BD4NRG
Big Data for Next Generation Energy

D2.2:

BD4NRG Open APIs and Data Assets Functional Specifications

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	www.bd4nrg.eu		
Project Coordinator	ENG		
Deliverable	D2.2: BD4NRG Open APIs and Data Assets Functional Specifications		
Work Package	WP2: System Requirements and Specifications		
Delivery Month (DoA)	M7	Version	1.0
Actual Delivery Date	31/08/2021		
Nature	Report	Dissemination Level	Public
Lead Beneficiary	Atos		
Authors	Javier Valiño (Atos), Diego Ruiz (Atos), Elisa Herrmann (Atos), Arturo Medela (Atos), Marzia Mammina (ENG), Diego Arnone (ENG), Elissaios Sarmas (NTUA), Elena Lazovik (TNO), Andrea Manzi (EGI), Francesco Saverio Nucci (ENG)		
Quality Reviewer(s):	Nikos Bilidis (ED), Diego Arnone (ENG), Alexander Pastor (RWTH)		
Keywords	Logical, Implementation, Development, API, functional module, UML, specifications, data, analytics, marketplace, architecture		

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 (EPES) 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

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

Who We Are

No	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	**RDN**	PT	

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

No	Participant Name	Short Name	Country Code	Logo
	ENERGIA REN - STATE GRID SA			
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	
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	

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

No	Participant Name	Short Name	Country Code	Logo
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 DOO SISTEMSKI OPERATER PRENOSNEGA ELEKTROENERGETSKEGA OMREZJA	**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

Tables

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.

Executive Summary

The work described in this deliverable (D2.2) was carried out in the framework of WP2 – "System Requirements and Specifications", and more specifically, in the framework of T2.3 – "Open APIs and Data Assets Functional Specifications (Data Asset Quality)". The report recapitulates the work partners conducted in this task, the main purpose of which consists in offering the analytical description of the structure and functionality of the BD4NRG components. Therefore, D2.2 defines and documents the operations that run internally and the structure of each component in the architecture, along with the APIs and interfaces that establish the way these components interact with each other.

All technical partners involved in this task collaborated and developed the appropriate tools to meet the objectives set out in the project. Every partner focuses on the individual modules that they are responsible for, during the implementation phase of WP3, WP4 and WP5 and supports the integration activities of WP2, while following the common Architecture framework that will be definitely set by WP2 in D2.6.

Deliverable 2.2 is structured as follows. Section 1 introduces the scope of this document and its relationship to other WPs and Tasks. Section 2 presents the methodology applied to perform this activity. Sections 3, 4 and 5 aggregate all the main outcomes of the task, present extensively the data governance and management strategy, as well as the data modelling and open modular Artificial Intelligence-based edge-level Analytics Toolbox that BD4NRG offers. To do so, the individual components of every layer in the architecture are defined. In addition, the marketplace for realizing an EU-level Energy Data Economy is covered by the task, by providing an extensive description of the corresponding functionalities and details related to the characteristics of these marketplaces and its interactions with the rest of components in the architecture of BD4NRG. Finally, Section 6 concludes the document.

All in all, the deliverable exposes the BD4NRG functional specifications related to T2.3 as well as the guidelines to conduct the related interactions and integration activities in the upcoming months.

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

Acronyms and Abbreviations

Acronym	Description
B2B	Business-to-Business
BDA	Big Data Analytics
CA	Certification Authority
CDM	Common Data Model
DaaS	Data-as-a-Service
DAPS	Digital Attribute Provisioning Service
DAT	Dynamic Attribute Token
DL	Deep Learning
DLT	Distributed Ledger Technology
GUI	Graphical User Interface
HTAP	Heterogeneous Data Capture, Cleansing, Integration, Polyglot Persistence
IAM	Identity and Access Management
IDS	International Data Spaces
IDSA	International Data Spaces Association
ISST	Institute for Software and Systems Engineering
JSON	JavaScript Object Notation
LSP	Large Scale Pilot
ML	Machine Learning
MLaaS	Machine Learning as a Service
MQTT	Message Queuing Telemetry Transport
OOP	Object-Oriented Programming
PAP	Policy Administration Point
PDP	Policy Decision Point
PEP	Policy Enforcement Point

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

RAM	Reference Architecture Model
SFTP	Secure File Transfer Protocol
SGAM	Smart Grid Architectural Model
SSO	Single-Sign-On
UC	Use Case
UML	Unified Modelling Language

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 this deliverable

The main focus of this task is to deliver a systematic description of the structure and functionality of the BD4NRG components as part of the system requirements and specifications activities. The goal here is to comprehensively and accurately define and document the diverse internal operations and structure of each component, as well as the interfaces that let all of them interact with each other in order to comply with the desired extensibility and scalability of the overall solution.

This task works closely with the one devoted to defining and specifying the functionalities related to the Big Data Value Chain within BD4NRG Framework, as well as the various relevant stakeholders in the project's environment along with their already available Big Data/IoT infrastructure. Therefore, there are contributions to achieve the goal of shaping an integrated data value chain.

This document addresses the main objectives of this task, which delves around the analytical description of the structure and functionality of the BD4NRG components, at least to the extent that is possible at this stage in the project execution; the internal operations and structure for each component will be defined. Thus, the report takes the initial steps on the desired goal to establish the BD4NRG APIs and interfaces, which in the end will provide the mentioned extensibility and assure the interconnection with systems outside BD4NRG. At this precise moment, partners involved in technical activities are just able to express the kind of interactions that should take place among modules in the BD4NRG architecture and provide hints that suggest how they would pan out in the long term.

This set of functional components will act as the foundation coming out of Task 2.3 and in the long term feed the various project pilots, where they are put to a test in real-life situations, according to the respective use cases.

1.2 Relationship with other BD4NRG tasks

Figure 1 gives an overview of the relations of this deliverable and the corresponding Tasks (T2.2 and T2.3) to other WPs and activities. The figure is focused on D2.2, so not all relations between the rest of the WPs and tasks are presented.

More specifically, Task 2.2 and Task 2.3 relate to the most technical work packages. For instance, the activities performed here feed into WP3 *"Data Governance and Management"* where BD4NRG-GOVERNANCE and BD4NRG-PROCESSING layers will be developed in order to make possible the diverse assets and systems exchange information in a common manner. They also impact on WP4 *"Data Modelling & Open Modular AI-based edge-level Analytics Toolbox"*, in charge of the development of various libraries and algorithms to perform the mining of information from big data, business intelligence and energy analytics, the so called BD4NRG-ANALYTICS layer. In addition, there is a link to WP5 *"Marketplaces for realizing EU-level Energy Data Economy"*, which looks to make use of the project results to trigger business opportunities for the actors involved via the BD4NRG-MARKETPLACE.

All in all, the results of this report are directly provided as input to those WPs which lead the overall integration activities. Thus, all of them together provide valuable input to the Large Scale Pilots (LSPs), with D2.2 acting as one of the main stepping stones providing information and specifications required to achieve the integrated solution ambitioned. Nevertheless, it is worth noting that the present report acts as a preliminary step in a living process that will reach its final stage in D2.6, where the final BD4NRG architecture will be presented.

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

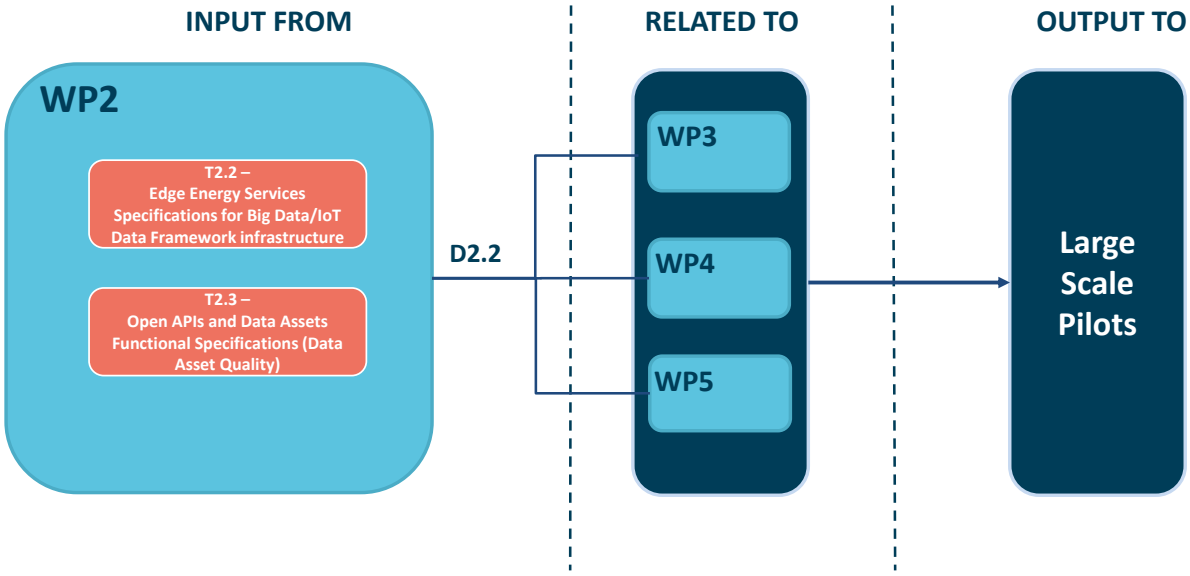

Figure 1. D2.2 relation to other WPs

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

2 Methodology

According to the International Standard ISO/IEC/IEEE 42010 [1], the application of architecting concepts, principles and procedures becomes essential to manage modern systems that grow in complexity. This approach also helps to understand the meaning and key features of the system's behaviour, structure and development, which has an impact on the system's feasibility, utility and maintenance. Thus, and to ease the description of architectures to promote collaboration and communication between stakeholders, architecture frameworks and description languages were designed to systematize the common practices and protocols of architecting within different contexts.

Within the conceptual model of an architecture description defined by ISO/IEC/IEEE 42010, architecture viewpoints appear. They are the elements that set up the conventions to build, interpret and analyse architecture to address potential concerns that may be raised by different stakeholders. Viewpoints are then specified as requirements for conformance with the standard.

2.1 Architectural model selection

In this section the methodology used in BD4NRG will be described to define a framework for architectural design. First, different models of architecture are described, then the most suitable model is selected and finally, the reasons for doing so will be detailed. Finally, the methodology followed to define each view will be explained.

The Smart Grid Architectural Model (SGAM) has been followed to define the Use Cases (UC) of the project. The SGAM provides an approach to develop the architecture of current and/or future implementation of the electrical grid. It allows for a common modelling methodology to capture smart grids into five layers.

However, the BD4NRG platform not only maps to SGAM, but also offers complex Analytic Tools and a wide variety of external systems (hardware devices, external APIs, interfacing modules, etc. …) requiring interaction with more than one layer for the development of the platform.

According to this, the standard ISO/IEC/IEEE/42010 has been selected as a reference to help us choosing a proper architectural model.

Some examples of architectural models are:

- **Documenting Software Architecture:** this is a 3 views approach (module, component/connector and allocation) with rules to understand the information to be provided for each one of them.
- **Software Process Architecture:** this is a model made of 9 viewpoints, namely: requirements, functional, deployment, validation, application, infrastructure, system management, availability, performance and security.
- **Architecture Based on Execution:** focused on software, this model describes the templates for 4 execution views, which are execution profiles, execution deployment, execution concurrency and resource usage.
- **Architecture Based on '4+1' views:** this is a model based on 4 views (logical, development, process, and physical). Those are integrated via a fifth view called scenarios.
- **Architecture Based on Software Systems:** this model takes up 6 points of view (functional, information, concurrency, development, deployment and operational) that can be expanded with optional viewpoints (security, performance and scalability).

Considering the need of addressing the concerns of all members involved in the project, managing the complexity of the system, as well as, testing the use cases in the pilot sites; the "4+1 view model" is positioned as the best choice for the BD4NRG project.

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

The 4+1 view model architecture is SW-oriented, framed in the Energy R&D framework and offers the 3 main benefits of a software system:

(1) supplies a framework for communication among the stakeholders

(2) exposes design decisions early on

(3) sets up a reusable and manageable model.

Furthermore, it allows for a complete architectural definition distributing few resources in its definition and implementation on pilot sites.

2.2 The '4 + 1' model of architecture

Philippe Kruchten describes the concept of 'view' in his '4 + 1' model as the representation of the whole software system from a certain perspective [2]. The term 'point of view' has been coined with the rise of the OOP (Object-Oriented Programming) paradigm, where a set of rules (or norms) are defined for making and understanding the views.

The architecture of the system is detailed in-depth from each view below:

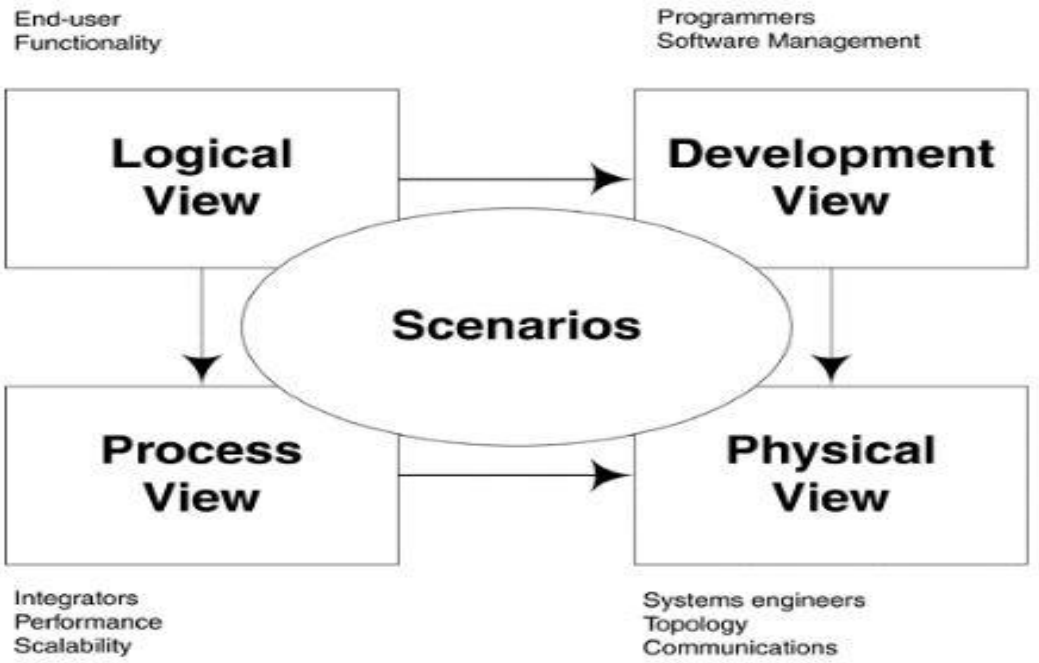

Figure 2. Krutchen's 4+1 model

- **Logical View or Structural View:** this view represents the functionality provided to end-users, giving a static approach of what the system and the offered services should be able to do.
- **Implementation View or Development View:** it depicts the system from a software developer perspective and takes care of software management. It shows the software components and their dependencies, stating in a clear manner all technical requirements and specifications.

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

- **Process View or Behavioural View:** this view shows the processes in the system and the way in which these processes communicate. This captures the dynamic aspects of the system, represented from the perspective of an integrator, it clearly defines the workflow followed by operations in the system.
- **Deployment View or Physical View:** this view takes a look at the system from the perspective of a systems engineer. Thus, it represents the physical components and the connections between them, and the different pieces of hardware contained in the system.
- **"+1" Scenarios View or Use Case View:** this view consists of the software use cases and acts as a junction between the other 4 views. This means it pictures how the other 4 views are linked in a typical use case. This scenario view allows for a traceability of components, classes, equipment, packages needed etc.

The current report covers the Logical View of the architecture, due to the fact that not all these views can be represented in the initial stages of the project execution. Thus, D2.2 provides an initial high-level Logical View with the functional description of the modules that each comprise one of the three architectural layers (Governance, Analytics, Marketplace). Further activities within the most technical side of the project will derive in the complete coverage of the 4+1 views and the definitive structure of the BD4NRG architecture, which will be presented in the upcoming deliverable D2.6.

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

3 Data governance and management

This section recaps a first approach on the Logical View and functional description of modules present in the data governance and management layer. At this stage of the project execution, what follows is just a preliminary proposal concerning the modules that may be present in this layer and how they could interact with each other. Further details will be provided in upcoming deliverables, mainly D2.6 when the BD4NRG architecture is consolidated.

Every component will be introduced via a table which summarizes its main characteristics (see Table 1).

Table 1. Table template for component definition

Module_Id	Module Name
General Description	What this module allows to perform.
Functionality	Describe the way this module works and what provides.
Input data	Describe the data required by the asset for operating (e.g. template, cost, data granularity, specification of the service, etc..)
Output data	Detail the output data
Interconnections	List the modules with which it interacts
Dependencies	Detail if there is any dependency with other modules (e.g. one services is able to start if another is already running)
UML diagram	Depict the UML diagram of the component

3.1 DLT & Smart contracts for B2B cross-stakeholder trusted off-chain data sharing and re-use

Table 2. Trusted off-chain data sharing and re-use module functional definition

Module_Id	Trusted data sharing and re-use
General Description	The Trusted off-chain data sharing module is dedicated to use the DLT/Blockchain technology to ensure an integrity and trustworthiness of the data shared in the BD4NRG project. The primary purpose of the blockchain technologies used in this module is to remove the need for intermediaries and replace them with a distributed network of digital users who work in partnership to verify transactions and safeguard the integrity of the ledger. Use of blockchain in the data sharing model of BD4NRG has these main key advantages: • Traceability and data storage: decentralised and distributed system that becomes a secure way to track changes in information and data over time. • Trust of data: the creation of trust among untrusted participants and among the other the possibility to maintain data trustiness during time • Peer-to-peer transactions: the absence of intermediaries promotes a more transparency data sharing.
Functionality	The proposed solution will combine the on-chain data storage together with distributed (or centralized) databases for off-chain data storage according to a hybrid approach. In order to minimize the amount of data to be stored on the blockchain, raw data from the data sources will be stored using a distributed database: the blockchain will then be used to periodically store a fixed-length hash of the data received in the last time interval. The hybrid approach combines the functionality scalability of a distributed NoSQL database

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

Module_Id	Trusted data sharing and re-use
	with the blockchain, making tampering attempts evident, enabling provenance tracking, non-repudiability, and the use of self-enforcing smart contracts. Each data source will be connected to a gateway, an embedded device which takes care of pre-processing, calculating digital fingerprints (hash codes) to assure data immutability via notarization, and storing the data off-chain. It is also included the possibility that handle data from the Data Storage. A private network based on Ethereum will be used as blockchain network. The Ethereum nodes provide the networking capabilities and the rules needed to determine consensus among them. The private network will be used as a platform for running smart contracts and register data with a timestamp. Thanks to dedicated APIs, every node in the network can be queried and it is possible to build dedicated applications on top of this stack for monitoring the process, validating the off-chain data against the data stored on-chain, provide limited access for auditing purposes, or provide even finer-grained access to specific users for specific purposes, interacting with smart contracts deployed in the network.
Input data	Data sources
Output data	--
Interconnections (Interfaces)	Connection with IoT devices will be implemented via MQTT, a lightweight protocol suitable for unstable connections and low-power devices. Additionally, connection with NoSQL Data Storage will be supported for off-chain data handling.

3.2 Data Access Policy Brokerage

Table 3. Data Access Policy Brokerage module functional definition

Module_Id	Data Access Policy Brokerage
General Description	The central idea behind the Data Access Policy Brokerage is the development and enforcement of mechanisms for the application of Data Access Policies in the BD4NRG datasets. These mechanisms will be used by the upper layer's BD4NRG components specifying user access rights and rules based on: • Volume of data access • Frequency of data access • Request Device (e.g., laptop, smartphone etc.) • Person or Group Role (e.g., location, IP, role in the company etc.) • Historic data regarding access patterns (e.g., frequency, usual dates or hours of access, usual duration of access, previously accessed data etc.) • Access Control Type (MAC, DAC or RBAC/ABAC)

Functionality	In more detail the Data Access Policy Brokerage component will be responsible for:
	• The development of appropriate context model for semantically describing associations between types of access depending on the data objects and circumstances under which this access should be allowed
	• Based on the above-mentioned context model the creation, management and maintenance of BD4NRG Data Access Policies
	• Enforcement of DAPs through specific User Access Rights
	• Support of the User Access Rights administration through specific software components which will provide the following functionalities:
	  ○ Identity and Access Management (IAM) for users, groups and devices management.
	  ○ Policy Enforcement Point (PEP) Proxy which will carry out or enforce policy decisions.
	  ○ Policy Decision Point (PDP) which will evaluate and issue authorization decisions.
	  ○ Policy Administration Point (PAP) which will manage access policies.

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

Module_Id	Data Access Policy Brokerage
	○ Data Usage Control. ○ DAP Annotation Management. ○ Validity checking of DAP annotations (e.g. spotting inconsistent annotations). ○ Tracking dependencies between DAP annotations. ○ Ensuring compliance with pertinent organizational policies, especially with respect to the available encryption and data distribution strategies.
Input data	Data sources
Output data	--
Interconnections (Interfaces)	--
Dependencies	--

3.3 Legal, Regulatory, Privacy and Cybersecurity Management & Compliance Tools

Table 4. Legal, Regulatory, Privacy and Cybersecurity Management & Compliance Tools module functional definition

Module_Id	Legal, Regulatory, Privacy and Cybersecurity Management & Compliance Tools
General Description	The central idea behind the Legal, Regulatory, Privacy and Cybersecurity Management & Compliance Tools is the adherence of the BD4NRG Data Governance and Management layer to the legal, regulatory and cybersecurity guidelines and principles especially based on the T2.5 outcomes. The aim of this process is to guarantee the security of the individual subsystems, in order to satisfactorily address security concerns in relation to receiving and sending information and sensitive data.
Functionality	This will be achieved by specific functionalities provided: 1. Data collection from network infrastructure level and endpoint level. 2. Transform and normalize above data (if needed). 3. Network traffic data analysis. 4. Application endpoints data analysis. 5. Classification of data flows using anomaly detection and machine learning techniques. 6. Development of automated rating and classification mechanisms. 7. Evaluation of Security mechanisms based on data analysis results. 8. Update Access control policies and identification of new ones. 9. Basic Audit The T3.3 outcome functionality-wise will be the provision of a proactive and exploratory network/application security tool via behavioural machine learning or other analytics techniques. Based on this, theBD4NRG platform will be able to identify events and at the same time prevent threats or at least decrease response times. The whole process should make use of data ingestion pipelines which through an analytics engine will provide data visualizations, alerts, or automated workflow triggers for threat remediation.
Input data	Data Sources
Output data	--
Interconnections (Interfaces)	--
Dependencies	--

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

This module plays a relevant part in the data cycle shown in Figure 3, depicting how to comply with Legal, Regulatory, Privacy and Cybersecurity policies.

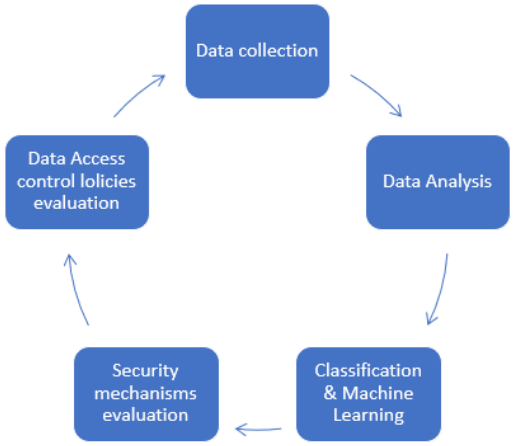

Figure 3. Data cycle to comply with Legal, Regulatory, Privacy and Cybersecurity policies

3.4 Data Quality Tools

Table 5. Data Quality module functional definition

Module_Id	Data Quality
General Description	This module allows to properly define the minimum requirements in terms of data quality (format, latency, continuity …) to not compromise the results delivered by the tools available in the BD4NRG-MARKETPLACE.
Functionality	Data sources will be pre-processed to apply the defined profiling rules; evaluate profiling routines; define cleaning rules; auditing the cleaning rules; check data standards and formats
Input data	Energy data assets
Output data	Data standardized to comply with the Marketplace requirements
Interconnections (Interfaces)	Data Sources Actors Energy Assets Data Access Layer
Dependencies	Interoperability and Homogenization (T3.5)

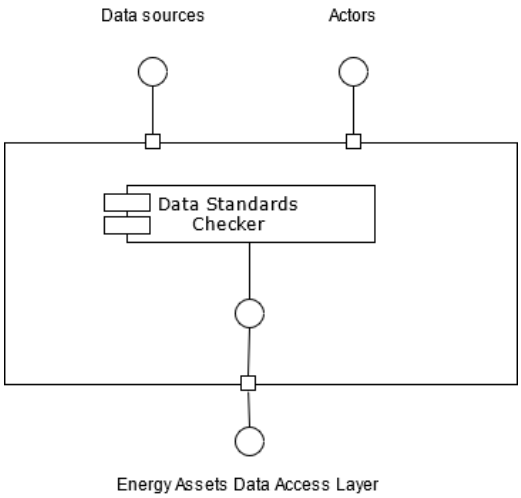

Figure 4. Data Quality Component Diagram.

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

3.5 Interoperability & Homogenization

Table 6. Interoperability and Homogenization module functional definition

Module_Id	Interoperability and Homogenization
General Description	The BD4NRG project will use the IDS data model as a starting point to have a common model for the data interoperability (as explained in D2.3). In addition, the Interoperability and homogenization module will be in charge of integrating and managing heterogeneous data from different sources and/or platforms in the energy ecosystem and beyond. Such data, historical or generated in real-time, will be provided by the BD4NRG data providers via different channels/technologies such as IoT devices, smart meters, SMX, sensors, etc. The Interoperability Service module will be based on specific data connectors developed as micro-services to manage the different data available under different formats and communication protocols (Rest APIs, SFTP, IoT protocols, Sensor Network). It will also provide interfaces to other energy and non-energy datasets/platforms of third parties willing to federate/integrate with BD4NRG, in order to allow the incremental deployment of the BD4NRG Data Hub. This Interoperability module will be ready to cover data exchange at EDGE level, offering in this case both local computing power and storage capacity to services and functions that may need ultra-low latency.
Functionality	The Homogenisation module will be built by different micro-services functionalities that constitute the pipeline to take care of the datasets that will be shared to the BD4NRG Energy Applications. This module can be organized in several stages to provide different functionalities, taking into account which task is being processed: • Pre-processing • Common Data Model • Semantic enrichment • Anonymisation *Pre-processing* service includes data curation, data cleansing, homogenisation and noise reduction. *Common Data Model* tasks transform Pilots' data into data models relying on exiting standardized data models. *Semantic enrichment* process will provide the data semantic annotation and establish an ontology to define appropriate semantic information in the energy sector. *Anonymisation* will allow only the usage of public and non-sensitive data.
Input data	Data sources
Output data	Data sources homogenized
Interconnections (Interfaces)	The idea is to develop the Homogenization module taking advantage from the Apache NiFi capabilities. Moreover, to facilitate the interfaces and interconnections, the JavaScript Object Notation (JSON) format will be used as an output format of the data pre-processing service because it is a lightweight data-interchange format easy to manage.

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

3.6 Seamless Elastic Distributed Data Management and HTAP

Table 7. Data Virtualization and Delivery Platform module functional definition.

Module_Id	Data Virtualization and Delivery Platform
General Description	This module offers *data consumers* a unified, abstracted, and encapsulated view for querying and manipulating data stored in a heterogeneous set of *data stores* that will be – thus – presented as one integrated data source. *Data Consumer*: refers to any application that retrieves, enters and/or manipulates data. *Data Store*: is any source of data that is relevant for the BD4NRG project. At this point the platform will be specified to provide a *virtual space* – during the run-time – that ensures that data will stay in place even if they appear as though they sit in a single place. This design choice can be changed if a physical data warehouse is needed. Therefore, the platform does not contain any actual data; it only contains the logic/intelligence for accessing the data sources.
Functionality	The Data Virtualization and Delivery Platform is responsible to integrate data from heterogeneous and distributed data sources including data-at-rest, machine data, external services while delivering a single access point to data consumers for accessing data. The platform will provide both run-time and design-time intelligence. The latter allows to transform data from data sources into virtual tables while specifying a *mapping*. The former includes all the necessary mechanisms to implement *on-demand* data integration by using the *mapping*.
Input data	Energy data assets
Output data	Pilot and Use-case specific dataset
Interconnections (Interfaces)	Data Sources Actors Data Querying for Business Intelligence Applications from Task 3.7
Dependencies	Seamless Elastic Distributed Data Management and HTAP (Heterogeneous Data Capture, Cleansing, Integration, Polyglot Persistence (T3.6))

In this case, no UML diagram is shown but the tentative structure and architecture of this Data Virtualization and Delivery Platform is instead in Figure 5. There, the lower layer contains diverse data sources, which connect to the Data Virtualization Server application where all that data combines and gets ready to be consumed by different users and applications.

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

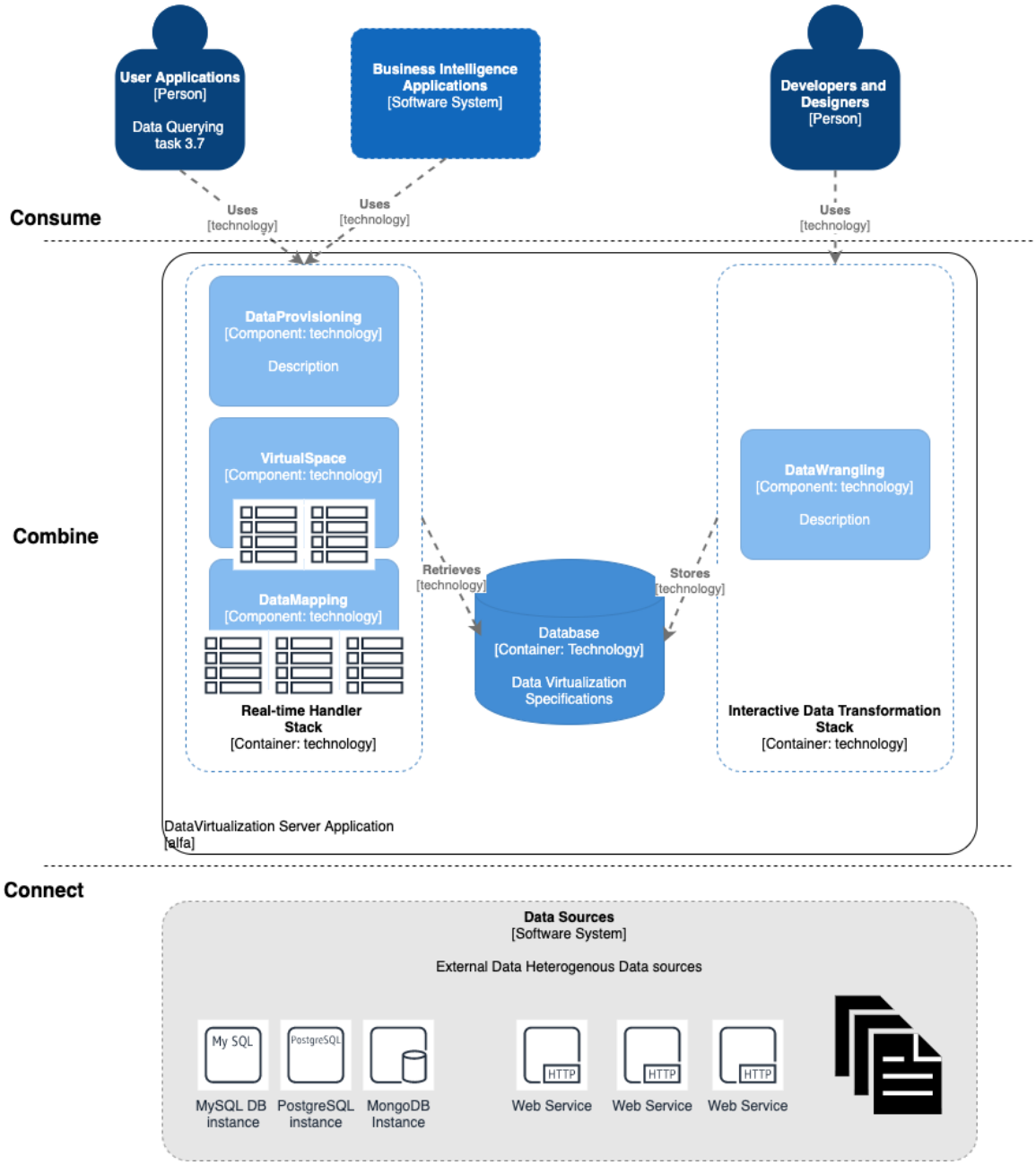

Figure 5. Data Virtualization and Delivery Platform tentative structure

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

3.7 Integrated Querying of Streaming Data and Data at Rest

Table 8.Integrated Querying of Streaming data module functional definition

Module_Id	Integrated Querying of Streaming data
General Description	This module enables the creation and efficient execution of queries to BD4NRG data sources, including combinations of different and heterogeneous data sources (e.g., joins). It also allows queries to near real time data. Moreover, it provides a useful interface to the users to enable them build queries without any prior knowledge of any programming language. In addition, it enables a user to save and retrieve queries in order to be used in the future.
Functionality	The querying module is responsible for querying different data sources that are built with different technologies, and the combination of them (e.g., real-time data with historical data or their metadata). It will be accessible not only through a user interface to the users, but also through connectors to other services and/or APIs.
Input data	Static and real-time data sources.
Output data	Query results.
Interconnections (Interfaces)	Data Sources Energy Assets Data Access Layer Users Other services
Dependencies	Interoperability and Homogenization (T3.5)

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

3.8 Automated Parallelization of Data Streams and Intelligent Data Pipelining

Table 9. Dynamic Data Provision module functional description

Module_Id	Dynamic Data Provision
General Description	Adapt data coming from diverse data sources to the BD4NRG common data model. This action is performed by the Dynamic Data Provider which through the Controller and Provenance Collector gets data into Distributed Database and Storage and delivers processed data to the Distributed Message broker.
Functionality	Adaptability and flexibility
Input data	Data sources
Output data	Distributed Message broker
Interconnections (Interfaces)	---
Dependencies	Data received from scenarios

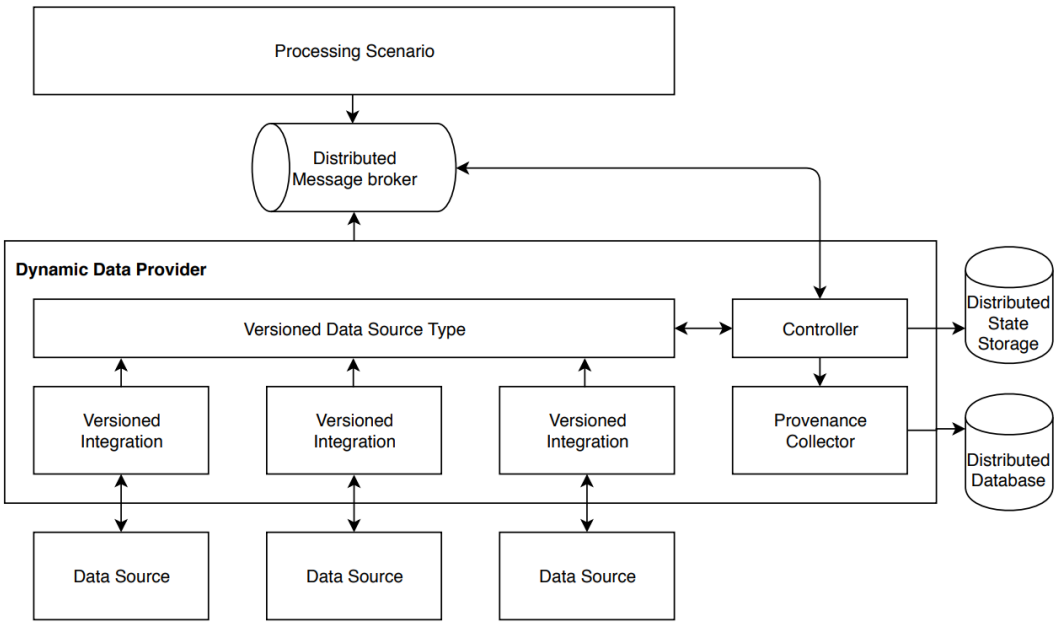

Figure 6. Dynamic data provision layer

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

Table 10. Data Pipelines Optimizer module functional description

Module_Id	Data Pipelines Optimizer
General Description	The module optimizes the distribution and parallelization of Data Pipelines on the infrastructure nodes, trying to exploit as much as possible data locality
Functionality	Optimization of data pipelines and parallelization of data streams
Input data	Data Catalogue, infrastructure status
Output data	Data streams allocation
Interconnections (Interfaces)	---
Dependencies	Module providing information about the infrastructure resources

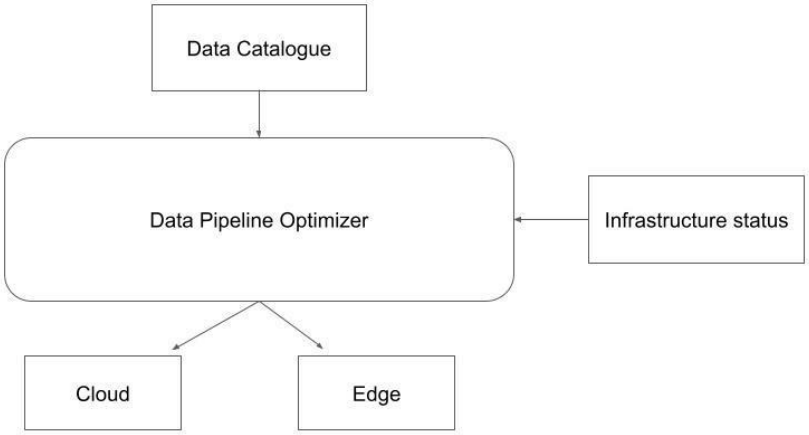

Figure 7 Data Pipelines Optimizer interactions diagram

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

3.9 International Data Spaces

3.9.1 IDS Connector

The IDS Connector is the core of the data space. It is the gateway to connect existing systems and their data to an IDS ecosystem. Its architecture and functionalities are defined by the IDS Reference Architecture Model (RAM) and specified by the certification criteria.

The IDS Connector allows to exchange data and enrich it with metadata. An important aspect of this are usage conditions, which can be defined, administrated, and implemented by the Connector. The metadata is described by the ontology of the IDS Information Model. The main advantage of the IDS reference architecture and the use of an IDS Connector is the decentralized data storage. This enables data integration from different data sources and allows data access exclusively through other IDS Connectors. Thus, a technical implementation of data sovereignty is achieved.

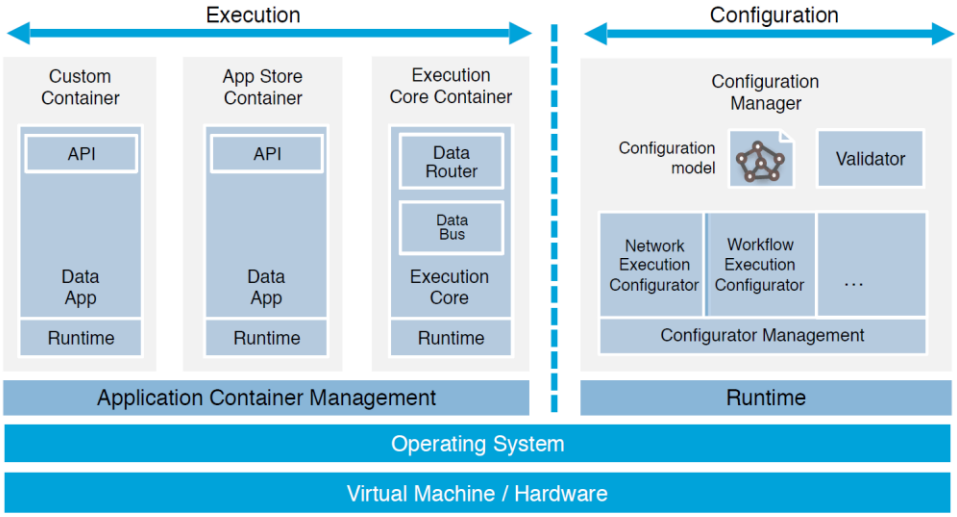

Figure 8. IDS Connector

An IDS Connector is composed of various system services:

- Execution core container with message systems (message router/bus)
- Configuration Manager to configure the Connector (execution core container, application container management, network, firewalls, etc.)
- Data Apps for data processing and handling
- Application container management
- Hardware/Operating system

Primarily to establish trust, the IDS Connector also addresses Application Container Management, Network, OS, Hardware, etc. Furthermore, trust between different IDS participants is ensured by certifying the software components used, i.e. the IDS Connector. Thereby, three different trust levels are defined: base, trust, and trust+. All profiles comprise IDS-specific requirements and some from Secure Development and the DIN standard for IT security for industrial automation systems (DIN EN IEC 62443-4-2). Details are listed here [3].

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

3.9.2 Reference Implementation

The Dataspace Connector is one of several implementations of an IDS Connector. It is the reference implementation of the data economy department of the Fraunhofer Institute for Software and Systems Engineering (ISST). It is maintained as an open source software project on GitHub, supported and promoted by the International Data Spaces Association (IDSA).

In addition to the project, other components are also being developed as open source. For example, the ConfigManager can be used to configure the connector and enable user-friendly interactions via a GUI (Graphical User Interface). Find more details here [4].

The Dataspace Connector is designed to provide an easy entry point into the IDS Ecosystem, enabling projects and companies to connect to data spaces and exchange data in a sovereign way. Thereby, the defined usage policies are not only transferred between the IDS participants, but also directly enforced. How the architecture is designed can be seen here [5].

Below there is a summary of the way the Dataspace connector documentation is structured under its current specification:

Repository with open source implementation

- Dataspace Connector Source Code
- Dataspace Connector Core latest docker file

Installation Guide

- Quick start (without valid Identity in IDS)
- Installation Guide
- Deployment Examples for Docker

Other interesting Documents

User API

- Dataspace Connector Communication Guide
- Dataspace Connector REST API

IDS API

- Dataspace Connector Messages
- IDS Message Flow

3.9.3 IDS Dynamic Attribute Provisioning Service (DAPS)

The infrastructure component "Digital Attribute Provisioning Service" (DAPS) in a given IDS ecosystem enables the enrichment of identities of organizations and connectors with additional attributes.

DAPS can be understood (like the Certification Authority [6], too) as building block of the construct of an IDS Identity Provider [7].

These attributes are embedded by the DAPS into a Dynamic Attribute Token (DAT) [8]. The advantages of this technique of *dynamic attribute provisioning* instead of embedding them into X.509 certificates [9] are:

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

- Attribute revocation does not enforce certificate revocation and re-issuance. This is also true if new attributes are added.
- By defining scopes or attribute sets, only the needed attributes can be included in the DAT. This limits information leakage, so only needed attributes are communicated.
- Baseline certificates can be issued to enable connector deployment in uncomplicated manner.
- More complex scenarios can be created as soon as attributes are assigned later on.

See also:

- Glossary "Dynamic Attribute Provisioning Service" [10]
- Shortcut: DAPS

3.9.3.1 Unique Identifier

Each connector in the IDS needs a valid, outlasting and unique identifier, never be re-used for any other resource inside the IDS ecosystem.

The architecture aims to be open to multiple Certification Authorities (CAs) issuing certificates. This means, a truly unique identifier needs to consist of the issuer of the certificate and the subject identifier. For an easy machine readable identifier, two ´X.509v3´ extensions will be used: the Subject Key Identifier (SKI) and the Authority Key Identifier (AKI).

3.9.3.2 Subject Key Identifier (SKI) & Authority Key Identifier (AKI)

The concatenation of ´SKI´ and ´AKI´ provides a valid unique identifier even if multiple CAs are able to issue valid certificates.

EXAMPLE (snippet from a X.509 certificate):

```
X509v3 extensions:
    X509v3 Subject Key Identifier:
        DD:CB:FD:0B:93:84:33:01:11:EB:5D:94:94:88:BE:78:7D:57:FC:4A
    X509v3 Authority Key Identifier:
        keyid:CB:8C:C7:B6:85:79:A8:23:A6:CB:15:AB:17:50:2F:E6:65:43:5D:E8
```

leads to a connectors unique identifier:

```
DD:CB:FD:0B:93:84:33:01:11:EB:5D:94:94:88:BE:78:7D:57:FC:4A:keyid:CB:8C:C7:
B6:85:79:A8:23:A6:CB:15:AB:17:50:2F:E6:65:43:5D:E8
```

In examples and for reasons of readability editors might use SKI:AKI.

3.9.4 IDS Metadata Broker

The IDS Meta Data Broker is a service for publishing and searching metadata of Connectors and resources between International Data Spaces Participants. In order to ensure the necessary interoperability and general interactions, an IDS Meta Data Broker (like the App Store) is also defined as a specialized IDS Connector. The communication between an IDS Connector and an IDS Meta Data Broker is based on the same principles as any other Connector-Connector communication within the International Data Spaces. Still, an IDS Meta Data Broker provides a collection of additional functionalities: - Indexing services in order to effectively and efficiently respond to queries and present known Connectors and other resources - Interfaces for Users or IDS-Messages to ensure access to the stored information.

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

3.9.4.1 Brokering

The operator of a Connector must be able to provide an interface for data and metadata access. Each Connector must be able to transmit metadata from its data sources to one or more brokers. Each participant must be able to browse and search metadata in the metadata repository, provided the participant has the right to access the metadata. Furthermore, each participant must be able to browse the list of participants registered at a broker.

The current specification for this *MetaData Broker* is available at IDS-G [11], while a white paper on its specification can be consulted as well [12].

In addition, GitHub offers a Repository with open source implementation [13].

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

4 Data Modelling & Open Modular AI-based edge-level Analytics Toolbox

Figure 9 represents the logical architecture of the Analytics Toolbox. The edge side is tasked with processing data in real time and exchanging information with the cloud. On the cloud side, the overall toolbox is accessible by means of the API gateway, the access point from external requests. It will be based on Kubernetes that orchestrates cloud resources. The services of the catalogue dedicated to BD4NRG are managed by a Big Data Analytics (BDA) orchestrator which leverages an existing tool by ENG (ALIDA [14]) to deploy, execute, monitor, schedule BDA applications. The data to be processed are saved in a secure storage. The BDA orchestrator will be supported by all the frameworks and libraries used within the BDA services (execution cluster).

The Edge-to-cloud communication must take place through a component (Data connector) that will manage the interaction of/with cloud-federated edges.

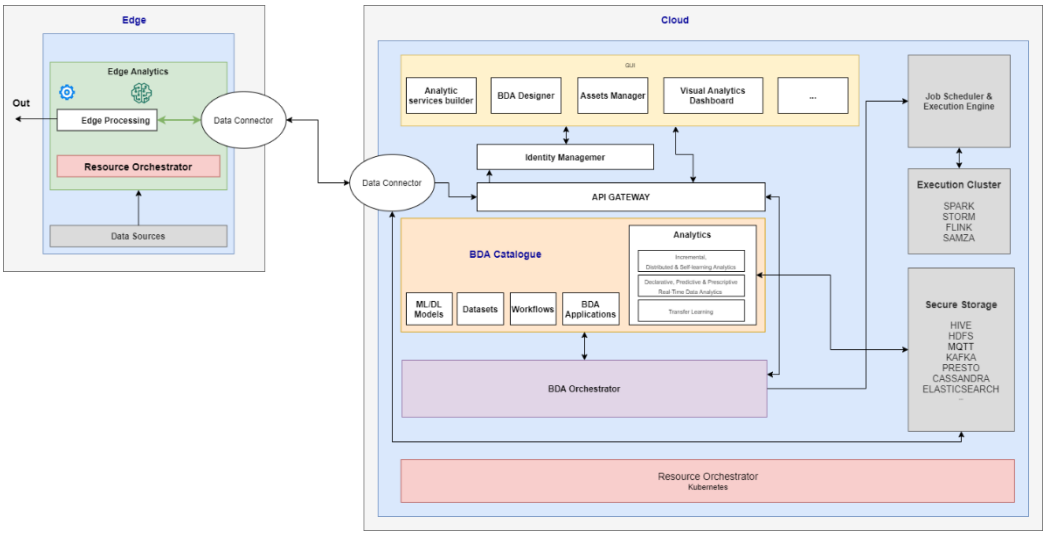

Figure 9. Analytics Toolbox Architecture

The list provided in Table 22 summarizes the modules that shape the architecture and the B4NRG consortium partner which owns the responsibility over its development.

Table 11. List of architecture modules and the responsible partner

#	Module	Partner
	Edge node	ENG
	Data connectors (edge and cloud side)	TNO
	Analytic service builder	ENG
	BDA designer and asset manager	ENG
	Visual analytic dashboard	NTUA
	Identity Manager	ENG
	BDA Orchestrator	ENG

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

#	Module	Partner
	Incremental, Distributed & Self-learning Analytics	NTUA
	Cross-stakeholder Transfer learning analytics	IMEC
	Declarative, Predictive & Prescriptive Real-Time Data Analytics	ENG

Every subsection in this chapter focuses on one of these modules and offers an initial description along with their intended functionality.

4.1 Edge Node

Table 12. Edge Node definition

Module_Id	Edge Node
General Description	This module allows the retrieval of data, offers processing algorithms and enables communication with the cloud.
Functionality	Each BD4NRG edge node must be able to process (edge processing) the data detected by the sensors and provide useful information to be sent to the cloud for further descriptive, predictive, etc. analysis. It will be supported by a suitable framework leveraging a resource orchestrator to ease the transformation of the information coming from sensors and devices and the delivery of it in formats and structured data that meet the needs of BD4NRG. The information will be sent to the cloud via the data connector for final processing. The Data connector will also be in charge of taking data from cloud and deliver it to the edge node for updates.
Input data	ML/DL Model Streaming Data
Output data	Streaming Data
Interconnections	Data Connector Data Sources
Dependencies	--
UML diagram	Not available yet

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

4.2 Data Connectors (Edge and Cloud side)

Table 13. Data Connectors (edge and cloud side) definition

Module_Id	Data Connectors
General Description	These assets establish the path that gets data from the cloud to the federated edges and vice versa.
Functionality	Manage the communication among the API Gateway in the Cloud to the Edge Processing module in the federated edge.
Input data	Data coming from the Resource Orchestrator in both sides
Output data	Data apt to conduct analytics and to feed services.
Interconnections	API Gateway Edge Processing
Dependencies	Secure Storage in the Cloud Data Sources in the Edge
UML diagram	Not available yet

4.3 Analytic Service Builder

Table 14. Analytic Service Builder definition

Module_Id	Analytic service builder
General Description	This component will allow the registration of new Big Data Analytics (BDA) services and the deletion and update of the existing ones.
Functionality	By means of a wizard, it must be possible to allow data scientists, IT operators, and in general BDA service providers to create, register, and manage BDA service as microservices. The aim of this tool is to automate the creation of Docker files, pushing the new images inside a repository and the registration into the BDA service catalogue.
Input data	BDA Service Properties Algorithm (URL of a Docker image)
Output data	BDA Service
Interconnections	Identity Management API Gateway BDA Catalogue BDA Manager

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

Module_Id	Analytic service builder
Dependencies	--
UML diagram	Not available yet

4.4 BDA Designer and Asset Manager

Table 15. BDA Designer and Asset Manager definition

Module_Id	BDA designer and asset manager
General Description	This component is responsible to enable the design, scheduling, execution, and monitoring of BDA applications.
Functionality	A BDA application must be created as a bundle generated from the design and successfully execution of a workflow of BDA services properly chosen to analyse specific data sets. Such a bundle includes several assets: data sets (input data set, intermediate data sets, resulting data sets), the set of BDA services composing the workflow, the workflow itself, and any models trained within it. The Asset Manager manages the assets for further usages.
Input data	BDA Service BDA Service parameters Datasets ML/DL Models
Output data	Datasets ML/DL Models
Interconnections	API Gateway Identity Management BDA Catalogue
Dependencies	--
UML diagram	Not available yet

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

4.5 Visual Analytic Dashboard

Table 16. Visual Analytics Dashboard definition

Module_Id	Visual Analytics Engine
General Description	This component provides data visualisation utilities, along with interactive dashboards and reports to the users.
Functionality	Visual Analytics Engine will use as input one or more datasets and will offer a wide variety of configurable visualisations to facilitate exploratory analysis. Moreover, it will provide the ability of saving visualisations and create dashboards and reports, that will combine different data visualisations and other content.
Input data	Any dataset from the harmonized data storages of the project, or even combination of different datasets.
Output data	A wide variety of data visualisations, from simple scatterplots to heatmaps, histograms, and regional maps visualisations. Also, combination of different visualisations, along with custom content. The result will be available through a web page as well as downloadable.
Interconnections	Available datasets, Access Control.
Dependencies	Available datasets, Access Control.
UML diagram	Not available yet

4.6 BDA Orchestrator

Table 17. BDA Orchestrator definition

Module_Id	BDA Orchestrator
General Description	This component is responsible to enable the design and the execution of Big Data Analytics (BDA) applications.
Functionality	The BDA Orchestrator must enable the design, deployment, execution, and monitoring of big data analytics workflows (both stream and batch). It must be cloud-native so as to scale computing and storage resources by leveraging the capabilities of Kubernetes for cloud resource management. The BDA Orchestrator will be developed by adoption and proper extension of ALIDA. ALIDA is a micro-service based platform for composition, deployment, optimization, execution, and monitoring of big data analytics workflows (covering ingestion, preparation, analysis and visualization). It is a research asset by ENG resulting from previous and ongoing research and innovation actions, designed and developed on top of the most advanced open-source Big Data technologies and frameworks.

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

Module_Id	BDA Orchestrator
	This module will be responsible for ensuring the consistency among the assets that make up a BDA Application. It will also be responsible for communication with Engine for the management of executions and BDA Services.
Input data	Execution Information BDA Application Details BDA Service Model
Output data	--
Interconnections	API Gateway Identity Management BDA Catalogue Execution Engine
Dependencies	--
UML diagram	Not available yet

4.7 Incremental, Distributed & Self-Learning Analytics

Table 18. Incremental, Distributed & Self Learning Analytics definition

Module_Id	Incremental, Distributed & Self-Learning Analytics
General Description	This module involves the exploitation of a large variety of technologies, algorithms and optimization techniques, aiming to efficiently train complex models and to solve computationally intensive problems.
Functionality	The incremental, distributed and self-learning analytics module is responsible for the necessary training and evaluation procedures, regarding the exploitation of the most suitable ML and DL algorithms per case. Considering the problem that is addressed in each case, the module will provide the most appropriate model of supervised or unsupervised learning, among several pre-trained algorithms for various problems, such as classification, regression and clustering, among others. Also, fine tuning will be performed, and the hyper parameters will be examined for each one of the considered algorithms, in order to optimize their performance, with respect to the size of the particular dataset. Finally, the performance of the delivered algorithms will be measured with the most appropriate performance measures per case, involving mean absolute and squared error, confusion matrix, classification accuracy etc. In practice, the developing team will have to construct: • Self- learning models • Models that exploit knowledge from large data sets (cross- and transfer-learning)

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

Module_Id	Incremental, Distributed & Self-Learning Analytics
	• Scalable services (effectively handle Big Data and deep learning applications) • Intuitive algorithms (transform black box ML models into actionable knowledge) Easily accessible and computationally cheap models real time evaluation frameworks and documentation.
Input data	Various streaming and historical datasets in order to train the models including: • Energy consumption (electricity, natural gas and oil) • Energy production (wind power, solar) • Weather forecasts (radiation, temperature, wind speed, wind direction, humidity) • Building information (orientation, age, floor area, type, year of construction) • Occupant behaviours (habits, operational hours, comfort)
Output data	ML/DL algorithms output (forecasts for certain variables, classification results) which will differ according to the expected solution of the model Analytics results (reporting and visualizations) for the respective services
Interconnections	BDA Orchestrator
Dependencies	Task 4.7. Open Modular Analytics Toolbox & Deployment: The developed algorithms and models will be integrated to the Analytics Toolbox, after the fine-tuning of hyperparameters has been conducted.
UML diagram	Not available yet

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

4.8 Cross-stakeholders Transfer Learning Analytics

Table 19. Cross-stakeholders Transfer Learning Analytics definition

Module_Id	Cross-stakeholder Transfer Learning Analytics
General Description	The learning process for new residential assets must start from scratch as there is no historical data available. This cold-start problem causes slow learning and a heavy cost on the system as time and effort are required to build up site-specific training data sets which in turn limits scalability. This module exploits the potential of transfer learning technologies to overcome the challenges of limited data availability and long training times.
Functionality	Residential flexibility that is valorised on multiple energy services is becoming increasingly interesting. Residential assets such as electric boilers, heat pumps, electric vehicles and batteries are an untapped potential of flexible energy to integrate renewable energy by providing balancing services and mitigate grid congestion. In order to control these assets efficiently, flexibility providers are building models of the assets to optimize their control actions. This is a labour-intensive and time-consuming process that limits scalability due to the aforementioned challenges. The goal of this module is to leverage data that is obtained from similar assets but that are deployed in different locations or different external conditions (user habits, grid constraints, climate). By leveraging this data, initial control policies can be developed for new assets which makes it possible to overcome the cold-start problem and not have to start the learning process from scratch.
Input data	• Weather data (temperature, solar irradiance, humidity, etc.) • Public LV-MV grid topologies or representations • State information of flexible devices, such as measurements of internal temperatures, state-of-charge, energy consumption of the assets, along with a description of the available building attributes when available (e.g., solar panels, assets that can be used for demand response such as heat pump, electric vehicles, boilers, storage tanks, batteries, etc.)
Output data	At each time step (to be defined if this will be every 15 min, every hour, etc.), the algorithm will output the optimal action to take for all demand response related assets of a corresponding household.
Interconnections	• BDA Orchestrator • BDA Catalogue • Data Connector • Secure storage
Dependencies	--
UML diagram	Not available yet

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

4.9 Declarative, Predictive & Prescriptive Real-Time Data Analytics

Table 20. Declarative, Predictive & Prescriptive Real-Time Data Analytics definition

Module_Id	Declarative, Predictive & Prescriptive Real-Time Data Analytics
General Description	It is a kind of BDA Service, they will be implemented as Docker-based microservices and registered within the catalogue.
Functionality	BD4NRG must provide a set of declarative, predictive and prescriptive data analytics for real time data processing to make stream data pipelines operational. In general, descriptive analytics describe the past. They are useful because they allow us to learn from past behaviours and understand how they might influence future outcomes. Predictive analytics provide insights based on data, estimating the likelihood of a future outcome. Prescriptive analytics attempt to quantify the effect of future decisions, in order to advise on possible outcomes before the decisions are actually made.
Input data	BDA Service Properties ML/DL Model Dataset
Output data	Dataset ML/DL Model
Interconnections	BDA Catalogue BDA Manager Execution Engine Secure Storage Identity Management
Dependencies	--
UML diagram	Not available yet

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

4.10 Identity Manager

Table 21. Identity Manager definition

Module_Id	Identity Manager
General Description	This component will verify a user's identity and their level of access to a system or a service.
Functionality	This component is responsible for the identity management and must offer features such as Single-Sign-On (SSO), Identity Brokering, User Federation. Moreover, it aims to guarantee the user's secure access to own assets (data sets, trained models, BDA applications).
Input data	Credentials information Token
Output data	Organization Details Role Details Token
Interconnections	API Gateway GUI
Dependencies	--
UML diagram	Not available yet

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

5 Marketplaces for Realizing EU-level Energy Data Economy

The integrated BD4NRG framework as depicted in Figure 10 will offer a marketplace that facilitates the exchange of data, a variety of energy assets, third party services, computing resources, storage resources and even machine learning models. It will integrate functionalities related to the publication and management of metadata, the Orchestration component will oversee the management of the access and the metadata matching among data/service providers and consumers. It will include a module in charge of tokenization which will allow the creation of a utility monetary and/or non-monetary reward/incentive metric/function.

The Marketplace will provide a library of reusable AI-based Machine Learning (ML) models offered as a service (MLaaS), with the focus on the promotion of the quick adaptation and reuse of machine learning models along different contexts. Also, the Marketplace will facilitate access to simulated data testbeds (smart grid data on the top of grid model-based hardware-in-the-loop), as well as to supplementary external off-domain data sets (e.g., weather, geographical data) via third party services (Data-as-a-Service, DaaS) through interoperable service level Open APIs.

Furthermore, the deployment of a Tools Compliance management will be performed, intended to manage, and assess the compliance of the third-party services and tools to the BD4NRG reference architecture.

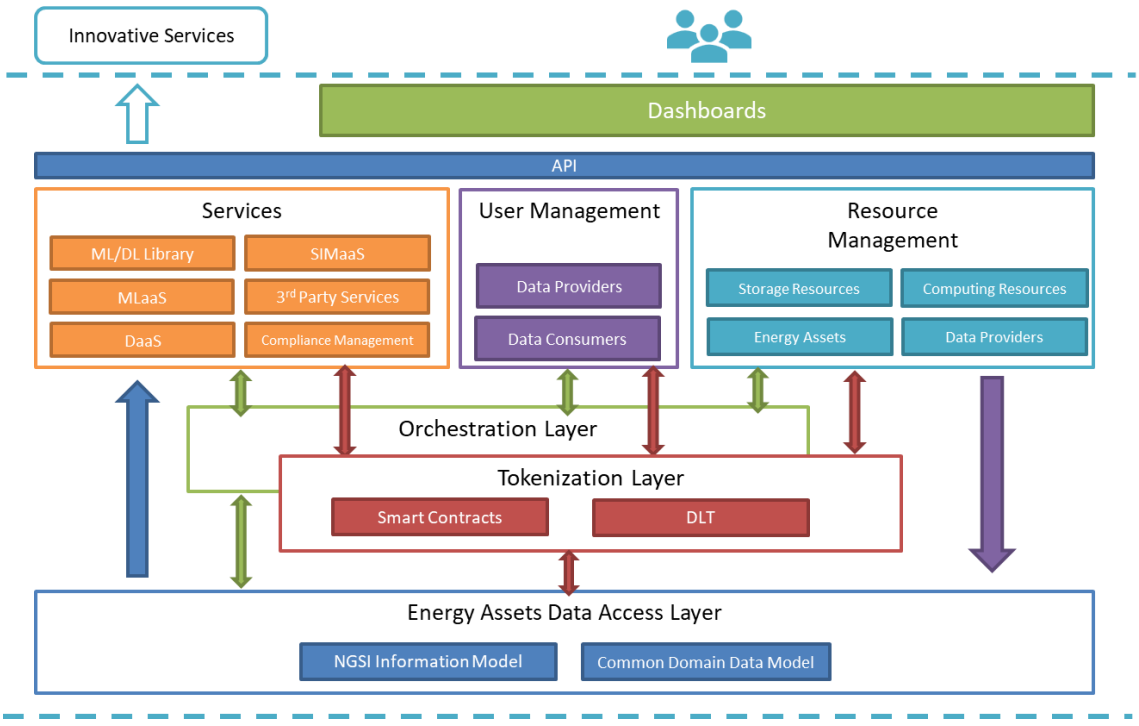

Figure 10. BD4NRG Marketplace Logical Architecture

Table 22 summarizes the list of modules that comprise this layer in the architecture. Each one of them is initially discussed in the different subsections of the present chapter.

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

Table 22. List of architecture modules in the marketplace and their responsible partner

#	Module	Leader
1	Energy Assets Data Access Layer	Atos
1.1	NGSI Data Model	Atos
1.2	Common Domain Data Model	Atos
2	Blockchain Based Asset Management Layer	ENG
2.1	Data Handling	ENG
2.2	Digital Asset Tokenization	ENG
3	Orchestration Layer	Atos
4	Service Layer	Atos
4.1	MLaaS – ML/DL Library	Atos – TS
4.2	DaaS	Atos
4.3	SIMaaS	RWTH
4.4	3rd Party Services	-
5	User Management	Atos
5.1	Data Providers	Atos
5.2	Data Consumers	Atos
6	Resource Management	Atos
6.1	Storage Resources	Atos
6.2	Computing Resource	Atos
7	Marketplace API	Atos
8	Marketplace Dashboard	Atos

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

5.1 Logical View of BD4NRG Marketplace architecture

5.1.1 Energy Assets Data Access Layer

Table 23. Energy Assets Data Access Layer definition

Module_Id	Energy Assets Data Access Layer
General Description	This module allows the ingestion of data from distributed sources, persistence, and streaming.
Functionality	The functionality allows to collect data in Common Data Model (CDM) format and in FIWARE Information Model (NGSI) natively from different sources.
Input data	Data access configuration. Data Query
Output data	NGSI and CMD data packages.
Interconnections (Interfaces)	Data Required Interface. Data Configuration Interface. Data Notification Interface.
Dependencies	Data Service Module. 3rd Party Module. Orchestration Layer. Resource Management Module.

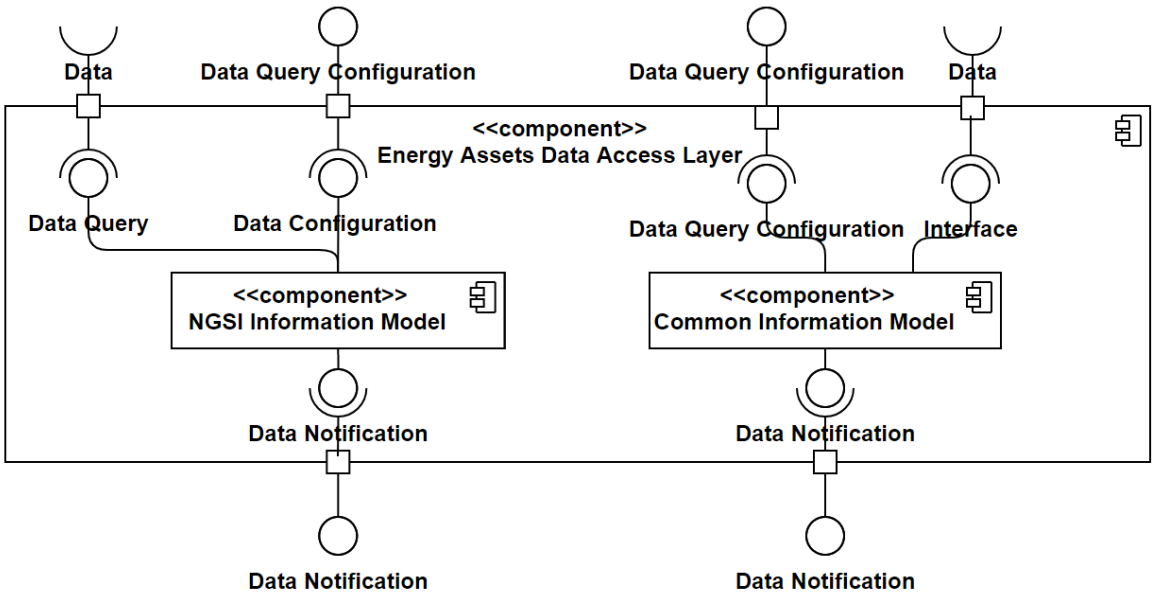

Figure 11. Energy Assets Data Access Layer Component Diagram

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

5.1.2 Blockchain Based Asset Management Layer

Table 24. Blockchain Based Asset Management Layer component definition

Module_Id	Blockchain Based Asset Management Layer
General Description	Handling the digital assets leveraging on blockchain tokens according to the loyalty model
Functionality	This layer encompasses two different modules, one concerning the mechanism of data handling and sharing and the other related to the digital assets trading. The latter manages the assets trading exploiting the tokenization model. Both modules rely on the blockchain smart contracts for trustiness.
Input data	Data from Energy Assets Data Access Layer (e.g., energy meters data, data on assets usage)
Output data	Users' tokens, Results of value transaction for assets exchange.
Interconnections	Service Configuration Interface Resources Configuration Interface User Permission Configuration Interface. Data Configuration Required Interface
Dependencies	Detail if there is any dependency with other modules (e.g., one service is able to start if another is already running)

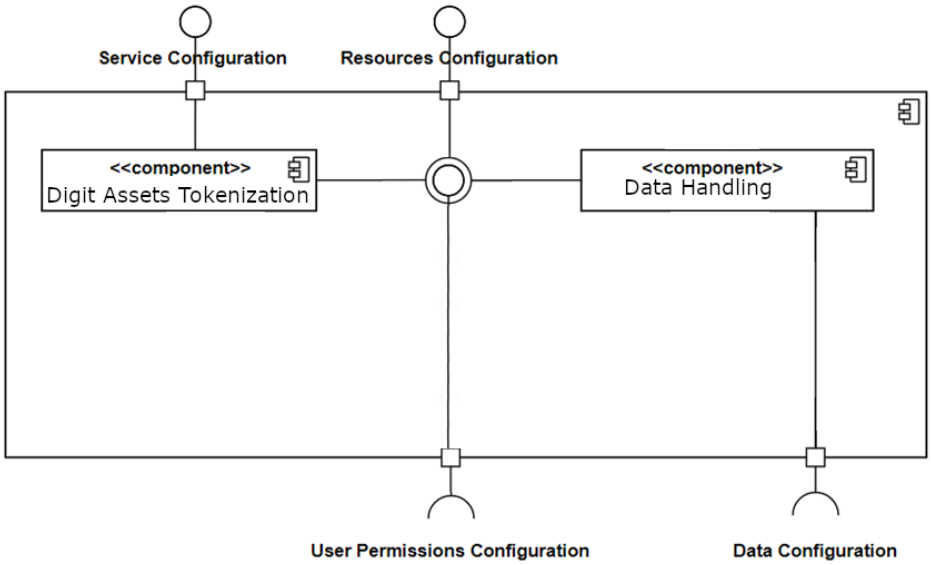

Figure 12. Blockchain Based Asset Management Layer Component Diagram

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

5.1.3 Orchestration Layer

Table 25. Orchestration Layer module definition

Module_Id	Orchestration Layer
General Description	This module allows to manage the service access to data and metadata resources.
Functionality	The module allows to configure services according with the resource configuration and the data providers and data consumers permissions.
Input data	Service Configuration Data. Resources Configuration Data User permission Data
Output data	Datasets Configuration
Interconnections	Service Configuration Interface Resources Configuration Interface User Permission Configuration Interface. Data Configuration Required Interface
Dependencies	Energy Assets Data Access Layer

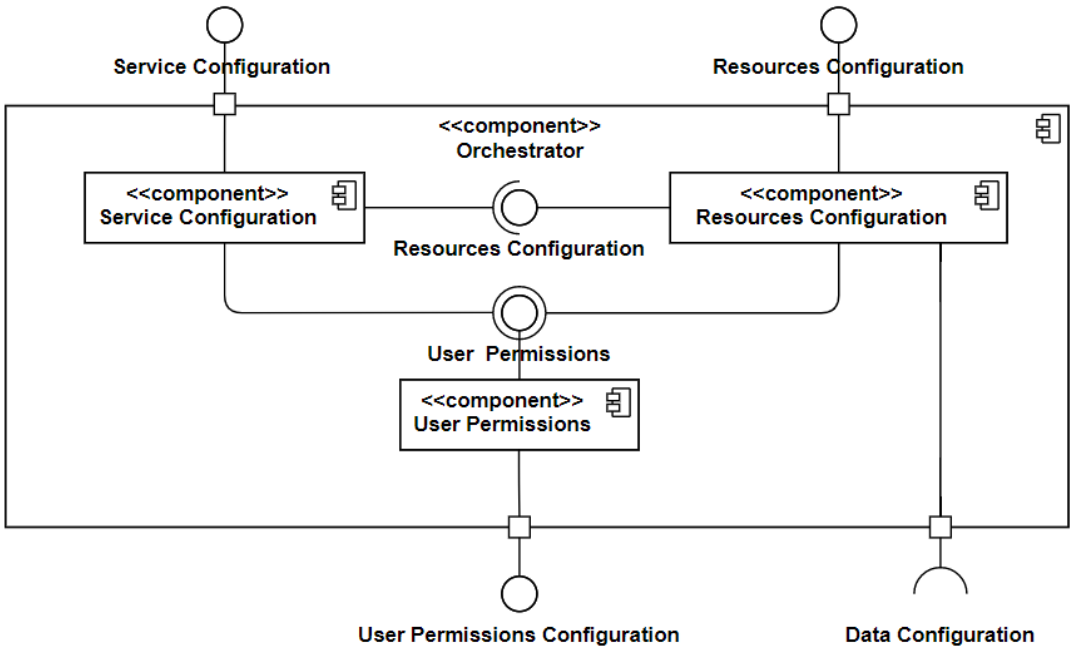

Figure 13. Orchestration Layer module diagram

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

5.1.4 Service Layer

5.1.4.1 MLaaS – ML/DL Library

Table 26. MLaaS – ML/DL Library component definition

Module_Id	MLaaS
General Description	This module allows to provide Machine Learning as a Service in a cloud-based environment
Functionality	The module will provide a set of trained Machine Learning libraries that will be deployed on demand in the Marketplace cluster
Input data	Dockerized trained models. Basic Service Configuration Prediction request to the service REST endpoint.
Output data	ML models information. Prediction response.
Interconnections	Service Configuration Interface. Service Query Interface. Model Provisioning Interface.
Dependencies	Blockchain Based Asset Management Layer Orchestration Layer.

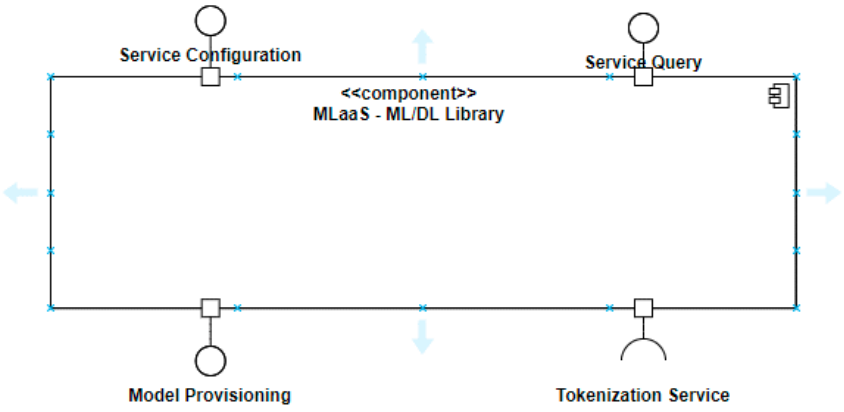

Figure 14. MLaaS Component Diagram

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

5.1.4.2 DaaS

Table 27. Data as a Service component definition

Module_Id	DaaS
General Description	This module allows to collect data from different sources to provide new services.
Functionality	The module allows to retrieve data from the sources configuring streaming or on-demand data consumption. The module will provide a discovery service to allow data consumer to identify new potential services from diverse data sources and types.
Input data	Data Configuration Resource Configuration Input Data
Output data	Metadata Output Data
Interconnections	Data Service Configuration Data Service Provision Discovery Service Resource Configuration Data Provisioning
Dependencies	Energy Assets Data Access Layer Resource Management Orchestration Layer.

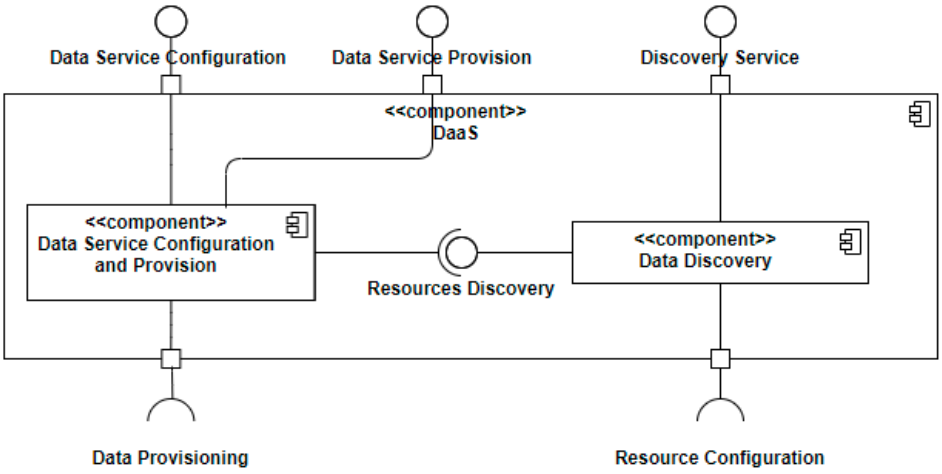

Figure 15. DaaS Component Diagram

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

5.1.4.3 SIMaaS

Table 28. SIMaaS module definition

Module_Id	SIMaaS
General Description	This module provides Simulation as a Service.
Functionality	The module consists of several independent but mutually complementary submodules. The module is structured in different layers for different timing requirements. The Web Interface Component offers facilities to plan, configure, control, execute and visualize distributed simulation projects through a web interface. The Data Gateway Component offers interfaces to simulation equipment, databases, and web services. The Control Component offers a set of command line tools for simulation control.
Input data	Infrastructure Configuration Simulation Configuration Simulation Visualization Configuration
Output data	Simulation Data Simulation Data Visualization
Interconnections	Simulation Service Configuration Simulation Service Provision Data Service Provision Resource Configuration
Dependencies	Energy Assets Data Access Layer Resource Management User Management Orchestration Layer

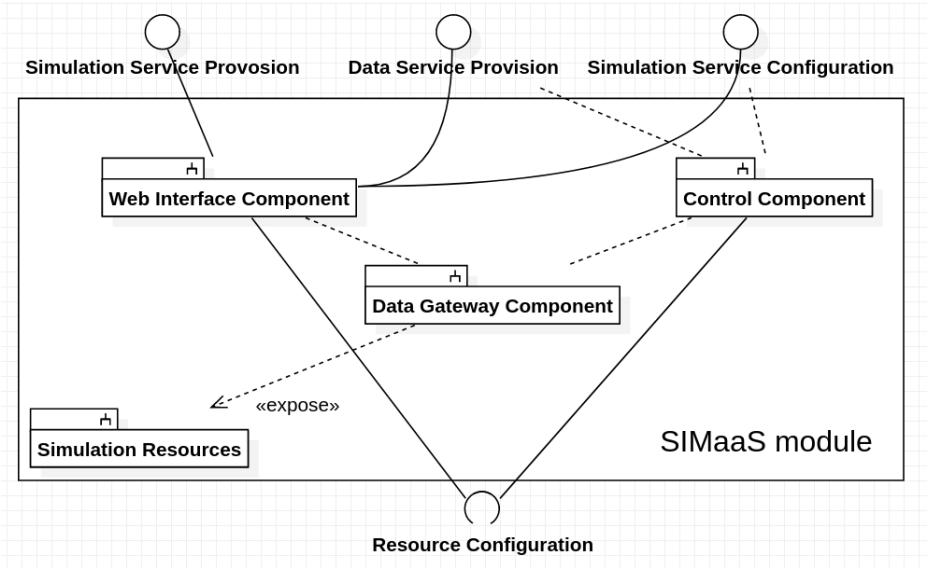

Figure 16. SIMaaS Component Diagram

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

5.1.4.4 3rd Party Services

Third-party services definition is an issue that has not been addressed yet, as it largely depends on the requirements gathered from the Pilot analysis in WP2. Therefore, its description is based on a generic description of service provision.

Table 29. 3rd Party Services module definition

Module_Id	3rd Party Services
General Description	This module allows to provide 3rd party services
Functionality	This module allows to configure and consume 3rd party services
Input data	Configuration Data Input Data
Output data	Service Output Data
Interconnections	Data Provisioning
Dependencies	Energy Assets Data Access Layer Resource Management Orchestration Layer

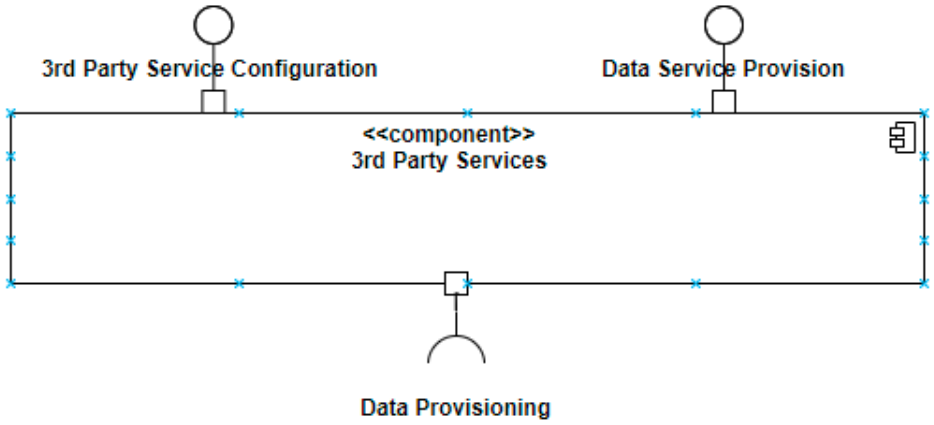

Figure 17. 3rd Party Services Component Diagram

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

5.1.5 User Management

Table 30. User Management module definition

Module_Id	User Management
General Description	This module allows to manage users and their privileges.
Functionality	The module allows to create, update, and delete Data Provider users and Data Consumer users. The component provides authentication and authorization functionality
Input data	User Data User Roles User Permissions Resource Configuration
Output data	User Data User Roles User Permissions Resource Configuration Authentication and Authorization
Interconnections	Data Providers Management Service Data Consumer Management Service Resource Configuration
Dependencies	Resource Management

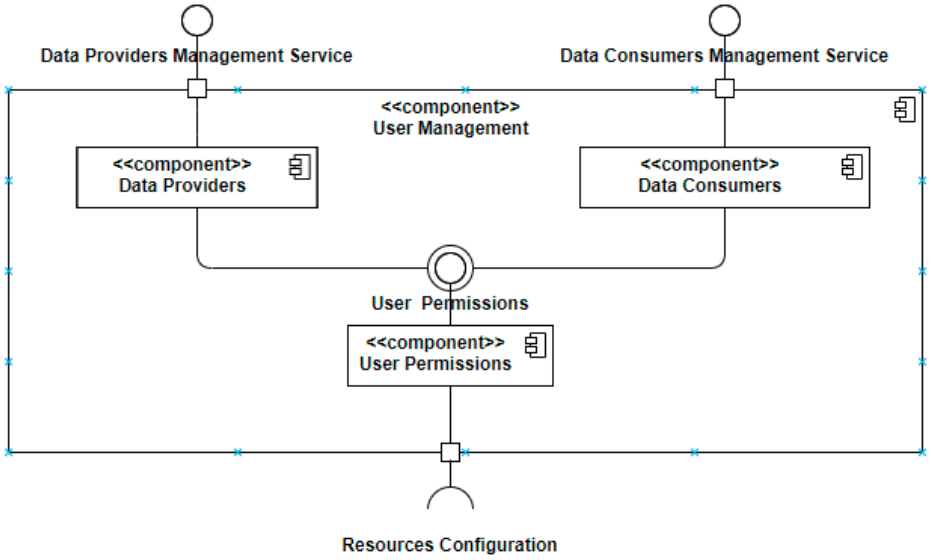

Figure 18. User Management Component Diagram

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

5.1.6 Resource Management

Table 31. Resource Management module definition

Module_Id	Resource Management
General Description	This module allows to manage storage and computing resources.
Functionality	The module will allow to create, update, and delete storage and computing resources.
Input data	Storage Resources Computing Resources User Permissions
Output data	Storage Resource Use Computing Resource Use
Interconnections	Storage Resource Service Computing Resource Service User Permission
Dependencies	User Management Component Orchestration Component

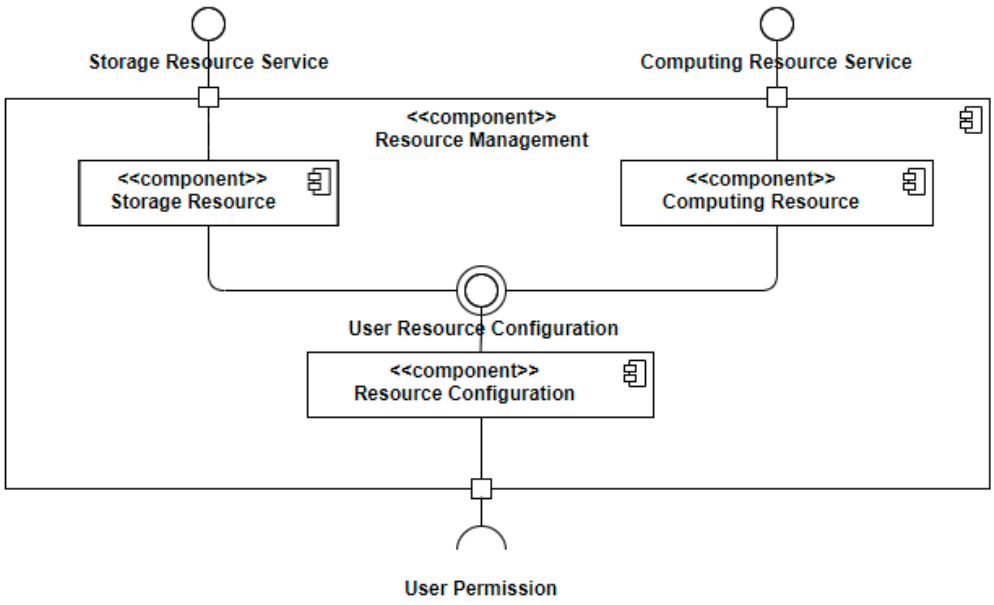

Figure 19. Resource Management Component Diagram

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

5.1.7 Marketplace API

The Marketplace API is a public available REST API that provides applications with access to the marketplace services like MLaaS, DaaS, SIMaaS, third party services, user management services, and resource management services.

5.1.8 Marketplace Dashboard

The Marketplace Dashboard is the GUI that Data Provider users and Service Provider/Data Consumer users can use to interact with the marketplace application. The dashboard serves as a metadata catalogue for the data provided by the data providers and a catalogue of data and services the service provider/data consumer can use to create new added value services. The dashboard is connected to the marketplace through the open API.

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

6 Conclusions

This deliverable offers an initial overview of BD4NRG platform as an interoperable and scalable architecture. It is built upon three main layers, presented under the 4+1 View Model. This first architectural release covers the so-called Logical View, while the rest of views will be discussed in upcoming reports.

The report is the result of a collaborative work between all partners involved, and one of its main objectives is to set the baseline for future iterations and implementation activities. As a matter of fact, this deliverable contains the contributions and vantage points from:

- *Pilots described in D2.1*. This report is one of the main inputs to BD4NRG architecture, since it provides an approach to the initial architecture the use cases will need to deal with. They focus on the development of energy marketplaces based on big data and advanced analytics, with innovative services and solutions in such domain. D2.2 uses this as a baseline and provides an initial description on the layers and their modules, as well as certain descriptive diagrams.
- *Initial BD4NRG architecture at proposal stage*. This draft proposal was employed as a guide and targeting at refining the proposed framework, trying to infer some of the interfaces and communication between the different modules that will be required.
- *Technical perspective*. BD4NRG marketplace includes forward-thinking and pioneering analytical solutions that entail the vision of component providers. This way, D2.2 captures the perspective of developers who may be involved in the architecture validation. This means partners from the technical work packages (WP3, WP4, WP5) have been involved in the preparation of the report.

These aforementioned contributions, along with criticism and advice received during this task, permitted the consolidation of this dedicated report which describes the logical design of the conceptual architecture addressing the concerns of all the stakeholders implicated in the *Logical vie*w, which provides a preliminary description of each of the three layers in the BD4NRG platform, detailing their logical connections and being useful for end users. The rest of views, namely Implementation, Process, Deployment and Use Cases views, will be provided in further deliverables.

With the release of this report, the BD4NRG architectural design team will switch the focus towards BD4NRG technical WPs, so that the exhaustive description and development of components can be aligned with the agreements already included in this report. Along these lines, those future implementations will be also straightforwardly mapped to the envisioned update of the architecture (D2.6), together with the aforementioned future inputs regarding the rest of architectural views.

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

References

[1] I. O. f. Standardization, «“ISO/IEC/IEEE 42010:2011 — Systems and software engineering — Architecture description”,» 2011-12-01.

[2] P. Kruchten, «“Architectural Blueprints—The “4+1” View Model of Software Architecture”,» de *IEEE Software*, November 1995, pp. vol. 12, no. 6, pp. 42-50.

[3] I. D. S. Association, «Data Space Connector Roadmap,» [En línea]. Available: https://international-data-spaces-association.github.io/DataspaceConnector/Roadmap/Ids-ready.

[4] I. D. S. Association, «Data Space Connector Features,» [En línea]. Available: https://international-data-spaces-association.github.io/DataspaceConnector/Features#libraries.

[5] I. D. S. Association, «GitHub Data Space Connector Architecture Documentation,» [En línea]. Available: https://international-data-spaces-association.github.io/DataspaceConnector/Documentation/Architecture.

[6] I. D. S. Association, «Certification Authority,» [En línea]. Available: https://github.com/International-Data-Spaces-Association/IDS-G/blob/master/core/CA/README.md.

[7] I. D. S. Association, «Identity Provider,» [En línea]. Available: https://github.com/International-Data-Spaces-Association/IDS-G/blob/master/glossary/README.md#identity-provider.

[8] I. D. S. Association, «Dynamic Attribute Token,» [En línea]. Available: https://github.com/International-Data-Spaces-Association/IDS-G/blob/master/core/DAPS.

[9] Wikipedia, «X.509 certificates,» [En línea]. Available: https://en.wikipedia.org/wiki/X.509.

[10] I. D. S. Association, «Glossary "Dynamic Attribute Provisioning Service",» [En línea]. Available: https://github.com/International-Data-Spaces-Association/IDS-G/blob/master/glossary/README.md#dynamic-attribute-provisioning-service.

[11] I. D. S. Association, «Meta Data Broker,» [En línea]. Available: https://github.com/International-Data-Spaces-Association/IDS-G/tree/master/core/MetaDataBroker.

[12] I. D. S. Association, «"Specification IDS Meta Data Broker" White Paper,» [En línea]. Available: https://internationaldataspaces.org/wp-content/uploads/IDSA-White-Paper-Specification-IDS-Meta-Data-Broker.pdf.

[13] I. D. S. Association, «Meta Data Broker Open Core,» [En línea]. Available: https://github.com/International-Data-Spaces-Association/metadata-broker-open-core.

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

[14] ENGINEERING, «ALIDA - Micro-service based platform for composition, deployment, optimisation, execution and monitoring of big data analytics workflows,» [En línea]. Available: https://home.alidalab.it/.

[15] International Organization for Standardization, «ISO/IEC/IEEE 42010:2011 — Systems and software engineering — Architecture description,» 2011-12-01.

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