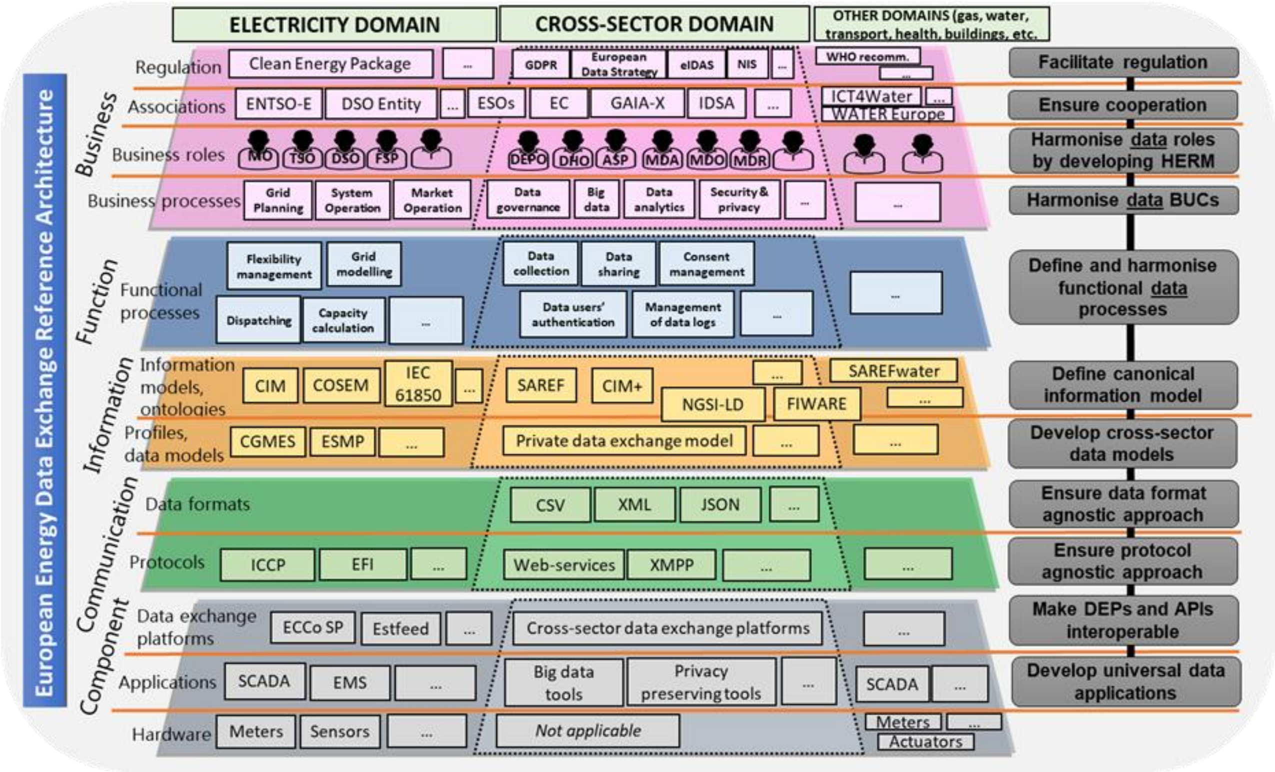


Figure 7 – The BRIDGE Reference Architecture

From the architecture models presented in this document the BRIDGE RA offers the most complete view of use cases in the electricity sector as well as cross-sectorial use cases. It extends the widely accepted SGAM. This makes it a suitable model for expressing the BD4NRG RA.

3.4 Mapping of BD4NRG Use Cases and WP2 efforts to the BRIDGE RA

Since the BRIDGE reference architecture as described in Section 3.3 is well-understood by the project partners, individual use cases are mapped to a variation of it in Appendix A. The aim is to help identify and conceptualize the BD4NRG RA blocks by providing an overview at a glance. Specifically, an understanding of the individual use cases' relation to BRIDGE will help in tailoring the envisioned reference architecture to the needs of BD4NRG.

However, this mapping is preliminary and subject to change. The main sources for these mappings are the use case descriptions from deliverable D2.1 and related pilot data documents. Information that is currently unknown is denoted by question marks. In the following the diagram structure is described. Refer to Figure 8 depicting LSP 7 UC 1 for an example of the mapping.

As this model mostly follows the BRIDGE RA, it is also divided into the three domains of "electricity", "cross-sector" and "other domains". The electricity sector encompasses energy generation, transmission, distribution and consumption. An example for a cross-sector application are electrical vehicles (EVs) which cover both the electricity and the mobility sectors. Other domains could be for instance the water sector. As for the horizontal layers, they are the same five as in the BRIDGE RA. The Business Layer is concerned with the (type of) organizational entities which conduct, support or regulate the given LSP. The Function Layer is about high-level services the LSP offers. The Information

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Layer deals with the representation of information. Information models are protocol-agnostic and focus on abstract entities and their relationships whereas data models are implementation-oriented. The Communication Layer includes data representation (Data Formats Sublayer) and data exchange protocols (Protocols Sublayer). The Component Layer is split into the sublayer that enumerates the actual hardware and the Applications Sublayer that lists software blocks that provide isolated service units. In a microservice architecture each of these blocks would correspond to a service.

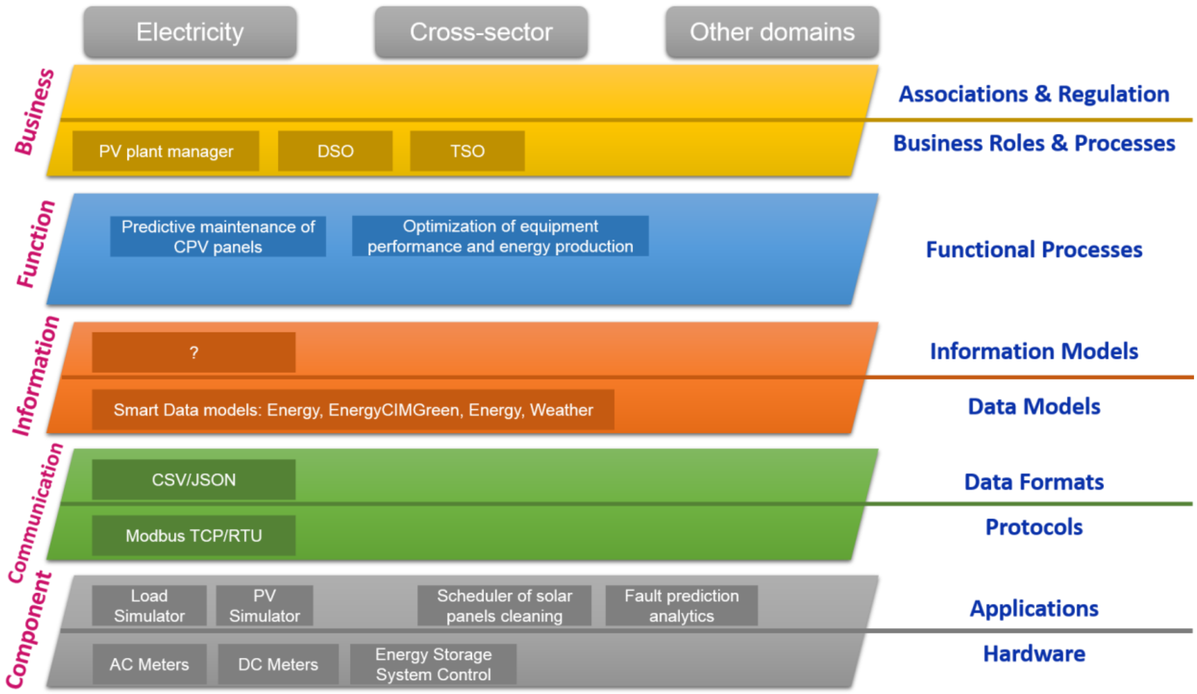


Figure 8 – Example mapping of LSP 7 UC1 to BD4NRG Reference Architecture

Besides the individual LSP use cases, also the WP2 deliverables D2.1 - D2.4 are mapped to the BRIDGE RA. This allows to see how the individual efforts contribute to the goal of reaching a common BD4NRG RA and how these efforts relate to each other. The mapping is depicted in Figure 9.

D2.1 *“Big Data Value Chain and Requirements”* identifies and presents the characteristics and functionalities of the BD4NRG LSPs and the respective use cases and defines their requirements. It therefore provides an overview of the involved hardware components and the envisioned functionality to be implemented during the project. D2.2 “*BD4NRG Open APIs and Data Assets Functional Specifications”* describes the three BD4NRG layers Data Governance, Analytics and Marketplace regarding their functional specifications and structure. This includes a list of the software models to be implemented and their interfaces. D2.3 *“BD4NRG Semantification & Interoperability”* concerns the data collected and used by the BD4NRG LSPs. The deliverable consolidates the provided data into clusters and analyses it regarding possibly useful data models. Hence, D2.3 is mostly concerned with the Information Layer of the BRDIGE RA. D2.4 *“Security, Privacy, Standards and Regulatory Compliance Specs”* defines requirements and compliance best practices regarding security and privacy of the system developed within BD4NRG. As a result, the content of D2.4 provides relevant information for the Business Layer.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

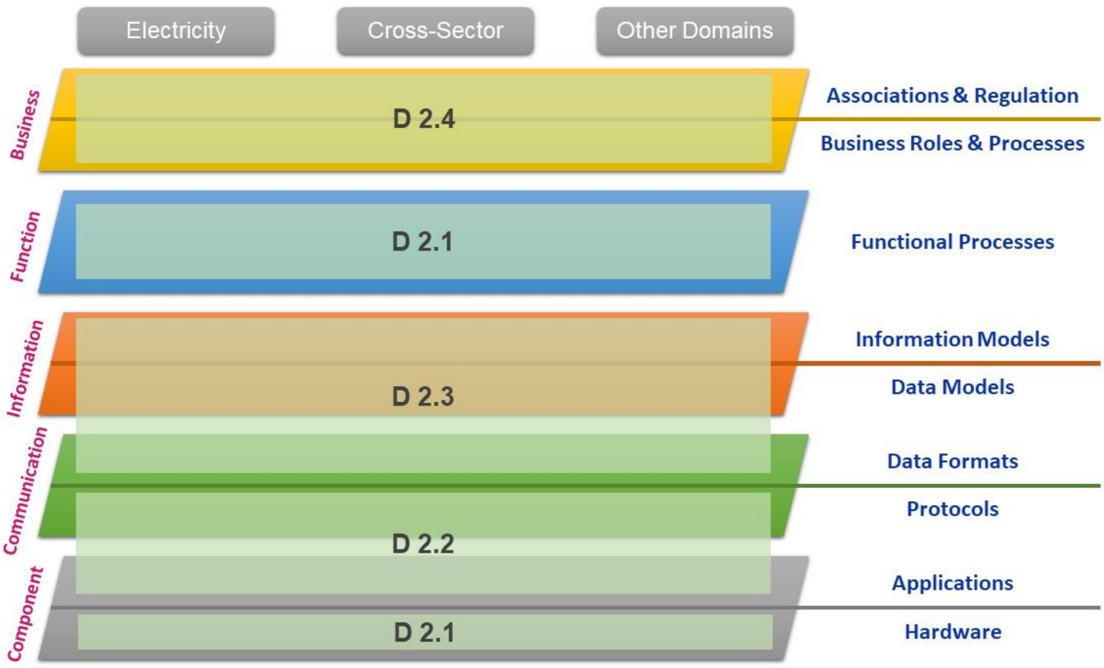


Figure 9 – Mapping of WP2 Deliverables to the BRIDGE Architecture Model

More information on the different efforts in BD4NRG and how they converge to the BD4NRG RA is provided in Chapter 5.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

4 Data Space Reference Architectures

The topic of Data Spaces is of great importance for the BD4NRG project. Data Spaces are concerned with the sharing of data across different stakeholders in a secure and sovereign way. Hence, they enable participants to create value from the data they offer or consume.

In the following we first present high-level architectures that have already been created by project coalitions and initiatives. Within BD4NRG these can serve as starting points, but also as dissemination opportunities for the contributions made in the project.

In a next step, this chapter presents concrete building blocks for the implementation of Data Spaces coming from IDSA and FIWARE. Also, a potential integration is discussed. Chapter 5 explains how these building blocks are integrated in the BD4NRG RA.

4.1 High Level Architectures

4.1.1 OPEN DEI

One of the main requirements for the digital transformation is the design and development of common data platforms based on unified architectures and established standards. The OPEN DEI project (Aligning Reference Architectures, Open Platforms and Large-Scale Pilots in Digitising European Industry) makes use of "Platforms and Pilots" to support the implementation of next generation digital platforms in four basic industrial domains: Manufacturing, Agriculture, Energy, Healthcare.

OPEN DEI[10] started in mid 2019 and will end in mid 2022. It aims at proving an important contribution to the implementation of Digitising European Industry policies: the project focuses on the necessary measures, channels and mechanisms to ensure cooperation between pilot projects so as to exploit synergies, share knowledge and maximise the impact.

All the research activities and work inside the OPEN DEI project are directed into the following topics:

- Platform building: Comparing reference architectures and open-source reference implementations, enabling a unified industrial data platform.
- Large scale piloting: Contributing to a digital maturity model, creating a set of assessment methods and a migration journey benchmarking tool.
- Data ecosystem building: Enabling an innovation and collaboration platform, forging a European network of DIHs, contributing to industrial skills catalogue and observatory.
- Standardization: Conducting cross-domain surveys, performing promotion and implementation, building strives alliances with existing EU and standard developing organizations.

One of the most important objectives of the OPEN DEI project is to harmonise and coordinate different Digital Transformation (DT) approaches under a common Reference Architecture Framework (RAF), which tries to combine knowledge and tools to assess and share experiences and lessons learned on how systems supporting DT can be architected, crossing the boundaries of specific applicative sectors. OPEN DEI proposes a conceptual model for integrated data-driven services for DT guidelines, to

[10] https://www.opendei.eu/

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

support their planning, development, operation and maintenance by adopting organizations. The model is modular and includes loosely coupled service components interconnected through a shared common data infrastructure. The OPEN DEI project has defined the approach for designing a common RAF able to describe the Cross Domain Digital Transformation.

A common scenario in the implementation of many DT solutions and industrial sectors is the extensive use of sensors and interconnected devices. The availability of this huge amount of information coming from different data sources can cover many business applications and scenarios. Data-driven pipelines and workflows management is crucial for data gathering, processing, and decision support. To handle this complexity, OPEN DEI has adopted the so called “6C Architecture”, adapted from the one suggested by the German Industrie 4.0 initiative[11], which is based on:

- Connection: sensors & networks
- Cyber: model & memory
- Computing: edge/cloud and data on demand
- Content/Context: meaning & correlation
- Community: sharing & collaboration
- Customisation: personalisation & value

The six levels mean connection of different data sources and networks, cyber modelling mechanisms, computing of data (in the edge or in the cloud), content/context correlation, community engagement, and customization tailored to the user perspective and expectations. The “6C Architecture” principles have led to the design of the OPEN DEI RAF, developed around the main concept of Data Spaces in which data are published, accessed and shared. With reference to Figure 10, three main different layers have been identified:

- Cloud Level Data Spaces, which includes data storage, data integration and data intelligence operations on the cloud. The cloud services will process Big Data, deploy algorithms, integrate different source platforms and services, provide advanced services such as AI prediction and reasoning.
- Edge Level Data Spaces, which defines the typical edge operations from the data acquisition (from the logical perspective) to the data processing through the data brokering. The edge services will play a key role also for data analytics (i.e., validating and improving models for data analysis).
- Field Level Data Spaces, which includes the Smart World Services able to collect data and support the interaction with the IoT Systems (configuration, calibration, data acquisition, actuation, etc.), Automation and Smart Assets (robots, machinery, and related operations) and Human Systems (manual operations, supervision, and control, etc.).

[11] https://www.plattform-i40.de/PI40/Navigation/EN/Home/home.html

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

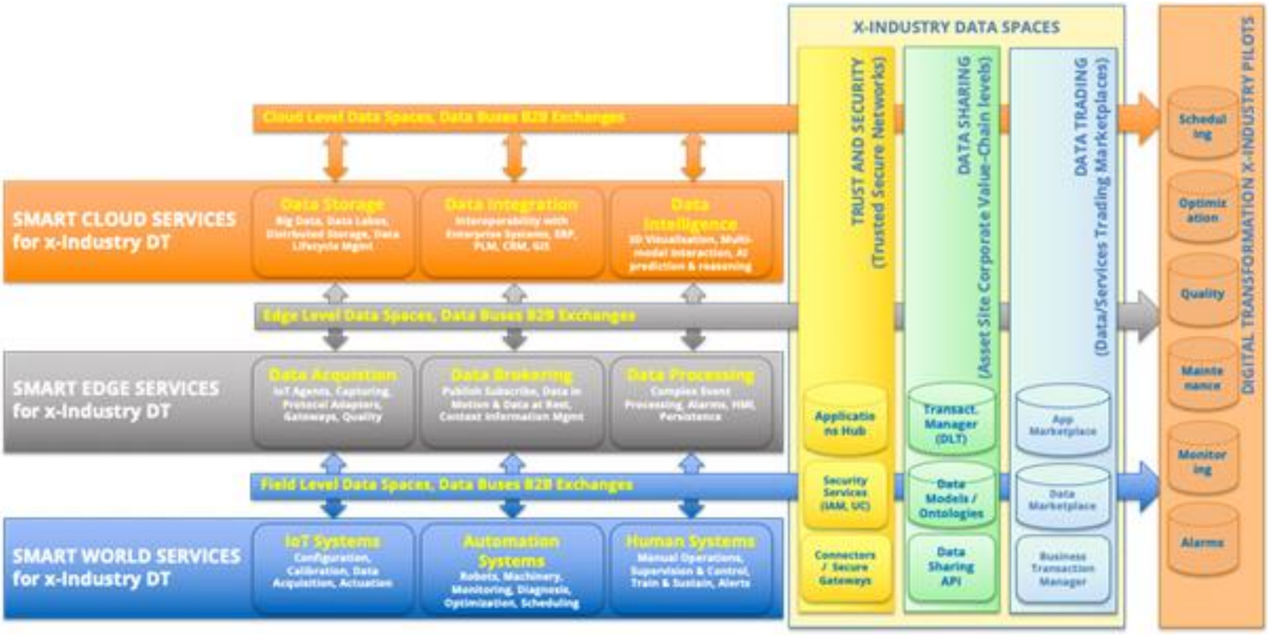


Figure 10 – OPEN DEI Reference Architecture Framework

All these horizontal Data Spaces layers will feed the OPEN DEI RAF in the main orthogonal dimension, named "X-Industry Data Spaces", characterized by following components:

- Trust and Security, incorporating technical frameworks and infrastructures that complements the previous to support trusted and secure exchange, which includes:
  - Applications Hub, an infrastructure which collects the recipes required for the provision of applications (e.g., deployment, configuration and activation) in a manner that related data access/usage control policies can be enforced.
  - Security Services, a technical framework to support Identity Access Management, Usage Control and other security services.
  - Connectors and Secure Gateways, a technical framework for trusted connection among involved parties.
- Data Sharing, incorporating technical frameworks and infrastructures for an effective and auditable data sharing, which more specifically includes:
  - Transaction Manager, a distributed ledger/blockchain infrastructure for logging selected data sharing transactions.
  - Data Models and Ontologies, to leverage common standard and information representations.
  - Data Sharing API, a technical framework for effective data sharing.
- Data Trading, incorporating technical frameworks and infrastructures for the trading (offering, monetization) of data, which includes:
  - App Marketplace, enabling the offering of applications and application building blocks which can be integrated plug & play to enrich existing data spaces.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

- Data Marketplace, enabling the offerings around data resources with associated terms and conditions including data usage/access control policies as well as pricing schemas.
- Business Support Functions, enabling data/applications usage accounting as well as implementing Clearing House, Payment and Billing functions.

All the mentioned layers serve the realization of Digital Transformation X-Industry Pilots, for enabling applications (sometimes sector specific) for supporting business scenarios from experiments.

The OPEN DEI RAF may drive the definition of the BD4NRG Data Space, providing a set of common models, principles and recommendations, to gather data and process them doing with the purpose of reaching the DT using a cross-sector mindset, at the same time ensuring the trust and secure behaviour during the exchange.

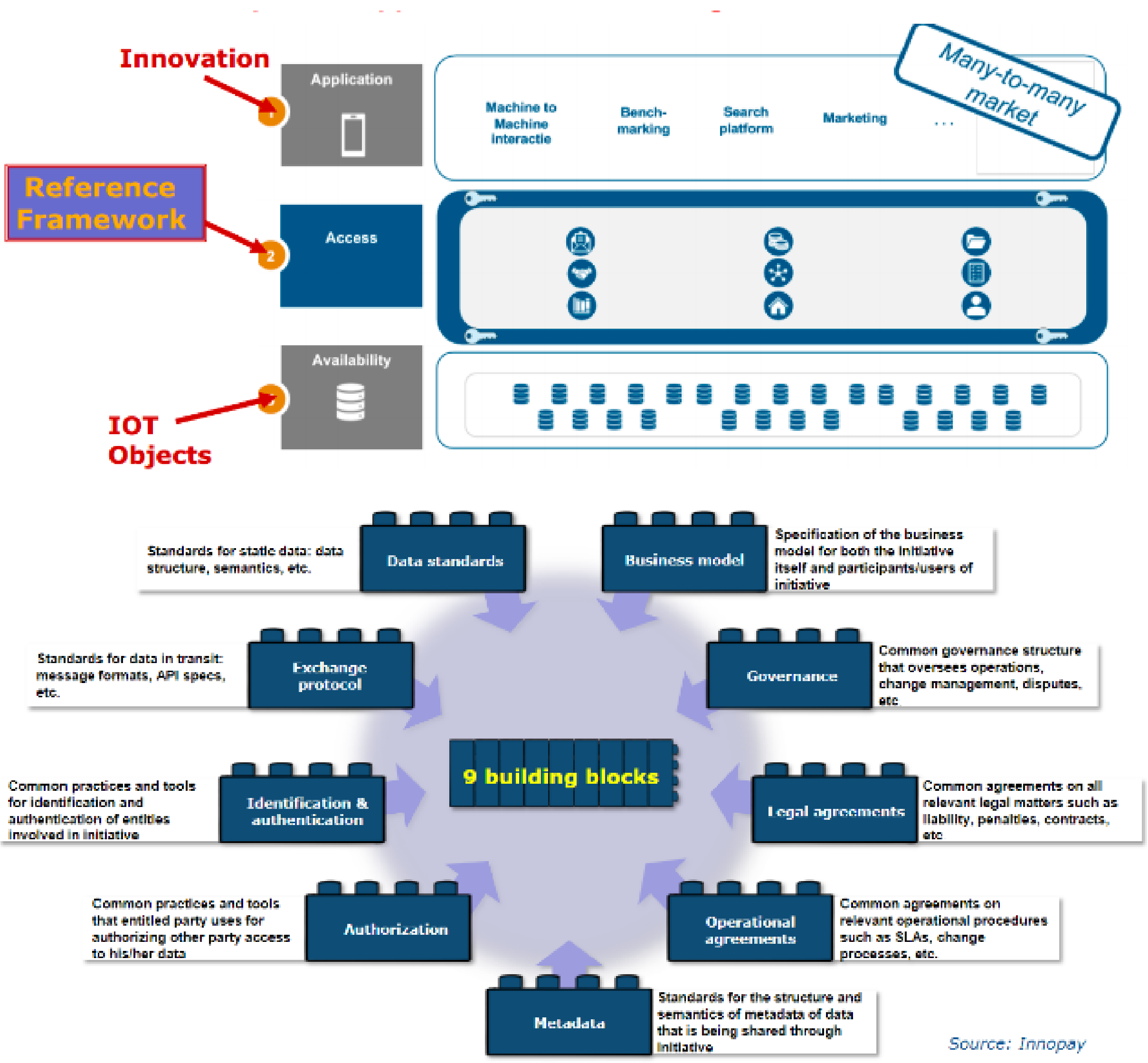


Figure 11 – Innopay Conceptional Architecture for Data Sharing and the underlying Building Blocks

Moreover, in the last years OpenDEI has set up a cross-fertilization effort at the interplay among ICT and Energy, where the new Working Group on Data Framework Architecture has been created under the auspices of EC DG CONNECT and DG Energy. The main objective of this working group, which is currently chaired by Dr. Massimo Bertoncini (ENG), is to understand how and to a what extent ongoing Smart Energy grids projects may converge towards the Innopay Data Conceptional Architecture, based on nine building blocks (see Figure 11).

4.1.2 BDVA/DAIRO

The emergent Data Economy relies on the availability of data as the basis for further innovation and exponential development of technologies. Data is the new currency underpinning the new data economy and catalysing the European economy for a faster and more effective development.

The European contractual Public Private Partnership on Big Data Value (BDV PPP) was signed in October 2014, marking the commitment of the European Commission, industry, and academia to build a data-driven economy across Europe, creating a competitive advantage for European industry, stimulating economic growth and job opportunities. Its goal is to enable the creation of a European Big Data Value economy, supporting, boosting, and mobilising stakeholders throughout society, industry, academia, and research, and delivering services and technologies, while providing highly skilled data engineers, scientists and practitioners along the entire Big Data Value chain.

The Big Data Value Association (BDVA) is the private counterpart to the EU Commission in implementing the BDV PPP programme. BDVA is an industry-driven international not-for-profit organisation, whose members from all over Europe represent large, small, and medium-sized industries, research, and user organizations.

The mission of the BDVA is to develop the Innovation Ecosystem that will enable the data-driven digital transformation in Europe delivering maximum economic and societal benefit and achieving and sustaining Europe’s leadership on Big Data Value creation and Artificial Intelligence.

A successful implementation and adoption of a pan-European data sharing space mark a milestone in the evolution of the new data economy. Macro-level space can incorporate existing vertical, cross-sectoral, personal, and industrial data spaces, offering services and opportunities for experimentation to all involved stakeholders.

Emerging data ecosystems enable large-scale data to be securely connected, valorised, and shared through the following three complementary technologies:

i. Data Spaces: any ecosystem including data storage, data models, datasets, ontologies, lifecycle management platforms, protocols, data sharing contracts and specialised management services together with soft competencies around it (for example, governance, social interactions, business processes), aiming to optimize data storage and exchange mechanisms, while preserving, generating and sharing new knowledge.
ii. Data Platforms: next generation data acquisition and processing platforms.
iii. Data Marketplaces: enabler of the Data Economy that in analogy with well-known product-service e-commerce portals offer a one-stop-shop where data providers (sellers) and data consumers (buyers) can meet, match, and trade their respective data assets and requirements, as well as data algorithms and processing approaches. Data Marketplaces addresses the disconnected and non-interoperable nature of multiple existing data spaces, allowing to maximize data value extraction. Data Marketplaces operate on top of Data Spaces and utilize Data Sharing and Data Exchange apps provided by Data Platforms to allow the interoperability and trading of data between different subjects, often but not

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

exclusively as B2B. Data are commercialized by using Open Data, Monetised Data and, Trusted Data sharing mechanisms.[12]

The realisation of a successful European-governed data sharing space able to generate economic value by broadening data access for AI, relies on a clear plan of iterative implementation strategies and on a coordinated effort of all relevant stakeholders. The key concept of trust has a central role in successful data sharing activities: from the validity of the data itself and the algorithms operating on it to the entities governing the data space, to the enabling technologies, as well as amongst the wide variety of users.

The BDVA identifies two additional opportunities that can materialise as a result of timely interventions to converge data sharing initiatives in Europe and realise its vision:

i. Achieve wider access to data to realise the full potential of emerging AI technology by designing and implementing of a common, trustworthy, decentralised data space able to enable secure and democratic data sharing, while boosting the European data economy.
ii. Achieve a European-governed data space, giving Europe a prominent role in steering international efforts to develop data and AI solutions reflecting and respecting European ethical values including democracy, privacy protection, and equality.[13] This vision for a European-governed data sharing space should be built around and in collaboration with the same wide variety of stakeholders that can exploit its benefits as users.

The BDVA position assert that data spaces, platforms, and marketplaces are key enablers to unlock the potential of Big Data. This position is in line with the European Commission's plan to promote the development and use of "made in Europe" AI technology, supporting the call for a joint action of Member States to establish "common European data space to make data sharing across borders seamless" and thus enable "large, secure and robust datasets" for AI technology to be developed.

Several challenges must be addressed to realize this vision:

- new business models for the industry for fata-driven AI-based solutions.
- trustworthiness in AI and its results.
- implementation of an AI and Big Data ecosystem, overcoming the lack of data interoperability.
- integration and development of several technologies, such as advanced data analytics, distributed AI, and hardware optimized for AI.

The BDVA Reference Model (European BDVA Strategic Research and Innovation Agenda v4.0, 2017[14]) is a reference framework made by the BDVA that describes logical components of a generic Big Data (BD) system. As shown Figure 12, the BDVA Reference Model is structured into horizontal and vertical concerns: horizontal concerns cover specific aspects along the data processing chain, while vertical concerns address cross-cutting issues, which may affect all the horizontal concerns and may also involve non-technical aspects. Section 3.2 of the deliverable D2.1 provides details on the BDVA Reference Model.

[12] https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpace%20PositionPaper_April2019_V1.pdf
[13]https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpaces%20PositionPaper%20V2_2020_Final.pdf
[14] https://bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

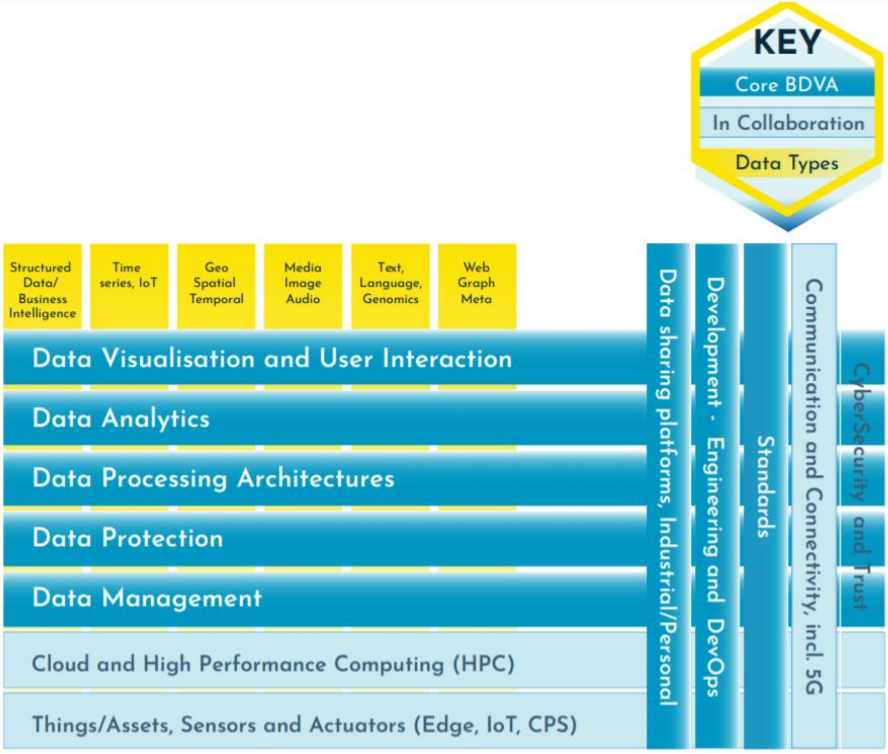


Figure 12 – BDVA Reference Model

In 2020, BDVA was strengthen by the members, with a new mandate, expanding its scope and activities. In 2021, BDVA assumed a new name to testify the ambition of the Association to collaborate closely with other communities at the intersection of the key disciplines of Data, AI and Robotics; thus, BDVA became DAIRO.

4.1.3 AIOTI

In the following, the High Level IoT architecture, release 5.0 considerations of the use of ISO/IEC/IEEE 42010 by AIOTI is presented and analysed for the purposes of the BD4NRG. In summary it

- Presents a Domain Model and discusses the "thing" in IoT
- Presents a Functional Model (Figure 13)
- Introduces the Identifiers for IoT
- Provides deployment considerations related to relevant IoT architectural matters such as cloud and edge computing, Big Data, virtualization, security, privacy and (platform) interoperability
- Links this work with the AIOTI WG03 Semantic Interoperability work and the Standard Development Organizations (SDOs) Landscape work
- Provides mapping examples to some existing SDO/Alliances' architectural work related to functional models: ITU-T, oneM2M, BDVA
- Establishes the link to other architectures and frameworks such as Big Data and IoT-enabled Data Marketplaces

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

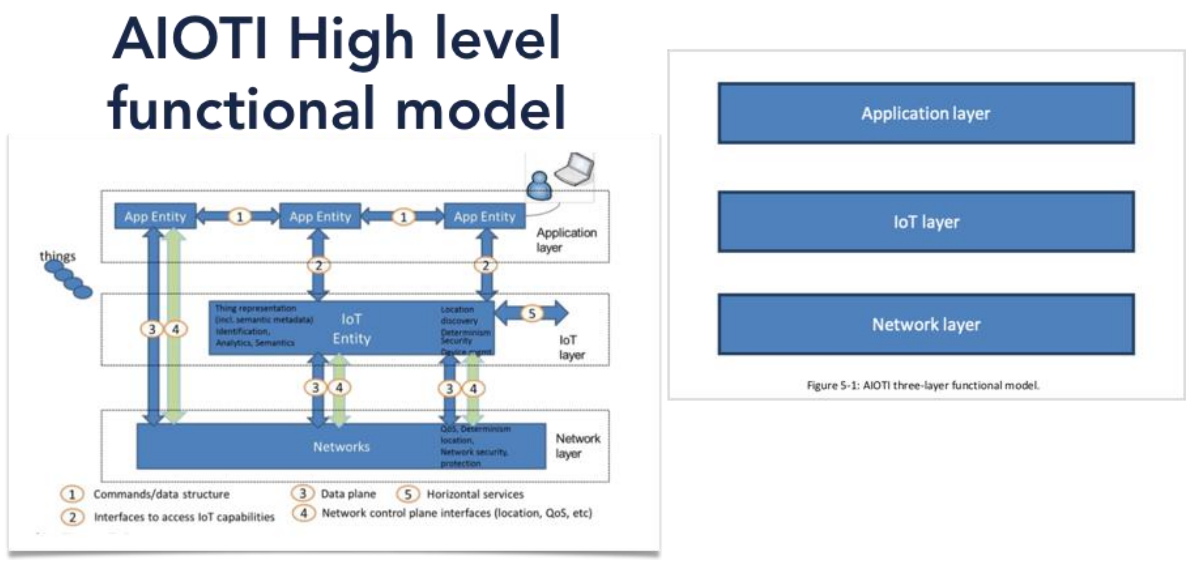


Figure 13 – AIOTI High-level Functional Model[15]

AIOTI HLA is typically deployed using cloud infrastructures. Cloud-native principles can be applied to ensure scaling and resilience for IoT. In certain use cases, deploying edge cloud infrastructures, will be beneficial to allow data processing locally. AIOTI designed its HLA to allow for distributed intelligence, it is therefore compatible with Cloud and Edge computing. It has the following objectives:

- Big Data: collecting, storing and sharing data is an integral part of IoT, therefore also for AIOTI HLA. Big Data can be seen as the set of disciplines, such as storing, analysing, querying and visualization of large data sets. Those disciplines are equally applicable to IoT data sets.
- Virtualization: ensuring flexibility and scale is one of the major challenges for deploying IoT. Virtualization would help scaling IoT for a large number of use cases.
- The 3D IoT Layered Architecture (aka the 3D model), is an approach to define, identify and co-relate multiple IoT system features, architectural characteristics and properties in Large Scale pilot (LSP) IoT systems.
- The principle of this Reference Architecture is to use a number of 2D views that are a projection of the 3D view on a specific plane. The Figure 13 depicts the three main views in the 3D Model (Layers, Cross-cutting functions, and Properties)
- The layers view in the 3D model, see Figure 14, characterises IoT Systems from a functional and operational perspective. It includes aspects from physical devices, networking, cloud infrastructures, data, services and applications but also collaboration and administration. The main aim of this view is to facilitate the identification of the necessary functional blocks for interoperability at the different layers in IoT systems.

[15] Source https://aioti.eu/wp-content/uploads/2020/12/AIOTI_HLA_R5_201221_Published.pdf

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

- Other objectives include extension to more than the functional layers, include extension to AI, DLT, Digital Twins. Among other objectives is depiction of physical, digital virtual and cyber integrations. It allows for implementations of data, information, knowledge continuums from physical nodes, deep edge, edge, cloud, data centre. Mapping of other reference architectures could be facilitated by this depiction as well.

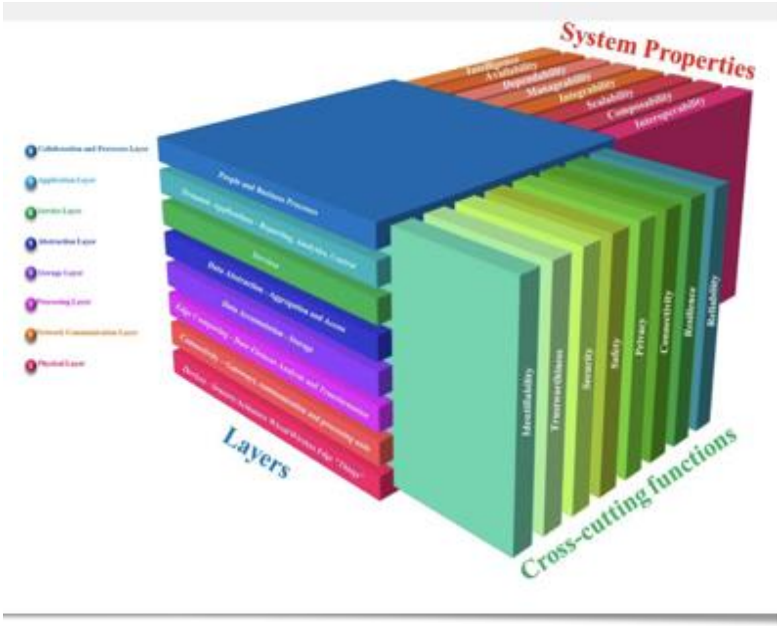


Figure 14 – AIOTI 3D HLA

- The Figure 15 depicts the layers components, cross-cutting functions example and properties mapping.

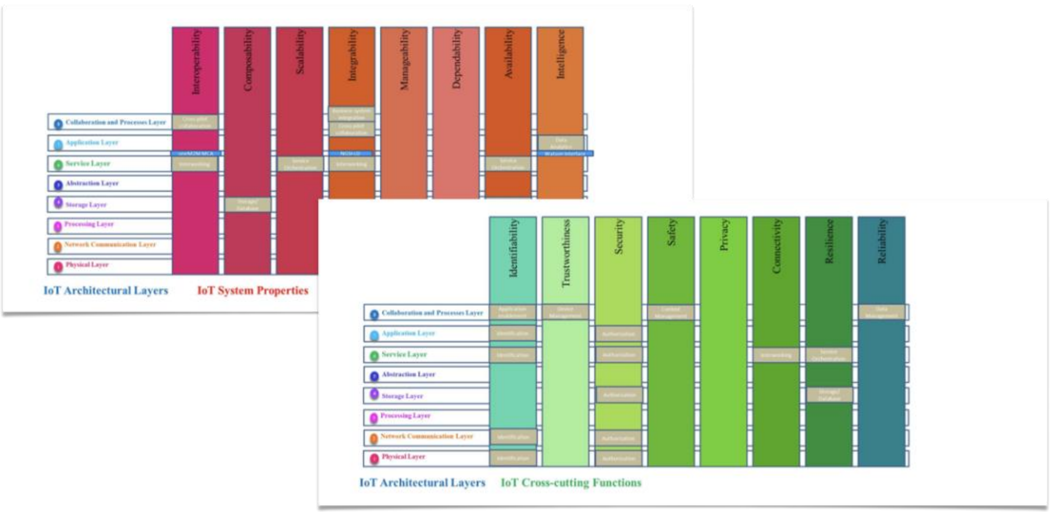


Figure 15 – Layers view of 3D

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

4.1.3.1 AIOTI Marketplace

As presented in the AIOTI Priorities Vision document 2019-2024[16], the main directions identified towards multidirectional data share and services development are along the following areas - leveraging IoT data, enabling cross-sectoral "data marketplaces", providing true interoperability, lowering regulation barriers, etc. to unlock the full potential of innovation IoT-based applications. This includes establishing a cyber-security strategy for safeguarding IoT technology and applications, whilst ensuring privacy by design.

AIOTI emphasises an ecosystem approach towards IoT-enabled Data Marketplaces development. It is recognised that the transformative journey from building the enabling infrastructures, which form a base towards local enablement to cross-domain marketplaces is underway across many domains and geographies. The ecosystem is fundamental for unlocking cross-domain data marketplaces and the related position paper provides the context and specific recommendations to guide the next steps across Europe.

The four stages of IoT Marketplaces presented in Figure 16 link the related technologies and challenges to each stage of infrastructure integration as well as business goals. Each stage links the related enabling technology to the stage of infrastructure and business model development. The physical layer of smart energy components is the first building block in cross-domain data valorisation.

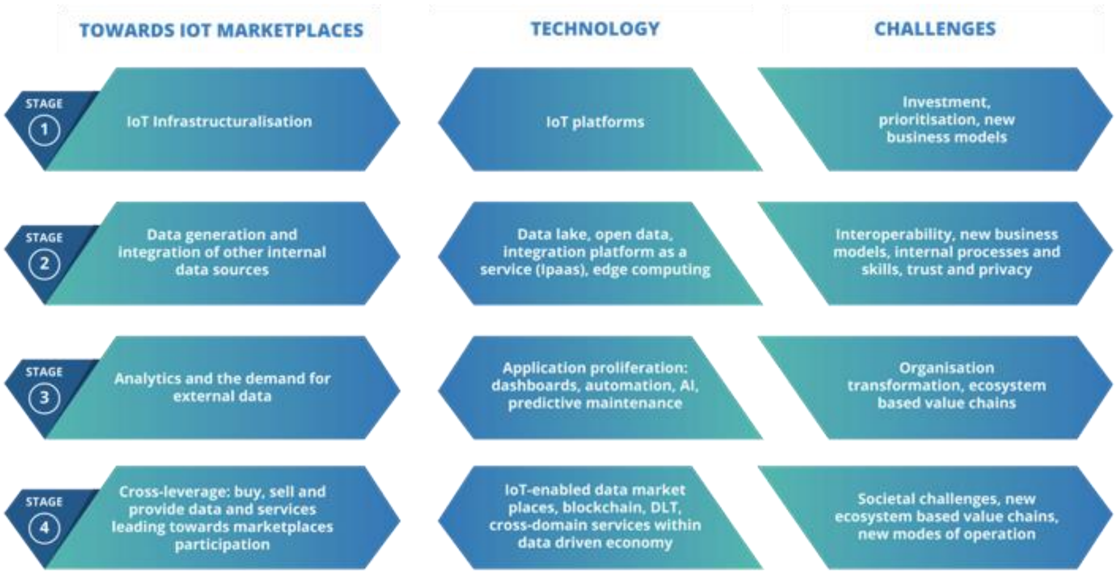


Figure 16 – The four stages of IoT Marketplaces

Main Concepts

The paper identified several general dimensions to activate marketplaces within the energy and cross-domain development relevant for BD4NRG:

[16] https://aioti.eu/wp-content/uploads/2019/09/AIOTI-Priorities-2019-2024-Digital.pdf

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

- Ensuring that the market design is fit to send the appropriate market signals for consumers to adopt and engage with these innovations – consideration for BD4NRG consider built-in incentives approach
- User experience and interoperability are key to scale-up the innovations - consideration for BD4NRG bottom-up from end use design and interoperability approach
- Fostering an innovation-friendly culture – consideration for BD4NRG agile innovation approach
- Public sector participation as an active energy prosumer through publicly owned buildings and sites that may contribute to fostering open energy marketplaces - consideration for BD4NRG non energy stakeholders' engagement approach
- The enabling technologies are at high TRL level, dynamically evolving - consideration for BD4NRG – edge solutions are ready for experimentations
- Governance is key for scaling up and integrations with market actors and grid operators - consideration for BD4NRG - governance by design and data spaces.

It is stated that marketplaces design and operations principles should ensure sustainability of the overall solution, as well as the ecosystem overall, security of supply within the energy sector (referencing integration of consumers) and competitiveness imperatives reflecting decentralisation drive.

4.1.3.2 Principles

The evolving bidirectional marketplaces enabled by communication platforms that expect to give rise to well-orchestrated multisided platforms will require careful considerations regarding operational principles.[17] The following are some principles for consideration in BD4NRG.

Main pillars

- Local to national to cross-border
- Cross-sector
- Integrating various scales and actors (TSO-DSO-aggregators-consumers)
- Integrating physical reality with market rules

Main drivers

- Coordination and orchestration Optimization
- Transparency
- Interoperability

Principles

- Trusted parties can access the resources
- Share data and visibility

[17] https://aioti.eu/wp-content/uploads/2021/03/Open-Energy-Marketplaces-Evolution-Published.pdf

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

In addition, the following considerations are of relevance.

- Operation and market perspective considerations as a holistic approach.
- An integrated system approach should be a shared vision, to assess and reduce the risk of market fragmentation.
- An integrated electricity system approach is the basis, in which all actors' roles and responsibilities, from local to cross-border, are recognized.
- Roles and responsibilities should be defined and shared, including clarity on market facilitator.
- A consistent model should be established to allow stacking value of flexibility across markets.

The position paper[18] further defines services and architectures principles, the following considerations can be taken into the account for architecture design purposes and governance related topics.

- Market processes should have sufficient coordination functions between them for economic efficiency and security of supply's sake, especially when the same assets can provide different services to different market processes, and when timeframes overlap. In that sense, fully separated market processes should be avoided.
- Liberalisation of energy markets and integration of renewables into the electricity distribution grid implies participation of new incumbents, e.g., distributed generation owners, storage providers, aggregators, or consumers participating in demand flexibility programs, in addition to historical operators in the energy sector. Planning processes and transactions between integrated utilities are being replaced by multi-stakeholder contractual relations, requiring information exchange across the different stages of the electricity supply chain.
- Furthermore, for energy marketplaces to stimulate innovation, new incumbents should have open access to data. At the same time, security and privacy requirements must be met to ensure sustainability of smart grids. Data sovereignty, i.e., rights of consumers to keep the ownership over individual data should be observed.

4.1.3.3 Governance Concepts

Defining governance of information management within energy marketplaces is identified as an imperative task. The paper further states that *"Governance of energy marketplaces is understood as a set of rules defining how data in smart grids is produced, stored, aggregated, verified, distributed and used for decision-making and business purposes. It encompasses physical flow of data, management of data exchange, roles and responsibilities of stakeholders, IT system and electricity infrastructure design."*[19]

The gaps in governance models within the implemented projects in a number of European countries identified in the areas of hyper local solutions, such as microgrids and peer to peer markets. *"However, the scale of regulated governance models may be ill-adapted for popping-up microgrid projects*

[18] https://aioti.eu/wp-content/uploads/2021/03/Open-Energy-Marketplaces-Evolution-Published.pdf
[19] https://aioti.eu/wp-content/uploads/2021/03/Open-Energy-Marketplaces-Evolution-Published.pdf

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

functioning as hyper-local markets of peer-to-peer energy exchange (also known as energy communities)."

The concept of *"collective self-governance rules such as reputation, transparency and accountability shall replace trusted intermediary services offered by centralised IT platforms"* could be considered for the BD4NRG toolbox.

In summary, the mechanisms for multi stakeholder approaches are essentially neutral and non-discriminatory ways of exchanging data while overcoming identified barriers, such as lack of data and means of observability, for example in case of certificates of origin *" easily accessible settlement data to service providers and regulated entities in charge of services should provide basic means to exchange data with national registries (example GOs). "*

Furthermore, *" instruments/funding programs etc for designing, testing and verifying business models (in the form of regulatory sandboxes while regulations are not clear) for smart energy communities "* platforms and enabling solutions and services, including platforms.

4.1.3.4 Future Pathways for Marketplaces Enablement

The following set of requirements identified in the chapter on Customer Centric Marketplaces could be considered for BD4NRG toolbox solutions supporting peer-to- peer energy marketplaces:

Openness principles: platforms should enable the development of new collaborative applications by preventing the privatization of data storage and management

Federated identity management: users should be able to interact with multiple centralized or decentralized services within and outside of the platform

Authorised Data access: personal data remains under the control of their respective owners and is available to community or to third parties on demand

Privacy-preserving management (indexing, aggregator & analytics): If a user (an individual or company) requests operations that need to access individual data, for example to compute statistical information, the platform executes in a decentralized and privacy-preserving fashion while respecting the rules set by data owners

Data and energy transaction support service: the platform should support peer-to-peer transactions without trusted intermediary and associated reputation service as a decentralised “soft” enforcement mechanism

Self-Governance service: platform should support collective negotiation and decision-making about collective resource/asset management without a central organizer

It is recommended a *" hybrid governance approach combining “top-down” regulatory conditions (e.g., governing data security, privacy policies, standards, non-discrimination and neutrality in data management) and “bottom-up” governance in decentralized open energy marketplaces. This approach could better address local specifications while providing a homogenous institutional environment for information management."*

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

In addition, BD4NRG solutions could take into design consideration separation of roles and responsibilities (governance, assets, data ownership, customer relation) as well as curative actions (feedback loops), sharing experience.

Marketplaces for data and services are expected to extend to smart sector integration topics. It is indicated that current activities in this field are very limited.

The relationship and the related business models driven roles are presented in the below diagram in Figure 17 [20]. Roles of data enricher, data buyer, data aggregator and data sellers are defined. The scope of AIOTI HLA is depicted in the lower block in relations to the Marketplaces architecture.

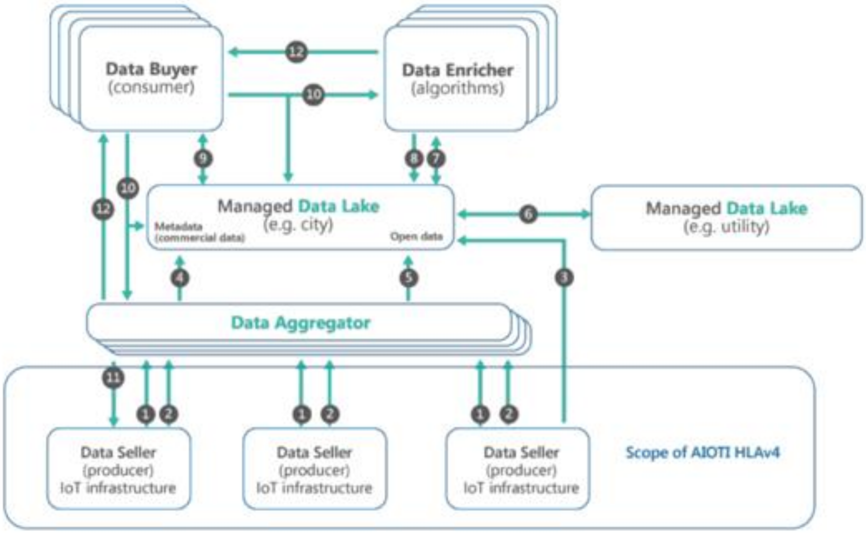


Figure 17 – Possible high-level architecture for IoT-enabled Marketplaces.

The three different conceptual layers for the definition of a digital energy marketplace for the purpose of Identifying fundamental regulatory dimensions for potential designs of future platforms are identified in Figure 18.

[20] Source: https://aioti.eu/wp-content/uploads/2019/02/IoT-data-market-places-drivers-and-architechtures-white-paper-Elloumi-De_Block-Samovich.pdf

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

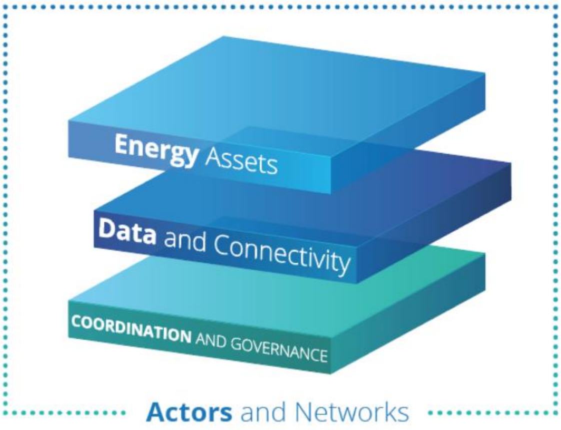


Figure 18 – Conceptual layers for digital marketplaces definition

The elements that need to be defined for the architecture of the digital energy marketplace are composed of energy assets layer, data and connectivity and coordination layer where actors and the related networks exchange data and services.

4.1.3.5 Mapping onto the BVDA and AIOTI

A corresponding mapping of the AIOTI HLA (entities) to the BDV Reference Model (technical areas and capabilities is depicted in Figure 19[21].

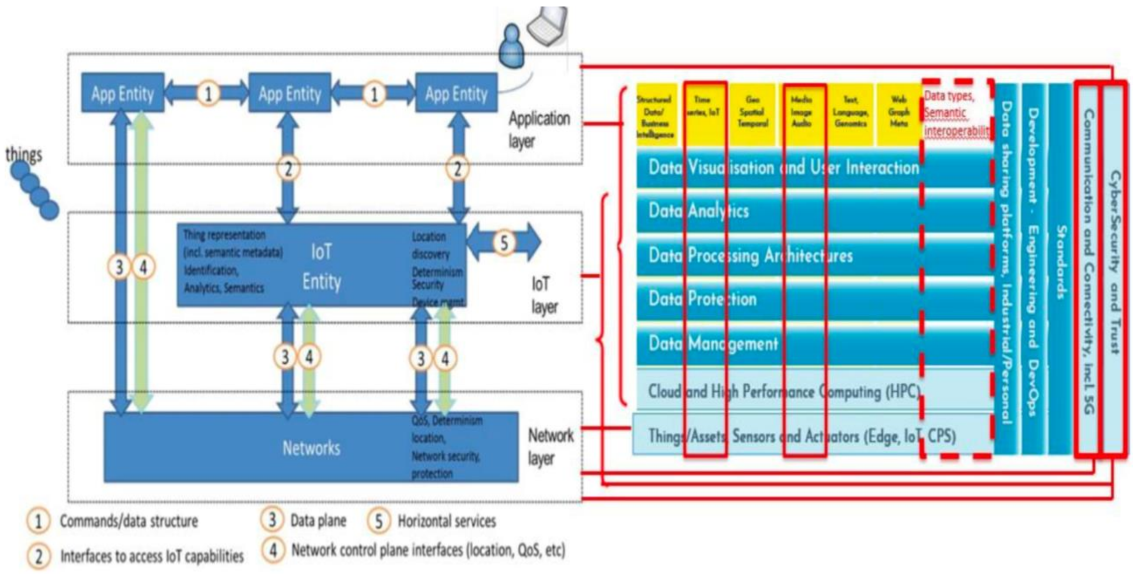


Figure 19 – BDVA Reference Model mapping to the AIOTI HLA

The App Entities of the AIOTI HLA provide application logic which may include data visualization and user interaction services, data analytics capabilities, various kinds of data processing capabilities, data protection support and data management logic, as well as support for cloud/HPC execution. In addition, the App Entities may include support for Cybersecurity and Trust.

[21] https://aioti.eu/wp-content/uploads/2020/12/AIOTI_HLA_R5_201221_Published.pdf

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

The IoT Entities of the AIOTI HLA may include access and management capabilities for sensors and actuators, but also support for data analytics (edge analytics), data processing, data protection and data management. In addition, the IoT Entities may include support for Cybersecurity and Trust.

The Networks of the AIOTI HLA are linked to the Communication and Connectivity area of the BDV Reference Model. In particular, they support short-range and long-range connectivity and data forwarding between entities, and both synchronous and asynchronous communication mechanisms, with appropriate QoS support. The Networks also include support for IoT devices' communication and connectivity. In addition, they may include support for Cybersecurity and Trust.

The BDV Reference Model areas of, respectively, "Data Sharing platforms, Industrial/Personal", "Development, Engineering and DevOps" and "Standards" are not mapped to the AIOTI HLA in the above figure. The first area might be relevant for IoT data management, the second area might be relevant for the total life cycle of IoT data, the third area is relevant for all areas represented by the layers in the AIOTI HLA.

An ongoing work is focused on identifying optimal ways to connect IoT and Data spaces. The related concepts and architectures impact are analyzed in the WG at AIOTI. The paper is expected towards November 2021 and will be discussed for purposes of BD4NRG considerations.

4.2 Building Blocks for Data Spaces

A data space can be defined as a decentralized data ecosystem built around commonly agreed building blocks enabling an effective and trusted sharing of data among participants. From a technical perspective, several technology building blocks are required ensuring:

Data interoperability - Data spaces should provide a solid framework for an efficient exchange of data among participants, supporting full decoupling of data providers and consumers. This requires the adoption of a "common lingua" every participant uses, materialized in the adoption of common APIs for the data exchange, and the definition of common data models. Common mechanisms for traceability of data exchange transactions and data provenance, are also required.

Data sovereignty and trust - Data spaces should bring technical means for guaranteeing that participants in a data space can trust each other and exercise sovereignty over data they share. This requires the adoption of common standards for managing the identity of participants, the verification of their truthfulness and the enforcement of policies agreed upon data access and usage control.

Data value creation - Data spaces should provide support for the creation of multi-sided markets where participants can generate value out of sharing data (i.e., creating data value chains). This requires the adoption of common mechanisms enabling the definition of terms and conditions (including pricing) linked to data offerings, the publication and discovery of such offerings and the management of all the necessary steps supporting the lifecycle of contracts that are established when a given participant acquires the rights to access and use data.

Besides the adoption of a common technology foundation, data spaces also require governance, that is the adoption of several business, operational and organizational agreements among participants. Business agreements, for example, specify what kind of terms and conditions can regulate the sharing of data between participants and the legal framework supporting contracts established through the

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

data space. Operational agreements, on the other hand, regulate policies that must be enforced during data space operation like, for example, compliance with GDPR (General Data Protection Regulation) or the 2nd Payment Services Directive (PSD2) in the finance sector. They may also comprise the definition of tools that operators of cloud infrastructures or global services supporting data spaces must implement, enabling auditing of certain processes or the adoption of cyber-security practices. Finally, organizational agreements establish the governance bodies (very much like ICANN for the Internet). They deal with the identification of concrete specifications that products implementing technology building blocks in a data space should comply with, as well as the business and operational agreements to be adopted. The complete taxonomy of building blocks required for creating data spaces is illustrated in Figure 20.

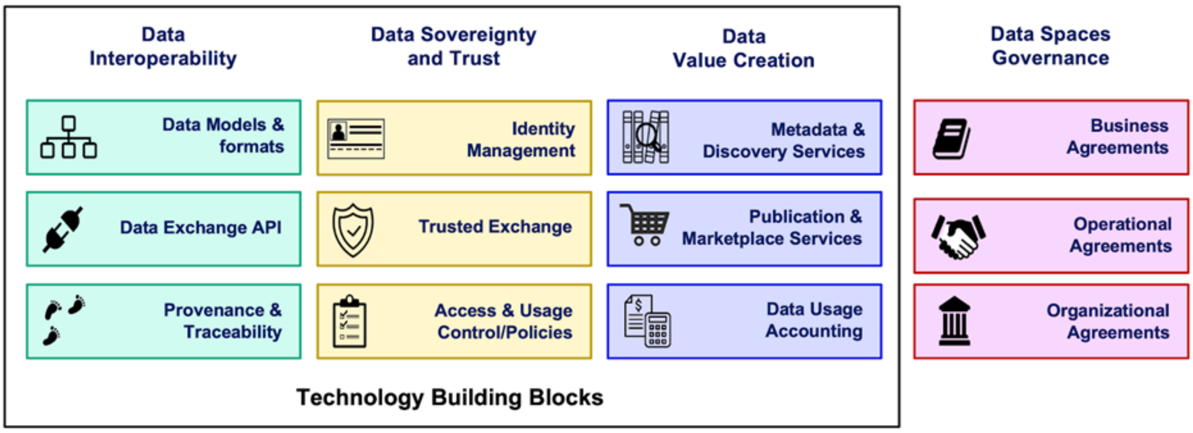


Figure 20 – Building blocks in a Data Space

Sharing of data within a given data space should not be limited to a single domain. This would severely limit the creation of new innovative services since individuals and organizations usually act in multiple domains at the same time and many opportunities will flourish when data generated within organizations operating in certain domain is shared for its exploitation in processes relevant to other domains. Therefore, technology building blocks for data spaces must be domain-agnostic. On the other hand, they should rely on open standards, allowing multiple infrastructure and global service providers to emerge and support data spaces, without getting locked by any provider. Given this, while making things work in living labs and pilots is relatively easy, the main challenge towards definition of successful data spaces is the decision of what concrete standards and design principles are adopted, since they must be accepted by all participants.

4.2.1 International Data Spaces (IDS)[22]

4.2.1.1 IDS Reference Architecture Model overview

The International Data Spaces initiative has defined a Reference Architecture Model to ensure Data Sovereignty, including requirements for secure and trusted data exchange in business ecosystems.

[22] IDS Reference Architecture Model 3.0 (https://internationaldataspaces.org/wp-content/uploads/IDS-RAM-3.0-2019.pdf)

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

The architecture is designed with the objective to overcome the differences between top-down approaches and bottom-up approaches. Figure 21 shows a typical architecture stack of the digital industrial enterprise.

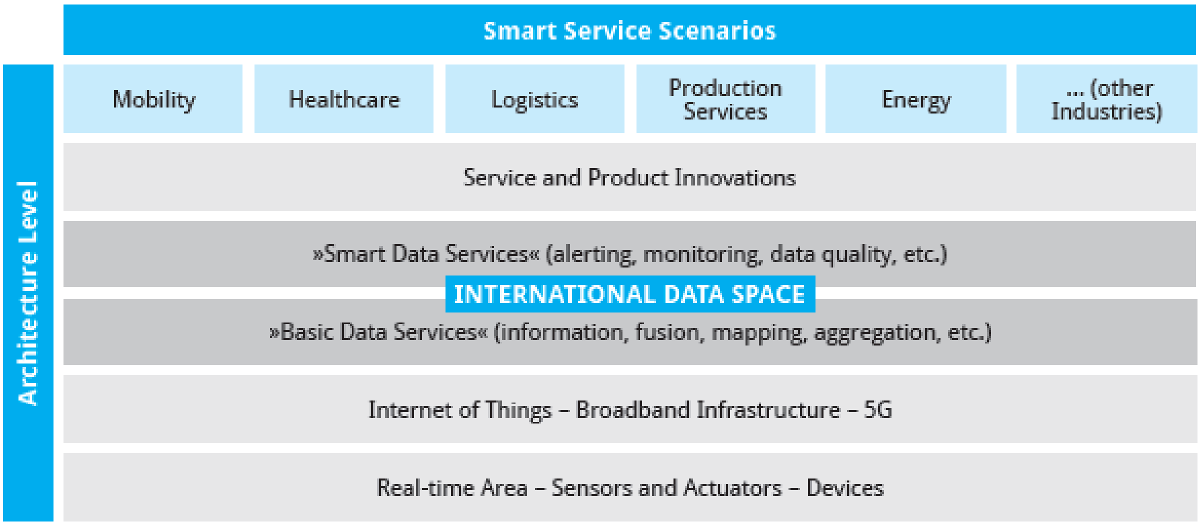


Figure 21 – IDS RAM 3.0: Typical enterprise architecture stack.

The IDS connects the lower-level architectures for communication and basic data services with more abstract architectures for smart data services. It therefore supports the establishment of secure data supply chains from data source to data use, while at the same time making sure data sovereignty is guaranteed for data owners.

As depicted in Figure 22, the IDS RAM is made up of five layers and three perspectives that need to be implemented across the five layers.

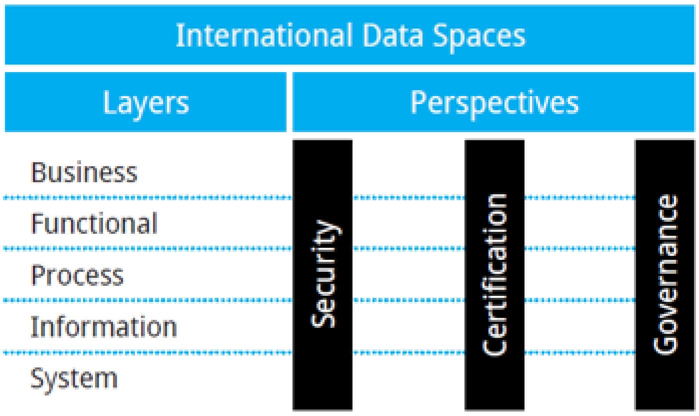


Figure 22 – IDS RAM Layers and Perspectives.

The IDS Reference Architecture Model (IDS-RAM) provides several elements, roles and interactions that constitute an infrastructure for sovereign data exchange (see Figure 23). There are four categories of roles:

- Category 1: Core Participant
- Category 2: Intermediary
- Category 3: Software / Service Provider

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

- Category 4: Governance Body

The Core Participants are Data Owner, Data Provider, Data Consumer and Data User. They are involved and required every time data is exchanged in the IDS. Intermediaries act as trusted entities. These are the Broker Service provider, an intermediary that stores and manages information about the data sources available in the IDS; the Clearing House, that provides clearing and settlement services for all financial and data exchange transactions; and the Identity Provider, which consists of a Certification Authority (managing digital certificates for the participants of the IDS), a Dynamic Attribute Provisioning Service (DAPS, managing the dynamic attributes of the participants), and a service named Dynamic Trust Monitoring (DTM, for continuous monitoring of the security and behaviour of the network. The 3rd category comprises IT companies providing software and/or services (e.g., based on a software-as-a-service model) to the participants of the IDS. The Certification Body, Evaluation Facilities, and the International Data Spaces Association are the Governance Bodies of the International Data Spaces.

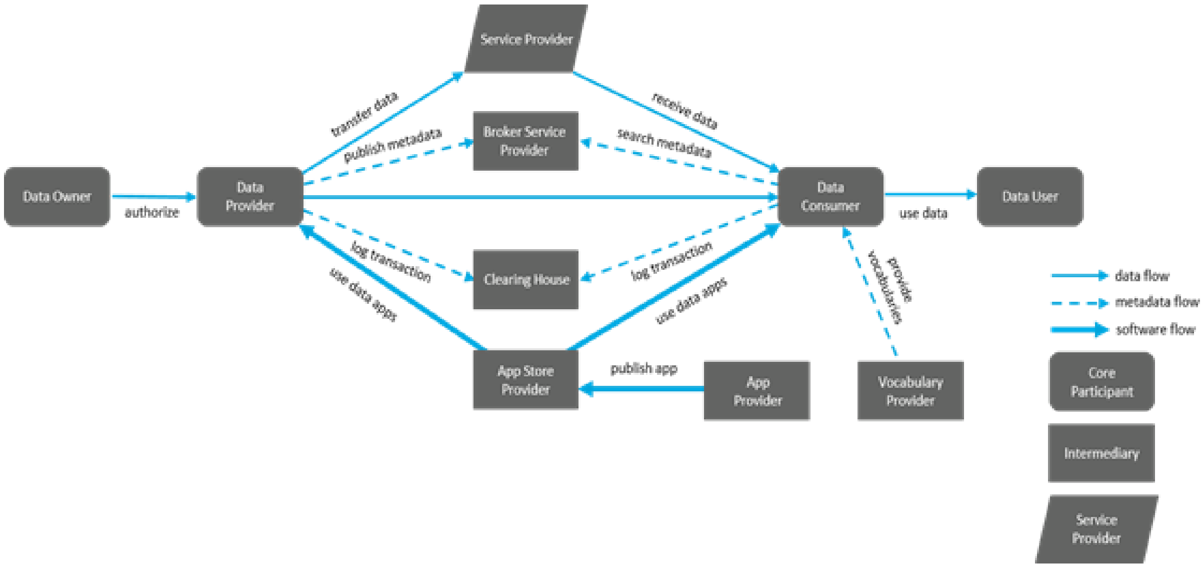


Figure 23 – IDS Roles and Interactions

The IDS connector is the core component of the IDS RAM as it is the access to a trusted ecosystem. It is responsible for the data exchange. In an IDS ecosystem, the data is transferred between the Connectors of the Data Provider and the Data Consumer (peer-to-peer network concept).

4.2.1.2 IDS Data Governance Model

Data Governance is a fundamental element in the International Data Spaces ecosystem, since it acts as an enabler for successfully engaging in a collaborative ecosystem. As depicted in Figure 22 Data Governance is one of the three perspectives of the IDS Reference Architecture Model.

The IDS-RAM distributes decision rights for data governance and management activities to the different roles in the International Data Spaces ecosystem. It thereby supports the requirements to be met by the actors within the ecosystem to achieve secure and reliable interoperability as well as desirable behaviour regarding the use of data.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

The IDS Data Governance Model defines a framework of decision-making rights and processes with regards to the definition, creation, processing, and use of data. While governance activities set the overall directive of the decision-making system, data management comprises three groups of activities with regards to the creation, processing, and use of data. In the IDS context, data governance comprises also usage rights of data shared and exchanged within the IDS ecosystem. The management of metadata specifies data about data and comprises both syntactical, semantic and pragmatic information.

This is of particular importance in distributed system environments that do not rely on a central instance for data storage, but instead allow self-organization of different heterogeneous databases. Additionally, data lifecycle management is concerned with the creation and capturing of data, including data processing, enrichment, storage, distribution, and use.

The following responsibility assignment matrix (RACI matrix, see Table 2) supports the allocation of these activities to enable a governance mechanism in the IDS ecosystem. RACI stands for “responsible”, “accountable”, “consulted” and “informed”. The focus lies on the „R“ and „A“ of the RACI matrix, supported by the notation „S“, which stands for „supported“.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Activity	Data Owner / Data Provider	Data User / Data Consumer	Broker	Clearing House
Management				
Determine data usage restrictions (execute data ownership rights)	R, A	-	S	-
Enforce data usage restrictions	-	R, A	-	-
Ensure data quality	R, A	-	S	-
Monitor and log data transactions	S	S	-	R, A
Enable data provenance	S	S	-	R, A
Provide clearing services	S	S	-	R, A
Metadata				
Describe and publish metadata	R, A	-	S	-
Look up and retrieve metadata	-	R, A	S	-
Data Lifecycle				
Capture and create data	R, A	-	-	-
Store data	R, A	S	-	-
Enrich and aggregate data	S	R, A	S	-
Distribute and provide data	R, A	-	S	-
Link data	S	S	R, A	-

Legend: R – Responsible; A – Accountable; S – Supporting.

Table 2 – Roles responsible, accountable and supporting in data governance.

4.2.2 GAIA-X[23]

4.2.2.1 GAIA-X Architecture Overview

The GAIA-X project was started in 2019 with the aim to enable a secure, open and sovereign use of data. It is motivated by challenges to the European digital economy, such as:

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

- Decentralised processing locations
- Lack of transparency and sovereignty over stored and processed data and infrastructure.
- Sector-specific data spaces and lack of ontology

As shown in Figure 24, the GAIA-X reference architecture comprises an infrastructure ecosystem, where infrastructure services are provided, connected or consumed, and a data ecosystem, that deals with data as the main business asset. Both ecosystems are connected via the federation services. They include Identity and Trust Services, Compliance Services, Federated Catalogues and Sovereign Data Exchange Services.

This architecture serves as an underlying framework common to all domains.

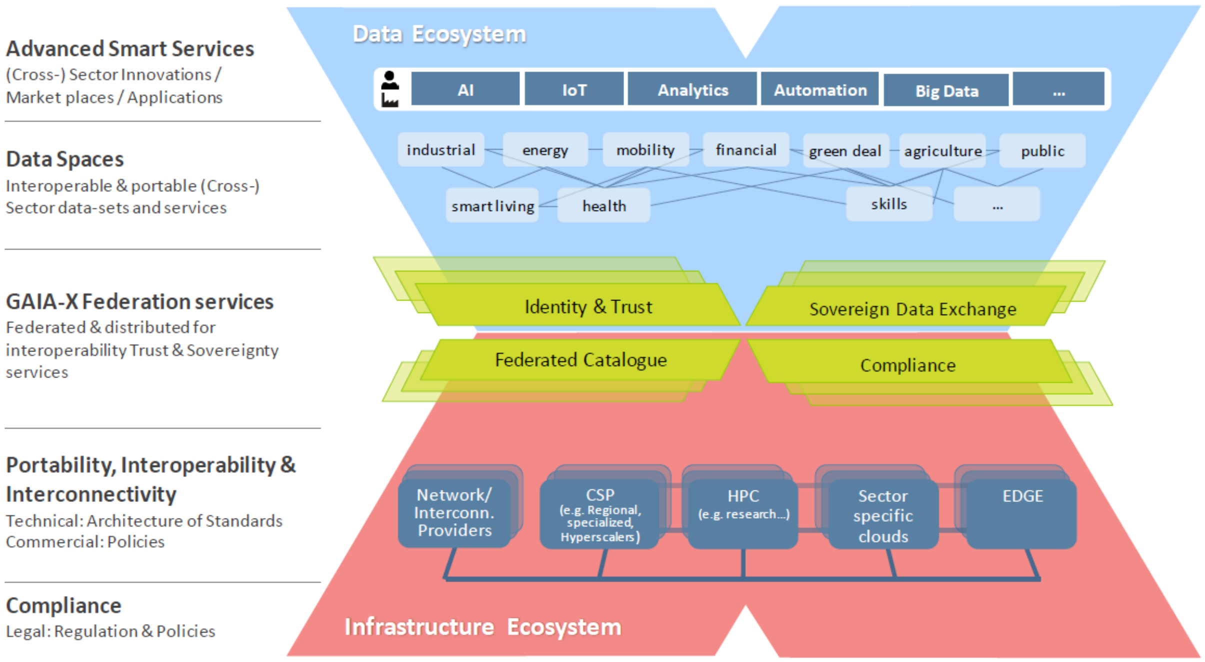


Figure 24 – GAIA-X Architecture

To ensure high-level data protection, security, transparency and portability within the GAIA-X ecosystem, a framework of Policy Rules is being developed.

GAIA-X defines three main roles which a participant can take on:

- Provider: Provides Assets and Resources in the GAIA-X Ecosystem.
- Consumer: Searches for Service Offerings and consumes Service Instances in the GAIA-X Ecosystem to enable digital offerings for End-Users.
- Federator: Federators oversee the Federation Services and the Federation which are autonomous of each other.

[23] GAIA-X: Technical Architecture (https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Publications/gaia-x-technical-architecture.pdf?__blob=publicationFile&v=5)

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

4.2.2.2 Data Sovereignty Services

Data Sovereignty Services provide Participants the capability to be entirely self-determined regarding the exchange and sharing of their data. They can also decide to act without having the Data Sovereignty Service involved if they wish to do so.

Informational self-determination for all Participants comprises two aspects within the Data Ecosystem: Transparency of data usages and control of data usages. Enabling Data Sovereignty when exchanging, sharing and using data relies on fundamental functions and capabilities that are provided by Federation Services in conjunction with other mechanisms, concepts, and standards. The Data Sovereignty Services build on existing concepts of usage control that extends traditional access control. Thus, usage control is concerned with requirements that pertain to future data usages (i.e., obligations), rather than data access (provisions).

Capabilities for Data Sovereignty Services

The foundation for Data Sovereignty is a trust-management mechanism to enable a reliable foundation for peer-to-peer data exchange and usage, but also to enable data value chains with multiple Providers and Consumers being involved. All functions and capabilities can be extended and configured based on domain-specific or use case-specific requirements to form reusable schemes.

The following capabilities are essential for Data Sovereignty in the GAIA-X Data Ecosystems:

Capability	Description
Expression of Policies in a machine-readable form	To enable transparency and control of data usages, it is important to have a common policy specification language to express data usage restrictions in a formal and technology-independent manner that is agreed and understood by all Gaia-X Participants. Therefore, they have to be formalized and expressed in a common standard such as ODRL[13].
Inclusion of Policies in Self-Description	Informational self-determination and transparency require metadata to describe Data Assets including Provider, Consumer, and Usage Policies as provided by Self-Descriptions and the Federated Catalogues.
Interpretation of Usage Policies	For a Policy to be agreed upon, it must be understood by all Participants in a way that enables negotiation and possible technical and organizational enforcement of Policies.
Enforcement	Monitoring of data usage is a detective enforcement of data usage with subsequent (compensating) actions. In contrast, preventive enforcement[14] ensures the policy Compliance with technical means (e.g., cancel or modify data flows).

Table 3 – Capabilities for GAIA-X Data Sovereignty Services

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Functions of Data Sovereignty Services

Information services provide more detailed information about the general context of the data usage transactions. All information on the data exchange and data usage transactions must be traceable; therefore, agreed monitoring and logging capabilities are required for all data usage transactions. Self-determination also means that Providers can choose to apply no Usage Policies at all.

The Data Sovereignty Services in GAIA-X implement different functions for different phases of the data exchanges. Therefore, three phases of data exchanges must be differentiated:

- before transaction
- during transaction
- after transaction

The services are depicted in Figure 25.

Before the data exchange transaction, the Data Agreement Service is triggered and both parties negotiate a data exchange agreement. This includes Usage Policies and the required measures to implement those.

During transactions, a Data Logging Service receives logging-messages that are useful to trace each transaction. This includes data provided, data received, policy enforced, and policy-violating messages.

During and after the transaction the information stored can be queried by the transaction partners and a third eligible party, if required. The figure below shows the role of mentioned services to enable sovereign data exchange.

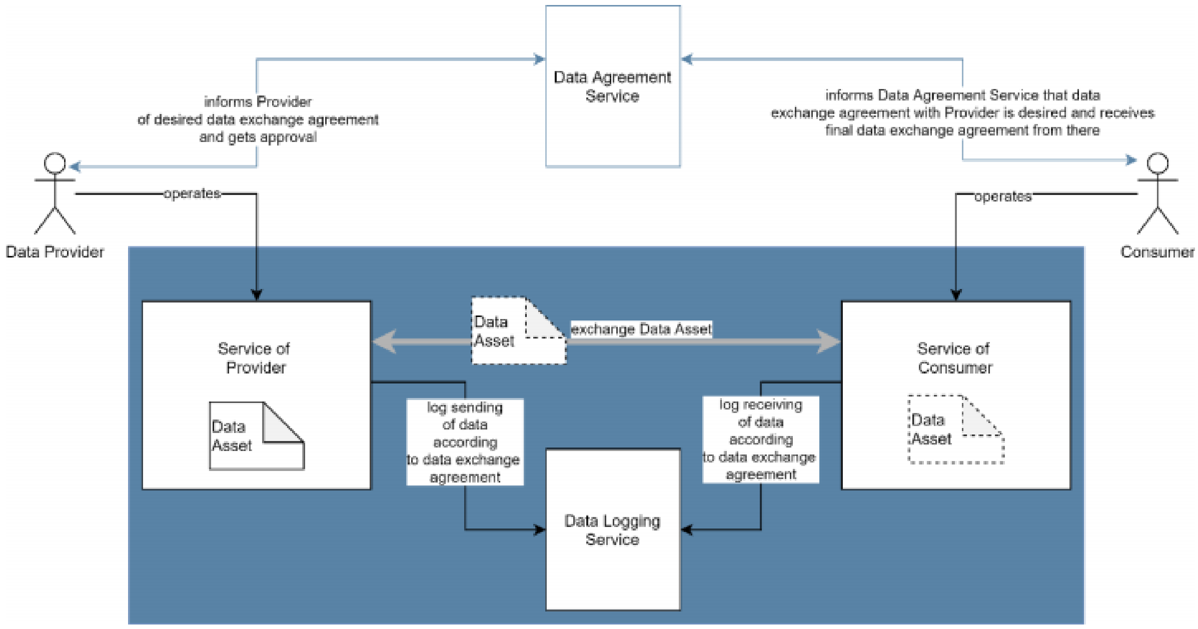


Figure 25 – Data Sovereignty Services Big Picture

The Data Agreement Service enables data transactions in a secure, trusted, and auditable way. It offers interfaces for the negotiation detailing the agreed terms for planned data exchange. The service is not meant to handle the transaction of data (that is described in the negotiated data contracts).

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

The Data Logging Service provides evidence that data has been (a) transmitted, (b) received and (c) that rules and obligations (Usage Policies) were successfully enforced or were violated. This supports the clearing of operational issues but also identifies fraudulent transactions.

The Provider can track if, how, and what data was provided, with the Consumer being notified about this. The Consumer can track if data was received or not, and, additionally, track and provide evidence on the enforcement or violation of Usage Policies.

4.2.3 Integration & Differences Between GAIA-X and IDS[24]

The combined architecture of GAIA-X and IDS supports and enables data spaces and builds advanced smart services in industry verticals. GAIA-X focuses on sovereign cloud services and cloud infrastructure, while IDS focuses on data and data sovereignty. The interaction of GAIA-X and IDS has three main tasks: self-sovereign data storage, trustworthy data usage and interoperable data exchange.

Figure 26 shows the mapping of the IDS components into the GAIA-X Architecture.

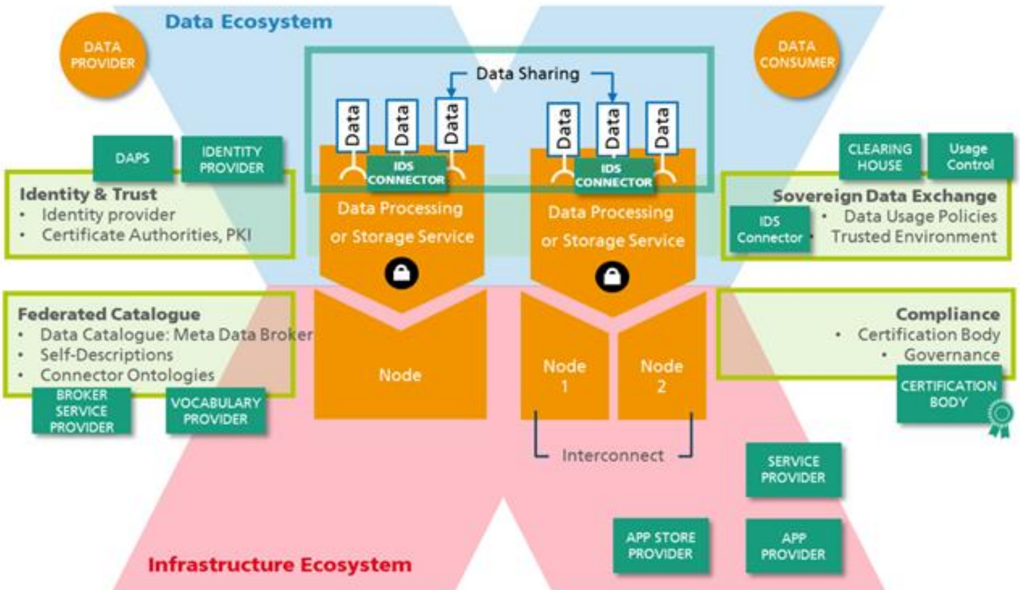


Figure 26 – Mapping of IDS Components into the GAIA-X Architecture

The Data provider and Data Consumer are mapped into the data ecosystem whereas the Service, App and App Store Providers are mapped into the infrastructure ecosystem.

The IDS connectors can be integrated in the GAIA-X Nodes, as they work as secure gateways. Other key elements of the IDS architecture can be mapped into the GAIA-X federation services. For instance, the DAPS and Identity Provider can cover the Identity & Trust federation services. The Federated Catalogue can leverage on the IDS Broker Service Provider, the Vocabulary Provider and the IDS Information Model. The sovereign data exchange service can benefit from the IDS Clearing House and

[24] Source: IDS Position Paper "IDS and GAIA-X" Version 1.0 January 2021

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

the IDS Usage Control concept. And the IDS Certification mechanisms could support the GAIA-X compliance services.

4.2.4 FIWARE

The following sections elaborate on the different components FIWARE brings materializing the different technical building blocks required for creation of data spaces.

4.2.4.1 Data Interoperability Building Blocks

Data providers joining data spaces must be able to publish data resources at well-defined endpoints knowing that data consumers, a priori unknown to them, will know how to retrieve and consume data through those endpoints. Data consumers, on the other hand, must know how data available through endpoints they discover can be consumed. This is a key principle which was observed in the design of the world wide web: content providers publish web pages on web servers (endpoints) knowing that web browsers will be able to connect to them and retrieve web pages whose content they can render and display to end users. It means that all participants in data spaces should 'speak the same language', which translates into adopting domain-agnostic common APIs and security schemas for data exchange (the way of constructing sentences) together with data models represented in data formats compatible with those APIs (the vocabulary used in constructed sentences).

The NGSI API is domain-agnostic. Many different systems use NGSI not only for smart energy but also for smart cities, smart manufacturing, smart water, smart agriculture, smart ports, or smart health, to mention a few. This facilitates data sharing because each system participating in a data space will be publishing data that simply enriches a Digital Twin data representation of the world that the rest of systems connecting to the data space will know how to access. Systems participating into the data space don't know a priori what other systems may consume the data they publish (although they will be able to set up concrete terms and conditions for accessing/using data as we will explain in the next section).

Figure 27 illustrates how different systems participating in data spaces "powered by FIWARE" will exchange data. Context Broker servers are the endpoints through which systems connected to the data space publish digital twin data, very much like web servers publish HTML content on the world wide web. Those systems can in turn connect to Context Broker servers to obtain information they need. Note that data spaces powered by FIWARE enable near real-time (right-time) exchange of digital twin data which is fundamental in the design of innovative value chains demanding a very dynamic exchange of data among participants. NGSI-LD brings quite simple and therefore easy to use operations for creating, updating, and consuming context / digital twin data but also more powerful operations like sophisticated queries, including geo-queries, or the subscription to get notified on changes of digital twin entities. On the other hand, data spaces "powered by FIWARE" can also support the exchange of large files using standard file transfer protocols, since this kind of file transfer may be required for certain scenarios like training of AI algorithms.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

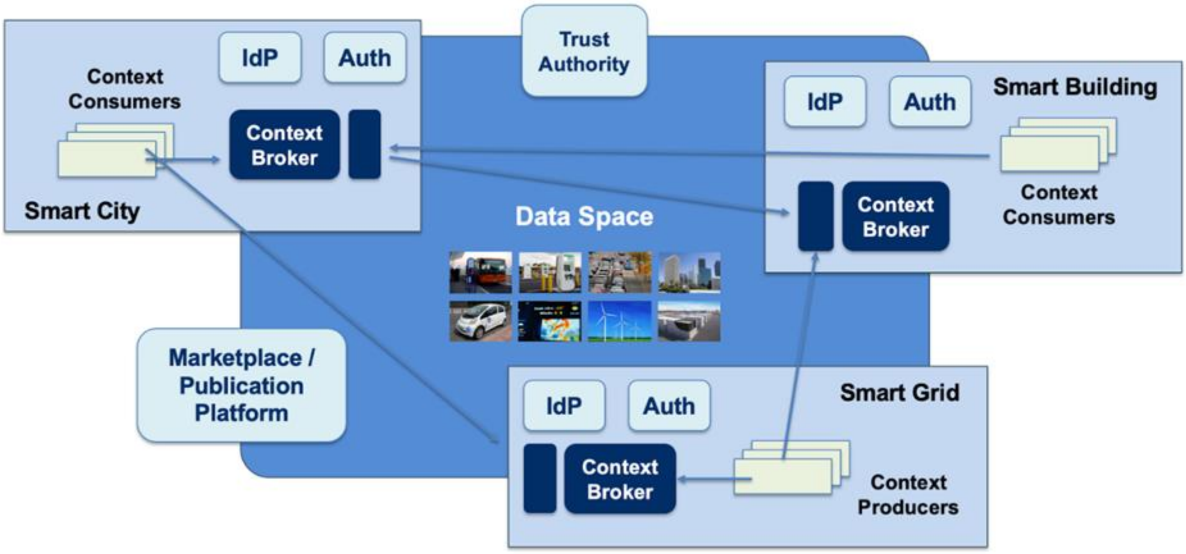


Figure 27 – Data Exchange in a data space "powered by FIWARE"

Note that systems participating in data spaces “powered by FIWARE” do not need to be themselves “powered by FIWARE.” Systems which have not been architected using FIWARE can still use the NGSI API to share data they produce and consume data they need in the form of data associated with attributes of digital twin entities which represent that part of the world they deal with. This can be done directly by the systems or through NGSI system adapters which implement a conversion between NGSI and the API that the system natively supports for managing data.

From a theoretical perspective, systems connected to a data space should be able to share data using the API they prefer. The specification (information model) of each API could be published as some kind of manifesto that certain components, integrated as part of the platform that systems should use to connect to the data space, can dynamically process to perform an automated adaptation from/to the APIs. However, such an approach faces important challenges. In the first place, such an approach has only been demonstrated in remarkably simple scenarios and involving quite simple APIs. Thus, the ability to exploit the sophisticated features of NGSI-LD to support for accessing digital twin data will be rather limited. On the other hand, creating a “common lingua” has proven to work for creating the kind of ecosystems we pursue: we shall ask ourselves whether the world wide web would have been as speedily of adopted if HTTP and HTML hadn’t been adopted as “common lingua” for web servers and browsers and each web server had the ability to choose a different protocol (as opposed to HTTP) or a different document format (as opposed to HTML). NGSI-LD has the advantages of being an open standard (defined by ETSI), has a strong open-source community behind (FIWARE) and, quite relevant within Europe, it will pave the way for alignment with developments in the Connecting Europe Facility (CEF) program. Creation of system adapters that transform from specific APIs a given system may still require to use from/to NGSI-LD has proven to be not a complex task.

As mentioned before, the NGSI-LD API is domain-agnostic, therefore it is designed to work for any type of digital twin. Consequently, achieving full interoperability also requires the adoption of common data models to be represented in formats compatible with the API. Here, the Smart Data Models initiative reaches great relevance. It brings a powerful resource for developers who can rely on the way data

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

model specifications are mapped into concrete JSON and JSON-LD structures under the initiative, compatible with NGSIv2 and NGSI-LD respectively.

Figure 28 illustrates how resources are organized within the Smart Data Models initiative on GitHub. Data models are grouped into "subjects" (weather, parking, aquaculture, etc) which in turn are referred to from repositories associated with the multiple application domains being considered (Smart Energy, Smart Cities, Smart Agriculture, Smart Manufacturing, Smart Water, etc). Note that there are subjects which are very specific to a given application domain (e.g., "street lighting" with regards to smart cities and communities) while others may be relevant to multiple domains (e.g., "weather" that is relevant to almost every domain or "sewage" that is relevant to the Smart Cities and Smart Water domains). Published data models are open to contributions and royalty-free.

The Smart Data Models initiative defines an open governance model specifying the lifecycle of data models comprising incubation of brand-new data models as well as curation of data models via harmonization of different contributions. Processes and procedures for management of the different activities follow best practices from open-source communities, guided by principles of transparency and meritocracy.

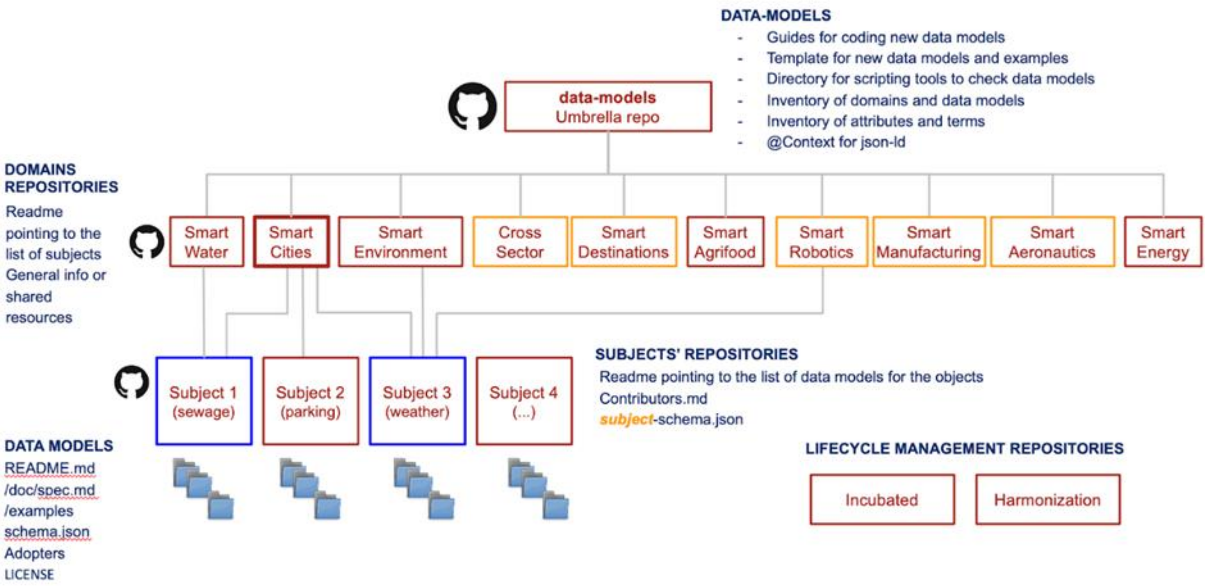


Figure 28 – Smart Data Models organization on GitHub

Completing the picture of building blocks for Data Interoperability, FIWARE brings components which provide the means for tracing and tracking in the process of data provision and data consumption/use. It provides the basis for several essential functions, from identification of the provenance of data to audit-proof logging of NGSI-LD transactions. For those data spaces with strong requirements on transparency and certification, FIWARE brings components (i.e., Canis Major) that ease recording of transaction logs into different Distributed Ledgers / Blockchains.

4.2.4.2 Data Sovereignty and Trust Building Blocks

Data spaces must provide means for guaranteeing organizations joining data spaces that they can trust the other participants and that they will be able to exercise sovereignty on their data. That requires

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

the definition of common building blocks, based on mature security standards that will be used by all participants in the data space.

A first fundamental building block to support within data spaces has to do with Identity Management (IM). This building block allows identification, authentication, and authorization of organizations, individuals, machines, and other actors participating in a data space. While this building block can be implemented based on third party open-source technologies like KeyCloak, just to mention one popular example, FIWARE brings the Keyrock component which supports OpenIdConnect, SAML 2.0 and OAuth2 standards. Quite relevant for data spaces deployed in Europe, Keyrock also resolves integration with eIDAS, a building block provided by the European Commission that enables the mutual recognition of national electronic identification schemes (eID) across borders, allowing European citizens to use their national eIDs when accessing online services from other European countries.

A second fundamental building block will be the one which facilitates trusted data exchange among participants, providing certainty that participants involved in the data exchange are who they claim to be, and that they comply with defined rules/agreements. Trust refers to the fact that data providers and data consumers can rely on the identity of the members of the data ecosystem and beyond that, between different security domains. Here, IDS Connector technology, as described in the IDS Reference Architecture Model (RAM) emerges as a solid basis and the FIWARE Community has incubated an open-source implementation of this technology that has already been tested on how it can integrate with the rest of core FIWARE components.

Figure 29 illustrates how an AI service provider application and an AI service consumer application, each hosted in different organizations and participating in a common data space "powered by FIWARE," may interact with each other. It shows the role of components that are part of the incubated FIWARE IDS Connector implementation and their integration with other FIWARE components. The AI service provided can be an advanced traffic prediction service while the AI service consumer application might be a traffic management system running in each city. The AI service consumer wishes to incorporate near real-time traffic predictions based on current traffic levels measured through IoT sensors deployed in the streets. The Context Broker component running on the consumer application sends notifications based on updates on attribute "current traffic" of a given street. These updates are notified to the AI service where the value is processed in real-time applying AI algorithms (e.g. using Apache Spark). The predicted traffic in about 30 minutes is calculated (generating, for example, values "low"," medium" or" high") and the value of attribute "traffic in 30 minutes" of the given street is updated if changes occur.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

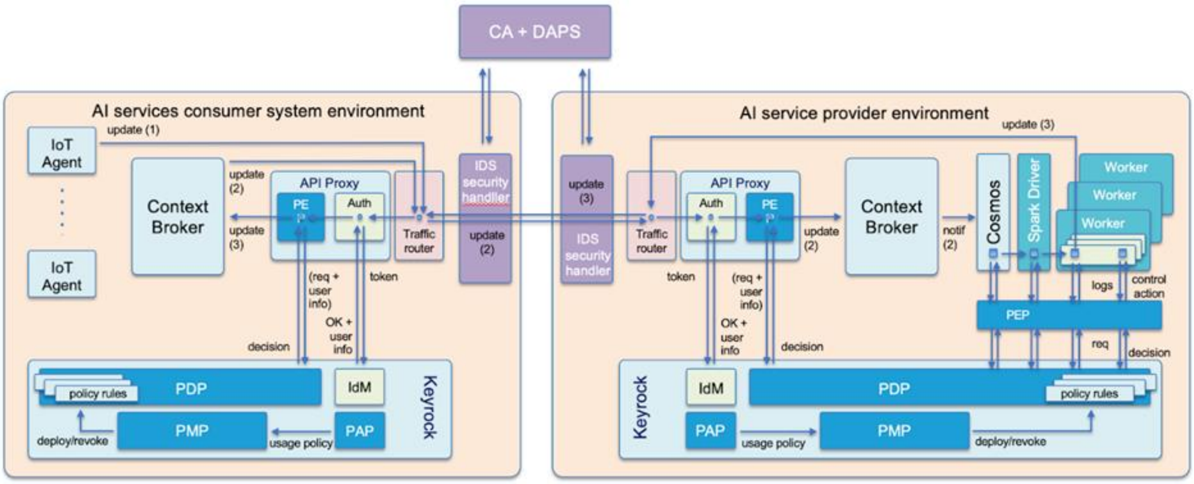


Figure 29 – Integration of FIWARE IAM and IDS Connector functionalities

In FIWARE, the implementation of IDS Connectors comes closely linked to the deployment of Kubernetes clusters. In this respect, the implementation of the concept of IDS connectors would not be limited to the implementation of the procedures used to exchange data among IDS connectors but also the monitoring and control of what is deployed in the associated clusters (e.g., guaranteeing that only certified components/applications are deployed) as well as the control of the flow of data within the cluster, ensuring enforcement of data access/usage control policies. For this last purpose, the usage of products supporting service mesh concepts like Istio are essential. In Figure 29, such components are identified as “traffic routers”.

In the IDS RAM, identity and access/usage control is currently performed at organization level. The Certification Authority (CA) and Dynamic Attribute Provisioning Service (DAPS) oversee verifying that organizations exchanging data are known and trustworthy. The IDS Connectors ensure that data exchanges are authorized at organization level. In Figure 29, for example, the FIWARE IDS Connector running at the AI service consumer side ensures that when an IoT agent updates context / digital twin attributes on the local Context Broker (1) forwarding of updates (2) are sent out only to authorized organizations like the AI service provider in this example. The FIWARE IDS Connector running at the service provider side will in turn verify that the organization from which updates are received is a trusted party and is authorized to make such updates to be then forwarded through the FIWARE Cosmos component to Spark where AI algorithms for traffic prediction purposes run. Similarly, when updates on the Context Broker deployed at the consumer side are invoked from the AI algorithms (3), the FIWARE IDS connector at the AI service provider side verifies that only requests authorized to be transmitted will be sent out. The FIWARE IDS connector at the AI consumer side verifies that the update request received comes from a trusted and authorized organization. All these verifications take place relying on the global CA+DAPS services of the data space.

On top of this authentication and data access/usage control procedures, performed at organization level via IDS Connectors, FIWARE security components enable an additional and finer grained authentication and access control at the end user level. Therefore, coming back to Figure 29, notifications received at the AI service provider side, same as update requests handled at the AI service consumer application side carry additional security tokens (namely, Json Web Tokens - JWT) linked to

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

users on behalf of which those notifications/requests have been issued. The API proxy components, both at the consumer and on the provider side, validate those tokens and obtain the info associated with the corresponding users relying on standard Identity Provider services supported by the FIWARE Keyrock component. For managing access control, a standard XACML process is implemented. The API proxy plays the role of the Policy Enforcement Point (PEP) and the FIWARE AuthZForce component, also alternatively Keyrock, implements the Policy Decision Point (PDP) functionality. That means that the proxy forwards user info of the particular user together with information about the concrete notification/request to the PDP which then checks whether users with those credentials are entitled to perform the given operation. FIWARE brings alternative implementations of the API proxy, namely Wilma, API Umbrella and CoatRack, all of them compatible with Keyrock or any third-party product implementing OpenID Connect, OAuth2 and XACML standards. Keyrock also implements Policy Administration Point (PAP) and Policy Management Point (PMP) standard XACML functions.

4.2.4.3 Data Value Creation Building Blocks

Loose coupling of participants is a fundamental principle in data spaces. Data providers and consumers do not necessarily know about each other. Therefore, it becomes essential to incorporate building blocks enabling the management of data resources as true assets with a business value. Assets which can be published, discovered and, eventually, traded. This way boosting the creation of multi-side markets where innovative services can be created.

FIWARE Business Application Ecosystem (BAE) components enable creation of Marketplace services which participants in data spaces can rely on for publishing their offerings around data assets they own. Distinct types of data assets can be defined via plugins, but three kinds of standard data asset types are supported by default, namely static data files, right-time data resources provided via NGSI-LD at well-defined endpoints, and data processing services, which typically have associated well defined endpoints for providing input data and publish results, both in right-time, using NGSI-LD. Marketplace services are accessible through a portal or via APIs. Users, either end users through the portal or applications via APIs, of the Marketplace service can perform the following main actions:

- define new data asset types via plug-ins
- register offerings around defined data asset types which typically means providing description of the asset, data models behind, endpoints (URLs) through with the data asset will be accessible and, particularly important, terms and conditions defined around the data asset, including SLAs, legal clauses, access control policies users of the asset must comply with and associated pricing schema, which may be based on:
    - Free access (open data): user pay no money
    - One-time payments: users pay only once.
    - Recurring payments: users pay periodically (monthly, yearly, etc) for getting access to some data. In addition, users will be able to cancel the subscription, but they won't be able to to access data anymore.
    - Usage payment: users pay per use. Their payments are based on the amount of information consumed.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

- ability to navigate and search/discover existing offerings based on selected criteria.

Following is the list of backend components and APIs associated to the FIWARE BAE Marketplace:

- Backend implementing standard TM Forum APIs supporting configuration of the marketplace:
    - Catalogue Management API
    - Product Ordering Management API
    - Product Inventory Management API
    - Party Management API
    - Customer Management API
    - Billing Management API
    - Usage Management API
- Rating, Charging, and Billing backend
- Revenue Settlement and Sharing System.
- Authentication, API Orchestrator, and Web portal

Figure 30 shows the FIWARE Data Marketplace portal deployed.

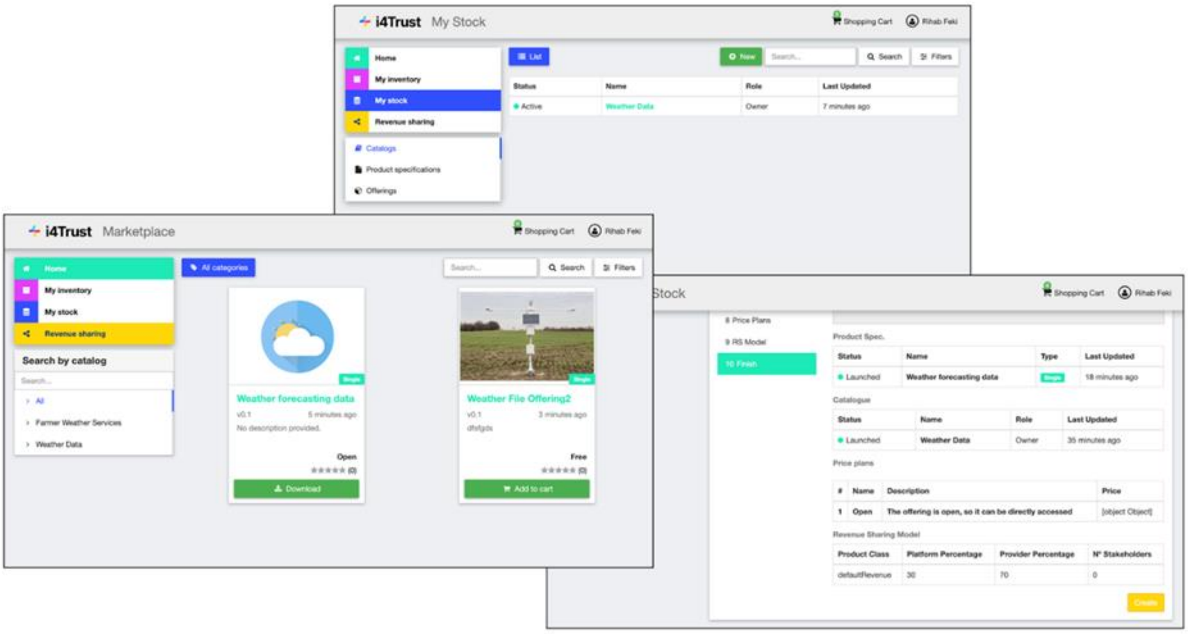


Figure 30 – FIWARE Data Marketplace portal

FIWARE also comprise components for publication of data resources linked to data assets around which offerings are managed through the FIWARE Data Marketplace. For this purpose, we have the Idra publication platform and an extended version of the CKAN open data platform, which is an open data publication platform widely adopted in the market. These extensions support enhanced data management capabilities and integration with FIWARE technologies including NGSI. In particular, data publication and discovery features provided by CKAN have been enhanced with the following features:

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

- Right-time (near real time) time data publication - thanks to this extension, CKAN is not limited to list data resources linked to static files as part of its catalogue but also data resources linked to NGSI-LD requests served by Context Broker components deployed in a data space. This brings the ability to discover data resources relying on DCAT capabilities publication platforms support,
- Identity Management, Authentication and Access Control functions based on Keyrock components - therefore supporting OpenId Connect, OAuth2 and XACML standards adopted at overall data space level.
- Publication of priced data resources - thanks to this extension, it is possible to mark data resources listed as part of the catalogue as linked to offerings visible in the Data Marketplace. Users can therefore click on those data resources and navigate to the Marketplace to proceed with the acquisition of access rights

Enhanced Data Visualization - thanks to this extension WireCloud (Configurable Dash-boards component) to create adaptive and incremental data visualization. This way, farmers (sellers) can decide the way they want their data to be shown.

4.2.4.4 Elements for a Reference Architecture with FIWARE components

The reference architecture of FIWARE platform for energy include quite different components that allow the integration with data spaces. The different elements can operate stand-alone or integrated into an ecosystem of solutions. The use of NGSI standard for accessing the information simplifies the data interchange. Eventually, any of the boxes describing a system can be replaced by other solutions, or eventually for the legacy systems already present in the pilots.

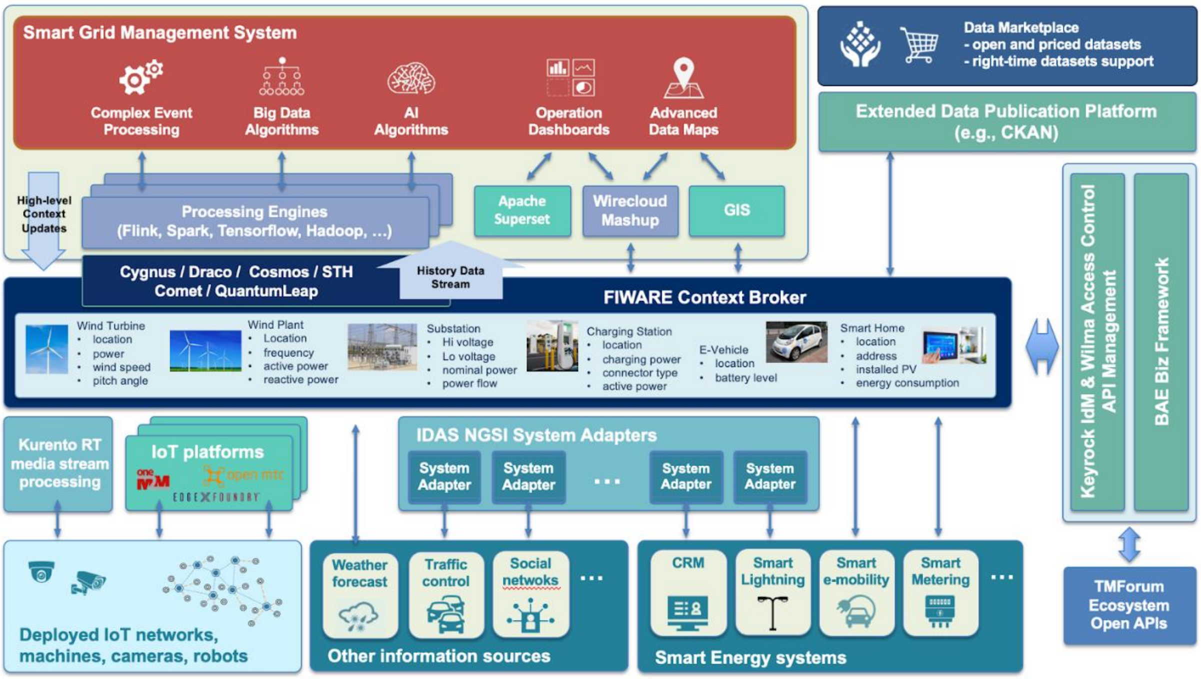


Figure 31 – FIWARE Reference Architecture

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Figure 32 – FIWARE Context Broker

The context broker is the core of the architecture. It allows the connection in right time with more than 800 updates a second which is useful for most of use cases in energy sector. It can integrate the data coming from different elements of the architecture and to define subscriptions to any event worth to be noticed. In example when a sensor measurement exceeds a specific value. Once the event happens an alert is sent to the connected system. The context broker allows to retrieve the information from heterogenous systems connected and defined as a data source. this is aq powerful mechanism, the federation, which allow the extension in the data sources but also in the processing capacity of the system. The information stored there complies with the NGSI standard. This standard, stores the data into entities composed of properties and relationships, as drafted in the example in the next image.

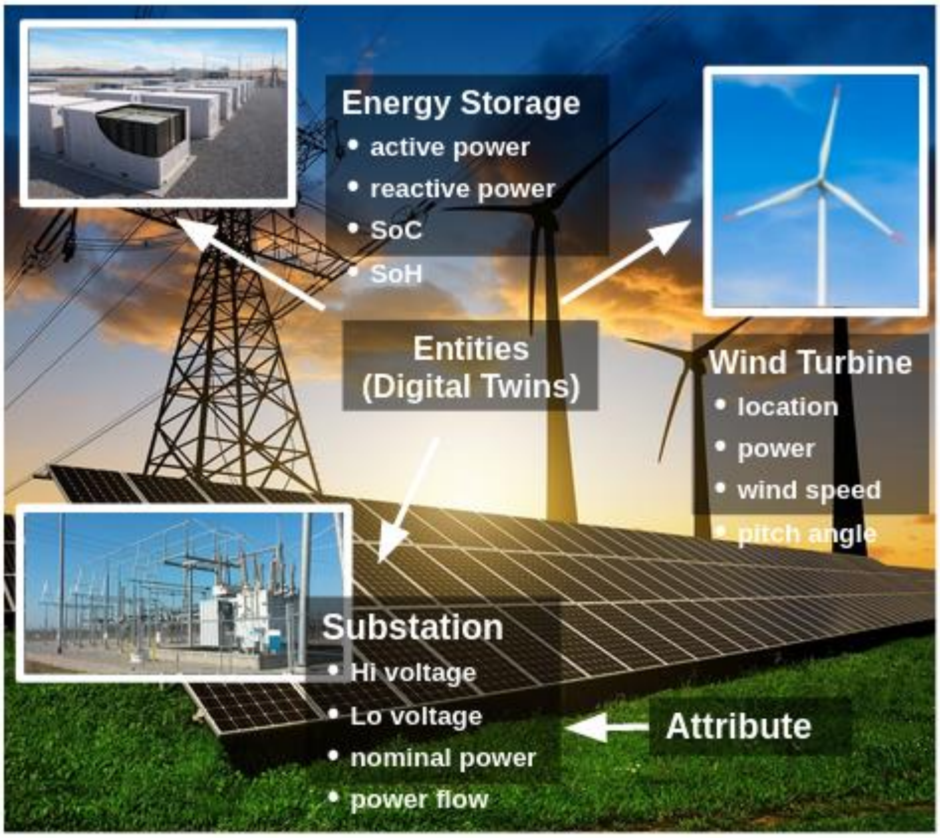


Figure 33 – Example of FIWARE Entities complying with the NGSI standard

These entities define the elements to be controlled and managed. A series of attributes provide details about these elements. This information can be updated on a real-time basis e.g., consumption, or rarely, e.g., maximum power produced).

To provide interoperability, the entities defined in some systems should be compatible with the same entity in other systems. The Smart Data Models Initiative has documented the different entities according to existing standards and actual experiences. Currently more than 400 data models are

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

defined including the Common Information Model for Energy[25], battery[26], other data models for elements Energy[27], including consumption[28], the use of devices for measuring the different magnitudes, but also some out-of-domain data models related like weather[29] or environmental impact[30].

4.2.5 Integration of FIWARE & IDS/GAIA-X

A fundamental building block will be the one which facilitates trusted data exchange among participants, providing certainty that participants involved in the data exchange are who they claim to be, and that they comply with defined rules/agreements. Trust refers to the fact that data providers and data consumers can rely on the identity of the members of the data ecosystem and beyond that, between different security domains. Here, IDS Connector technology, as described in the IDS Reference Architecture Model (RAM)[31] emerges as a solid basis. FIWARE Community has incubated an open-source implementation of this technology that has already been tested how it can integrate with the rest of core FIWARE components.

GAIA-X aims at creating a federated form of data infrastructure in Europe which strengthens the ability to both access and share data securely and confidently. Since its creation, the initiative has raised a lot of awareness because of the opportunity it brings to join forces and, leveraging existing assets, accelerate delivery of solutions to the market.

IDSA is providing several architecture elements which are necessary to create data spaces with trust and data sovereignty being some of the most important ones. These are necessary but not sufficient to create real living data spaces which are accepted and adapted on the market. Standard APIs, standard data models as well as marketplace functionalities are also necessary. The complementarities of both initiatives regarding technology building blocks are illustrated in Figure 34. As the FIWARE Context Broker technology was adopted in 2019 as one of the Building Blocks of the Connecting Europe Facility (CEF) program and has proven to integrate smoothly with other relevant CEF Building Blocks for the creation of Data Spaces (e.g., eIDAS, EBSI), this allows the alignment between these initiatives. As a result, the close partnership between FIWARE and IDSA is guaranteeing the aligned contribution of both ecosystems to the success of data spaces based on GAIA-X.

---

[25] https://github.com/smart-data-models/dataModel.EnergyCIM
[26] https://github.com/smart-data-models/dataModel.Battery
[27] https://github.com/smart-data-models/dataModel.Energy
[28] https://github.com/smart-data-models/dataModel.Device/tree/master/SmartMeteringObservation
[29] https://github.com/smart-data-models/dataModel.Weather
[30] https://github.com/smart-data-models/dataModel.Environment
[31] https://internationaldataspaces.org/download/16630/

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

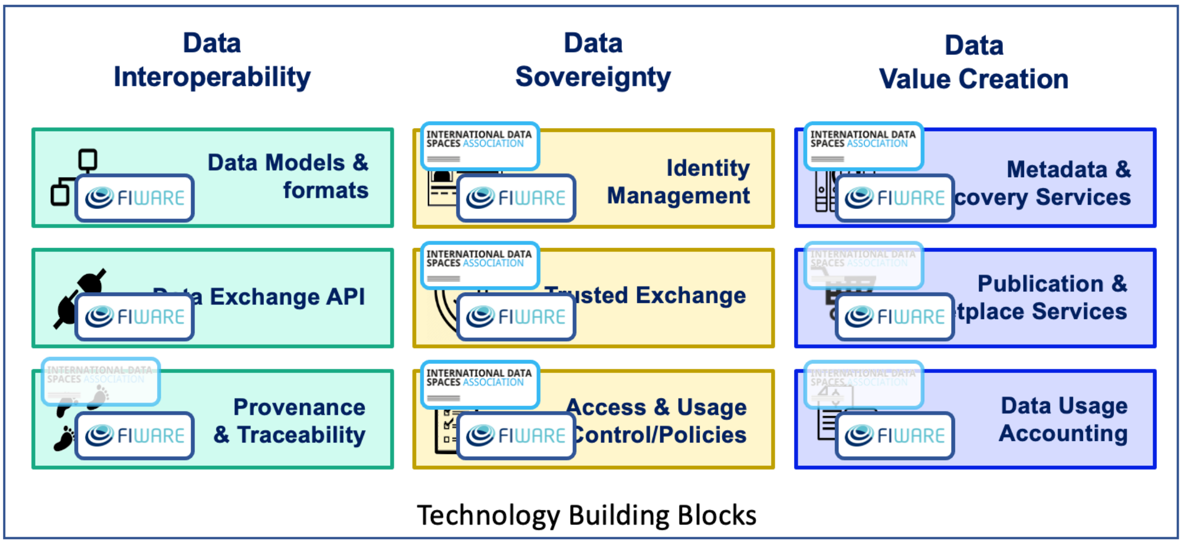


Figure 34 – IDSA and FIWARE positioning to create Data Spaces.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

5 Convergence Into BD4NRG Reference Architecture

This chapter introduces the first version of the BD4NRG reference architecture. Its first section presents the RA, its individual components and the respective responsible Work Packages and Tasks within the BD4NRG project. The chapter then goes on to discuss relevant aspects of the architecture, which are also found as layers in the architecture pictured in the project's Grant Agreement: Analysis, Governance, and Marketplace. Specifically, existing approaches to each of these aspects are highlighted, based on the efforts presented in chapters 3 and 4. Adapting from the mentioned approaches, each section finally describes the chosen course of action towards a BD4NRG-specific system addressing the respective aspect of the RA. Note that these procedures are presented as they are envisioned at time of submission of this deliverable, and may be adjusted as needed towards deliverable D2.6 *"BD4NRG Reference Architecture Final Version"*.

5.1 BD4NRG Reference Architecture

The ever-increasing digitization of energy systems accompanied by large volumes of data being produced by energy sensors, devices and systems provides an enormous opportunity for energy stakeholders to leverage state-of-the-art Big Data and AI technologies. This allows to improve business operational efficiencies, optimize energy production, efficiently manage and maintain energy grid resources and support smart decisions in energy consumption, while also considering environmental aspects. In this landscape decentralized energy ecosystems are emerging that need to cooperate in a federated manner, exchange data and take into consideration and exploit cross-domain information and related Big Data coming from other sectors, which can be combined in valued added energy applications.

The BD4NRG reference architecture (RA) aims to specify the building blocks and main elements that need to be considered by smart and decentralized data-driven energy ecosystems to create value out of Big Data by exploiting AI technologies. Through a concise and detailed RA, BD4NRG foresees to overcome barriers hampering the exploitation potential of energy Big Data, offering a standardized way of managing, sharing and using data and related AI services.

The proposed RA, depicted in Figure 35, extends the conceptual architecture presented in the BD4NRG Grant Agreement which has been revised in order to consider recent developments and architectural initiatives and models as presented in Chapter 3, including BRIDGE, COSMAG and RAMI, while highlighting the aspects that need to be taken into account for creating decentralized energy data spaces, following initiatives and approaches provided by IDSA, OPEN DEI, AIOTI, FIWARE and BDVA/DAIRO as presented in Chapter 4. Note that the BRDIGE RA greatly influences the proposed RA The BD4NRG RA adapts the BRIDGE RA and aligns it to the needs of Big Data analytics value chains. Under the BD4NRG view, the focus lies on data driven services in all aspects of energy ecosystems, including smart management of building environments as well as energy performance contracts.

The architecture considers four main layers of the data value chain and a vertical pillar focusing on data space enablers that allow decentralized energy ecosystems to make their data available, in a trusted and sovereignty-preserving manner, to analytics services providers so that value can be generated. In the following we provide more details for each of the building blocks of the BD4NRG RA, describe how these are mapped to the different tasks and work packages of the project and provide

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

references to other project deliverables and sections of this deliverable where more details on the different layers and components can be found.

The **Data Sources** Layer focuses on identifying and understanding the Big Data provided by numerous sources, including energy sensors and meters, data monitoring and acquisition platforms such as SCADA systems or Building Energy Management systems, databases with historical or real time data, smart grid data exchange platforms and cross-domain information such as environmental information and data coming from public administration services (e.g., related to energy performance contracts). Essentially, this layer includes all the data generating hardware, applications and platforms as described in the Component Layer of the BRIDGE architecture.

Within BD4NRG the consortium has worked from the start of the project to identify the available data sources and related data within WP2. Specifically relevant are T2.1 *"User Stories' Definition and Requirements Analysis"* where the data sources were identified, and T2.4 *"Semantic Data Models & Service and Platform Interoperability Specifications",* which analysed the available data. The data to be used in BD4NRG is described in detail in deliverable D2.3 *"BD4NRG Semantification & Interoperability"*.

The **Data Interoperability** Layer focuses on identifying the communication interfaces with the data sources and corresponding data formats, as well as on providing a set of data and information models which could be followed or used for data transformation, aiming to ensure interoperable data exchange and use. This layer integrates the "Communication" and "Information" layers of the BRIDGE architecture.

Within BD4NRG, the work towards identifying the communication interfaces and data formats has begun within WP2, specifically within tasks T2.3 *"Open APIs and Data Assets Functional Specifications"* and T2.4 *"Semantic Data Models & Service and Platform Interoperability Specifications".* The work in these tasks focuses on identifying and providing the directions for data interoperability while identifying relevant data and information models. The BD4NRG approach relies on Agile Standardization whereby smart data models are used and are extended based on end user and application requirements. More details are provided in D2.3 *"BD4NRG Semantification & Interoperability"*. Note that the work continues in T3.5 *"Interoperability & Homogenization",* where the data Integration & homogenization components are being developed, focusing on the communication interfaces and in task T4.2 *"Semantic Business and Platform Interoperability Management"* focusing on data and information models.

The **Innovative Data Analytics Services** Layer comprises four sublayers and specifies

- the processing infrastructure required for Big Data processing,
- different types of analytics that are employed depending on user requirements and data analytics use cases
- the analytics services that take as input the energy and cross-sector Big Data
- the functional process which are supported by the analytics services and receive the value extracted from Big Data. This layer represents the "Function" layer of the BRIDGE architecture by consolidating aspects related to data analytics and correspondingly supported functions.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Within Tasks T2.1 User Stories Definition and Requirements Analysis and T2.2 "*Edge Energy Services Specifications for Big Data/IoT Data Framework infrastructure*" the functions to be supported and corresponding analytics services have been identified and documented in deliverables D2.1 "*Big Data Value Chain and Requirements*" and D2.2 "*BD4NRG Open APIs and Data Assets Functional Specifications*".

The work related to the processing infrastructure is part of WP3 and specifically Tasks T*3.6 "Seamless Elastic Distributed Data Management and HTAP (Heterogeneous Data Capture, Cleansing, Integration, Polyglot Persistence)",* T3.7 *"Integrated Querying of Streaming Data and Data at Rest", T3.8 "Automated Parallelisation of Data Streams and Intelligent Data Pipelining", and* T4.1 *"Distribution Enablers & Open Fog Flexible Management at the Edge".*

The definition and implementation of the BD4NRG analytics services is performed in WP4. The work is distributed in the following tasks which deal with different value-added services, aiming to generate value from the available data: T4.3 *"Incremental, Distributed & Self-learning Analytics",* T4.4 *"Cross-stakeholder Transfer learning",* T4.5 *"Declarative, Predictive & Prescriptive Real-Time Data Analytics",* T4.7 *"Open Modular Analytics Toolbox & Deployment".*

More details regarding the BD4NRG approach in terms of data analytics are provided in Section 5.2 of this deliverable.

The **Business Actors and Ecosystems** Layer identifies stakeholders who participate in the data analytics ecosystem and corresponding energy data spaces. They include data providers and analytics services users / data consumers, analytics applications providers and providers of data space enablers and related platforms. Note that an organization or business role in the energy ecosystem can have one or more roles in data space, for example a TSO can be a data provider and a consumer of analytics services at the same time.

The actors and ecosystems relevant for BD4NRG are identified in D2.1 *"Big Data Value Chain and Requirements".* Furthermore, WP7, WP8 and WP9 deal with the different Large Scale Pilots (LSPs) which are organized in three main clusters: BD-4-NET focusing on showcasing the value of data analytics for increasing the efficiency and reliability of the electricity network, BD-4-DER focusing on optimizing the management of assets (DER-Distributed Energy Resources) connected to the grid and BD-4-ENEF demonstrating the de-risking of investments in energy efficiency and increasing the efficiency and comfort of buildings. Note that WP6 will analyse the impact the of the BD4NRG approach and validate the proposed approach by gathering and analysing stakeholders' feedback.

The vertical pillar of **BD4NRG Data Space Enablers** identifies the components and functions required to realize the distributed BD4NRG data space and is aligned with the design principles specified by OPEN-DEI[32].

A main element for smooth data space operation is specification and provision of a set of rules or agreements which should be followed by all stakeholders. This set of rules essentially define how the data space is governed and clarify how stakeholders interact within it on organizational, operational

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

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

and business levels. IDSA has already provided a rule book[33] specifying blueprints of agreements which should be considered when implementing a data space. The IDSA rule book will be a starting point for BD4NRG, whereas a detailed framework for trusted data sharing will be defined within T3.1 *"DLT & Smart contracts for B2B cross-stakeholder trusted off-chain data sharing and re-use".*

The data space governance is complemented by a set of technical components for data governance that need to be in place in order to ensure data sovereignty and trust, track the origin (provenance) and usage of data, and enforce the interoperability specification by applying measures for data quality compliance and/or certification (note that in the BD4NRG RA, interoperability aspects related to communication interfaces and data/information models are part of the Data Interoperability Layer). The work related to data sovereignty and trust is being performed in WP3 and tasks T3.2 *"Data Access Policy Brokerage"*, T3.3 *"Legal, Regulatory, Privacy and Cybersecurity Management & Compliance Tools".* Data provenance and traceability is handled in task T3.1 *"DLT & Smart contracts for B2B cross-stakeholder trusted off-chain data sharing and re-use".* Data compliance is part of the work of tasks T3.4 *"Data Quality Compliance & Certification Tools".*

A more detailed view of the BD4NRG approach for dataspace and data governance is provided in Section 5.3.

Data value creation within a data space is accomplished through data publication and marketplace services, which facilitate the exchange or transaction of data and the use of data analytics services. Marketplaces are spaces where data producers, analytics services providers and users of the services meet with the purpose of creating value from the available data. An efficient data marketplace is supported by services that allow data and metadata discovery as well as data usage accounting. Moreover, analytics services are provided in "as a service" manner (e.g., machine learning as a service). More details on the BD4NRG marketplace are provided in Section 5.4, whereas the related implementation work is being executed within WP5 where tasks T5.1, T5.2 and T5.3 provide specifications and components for Flexible Market Platforms, a Big Data/AI WorkBench, DaaS / MLaaS utilities.

Finally, the BD4NRG RA considers the various dataspace levels as specified by OPEN DEI (see Section 4.1.1), which can be:

- Cloud: data storage, integration and analytics services operate on the cloud. Scalable cloud resources provide the means for advanced Big Data driven services and the provision of DaaS/MLaaS.
- Edge: mission critical processing is performed at the edge or in the organizations' premises, whereas the cloud is used for long term Big Data storage and/or for offloading analytics tasks than can be executed on the cloud.
- Field: data analytics services are deployed at the user premises and are used in locally deployed infrastructure. Such an approach may be advantageous for organizations that are reluctant to make the transition to the cloud due to business, organizational or other restrictions.

[33]https://internationaldataspaces.org/rule-book-on-structures-and-processes-for-implementing-ids-in-the-real-world/

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

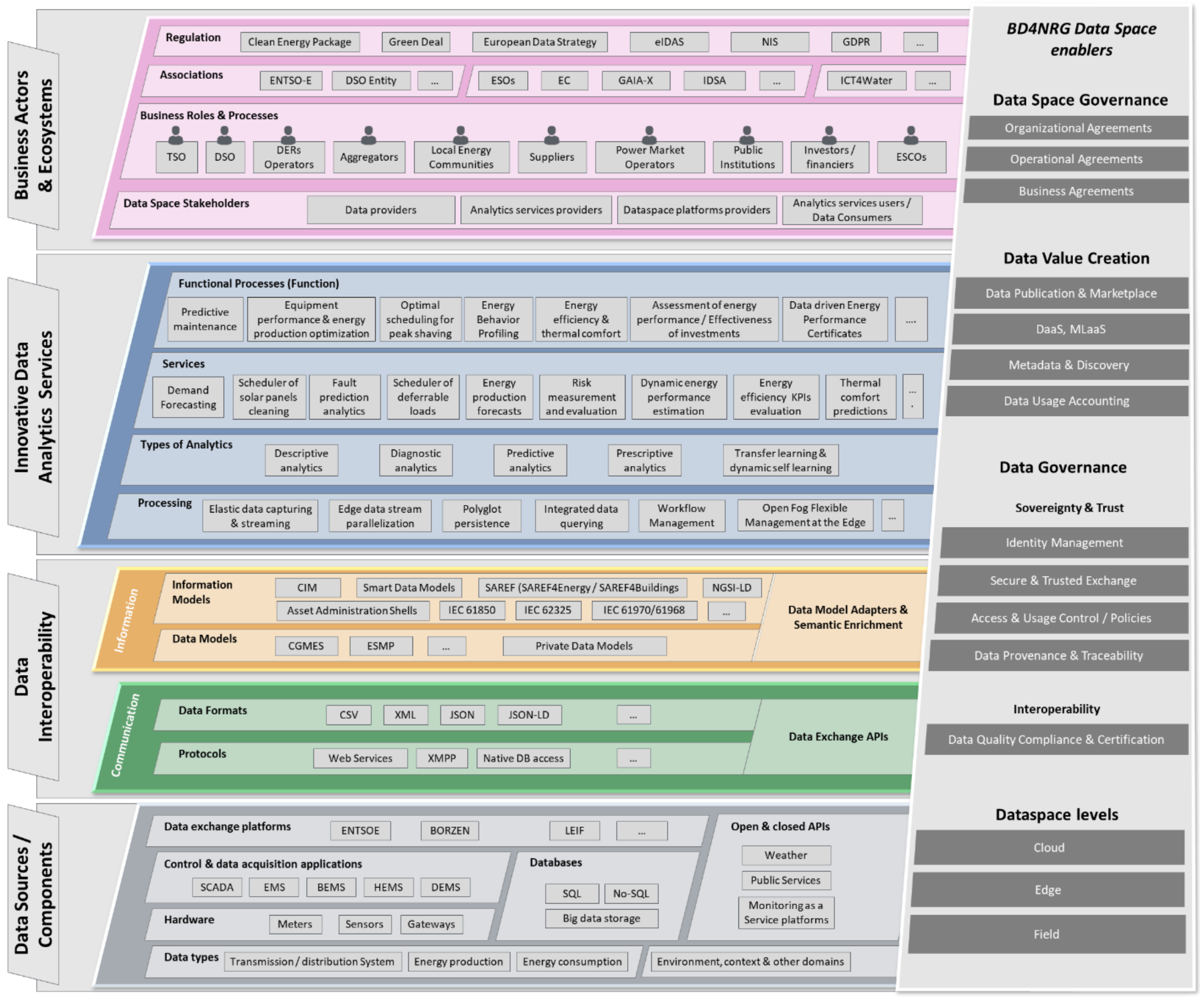


Figure 35 – Overview of the BD4NRG Reference Architecture

5.2 BD4NRG-Analytics Services

Big Data Analytics (BDA) provide opportunities for extracting unused data and combine them in computational models to be used to support real-time decision making, to identify patterns and trends, to gain insights, to make forecasts, to tackle complex problems.

Beside advanced algorithms to process data, Data Analytics requires tremendous computing power, and are therefore usually performed on the cloud (cloud computing). Cloud computing shares computer resources on the Internet instead of using software or storage on local computers. Computer resources are placed in a number of locations where those resources are run simultaneously in a computer group. This method is used to create analytics that run rapidly and are capable of performing time- and power-consuming data processing. Cloud Computing technologies provide global coverage used for monthly and yearly data storing; they are modelled for Big Data storage, where the data are stored in logical pools, that can be accessed remotely by end uses.

Nevertheless, increasingly attention is devoted to performing computations as close to the device as possible, relying on the so-called Edge Computing technologies, because of the security advantages over cloud computing. The Edge Computing represents the decentralized computing service for storage, processing and applications. It takes place on the network edge and acts as a middle layer

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

between end user and cloud data centres, thus reducing the distance data must travel on the network while producing minimal delays. The main barrier for the Edge Computing is related to limited capability to store data and nowadays data are in constant growth. This implies that besides Micro Data storages, the Edge Computing technology must support several types of storage, from ephemeral to semi-permanent, covering the wider range of local geographical area for longer period of time.

BD4NRG aims to develop data analytics services capable to analyse data in a seamless and holistic fashion, across multiple data sources and in different languages, delivering libraries and algorithms for Big Data knowledge extraction, business intelligence, and energy analytics. Three releases of the Data Analytics Toolbox and the respective BD4NRG integrated web-based platform are scheduled to be delivered.

In view of this, existing efforts for Big Data Analytics (BDA) have been analysed to evaluate if those efforts can converge into BD4NRG reference architecture. Among the efforts analysed, ALIDA Big Data analytics platform[34] has been selected as a starting point for BD4NRG and it will be extended and validated in the energy domain, according to the needs that emerged in the requirements elicitation phase of the project. ALIDA embraces the paradigm of Big Data Analytics-as-a-service (BDAaaS), which represents the next evolution step of Big Data. Beside relevant cost savings, the provision of Big Data analytical capabilities using the cloud delivery model could ease the adoption of the toolkit and could simplify useful insights with different kinds of competitive advantage. BDAaaS consists of as a set of automatic tools and methodologies that allows users lacking Big Data expertise to manage BDA and deploy a full Big Data pipeline addressing their goals: taking as input users' Big Data goals and preferences BDAaaS, it returns as output a Big Data pipeline applications ready-to-be-executed on premise/in the edge. ALIDA supports full orchestration of BDA application workflows and allows for composition, deployment and execution of workflows (either batch or stream) of BDA applications. Moreover, ALIDA addresses declarative, predictive, and prescriptive real-time data analytics as requested by the project.

Below is a review of the efforts analysed.

IDSA

The International Data Spaces Association (IDSA) aims at promoting the International Data Spaces Reference Architecture Model to establish an international standard. The International Data Spaces (IDS) is a virtual data space leveraging existing standards and technologies, as well as governance models well-accepted in the data economy, to facilitate secure and standardised data exchange and data linkage in a trusted business ecosystem.

Among its strategic requirements, IDS intends to meet value adding apps. The IDS allows to inject apps into the IDS Connector. Data apps can be used by Connectors for specific data processing or data transformation tasks. Those apps include services for data exchange protocols as well. Data analytics services can be provided by remote execution of algorithms.

[34] https://home.alidalab.it/

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

The App Store provides the Data Apps that can be deployed inside the Connector. The App Store is responsible for managing information about Data Apps offered by App Providers and provides interfaces for publishing and retrieving Data Apps plus the corresponding metadata.

On a conceptual level, Data Apps can be treated the same way as data offerings in the International Data Spaces. Therefore, just as data is provided by a Data Provider using a Connector and registering this Connector at a Broker, Data Apps are created by an App Provider and registered at an App Store. Consequently, App Providers also need to undergo the Onboarding process. However, instead of registering their Connector at a Broker, App Providers register their Data Apps at the App Store.

To be published, certain Data Apps require certification from the Certification Body.

GAIA-X

GAIA-X as a digital ecosystem strives for setting the Rules and Standard for Secure (European) cloud services, which satisfies European Data and Intellectual Property. That means providers must ensure their customers can easily switch their data from one cloud provider to another without losing any of their data. And GAIA-X also wants to offer their customers mutual use of different vendors' services, e.g., analytics. Various Working Groups are working intensively to define the GAIA-X standard procedures, rules and requirements, on the basis of which services can be developed and provided. GAIA-X is in early-stage development. Therefore, so far GAIA-X has made no special effort in terms of analytics services.

Knowage

In the FIWARE ecosystem, Knowage is the open-source suite for business analytics over traditional sources and Big Data systems. FIWARE has removed the Knowage Generic Enabler from its catalogue and added to the FIWARE Marketplace as it wishes to follow a different commercial business model.

It provides comprehensive support services for multi-source data analysis, such as data federation, mash-up, data mining, and advanced data visualization.

The suite is composed of several modules designed for a specific analytical domain and they can be used either individually, as complete solution for a certain task, or in combination with one another.

The modules of the suite are:

- BD (Big Data), which allows to analyse data stored on Big Data clusters or NoSQL databases
- SI (smart intelligence), which is oriented to self-service capabilities and agile prototyping and provides the features for business intelligence on structured data
- ER (enterprise reporting), which gives the opportunity to work with traditional data sources to produce static reports, usable with a full set of parameters and producing different outputs
- LI (location intelligence), which allows to relate business data with spatial or geographical information
- PM (performance management), which provides functionalities to measure and evaluate business performances

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

- PA (predictive analysis), which gives the opportunity to work with traditional data sources or external files, processing them with advanced algorithms written using R/Python and other scripting languages for more advanced analyses

Predictive Analysis refers to the ability to perform advanced processing using data mining techniques for forecasting and prescriptive purposes. It also concerns the ability to simulate actions and evaluate their effects on different assets. For what-if purposes, Knowage PA provides an On-Line Analytical Processing (OLAP) based solution that allows the interactive simulation process over measures and dimensions. It allows to simulate new values at any level and evaluate the effects through every dimension, by comparing different scenarios and versions and to explore and navigate data on different detail levels and from different perspectives and to. Knowage gives autonomy to the end-user by providing new self-service capabilities: the user can build his own analysis and explore his own data space, also combining data that come from different sources.

ALIDA[35]

BD4NRG aims to develop data analytics services capable to analyse data in a seamless and holistic fashion, across multiple data sources and in different languages, delivering libraries and algorithms for Big Data knowledge extraction, business intelligence, and energy analytics. Three releases of the Data Analytics Toolbox and the respective BD4NRG integrated web-based platform are scheduled to be delivered. In that sense, the ALIDA is a Big Data analytics platform has been selected as a starting point for this activity and it will be extended and validated in the energy domain, according to the needs emerged in the requirements elicitation phase of the project. ALIDA, which is based on advanced and open-source technologies, allows for the design, deployment, optimisation, execution, scheduling, and monitoring of Big Data analytics batch and/or stream workflows. Figure 36 provides an overview of ALIDA functionalities, that is:

- Streaming and Batch data workflow processing
- Catalogue of Big Data Analytics (BDA) services (covering ingestion, preparation analysis, visualization) for various data analytics scenarios
- Graphical editor for building data workflow
- Data pipeline deployment by means of modern resource orchestrators such as Kubernetes
- Workflow execution monitoring
- Graphical User Interface to easily register and manage new custom BDA services

[35] *https://home.alidalab.it/*

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

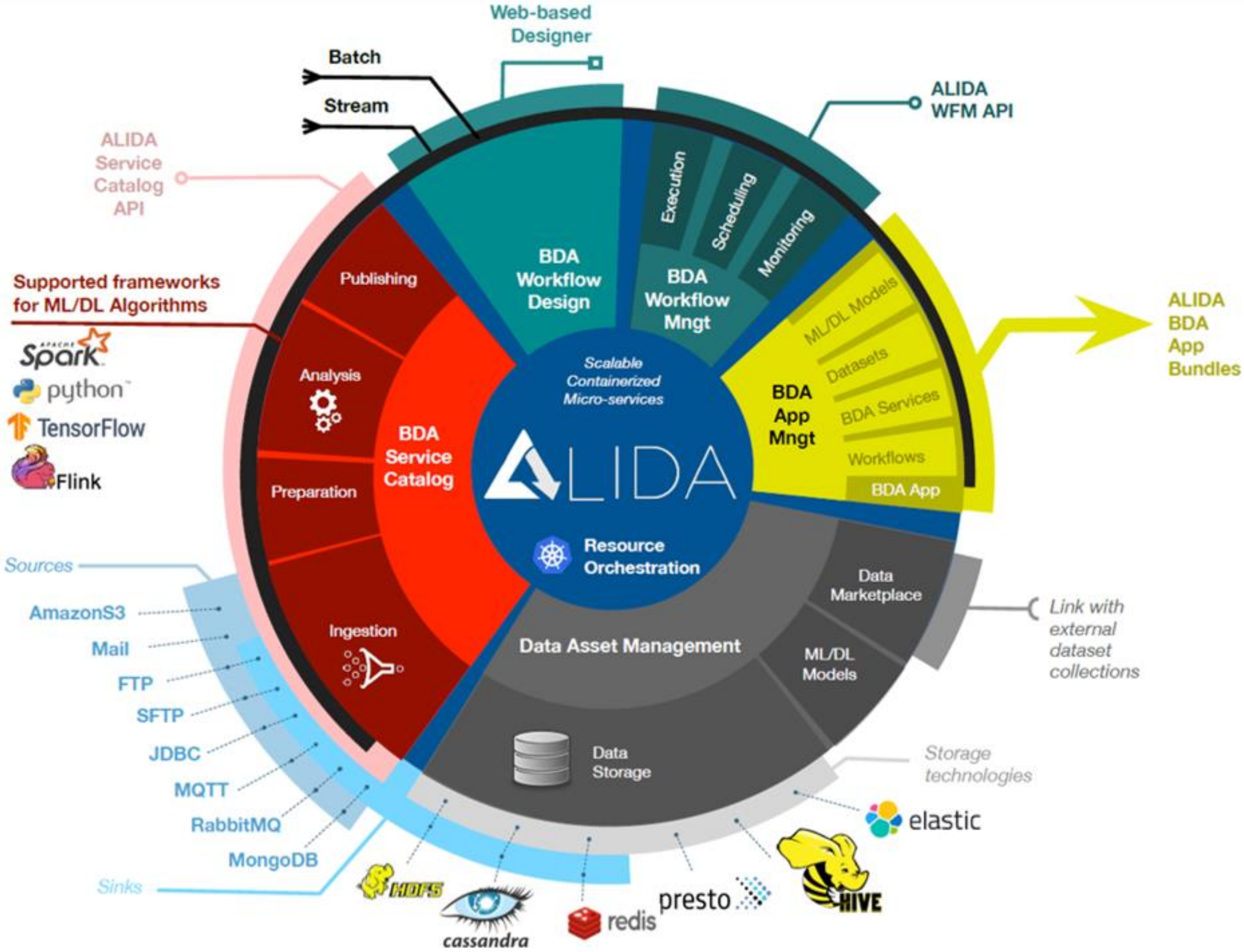


Figure 36 – ALIDA functionalities overview.

Figure 37 depicts the ALIDA architecture. ALIDA is cloud-native and able to scale computing and storage resources thanks to the workflow orchestration engine based on Kubernetes.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

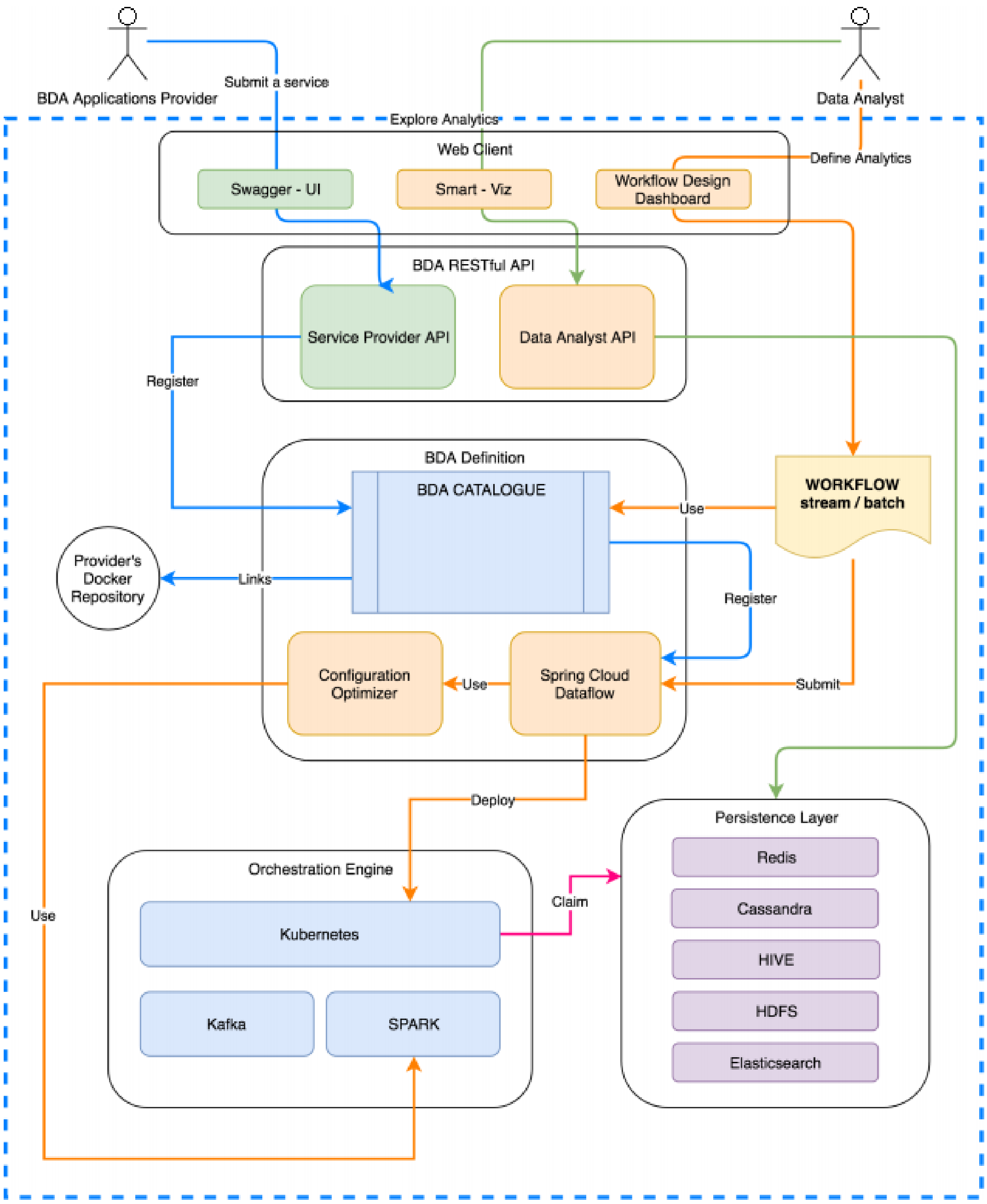


Figure 37 – ALIDA general architecture.

To facilitate the design of Big Data analytics workflows, ALIDA provides an extendable catalogue of BDA services (the building blocks of a BDA workflow). Such services may cover all phases of Big Data processing, from ingestion, to preparation, filtering, merging, cleaning, etc. to the analysis (mainly machine learning and deep learning analysis) and publishing (preparation for the visualisation of data resulting from the workflow). Data may be streamed from multiple sources (FTP, JDBC, MQTT, mail, etc.) trough sink connectors, such as Cassandra, Redis, HDFS, etc. Several frameworks are supported by ALIDA for the analysis, such as Spark, Flink, TensorFlow. The catalogue may be used either through a web interface or through specific APIs (allowing the registration of new BDA services, execution, scheduling, and monitoring). Once a workflow has been created and executed, a whole set of assets is generated in the ALIDA environment. The workflow and the related generated assets take the name

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

of BDA Application bundle. ALIDA also provides features for the management of these assets. A BDA App designed for BD4NRG is a set of assets user may manage, composed of: workflow(s), a set of BDA services composing the workflow, input dataset, intermediate dataset and resulting dataset, and trained ML model, if requested in the workflow.

Moreover, it offers the Configuration Optimizer to improve BDA applications metrics and the Smart-Viz for result visualization.

5.3 BD4NRG-Data Governance

Generally, we can identify Data governance as the process of managing the availability, integrity, usability & security of the data, based on data standards and policies that also control data usage. DG ensures that data is secure, consistent and trustworthy[36]. It's increasingly important as organizations face new data privacy regulations and rely more and more on data analytics to help optimize operations and drive business decision-making. For BD4NRG and mainly in WP3, the Data Governance (BD4NRG-GOVERNANCE) and Big Data Management and Modelling (BD4NRG-PROCESSING) Layers will be developed, as well as all of their subcomponents, to be used as the middleware layer, on top of the common IP infrastructure, to enable the exchange of information between assets, systems, data hubs and actors in a common manner. It is clear that this effort will be greatly assisted and influenced by the concept of Data Spaces from a theoretical but also an architectural/implementation point of view.

According to Design Principles for Data Spaces position paper[37]: “The centrepiece of the European Data Strategy is the concept of “data spaces”, for which the Commission defined nine initial domains, all driven by sector-specific requirements. From a technical perspective, a data space can be seen as a data integration concept which does not require common database schemas and physical data integration but is rather based on distributed data stores and integration on an “as needed” basis on a semantic level. Abstracted from this technical definition, a data space can be defined as a federated data ecosystem within a certain application domain and based on shared policies and rules. The users of such data spaces are enabled to access data in a secure, transparent, trusted, easy and unified fashion. These access and usage right can only be granted by those persons or organisations who are entitled to dispose of the data.”

By the above we can surmise that any Data Space initiative (e.g. IDSA or IDSA in conjunction with FIWARE) has as a mission to leverage existing standards and technologies, as well as Data Governance Models, in order to facilitate and establish a secure, standardized and sovereign system of data exchange in a trusted Business Ecosystem. As an example, IDS form a basis for creating Smart Service Business Processes and at the same time helps participants realize the full value of their data. In the BD4NRG context, a data space by defining and delivering a reference architecture for a data ecosystem, will guarantee a correct BD4NRG Data Governance implementation based on the security, sovereignty and trust principles. A Data Space realizes an envisioned Data Governance model by connecting data clouds/hubs, enterprises, business applications, IoT devices, data owners, data consumers and data providers thus creating the data ecosystem. Based on the above, BD4NRG Governance principles as they will be defined by functional specification and user requirements can be developed as a clear set of shared rules in a data space architecture, which will determine at any given time the rights and obligations for data management.

[36] https://searchdatamanagement.techtarget.com/definition/data-governance
[37] https://design-principles-for-data-spaces.org/

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Integration of IDSA, GAIA-X and FIWARE approaches to Governance is an ongoing topic of conversation between the respective organizations themselves (see Section 4.2.5). Unfortunately, at the time of submission of this document, no comprehensive result is available yet. It will be integrated in BD4NRG deliverable D2.6 (“BD4NRG Reference Architecture Final Version”).

5.4 BD4NRG-Marketplace

This section will describe the relation between existing marketplace layer efforts and their position within BD4NRG. Specifically, the Marketplaces designed by ENG and ATOS (AGORA) come into focus. While ENG has designed the tokenized marketplace specifically from the very beginning to support the energy and the flexibility trading, the AGORA marketplace was originally not tailored to the energy domain and is being adjusted accordingly at the time of submission of this document.

ENG and AGORA marketplaces’ relation with the other existing efforts will be discussed here, with the goal of encouraging interoperability by pointing out commonalities, differences and identifying possible synergies.

BD4NRG deliverable D2.6 will contain the final Reference Architecture, including the final BD4NRG-Marketplace Layer, which will combine elements and innovative aspects related to heterogenous energy and non-energy assets trading and reciprocal compensation via adaptation and maturation of multilateral tokens and smart contracts. Insight from this document and the coming developments within and outside of the BD4NRG project will be used to converge towards this final version.

5.4.1 ENG Tokenized Multi-sided Blockchain-based Marketplace

One of the technology assets which will be leveraged on, matured and integrated withing BD4NRG is the ENG blockchain-based multi-sided flexible marketplace. ENG has been one of the pioneers to work on DLT/blockchain and smart contracts to support both decentralized data handling and exchange as well as decentralized marketplace to coordinate peer-to-peer energy and/or flexibility trading within a variety of H2020 projects (eDREAM, SOFIE, Coordinet, PlatONE). Such marketplace technology asset will be operated at the local grid level by any grid network operator (e.g. DSO), third party market operator, or aggregator in the case of flexibility trading and optimal aggregation, allowing any prosumer to directly participate in the market trading sessions. This is of great importance to integrate many small-scale distributed energy prosumers to exploit cooperative procurement models. The market will match consumers with energy producers and will rely on energy tokens for representing the energy as a digital asset in the energy market. It will enable controlled access to data from different sources offering a data catalogue and a discovery service to support Service Providers or Data Consumers an easy access and detection of needed data.

Energy tokens are generated at a rate proportional with the forecast energy production at the micro-grid level, transforming the energy into a transactable digital asset. The producers and consumers will then use tokens to participate in the electricity market sessions, leveraging on self-enforcing smart contracts. The digital marketplace leverages on self-enforcing smart contracts to manage, in a programmatic manner, the P2P energy trading between distributed energy prosumers (see Figure 38). During a market session, each prosumer will submit bids and offers (i.e. from their contracts) representing the amount of energy they are willing to buy or sell. The use of smart contracts will allow

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

the participants to automatically submit bids and offers; validity checks ensure that market session rules are not violated; the evaluation is performed by the smart contracts.

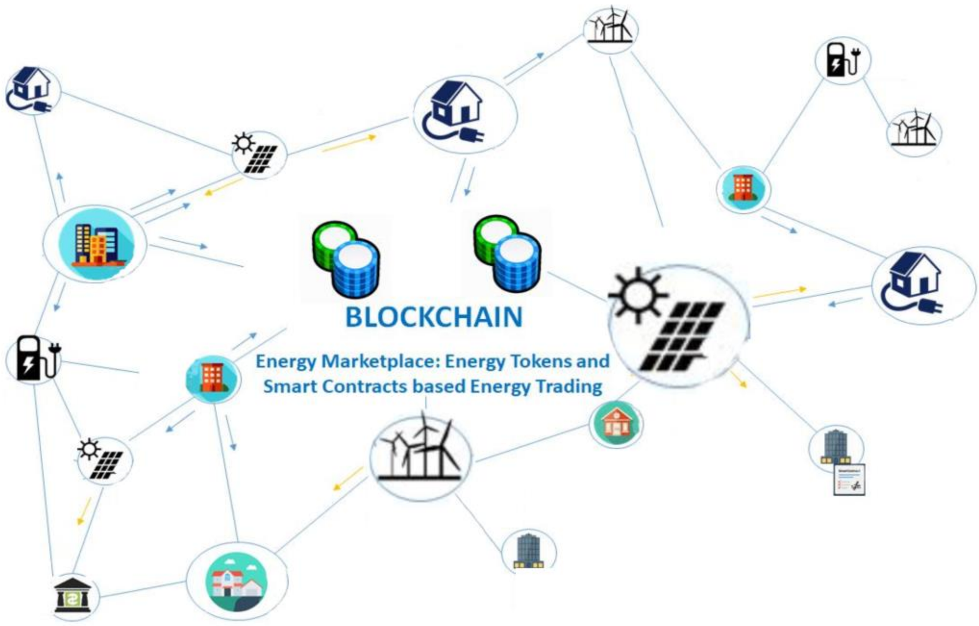


Figure 38 – Marketplace ecosystem

The architecture of the component and the flexibility of the model allows to extend it to capture beyond economic dynamics of energy consumers/prosumers and allow them to lay the foundations for facilitating and enabling a multi-dimensional trading off and reciprocal value compensation of energy versus data and computing resources. Hence, we envisage that the innovative maturation of ENG marketplace will incorporate innovative yet configurable multi-dimensional energy consumers preferences and incentives typologies within smart contracts and new different domains like the energy domain.

5.4.2 AGORA Marketplace

The Atos Marketplace consists of a cloud-based concept for a Big Data Cloud Storage and a Big Data Marketplace, derived from the results of the CROSS-CPP project, which enables controlled access to data from different sources offering a data catalogue and a discovery service to support Service Providers or Data Consumers an easy access and detection of needed data. The Cloud Storage and the Marketplace were conceived to support the brokering of brand-independent data from the different IoT devices in the domain of the Connected Vehicle and Home Building platforms but is flexible enough to incorporate data coming from various industrial sectors.

The Big Data Marketplace data flow is shown in Figure 39. The Marketplace acts as a mediator for all actors involved in the flow of data. Data Owners and Data Providers manage their data resources and control the data sharing. Service Providers or Data consumers configure their data set requests to collect diverse data to provide new services.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

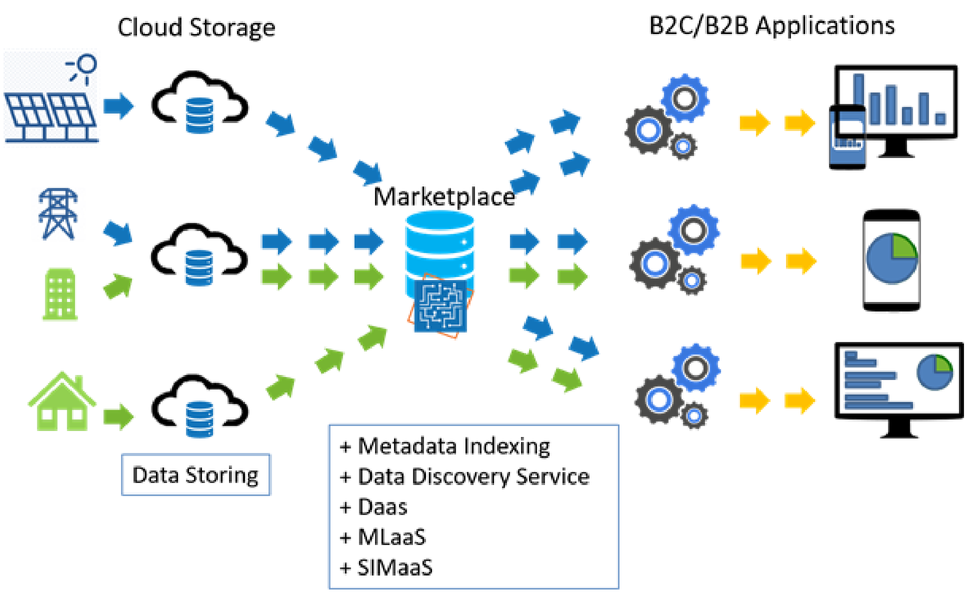


Figure 39 – Marketplace Data Flow.

The Big Data Cloud Storage enables data gathering from diverse sources and notifies the Marketplace when new data arrives, prevents data from unauthorized manipulation, ensures data completeness, as well as the trustworthiness of the data sources and the protection of the privacy of data owners, data sovereignty, etc. The storage infrastructure stores data in a brand-independent, open, and transparent data model represented by the Common Industrial Data Model (CIDM)[38]. A Storage Provider needs to setup and offer a certified data storage that consist of a storage infrastructure and application programming interface (API) to enable data reception from the data sources. The Big Data Marketplace collects data notifications to index data-package metadata and its location to provide the discovery service and enforces data privacy and security of the data sharing. The Marketplace also has to handle the certification of the various storage providers as well as to administrate the certified data catalogue.

The architecture of the component and the flexibility of the model allows to extend it to new different domains like the energy domain.

5.4.3 BD4NRG Marketplace Architecture

BD4NRG will integrate and combine ENG tokenized mechanisms managed via DLT/blockchain and smart contracts with some of the functionalities offered by the AGORA marketplace. BD4NRG Marketplace also seeks to support other existing data models such as “smartEnergy” FIWARE model, supporting NGSIv2 and NGSI-LD. The proposed architecture of the BD4NRG Marketplace is described in D2.2 Section 5. It comprises

- Energy Assets Data Access Layer that allows index metadata in Common Data Model (CDM) format and in FIWARE Information Model natively from different sources,

[38] https://www.cross-cpp.eu/cidm

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

- the Orchestration Layer that will oversee the management of the access and the metadata matching among data/service providers and consumers,
- the Blockchain Based Asset Management Layer will be in charge of tokenization functionality to create a utility monetary and/or non-monetary reward/incentive metric/function,
- the Service Layer includes the MLaaS that will provide a library of reusable AI-based Machine Learning models offered as a service, the SIMaaS that will facilitate access to simulated data testbeds, the DaaS that will include energy domain data as well as to off-domain data sets like weather and geographical data, and third-party services through interoperable service level Open APIs.

The User Management and Resource Management component allows to manage user and resources, and the permission configuration of the owned resources and data.

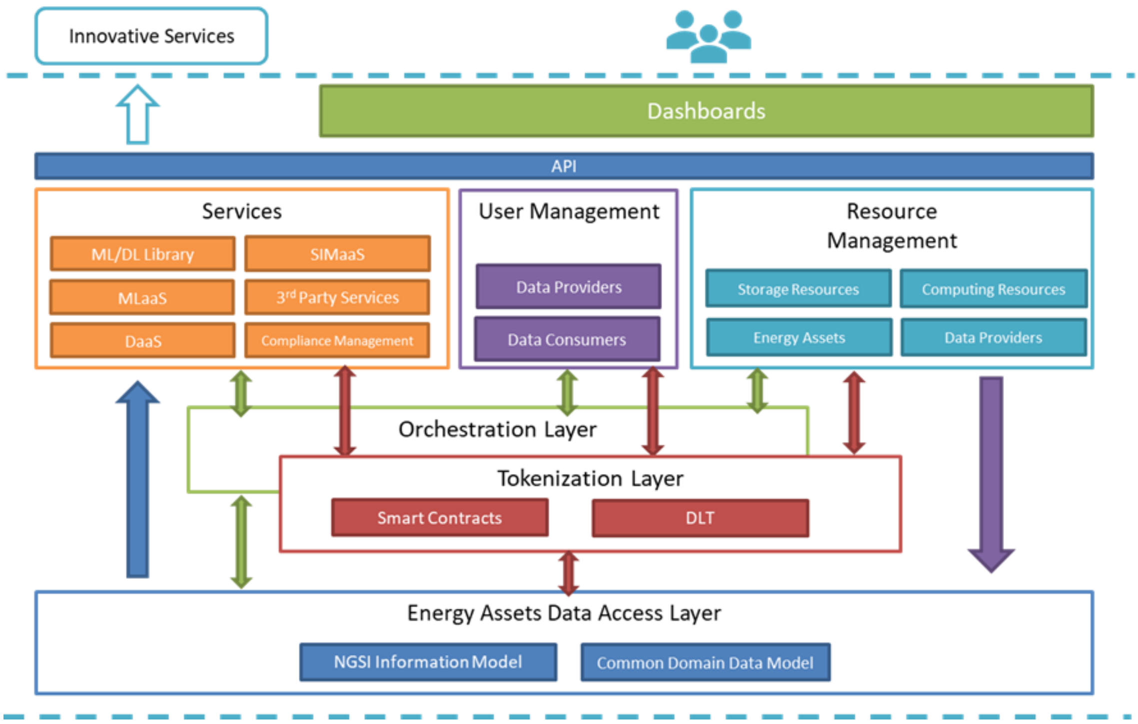


Figure 40 – BD4NRG Marketplace Architecture.

5.4.4 Relation to Other Approaches

The marketplace architecture is a work in progress. It is based on a proven and consolidated platform for sharing data which is now focused on the energy domain and its new requirements. This architecture seeks its alignment with the BD4NRG reference architecture and for this we investigate different aspects of it and its relationship with the different approaches previously analysed.

ENG and ATOS data marketplace architecture is decentralized in terms of the data cloud storages and centralized in that the marketplace provides a data discovery service based on data stored in cloud data storages. The cloud-based architecture allows the marketplace to be federated by adding federation services to the marketplace functionality.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

In terms of interoperability, the initial design is based on a common data model and a standardized API that ensures the interoperability of data that comes from various sources. The rise of FIWARE, its context broker and its “smartEnergy” model that supports NGSIv2 and NGSI-LD suggests the use of a context broker as a valid implementation of cloud storages that complements the implementations provided by the original environment as a good alternative.

From the governance point of view the data-driven services require enhanced data management capabilities to regulate policies that must be enforced during data space operation. The management of data-related resources is assured by means of the data sharing configuration that allows data owners give explicit permissions to data consumers to use their data and resources. To support the data governance and data sovereignty the marketplace includes accountability functionalities (Blockchain Based Asset Management Layer) and usage policies and contracts (User and Resource Management components).

The roles of the BD4NRG marketplace are: Data Owner, Data Provider, Cloud Storage Provider, Marketplace Operator, Software provider, and Service Providers/Data Consumer. The Data Owner is aligned with IDS Data Owner role as it is a legal entity or natural person creating data and/or executing control over it. The Data Provider role is aligned with the IDS Data Provider since it makes data available for being exchanged between a Data Owner and a Data Consumer. The Cloud Storage Provider makes available decentralized data storages for the Data Owners and it is aligned with the third-party data apps. The Marketplace Operator role is aligned with the IDS Governance Body since it takes care of the certification process, issuing certificates regarding organizations that want to participate and with regard to software components that are going to be used. Also, the Marketplace has the responsibility to provide the Information Model of the Marketplace and acts as the Broker Service Provider and is an intermediary that stores and manages information about the data sources available. The Identity Provider of the Marketplace offers a service to create, maintain, manage, monitor, and validate identity information of and for participants in the marketplace. The Software Provider role is in charge of providing software services for implementing the functionality required and orchestrated by the Marketplace. The Service Provider/Data Consumer roles are aligned with the IDS Data Consumer and Data User roles, as they receive data from Data Providers according with the data sharing configuration assuring data sovereignty.

The development approach of the marketplaces includes an agile development process that includes Continuous Integration and Continuous Deployment (CI/CD) practices to narrow the gaps between development and operation activities. The practices include the use of Docker and Kubernetes to deliver cloud native applications at high velocity.

Data Quality aspects will rely on the Data Quality Compliance and certification managed by the data governance layer (BD4NRG-GOVERNANCE), with a view to manage the compliance of data assets to be federated to the adapted COSMAG-compliant FIWARE Context Broker, as specified within the Data Integration & Homogenization Layer.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

6 Implementation Recommendations Based on LFE Projects

In this section a suitable selection of Linux Foundation Energy (LFE) projects is briefly summarized in terms of their use cases as well as their functional and implementation architecture.

The LFE itself is an open-source foundation focused on the digitalization of power systems by providing frameworks and reference architectures. Architectural concerns include security, interoperability, control, automation, virtualization, flexibility, orchestration, and balancing supply and demand.

With these aspirations in common with the BD4NRG reference architecture, implementation recommendations in furtherance of the latter are derived in the last subsection.

6.1 SOGNO[39]

SOGNO (Service-based Open-source Grid automation platform for Network Operation of the future) is a platform for development of data-driven services for DSOs, with the goal of open-source implemented services that are available not only for the network operator, but also for the actors that will be part of the new economy driven by the smart grids and future networks.

In SOGNO the automation functions can be developed both as part of the core components of the project or as separate projects that aligns with the requirements established for the interfaces, that are run as containerized microservices. The user can then decide how the SOGNO services fit their use cases and select the ones which wants to integrate to an existing platform (or a new development) or use the full suite.

SOGNO uses orchestration with Kubernetes to deploy the microservices in containerized environments. Such a practice brings the benefits of cloud environments and containers:

- Scalability and easy replication
- Optimization of use of hardware resources
- Isolation of the environment
- Portability and reusability

At first glance, the adopted premises give the automation system a flexible access from the cloud or in an on-premises, allowing integration of multiple suppliers of measurement devices and algorithms. This allows higher interoperability and greatly reduces the risk of vendor lock-in.

[39] Source: https://www.lfenergy.org/projects/sogno/

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

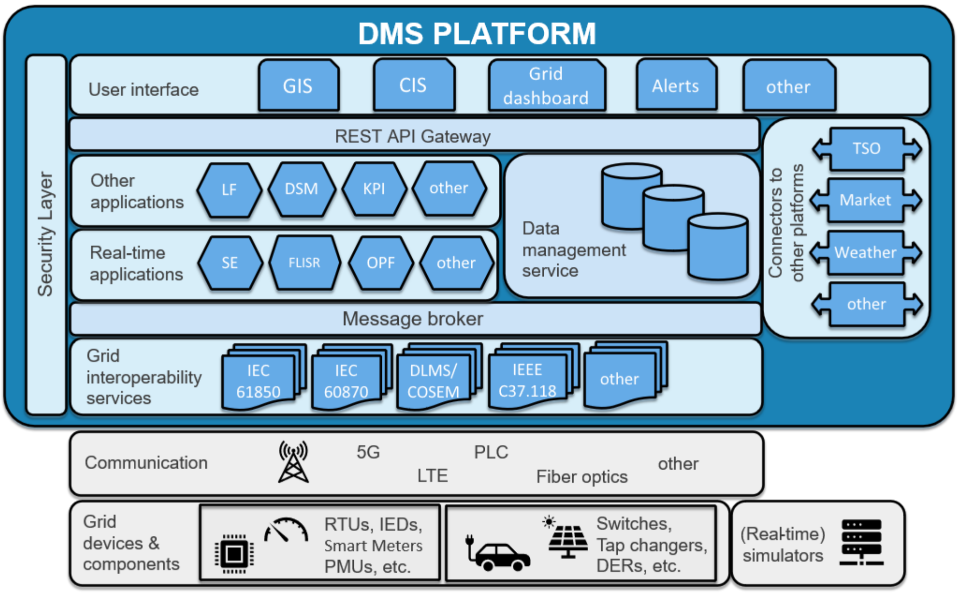


Figure 41 – The SOGNO DMS Platform.

With the requirements for decarbonization for the new grids, a new DMS platform, as seen in Figure 41 is envisioned as the most suitable answer for the challenges ahead. Such a solution will allow the DSOs to adapt faster, as the components do not need to be updated all at once. Industry standards that are widely adopted like IEC 61850 and CIM IEC 61970 are a core part of the driving force of the vision of integrating today's SCADAs gradually with SOGNO and migrating the functionalities into more flexible and powerful microservices.

Moreover, SOGNO does allow the use of real-time simulation. Hence, a seamless integration of development, testing and deployment are possible within the platform. New automation ideas and functionalities can be tested against a virtual representation of the power system and against those same microservices that are being used in the production environment, saving costs and ensuring high quality products.

6.2 Grid Exchange Fabric[40]

Grid eXchange Fabric (GXF), formerly Open Smart Grid Platform (OSGP), provides a web-based IoT platform allowing the control and monitoring of smart objects in the smart grid. Platform services are divided into functional domains (among others public lighting, smart metering, power quality, traffic control) that can be extended by third parties.

As depicted in Figure 42, GXF has a layered architecture and follows the design principles of Domain-Driven Design (DDD), the Dependency Inversion Principle (DIP) and Behaviour-Driven Design (BDD). In the following the layers are briefly summarized. Thereafter, the design principles and their underlying rationale are analysed.

[40] Source: https://documentation.gxf.lfenergy.org/index.html

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Layers

The topmost web services layer interfaces with external applications with authorization per functional domain. Below is the Domain Layer where the business logic of the functional domain is translated to one or more generic commands. For instance „turn traffic light on“ would be translated to „set switch 1 to closed position“. The next layer is the OSGP core layer that implements command functionality (device management, scheduling, and more). The last and lowest layer contains protocol adapters that translate the intermediate and generic format of OGSP commands to a format that connected devices understand.

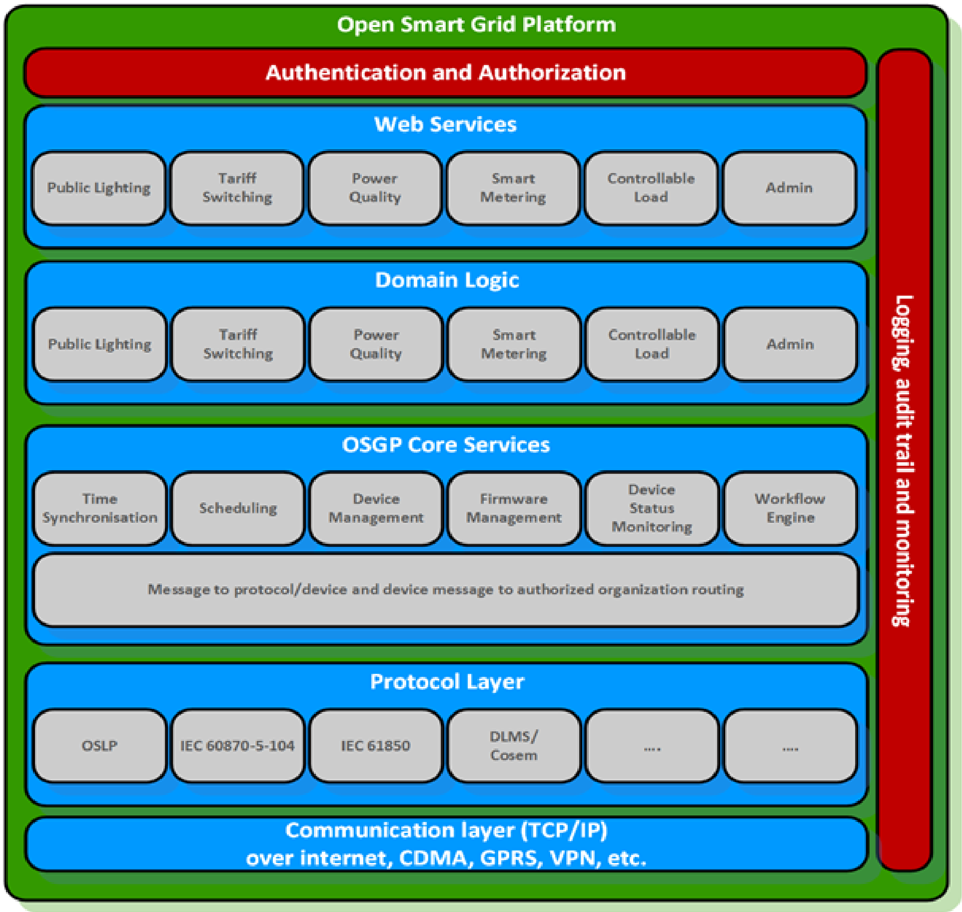


Figure 42 – The layers of the Grid Exchange Fabric (formerly known as Open Smart Grid Platform).

Design Principles

The core idea of DDD is close cooperation between technical and domain experts such that the structure and API of the program are well-understood by both businesses and developers. Apart from usability, further technical advantages can be attained for instance high maintainability due to typically fine-grained modularity and through the use of DDD building blocks. However, increased cost by inter-module communication and linguistic processing must be justified by adequate domain complexity.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

The essence of the Dependency Inversion Principle is that modules up higher in the hierarchy must not depend on ones lower in the hierarchy. This can be achieved by implementation against interfaces instead of against non-abstract classes.

Behaviour-driven development is about capturing the user stories and scenarios in the specification and translating this into code. User requirements formulated as acceptance criteria can be used for automated testing which also supports Test-Driven Development.

Design Choice: SOAP vs REST[41]

The web services layer uses SOAP and the following rationale for the choice is given. REST is data-oriented whereas SOAP and OSGP are function-oriented. SOAP features contracts in the form of web service definitions in the Web Services Description Language (WSDL). SOAP offers integrated security and compliance features. SOAP is older than REST and thus viewed as more reliable by less progressive companies.

Advantages of REST APIs include that they are more swiftly developed as REST guidelines provide more flexibility. SOAP responses are only available in XML whereas REST responses can be in HTML, XML, plain text, or JSON which is, in fact, the preferred format. REST web services are less complex to build and many publicly accessible APIs such as Google Maps are often RESTful.

6.3 RIAPS[42]

The Resilient Information Architecture Platform for Smart Grid (RIAPS) is a component-oriented software platform for building real-time embedded applications. Please refer to Figure 43. Apart from run-time components it includes a design-time software environment. Among others, the applications are monitoring and control, data collection and processing, microgrids, energy management, transactive energy and safety.

Architectural Principles

The main architectural characteristic is the distributed computing model as software applications are distributed across computing nodes on a communication network where each nodes accesses local sensors and actuators. Each application features multiple actors that interact with other actors in the communication network thereby realizing functionality.

Each actor encapsulates a multitude of RIAPS run-time layers including the component framework that facilitates distributed applications, the resource management framework, the fault management framework, the security framework, the time synchronization service, the coordination framework.

[41] Sources: https://www.soapui.org/learn/api/soap-vs-rest-api/ and https://www.soapui.org/learn/api/soap-vs-rest-api/

[42] Source: https://wiki.lfenergy.org/display/HOME/RIAPS+Architecture

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

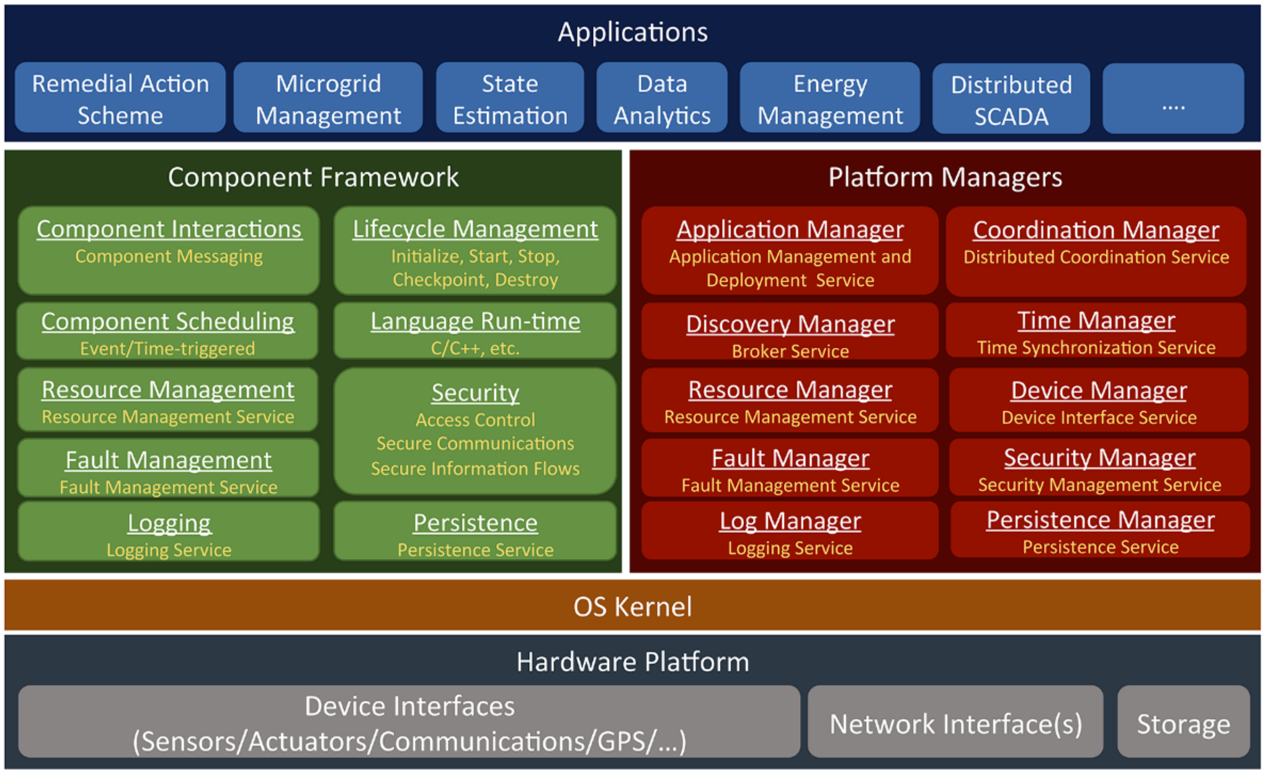


Figure 43 – RIAPS Architecture.[43]

The RIAPS communication network (see Figure 44) consists of a number of RIAPS Nodes and a RIAPS Control Node. Each RIAPS Node is a multi-tenant computing node, i.e., hosts several actors and can call the discovery/registration and deployment services. The “Control Room” on the RIAPS Control Node hosts the deployment control for scalability and administration. The nodes are firewalled from each other. Coordinated time-synchronized action is agreed upon in logical dynamic groupings of nodes.

[43] Source: https://wiki.lfenergy.org/display/HOME/RIAPS+Architecture

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

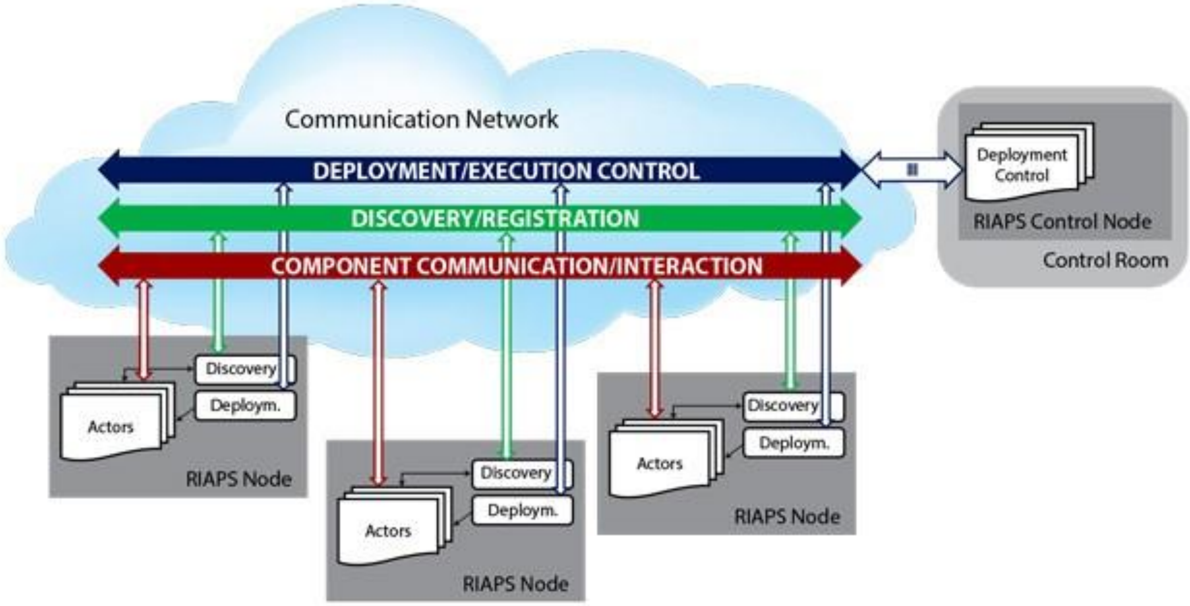


Figure 44 – Communication scheme in a RIAPS network.[44]

Secondly, precise timing is important for several use cases of this projects. These are synchronized distributed action (e.g., breaker activation), detecting network bottlenecks or Denial of Service (DoS) attacks, distributed coordination (group membership, leader election, time-synchronized coordinated action), and communication profiling and tracing. Thus, the nodes are aware of any delays introduced by network and software layers.

Design Principles

RIAPS provides a Model-driven Development Environment (MDE) that allows the developers to model their application in a Domain-Specific Modelling Language (DSML). The declarative specification of models in a DSML enables hiding complexity from developers and inserting additional validation and verification steps automatically. Leveraging this, developers can concentrate on solving power instead of software problems.

To be resilient against faults RIAPS follows the Detection, Isolation, Recovery Paradigm. In the detection phase an anomalous situation is detected. Subsequently, in the isolation phase the root cause is identified. Lastly, in the recovery phase action is taken to mitigate the fault effects. However, mitigation is within the responsibility of the application developer. Among the ten detected faults are (un)reported application process termination, operating system crash, network node/link failure et cetera.

Importantly, a number of security measures are taken to guard against security threats. All network communication is encrypted and tamper-protected. Strict access control and resource moderation

[44] Source: https://wiki.lfenergy.org/display/HOME/RIAPS+Architecture

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

further increase availability and yield additional protection against faulty and malicious code. Applications are deployed remotely and signed with cryptographic signatures.

6.4 SEAPATH[45]

SEAPATH (Software Enabled Automation Platform and Artefacts Therein) is a reference platform for real-time Digital Substation Automation Systems (DSAS). Virtualization is seen as the key innovation feature to provide more complex, dynamic, adaptive, flexible and scalable automation functions. Figure 45 shows the virtualization architecture. Such innovative DSAS would be able to handle infeed and elevated dynamics in flow patterns caused by distributed energy resources at lower grid levels. Key architectural concerns are reliability, performance and high availability of the platform.

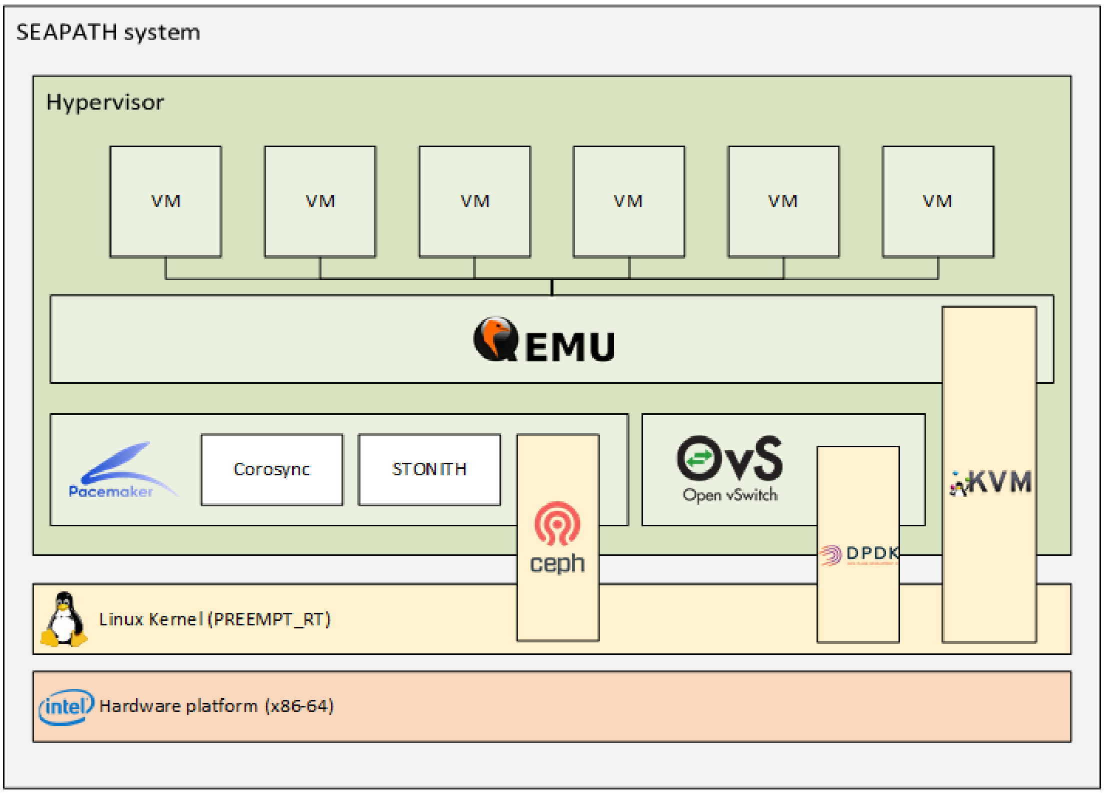


Figure 45 – SEAPATH Virtualization Architecture.[46]

The virtualization is done using QEMU (the well-known Quick Emulator) and KVM (a Linux kernel module that acts as hypervisor). Pacemaker is a high availability resource manager with clustering functionality. Optionally, it can be combined with Corosync (a cluster management system) and STONITH (a fencing tool, i.e., for isolation of failed node such that it does not further disrupt the cluster). Ceph is a distributed-storage tool for persistent storage within the cluster (used here to synchronize and replicate virtual machines (VMs) to guarantee availability and integrity). Open vSwitch and DPDK offer network automation and virtualization features, respectively.

[45] Source: https://wiki.lfenergy.org/display/HOME/SEAPATH
[46] Source: https://wiki.lfenergy.org/display/HOME/SEAPATH

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Design Principles

SEAPATH has the following security features: OS hardening, disk and connection encryption, storage and protection of secrets, process isolation, privilege management policies and user authentication processes.

To meet real-time requirements SEAPATH uses a fully preemptable Linux kernel. Additionally, the cyclictest tool is used to provide system latency statistics. Still, care has to be taken designing the applications and configuring networks to keep real-time promises.

To provide full configurability of the Linux kernel the Yocto project is leveraged. Once the SEAPATH Yocto images are deployed to the machines the automation and orchestration tool Ansible can be used to configure the cluster.

6.5 Fledge[47] & FledgePOWER[48]

With the ongoing evolution of modern power systems to include more and more renewables, electrical vehicles, edge and behind-the-meter assets network operators are put into the position to interact with high-volume, and volatile data. The proliferation of protocols used by the vast assortment of assets is a substantial obstacle in achieving this objective. FledgePOWER is a multi-protocol translation gateway for power systems that attempts to address this problem. Furthermore, FledgePOWER provides a toolbox for simulation, data configuration, and checking focused uniquely on power systems' protocols translation and power systems' use cases.

The amount of publicly available information on this project is rather limited. However, this is different for the Linux Foundation Edge project Fledge that FledgePOWER embeds. Therefore, the following description pertains to Fledge. Fledge is a platform framework to deal with fragmented and distributed data at the Industrial Edge. With Fledge any industrial IoT device, sensor or machine can be integrated with the cloud and existing "brown field" systems like Distributed Control Systems (DCSs), Programmable Logic Controllers (PLCs), SCADA (Supervisory Control and Data Acquisition). Target audiences include industrial equipment vendors, system integrators and operators. Some selected use cases are the development of learning and self-maintaining machines, eliminating data complexity and fragmentation as well as predictive maintenance. In fact, an electric utility company and Fledge client uses the framework for predictive maintenance of power transformers[49].

Fledge Architectural Principles

The Fledge platform as presented in Figure 46 has a microservice architecture where individual microservices are represented by green boxes in Figure 2. The blue boxes represent plugins that may be included to extend the functionality of Fledge. Notably, plugins can be written either in Python for swift development or in C++ for high performance. In the following a brief description of the microservices is given.

---

[47] Source: https://fledge-iot.readthedocs.io/en/latest/fledge_architecture.html
[48] https://wiki.lfenergy.org/display/HOME/FledgePower
[49] https://wiki.lfedge.org/display/FLEDGE/Use+Cases

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

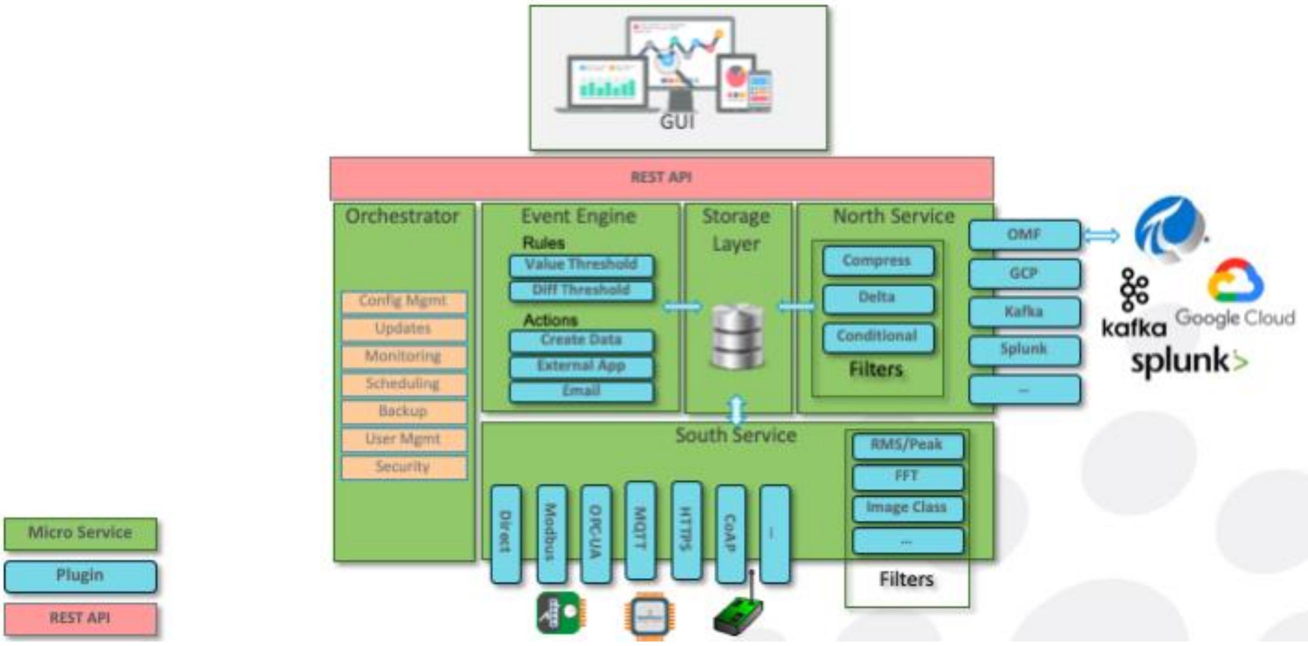


Figure 46 – Fledge architecture.

The Fledge Core (“Orchestrator” and “REST API” in the picture) is a microservice that coordinates all Fledge operations and of which only one instance can be active at any time. Its functionalities include the scheduler, configuration management, monitoring (including self-healing capability in case of unresponsive microservices), the REST API (exposing external management and data APIs for cross-component functionality), backups, audit logging, certificate storage, user management and asset browsing.

The storage layer is responsible for the maintenance of the Fledge configuration and run-time state, as well as the storage and buffering of asset data. Different plugins can be used as storage engine, for instance SQLite for minimal footprint or PostgreSQL with a larger footprint. On Edge devices, in-memory temporary storage may be used instead.

Southbound microservices offer bi-directional communication of data and metadata between Edge devices and Fledge. Northbound microservices offer bi-directional communication of data and metadata between the Fledge platform and larger local or cloud platforms. These larger systems could be private or public (and proprietary) cloud services.

Filters are plugins that modify data streams. These filters may be deployed at the ingress (in southbound services) or egress (in northbound services). In the former case, data is typically enriched, in the latter is typically reduced. Multiple filters may be placed in series to create a filter pipeline.

The event engine holds rule/action pairs. Each rule is subscribed to the specified data asset and triggers an action if the condition of the rule is met.

The Graphical User Interface (GUI) is an alternative means to administer Fledge leveraging the REST API. The GUI contains a number of pages from which the services may be managed and monitored (for the most part providing graphical access to “Orchestrator service” functionality).

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

6.6 CoMPAS[50]

CoMPAS (Configuration Modules for Power Industry Automation Systems) is concerned with the automation of substations developing software components around the IEC 61850 model for Protection, Automation and Control Systems (PACS). See also Figure 47. It features a cross-platform Docker-based microarchitecture. It is in a relatively early development phase.

The services offered by CoMPAS cover specification, configuration, deployment, and testing of substations accessible through a REST API. To facilitate compatibility to the power industry standard CIM, services mapping it to IEC 61850 System Configuration Language (SCL) exist. A detailed discussion of the differences between these is beyond the scope of this document. However, notably the former emphasizes actual equipment topology while the latter provides a functional view. Furthermore, SCL is limited to data exchange within substations whereas CIM covers the entire domain of power transmission and distribution.

Architectural Principles

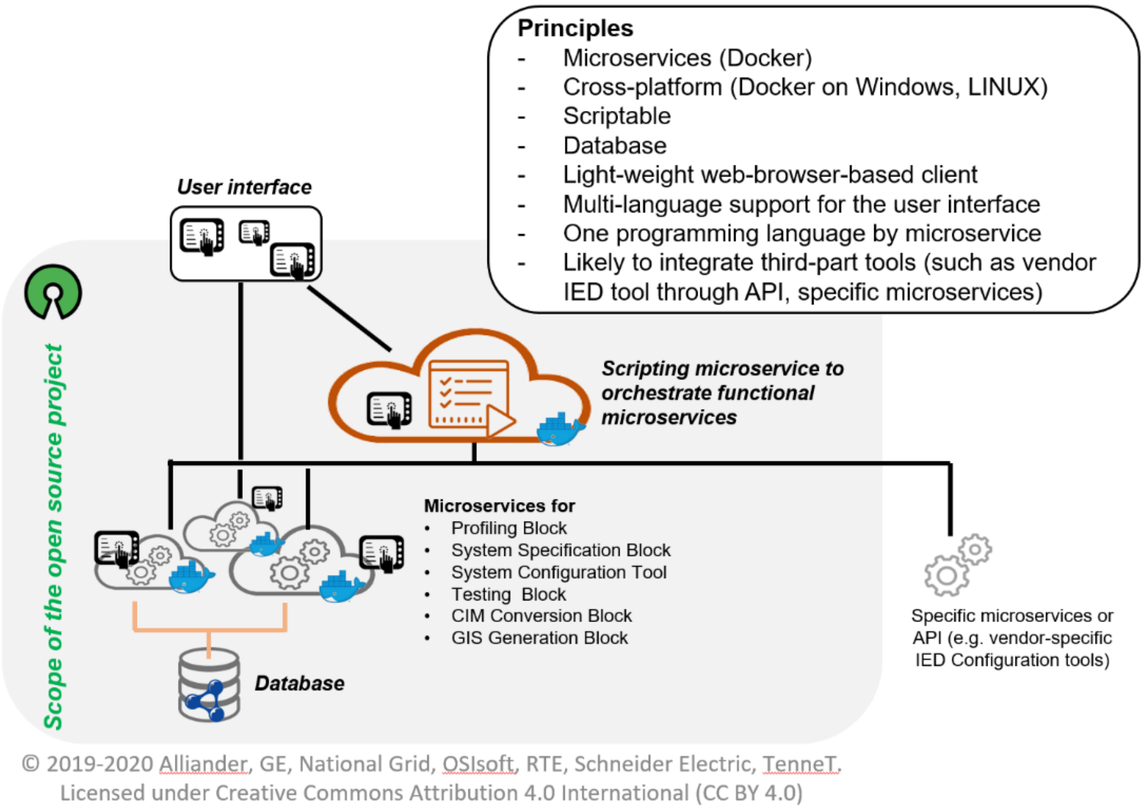


Figure 47 – CoMPAS project scope and principles.[51]

[50] https://com-pas.github.io/compas-architecture/ and https://wiki.lfenergy.org/display/HOME/CoMPAS
[51] Image source: https://www.lfenergy.org/projects/compas/

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Geared toward reaching the broadest possible audience, CoMPAS can be used cross-platform (Windows & Linux), has a light-weight browser-based client and supports multiple languages in its user interface. CoMPAS microservices are Docker container images and built in a CI/CD pipeline that includes security checks using the Gradle build process. The user uses the browser to edit the SCL (or converted) files. A REST (XML) plugin in the interface of the editor to save their work in the database such as, for instance, BaseX. Making use of Docker volumes they are saved persistently outside the database.

6.7 Recommendations

In this subsection non-functional recommendations based on the selection of LFE projects that were briefly presented in the preceding subsections are derived.

Platform Architecture Layout

Properly designed microservice architectures have several advantages compared to monolithic architectures. Firstly, microservice architectures are naturally modular, therefore extensible and understandable as each function is wrapped into a minimal unit called service. As a consequence, each microservice may be developed, maintained, deployed, and tested by independent teams in a programming language of their choice. Additionally, applications using microservices scale better leveraging advanced orchestration tools. As microservices use APIs and communication protocols to communicate with each other, but not otherwise depend on each other they can also easily integrate into federated data ecosystems.

On the downside, microservices rely heavily on messaging which may cause network congestion or other performance-related problems. To mitigate such problems, a number of tools (and personnel) might be necessary such as CI/CD servers, application performance monitoring tools or configuration management platforms. Also, security can be an issue as the distributed communication offers additional attack vectors. Mitigation measures are suggested in a later part of this subsection.

Furthermore, layered architectures from most abstract functional layers at the top to least abstract details-oriented layers at the bottom are recommendable following the principle of modularity for the most flexible and maintainable software.

A third general recommendation is making use of virtualization as another puzzle piece toward increased security, scalability, and availability. An immediate option that comes to mind is application containerization with Docker to be used together with orchestrators such as Kubernetes or Docker Swarm. Secondly, for even better security virtual machines can be leveraged using QEMU and KVM. With the recent advances especially in KVM, the slight performance hits can be tolerated in almost all scenarios. Another advantage of virtualization is that copies of application states may be kept for backup and rollback.

API Development

Platform developers should make life as easy as possible for application developers. One way to do this is to make sure that they can develop in a programming language of their choice. The SEAPATH

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

project for instance allows plugins to their services to be developed in either Python for swift development or C++ for high performance. A way to retroactively add such functionalities might be to use some glue frameworks such as SWIG.

Once a platform is established, applications must be developed for it. In the following some suggestions on how to make life as easy as possible for users of such applications are made. The core idea of this paragraph is that it is much easier to describe **what** a problem is than it to state the complex, error-prone way of **how** to solve it. Thus, the user should be able to state their concern in a domain-specific way, e.g., by using a domain-specific modelling language (DSML) and let an engine or simulator solve their problem for them. Another intuitively understandable and powerful technique is to allow the user to specify pairs of rules and actions.

A third topic is the design of APIs. A widespread type of APIs are REST APIs. They use a subset of HTTP verbs for requests. Responses may be received in different formats such as HTML, XML, JSON, or even plain text. Following the REST software architecture guidelines and developing a RESTful API has the advantage of having a flexible and swift API in the end. SOAP is an older and slower alternative that has some specific requirements such as XML-based messaging but on the upside, it has some incorporated security and compliance features. Still, the use is discouraged. A modern alternative to REST is GraphQL. Its main appeal is that only one request is needed to obtain desired data instead of multiple HTTP requests with REST. As a result, GraphQL causes less network load and is faster. GraphQL supports JSON and its own format for requests and JSON only for responses. GraphQL and REST can coexist in the application software stack.[52]

Data Representation and Translation

Another concern is data representation. An example for this would be the CoMPAS use cases where data relevant to substation management is represented in different ways that emphasize different points of view, i.e., the equipment topology in CIM and the functional view in SCL. To deal with both at the same time modules for model-mapping or translation gateways could be offered.

Instead, and to make the data usable for all participants of a Big Data platform incoming data should be converted into an intermediate representation such as NGSI-LD. A general benefit of intermediate representations is that fewer translations must be made. If, for instance, a set of m languages need to be translated to another set of n languages, then m*n translation mappings would have to be implemented. With an intermediate representation, all of the m input languages are translated into said intermediate representation and subsequently translated into n output languages. That makes m+n mappings instead of m*n in total. Furthermore, NGSI-LD itself is a powerful way to represent context data and paves the way for the vision of a truly Semantic Web.

The third consideration in this category is that data should be represented or representable as stream where it makes sense. This way, powerful stream-filtering techniques allow the enrichment or de-noising of data.

[52] https://graphql.org/learn/

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Security

Security is a key requirement that may be split into several concerns of their own. Most importantly, it must be made sure that only authorized users have access. To ensure this a number of measures should be taken. In some cases, access could be granted very granularly in explicit smart contracts, or less granularly based on role and/or domain. In any case, identification should not be on a per entity basis but on a per device basis (e.g., unique private keys per device). Additionally, each entity should be certified by a trusted certification authority.

Secondly, platforms need to be equipped against malicious attacks that can come in various forms. They need to be robust enough to fend off DDoS attacks making use of mitigation techniques such as connection tracking, IP reputation lists, deep packet inspection, blacklisting, or rate-limiting. The network connection itself should be secure and encrypted using TLS and VPNs. The platform servers should run hardened OS kernels to close some additional security holes. For the containment of threats by compromised components, it is recommended to put up firewalls between all platform servers and layers.

Documentation

An important factor that must not be underestimated is up-to-date, well-structured, and clear documentation. A number of diagrams with varying degree of abstraction explaining the functional and the technical architecture, respectively are tremendously helpful. They should be available per component and for the final architecture.

Textual descriptions of the functions of each block should be available in plain English. Component communication should be depicted in the form of message sequence charts (MSCs).

All third-party software used should be listed. This is also important for security.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

7 Conclusion

The aim of this Deliverable was to define a first version of the Reference Architecture for the BD4NRG project. To that end, existing efforts were classified, presented and considered with regards for their relevance to the project and its use cases.

The BD4NRG RA builds upon the BRDIGE RA. Since the BRDIGE RA reuses concepts introduced in the widely accepted SGAM and already introduces a cross-sectorial viewpoint, it is well suited for this purpose. The BD4NRG RA extends BRIDGE with relevant components, which are identified by mapping the individual uses cases and WP2 efforts to the BRDIGE RA. This mapping revealed that due to the early stage of the project, the information provided for the use cases is still incomplete. Therefore, an updated version of the RA will be presented in the deliverable D2.6.

In addition, this document reviews Data Space efforts. Existing building blocks for Data Spaces, provided by IDS/GAIA-X and FIWARE, are of great importance for the project and will be integrated in the work of the technical Work Packages 3-5.

Finally, first implementation recommendations were given, based on existing LFE projects.

During the creation of this document, it became clear that several questions regarding the structure of the BD4NRG system cannot be answered at this early stage of the project. To define a reliably useful and complete reference architecture, more information about the ongoing work within the LSPs as well as the technical Work Packages is necessary. As the project progresses, it will become possible to define a more detailed description.

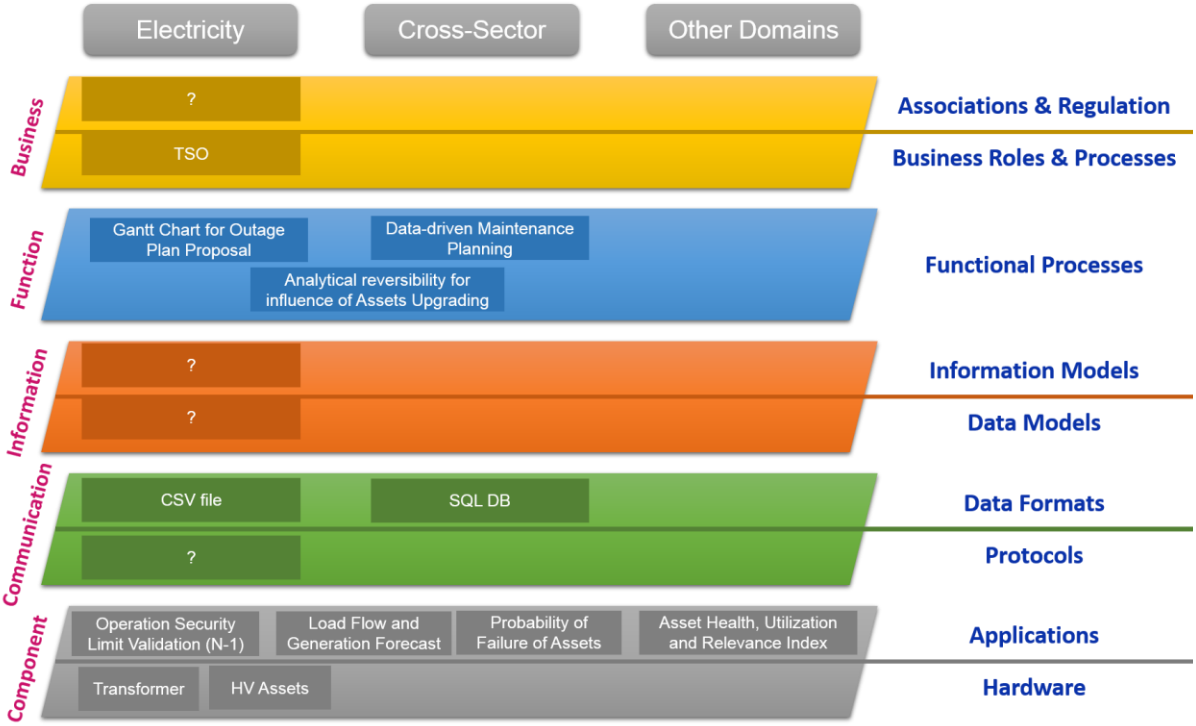
The coming deliverable D2.6 will therefore contain a full and more specific reference architecture. It will bring together the combined efforts of all tasks within Work Package 2 and identify final specifications for the realization of the project’s goals. To this end, close collaboration with the project’s technical Work Packages will be invaluable.

BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

Appendix A: Use Cases to BRIDGE Reference Architecture Mapping Diagrams

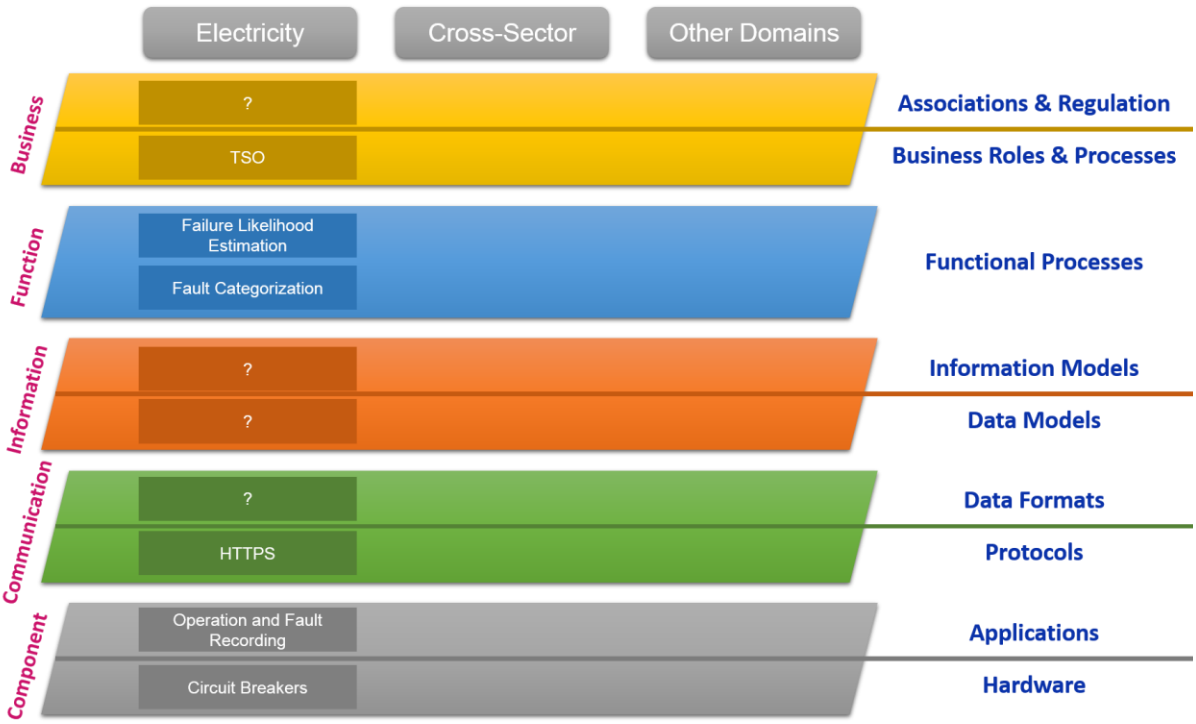
BD4-NET Cluster (LSPs 1-4)

LSP1 UC1 – AI Methods Applied to Condition-Based Monitoring of Circuit Breakers (REN & RDN)

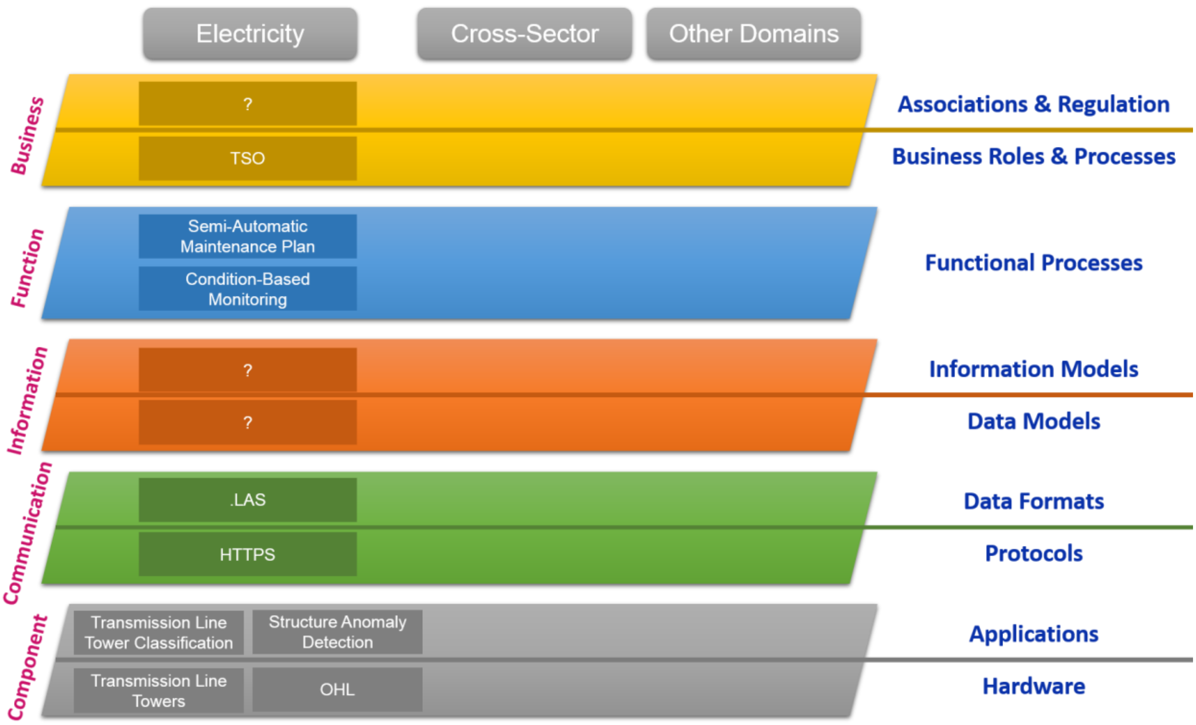


BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

LSP1 UC2 - Semi-Automatic Maintenance Plan for Right-of-Ways in Overhead Lines (REN & RDN)

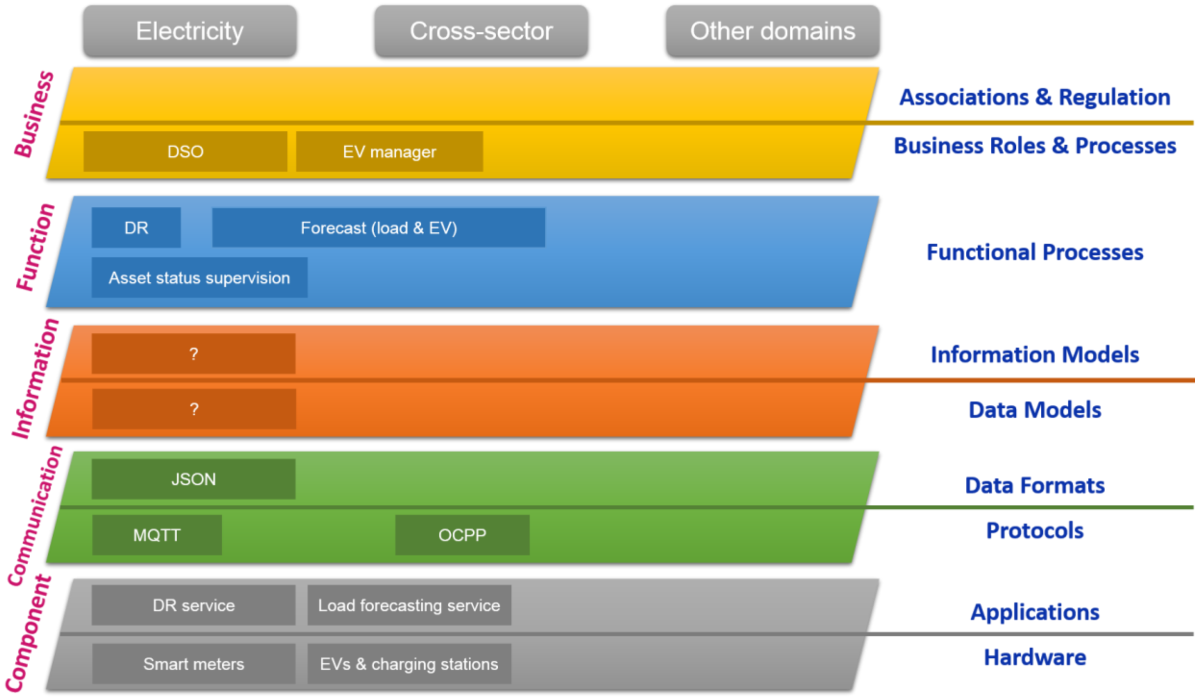


LSP2 UC1 - Maintenance Plan Optimization (ELES)

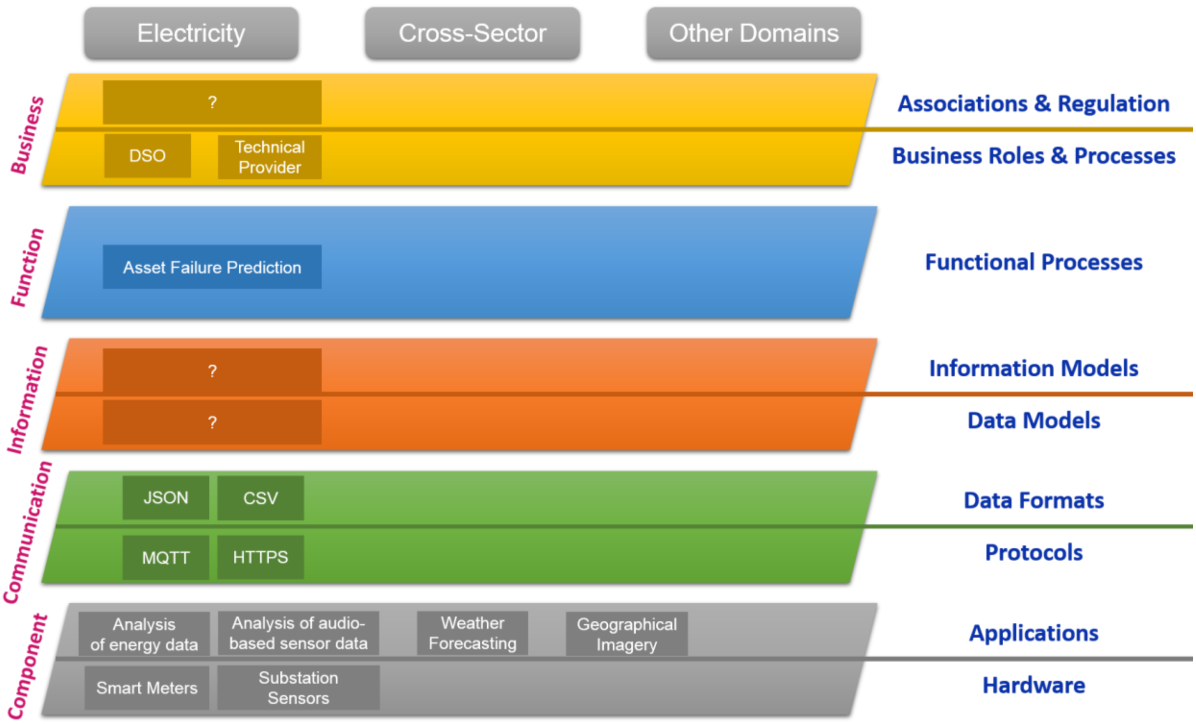


BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

LSP3 UC1 - Efficient Management of Flexibility Resources for Supporting Grid Operation (ASM)

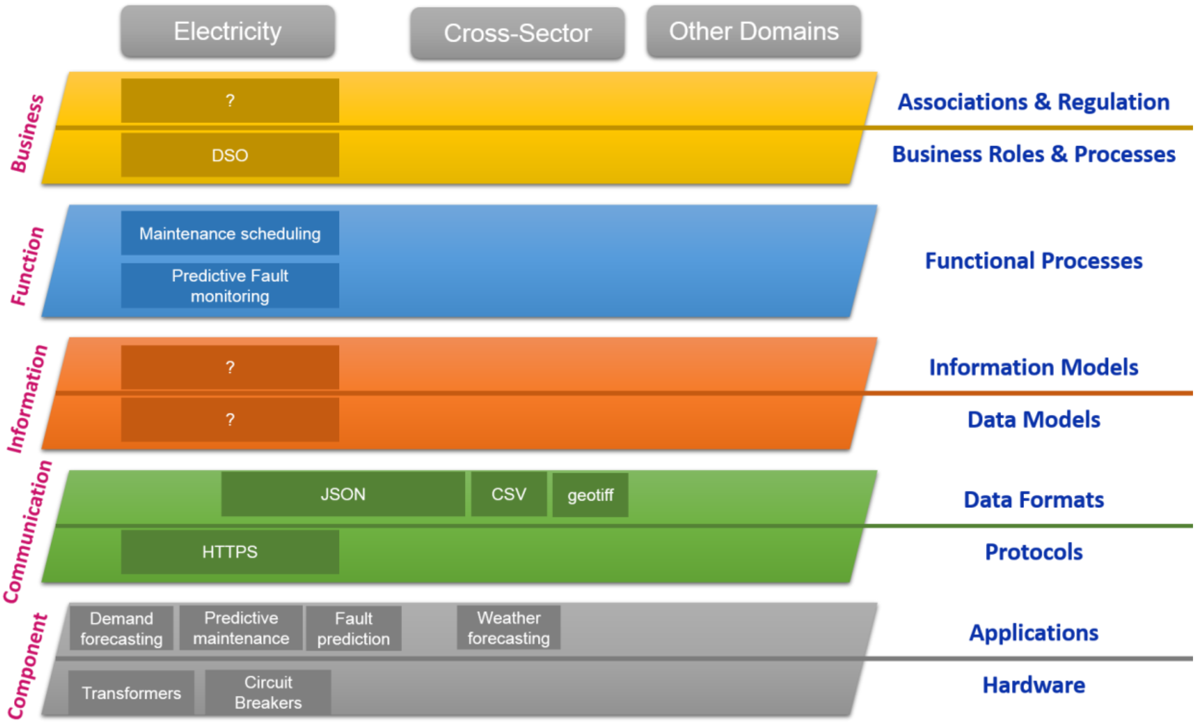


LSP3 UC2 - Anticipated Prediction of Asset Failures (ASM)



BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

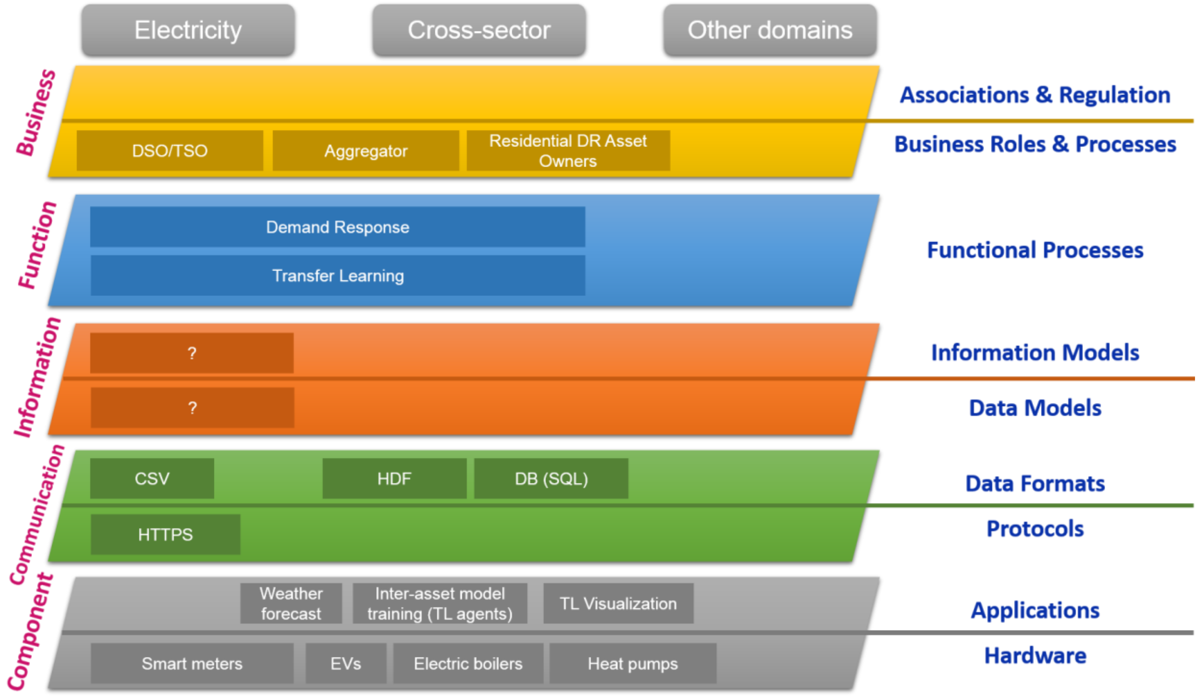
LSP4 UC1 - Predictive Maintenance and Fault Prediction for MV Grid and Cross-Functional Predictive Asset Management (OEDAS)



BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

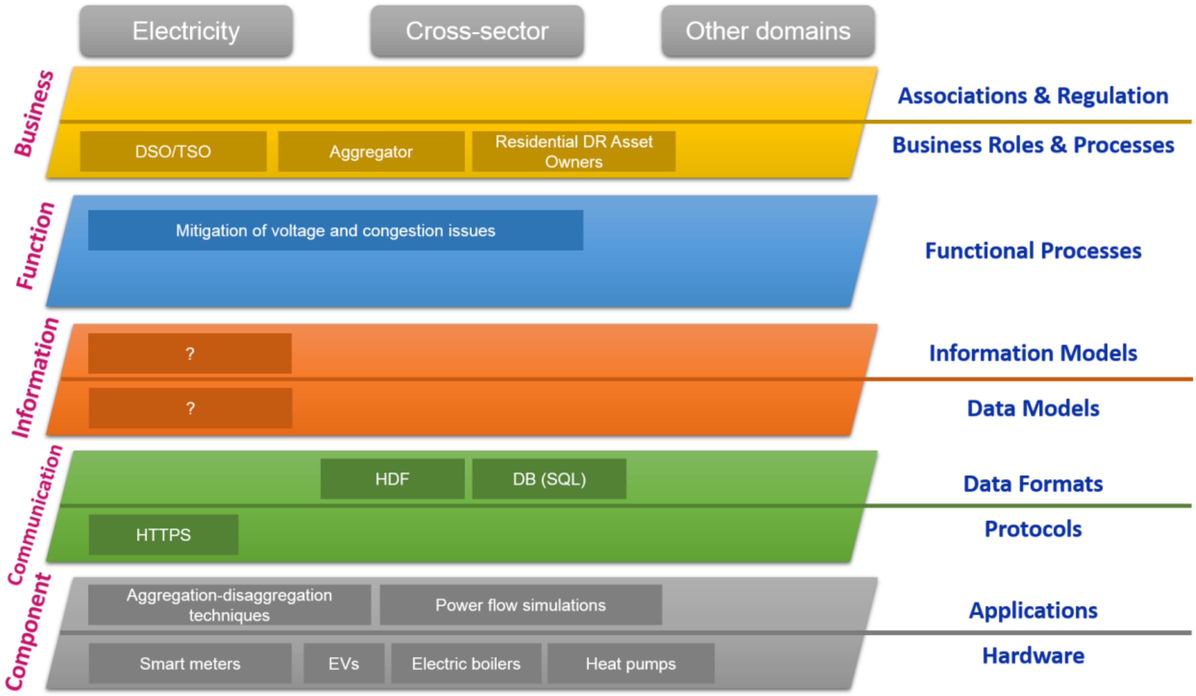
BD4-DER Cluster (LSPs 5-9)

LSP5 UC1 - The agents responsible for controlling residential demand response assets can benefit from the transferred knowledge from other agents in an MDP setting (CENTRICA)

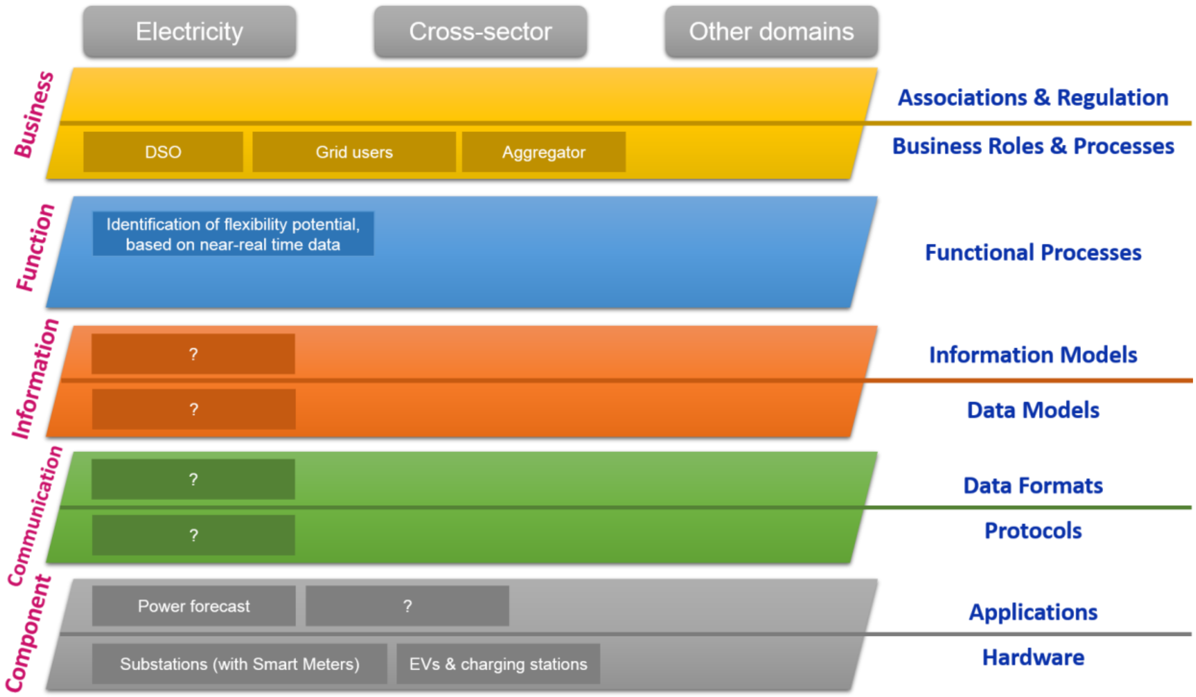


BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

LSP5 UC2 - Using residential DR assets to retrieve local grid information and use this information in the aggregation activations (CENTRICA)

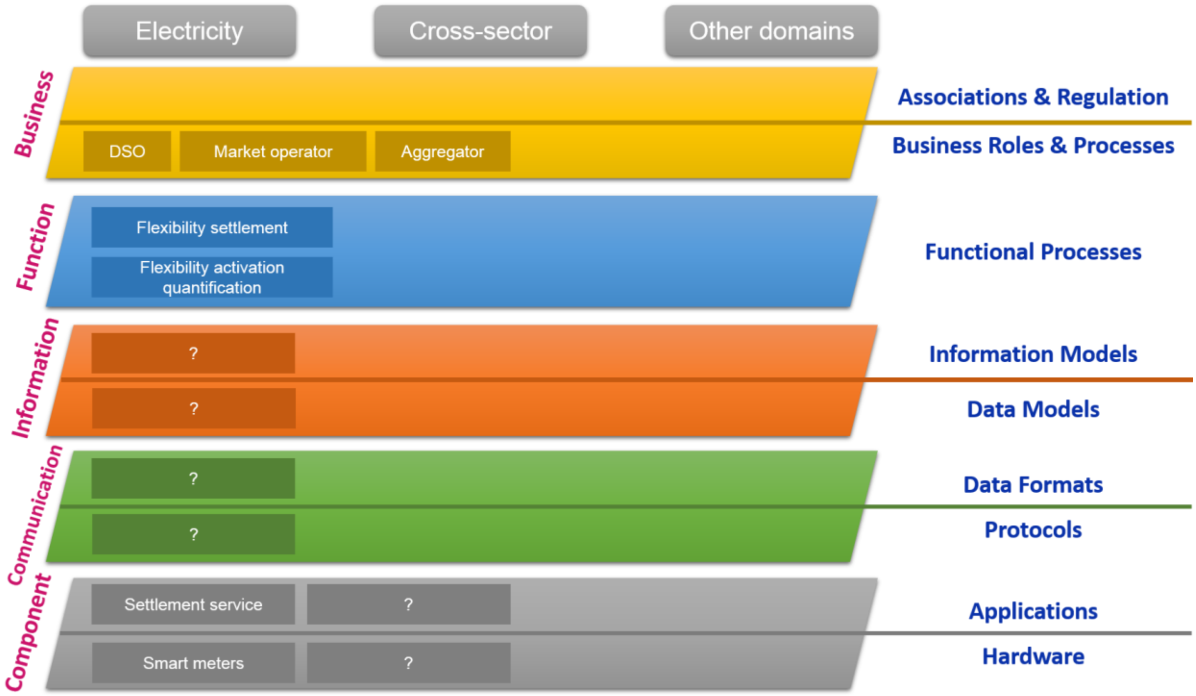


LSP6 UC1 - Identification of Flexibility Potential Based on Near Real-Time Data (EKL & Borzen)

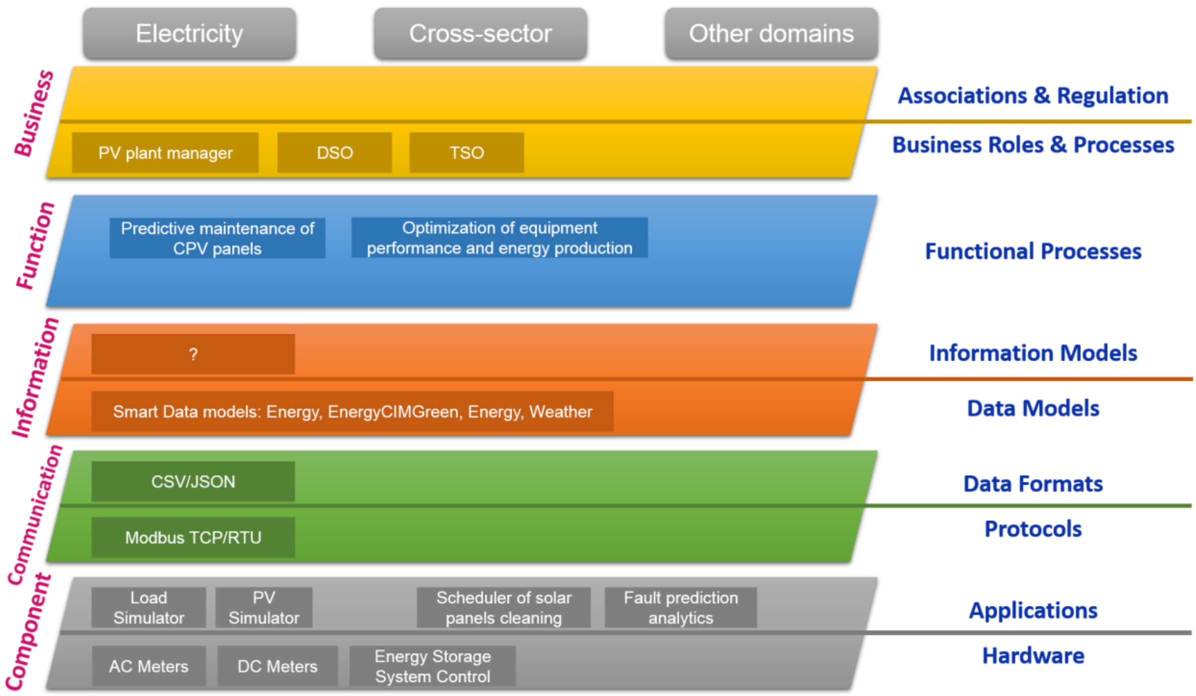


BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

LSP6 UC2 - Independent Aggregator Flexibility Activation, Quantification and Settlement (EKL & Borzen)

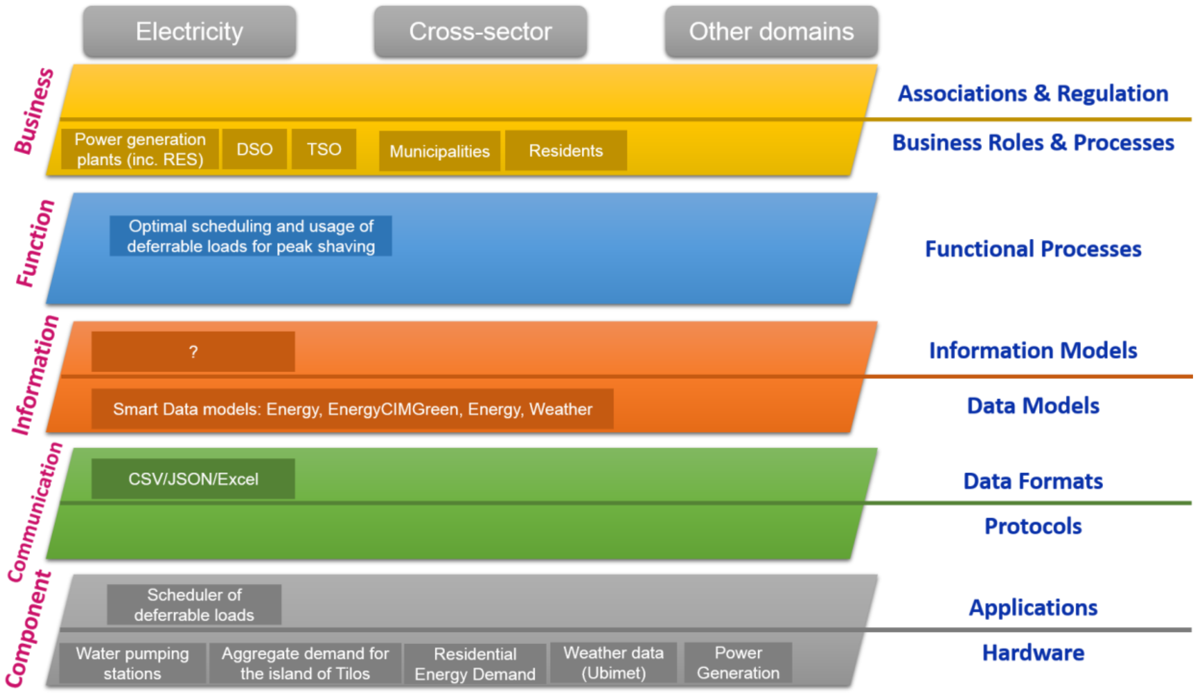


LSP7 UC1 - Predictive Analytics Forecasting of storage optimized operation for improved market reserve participation (ENELX)

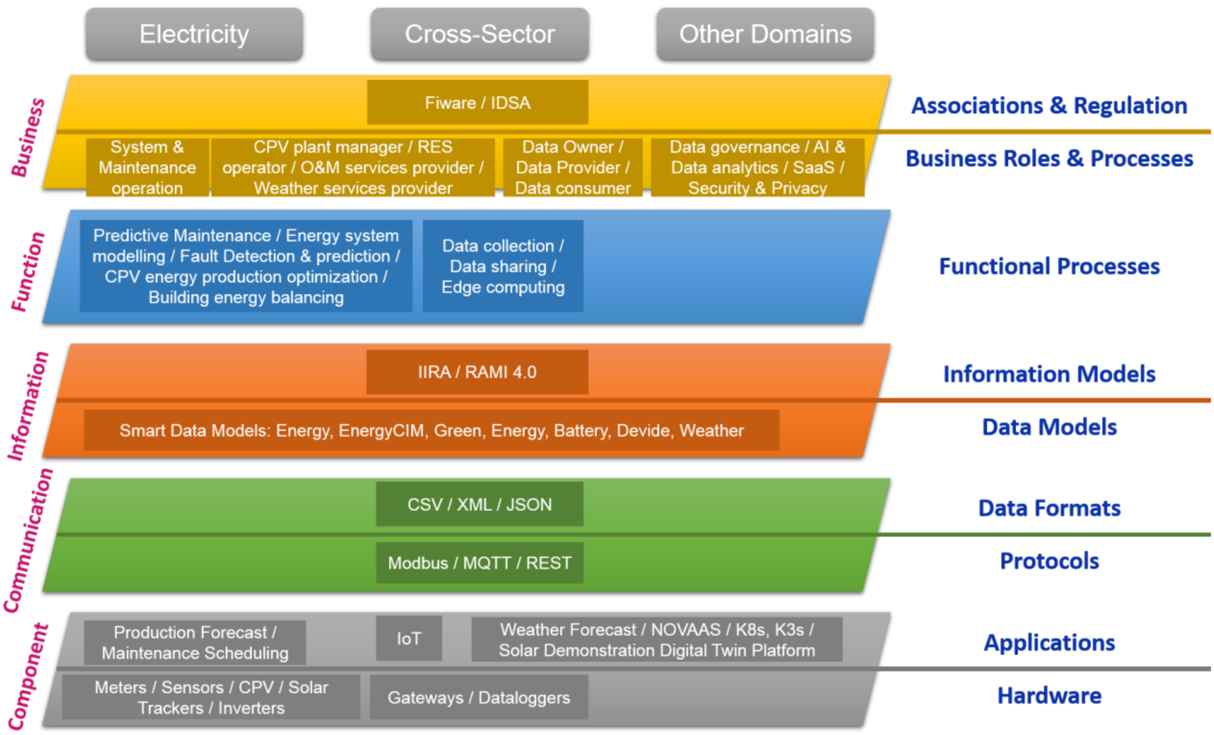


BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

LSP8 UC1 Predictive Demand and Generation Forecasting for Optimized Local Energy Community Management (UNIWA)

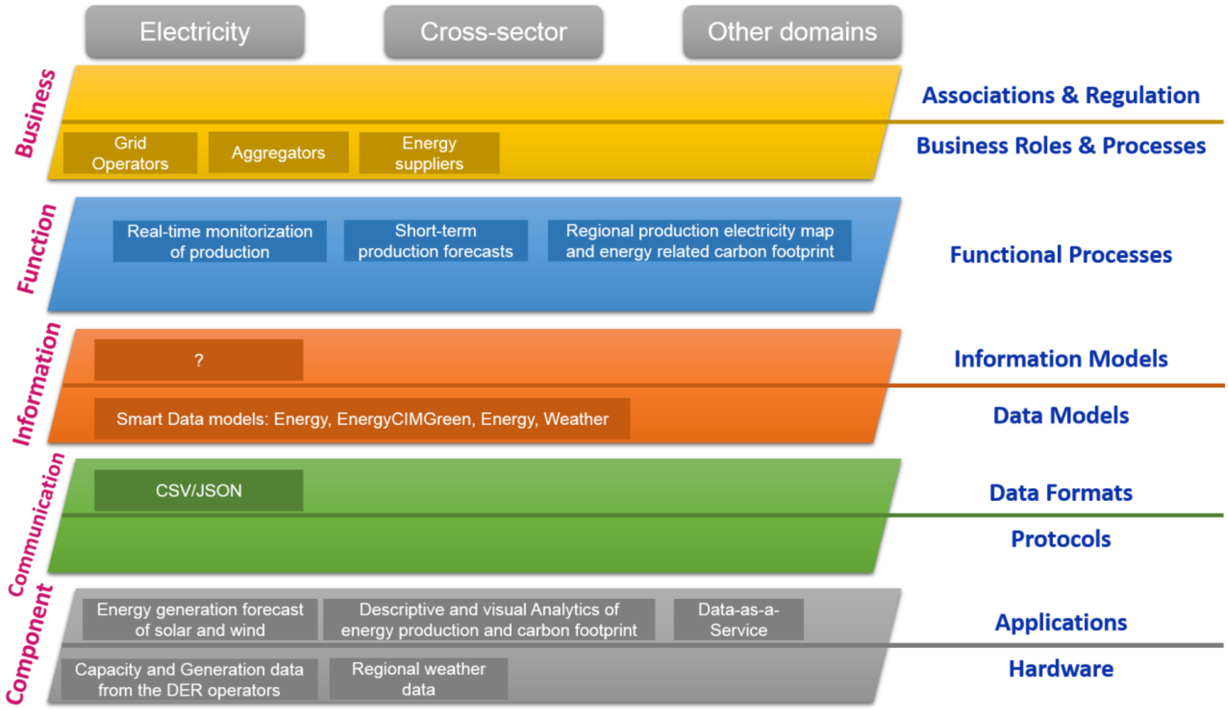


LSP9 UC1 - Optimizing and Improving Management and Prediction of DER Production (ENERC)

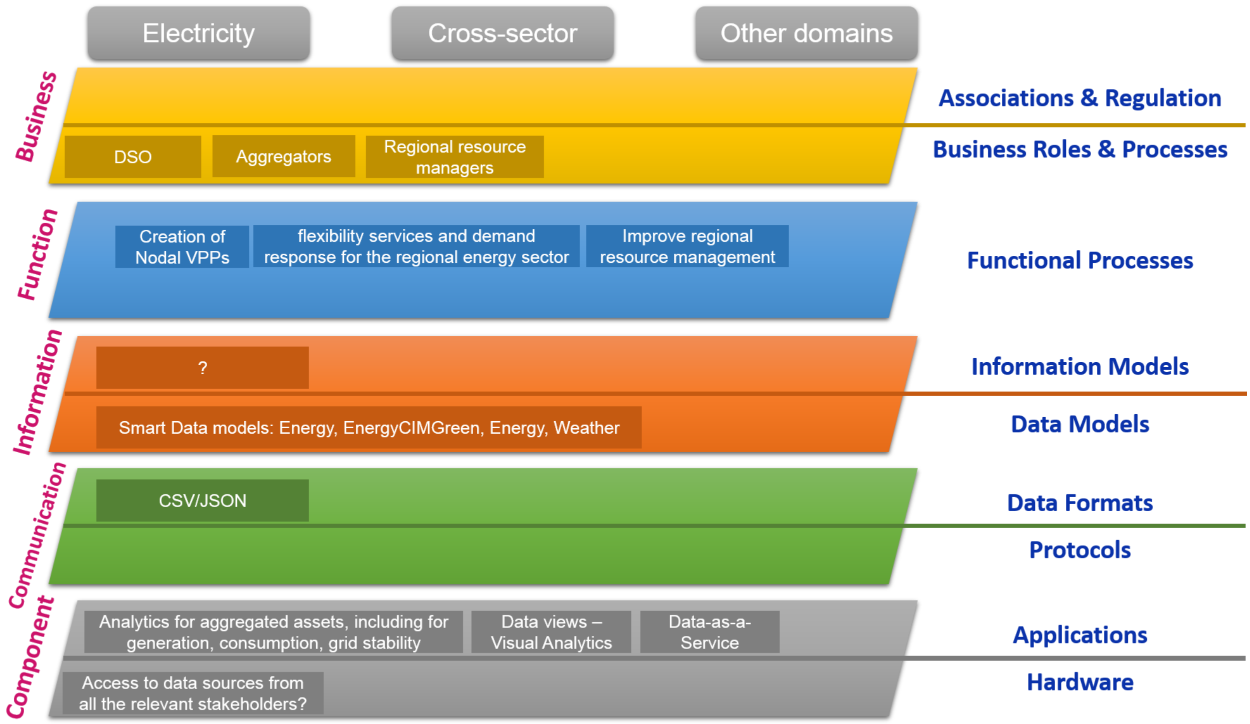


BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

LSP 9 UC 2 - Regional Energy Sector Aggregation Digital Twin (ENERC)



LSP 9 UC 3 - FlexATwin Regional Digital Twin With Multiple Interfaces for Flexibility Across Sectors (ENERC)

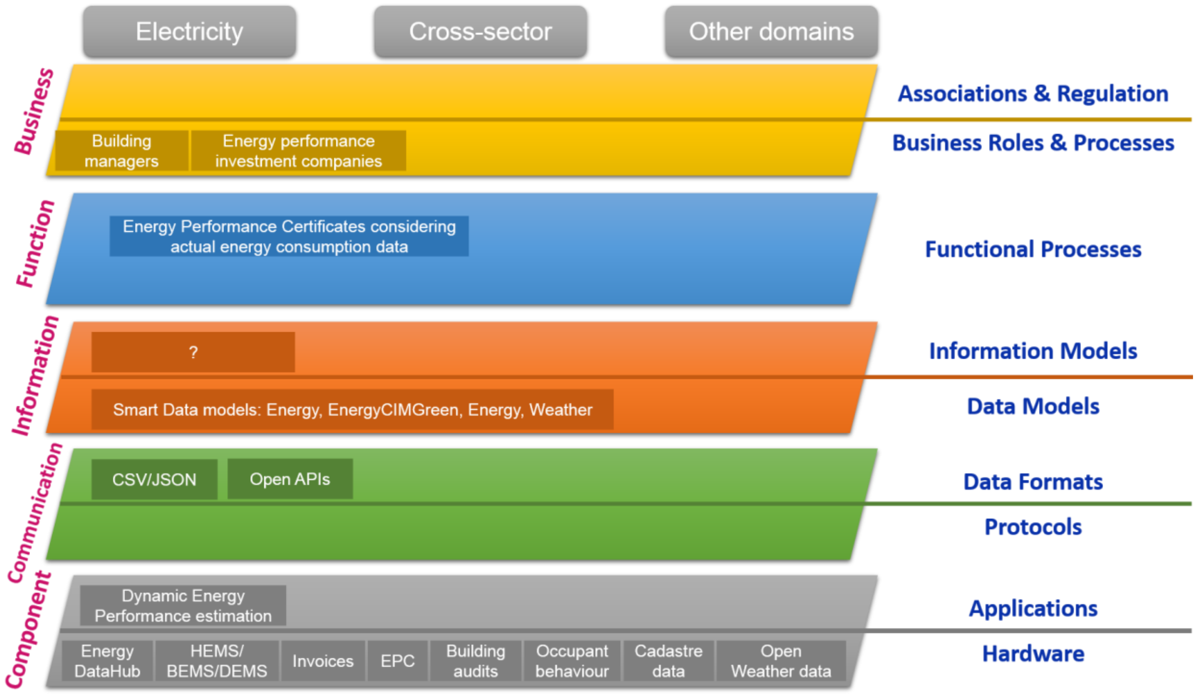


BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

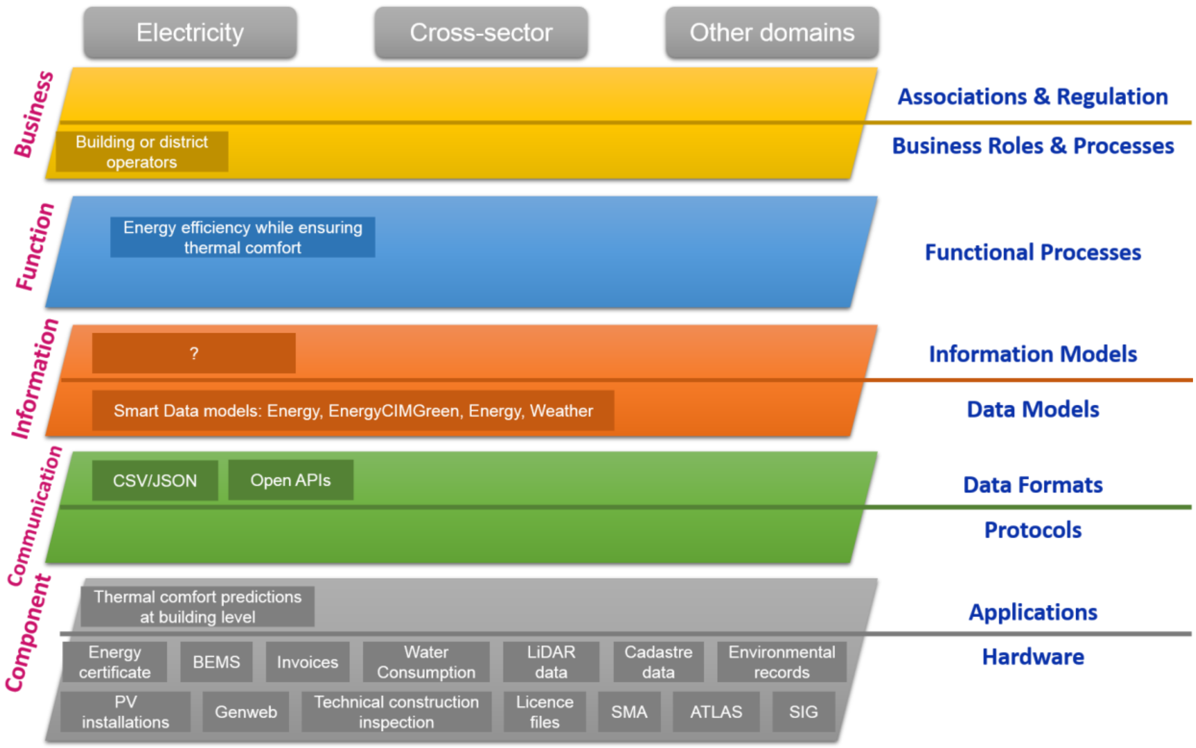
BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

BD4-ENEF Cluster (LSPs 10-12)

LSP10 UC1 - Improved Energy Performance Certificates Reliability (CARTIF & VEOLIA)

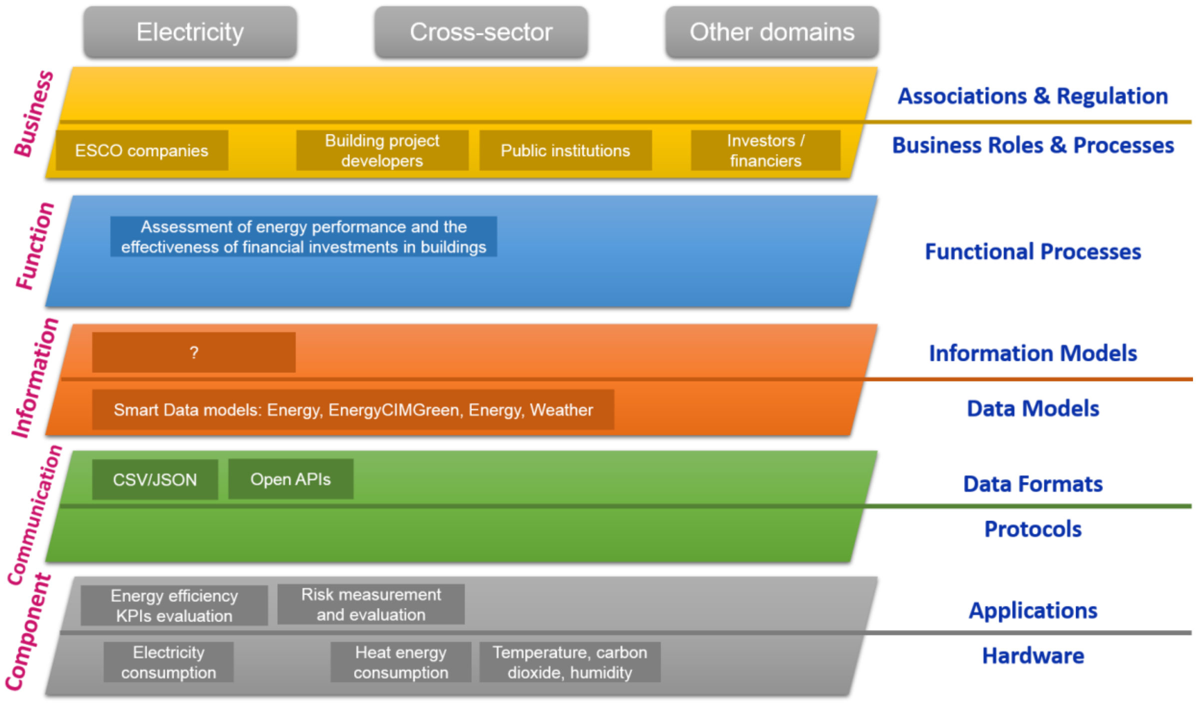


LSP11 UC1 - Weather-enhanced Building Thermal Comfort Prediction to Enable Energy Efficiency Reliable Forecasting (AJSCV & VEOLIA)



BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.

LSP12 UC1 - De-risking Energy Efficiency Investments (LEIF)



BD4NRG project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 872613.