AI, Big Data, IoT Enablers and FIWARE compliant interoperable interfaces for grid services

D5.4

Authors:

Ferdinando Bosco (ENG)

Denisa Ziu (ENG)

Sonja Happ (RWTH)

Malileh Haghgoo (RWTH)

Joao Campos (ENERC)

Tiago Alves (ENERC)

Arsim Bytyqi (ENTSO-E)

Responsible Partner	ENG
Checked by WP leader	**Ferdinando Bosco (ENG)** - Date: 27/01/2022
Verified by the appointed Reviewers	**Apostolos Kapetainos (ED)** - Date: 27/01/2022 **Konstantinos Kotsalos (ED)** - Date: 27/01/2022 **Antonello Monti (FhG)** - Date: 27/01/2022
Approved by Project Coordinator	**Antonello Monti (FhG)** - Date: 28/01/2022

Dissemination Level		
PU	Public	X
CO	Confidential, only for members of the consortium (including the Commission Services)	
CI	Classified, as referred to in Commission Decision 2001/844/EC	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

Issue Record

Planned delivery date	31/01/2022
Actual date of delivery	28/01/2022
Status and version	V1.0

Version	Date	Author(s)	Notes
0.1	14/10/2021	Ferdinando Bosco (ENG)	Table of Contents
0.2	11/11/2021	Ferdinando Bosco (ENG)	Structure of the deliverable consolidated and first draft
0.3	30/11/2021	Denisa Ziu (ENG) Sonja Happ (RWTH) Malileh Haghgoo (RWTH) Joao Campos (ENERC) Tiago Alves (ENERC)	First draft of Ch.2.3, 3.3, 4.1
0.4	18/01/2022	Ferdinando Bosco (ENG)	Consolidated version ready for internal review
0.5	27/01/2022	Apostolos Kapetainos (ED) Konstantinos Kotsalos (ED) Antonello Monti (FhG) Ferdinando Bosco (ENG)	Version reviewed. Comments and feedback from reviewers accepted and integrated. Final version ready for quality check.
1.0	28/01/2022	Stephan Gross (FhG) Ferdinando Bosco (ENG)	Quality Check concluded. Deliverable final version.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

About OneNet

OneNet will provide a seamless integration of all the actors in the electricity network across Europe to create the conditions for a synergistic operation that optimizes the overall energy system while creating an open and fair market structure.

The project OneNet (One Network for Europe) is funded through the EU's eighth Framework Programme Horizon 2020. It is titled "TSO – DSO Consumer: Large-scale demonstrations of innovative grid services through demand response, storage and small-scale (RES) generation" and responds to the call "Building a low-carbon, climate resilient future (LC)".

While the electrical grid is moving from being a fully centralized to a highly decentralized system, grid operators have to adapt to this changing environment and adjust their current business model to accommodate faster reactions and adaptive flexibility. This is an unprecedented challenge requiring an unprecedented solution. For this reason, the two major associations of grid operators in Europe, ENTSO-E and EDSO, have activated their members to put together a unique consortium.

OneNet will see the participation of a consortium of over 70 partners. Key partners in the consortium include: already mentioned ENTSO-E and EDSO, Elering, E-REDES, RWTH Aachen University, University of Comillas, VITO, European Dynamics, Ubitech, Engineering, and the EUI's Florence School of Regulation (Energy).

The key elements of the project are:

1. Definition of a common market design for Europe: this means standardized products and key parameters for grid services which aim at the coordination of all actors, from grid operators to customers;
2. Definition of a Common IT Architecture and Common IT Interfaces: this means not trying to create a single IT platform for all the products but enabling an open architecture of interactions among several platforms so that anybody can join any market across Europe; and
3. Large-scale demonstrators to implement and showcase the scalable solutions developed throughout the project. These demonstrators are organized in four clusters coming to include countries in every region of Europe and testing innovative use cases never validated before.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

Table of Contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

List of Abbreviations and Acronyms

Acronym	Meaning
AI	Artificial Intelligence
API	Application Programming Interface
CDPSM	Common distribution power system model
CEF	Connecting Europe Facility
CGMES	Common Grid Model Exchange Standard
DoA	Description of Action
DSO	Distribution System Operator
ENTSO-E	European association for the cooperation of transmission system operators (TSOs) for electricity
ESMP	European style market profile
ETSI	European Telecommunications Standards Institute
GUI	Graphical User Interface
HLA	High Level Architecture
IDS	Industrial Data Space
IoT	Internet of Things
IUDX	Indian Urban Data Exchange
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation – Linked Data
NGSI	Next Generation Service Interface
NGSI-LD	Next Generation Service Interface – Linked Data
REST	Representational state transfer
SLA	Service Level Agreement
TSO	Transmission System Operator
UI	User Interface
W3C	World Wide Web Consortium
WP	Work Package

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

Executive Summary

The OneNet Solution and Architecture have a big focus on the implementation of standardised interfaces and standardised data models adoption for enabling a seamless integration and cooperation among for cross-platform market and network operation services.

Analysing the Requirements, the OneNet Reference Architecture and the Functional specification of the Cross-Platform services is evident as the FIWARE adoption, together with IDS reference model, assumes fundamental importance for ensuring a standardised, secured, and trusted data exchange, maintaining the data control and ownership.

Another aspect of fundamental importance is that the adoption of innovative technologies related to Big Data, AI and IoT in the context of Smart Grid Management, are increasingly current and more used.

This led us to analyse, first of all, how these technologies impact on Smart Grid Management and on cross-platform services related to network and market operations, and above all how these services can be easily made available and integrated into the OneNet solution.

As already described, OneNet offers a decentralized solution for the integration of the platforms, tools and services that require secure and trusted data exchange, but also offers a series of tools and components for easy data integration with the services offered by Service Providers.

OneNet Orchestration Workbench and OneNet Monitoring and Analytics Dashboard offer the possibility not only to integrate data sources through the OneNet Decentralised Middleware and to use a series of tools for monitoring and data analytics, but also to integrate external services that can be orchestrated and evaluated within the OneNet Orchestration Workbench.

This document is therefore focused on the technical specifications of these components and on how the adoption of standardized data models and FIWARE can facilitate the decentralized and interoperable approach in in the implementation of the OneNet Solution.

The results described here is therefore the conclusion of a path of analysis, design, and identification of technical and functional specifications, started at the beginning of the project and that at this point, produced a consolidated output.

The OneNet Decentralized Middleware through the adoption and integration of the OneNet Connector, will be implemented following the IDS Reference Model and including the FIWARE Context Broker, in the updated version that follows the NGSI-LD standard, enabling and facilitating a seamless cross-platform integration and cooperation. In fact, the OneNet Connector will be made available for each OneNet Participants and will be deployable and pluggable in the already existing platforms in a standardised way.

In addition, the OneNet Orchestration Workbench and OneNet Monitoring and Analytics Dashboard will be integrated with the OneNet Decentralised Middleware, empowering the adoption of the OneNet ecosystem, and providing additional features to the OneNet Participants.

1 Introduction

This chapter describes the context in which the activities of WP5, and more specifically the T5.4, activities are placed and how they are coordinated and linked within the other project activities. In addition, a detailed description of the structure and objectives of this document is provided.

1.1 Scope

OneNet will develop an open and flexible architecture to transform the actual European electricity system, which is often managed in a fragmented country- or area-level way, into a pan-European smarter and more efficient one, while maximizing the consumer capabilities to participate in an open market structure. According to OneNet Description of Action (DoA), WP5 contributes to the direction of fulfilling the OneNet envision by striving to attain two objectives; First, to design an open conceptual architecture for effective yet seamless operation of a smarter pan-European electricity system where market and network technical operations are coordinated closer to real-time across countries, and second to provide requirements, functional and technical specifications, together with interoperable and standardisable interfaces for an open scalable decentralized interconnection of platforms, technology agnostic adaptable and flexible IT reference architecture which fully support the OneNet concept and provides the necessary backbone for the WP6 subsequent implementation of the OneNet data sovereignty-preserving working space.

The WP5, together with WP6, act as IT pillar of the overall OneNet project. The IT pillar it is closely linked to all the other pillars of the project, as shown in Figure 1. It takes into consideration all the results provided in the Market Pillar (WP2 and WP3) as well as the Operation Pillar (WP4). In addition, the OneNet Solution, implemented in WP6 will be tested and evaluated in 4 Demonstration Clusters and the results of the evaluation will be used for adapting, improving, and enhancing the OneNet Solution.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

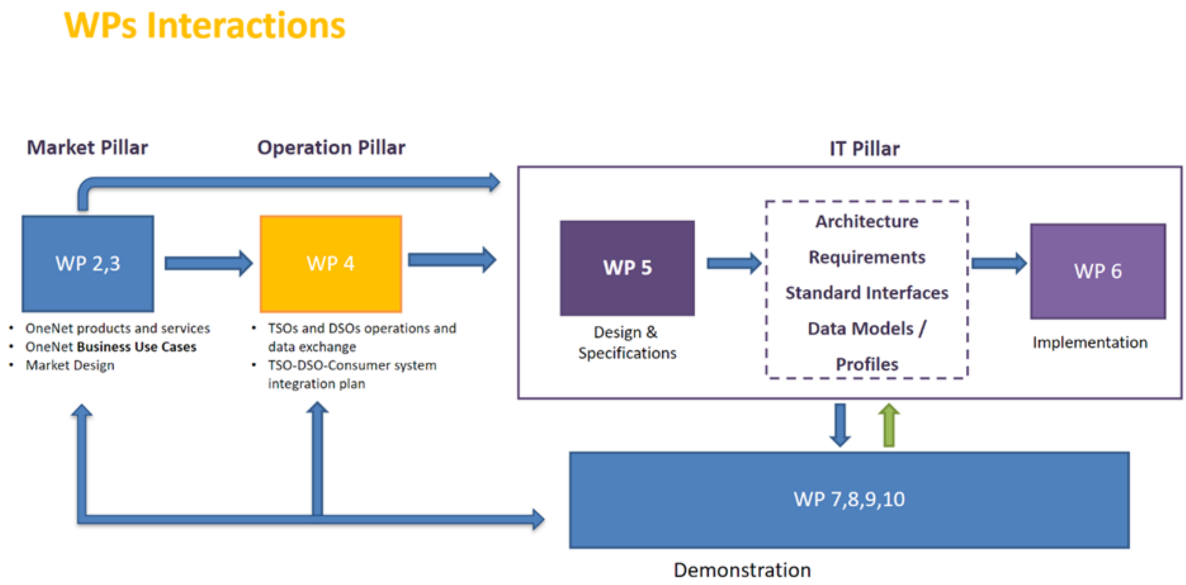

Figure 1: WP5 interactions

1.2 Task 5.4

Within the context just described in Section 1.1, the main goal of the Task T5.4 is dual:

- define open interoperable interfaces for the **OneNet Orchestration Workbench and OneNet Monitoring and Analytics Dashboard**, supporting the integration and evaluation of Big Data, AI models and IoT for enabling cross-country, cross-domain and cross-platform services for network and market operations;
- provide a standardised data gates relying on **standardised data models** and FIWARE adoption and more in detail the evolution of the **FIWARE Context Broker with the ETSI NGSI-LD standard APIs**

The architecture will have to consider the specifications of existing projects, architectures and initiatives, to facilitate its adoption and dissemination.

Task 5.4, started at M4 of the OneNet projects, collects all the necessary information provided from other WP5 tasks and in particular: the requirements and use cases (from T5.1), the OneNet Reference Architecture (T5.2) and the Cross-Platform Services and Functional Specification (T5.3). All these inputs are the base for the definition of the OneNet Orchestration Workbench and OneNet Monitoring and Analytics Dashboard as well as the integration and evolution of the FIWARE Context Broker.

Finally, the T5.4 foresees two important results:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

- the release of the report on the **AI, Big Data, IoT Enablers and FIWARE compliant interoperable interfaces for grid services** in the Deliverable D5.4 (this document) in January 2022, M16 of the project
- the release of the **final report on the AI, Big Data, IoT Enablers and FIWARE compliant interoperable interfaces for grid services**, in the Milestone MS10 in September 2022, M24 of the project

1.3 Outline of the deliverable

This deliverable is structured in 6 different chapters.

Chapter 1 is the introductory one where the deliverable is contextualized within the WP5 and task T5.4.

Chapter 2 analyses the state of the art on the topics related to Big Data, AI, IoT and Analytics for Smart Grid Management.

Chapter 3 reports the results of the previous tasks of the WP5 (T5.1, T5.2 and T5.3) useful for the analysis and the work conducted in T5.4

Chapter 4 describes the FIWARE initiative and in particular the FIWARE Context Broker and its evolution following NGSI-LD standard as well the Standardised Data Models to be adopted in the OneNet System.

Chapter 5 includes the technical description and specification of the OneNet Orchestration Workbench and the OneNet Monitoring and Analytics Dashboard as well as the standardised interfaces for the integration in the overall OneNet System.

Finally, chapter 6 concludes the document.

To facilitate the readability of D5.2, it might be useful to refer to D5.1, D5.2 and D5.3 as well.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

2 AI, Big Data and IoT – Technological enablers for Smart Grid Management

2.1 State of the Art of the AI, Big Data and IoT technologies for smart grid management

Smart grids have been gradually replacing the traditional power grids during the last decade. Such transformation is linked to adding many of smart devices and other data sources which produce a large amount of information. This provides various opportunities associated with the collection and the exploitation of big data [1].

Big Data take the role of game changer for the implementation of smart energy services, such as intelligent energy management, energy consumption prediction and exploitation of Internet of Things (IoT) solutions.

In addition, the ever-increasing use of algorithms based on Artificial Intelligence (AI) and Machine Learning has created a real boost for Big Data Analytics.

2.1.1 Big Data Concept and Characteristics

The definition of big data is not very clear and uniform at present. But there is a consensus among different descriptions: this is an emerging technical problem brought by a dataset of large volume, various categories and complicated structures which needs novel framework and techniques to excavate useful information effectively [2].

The characteristics of big data are universally identified in the 5 V big data model [3]:

1. Volume, it refers to the size of Big Data. Data can be considered Big Data or not is based on the volume. The rapidly increasing volume data is due to cloud-computing traffic, IoT, mobile traffic etc.
2. Variety, it refers to Structured, Semi-structured and Unstructured data due to different sources of data generated either by humans or by machines. Structured data: It's the traditional data which is organized and conforms to the formal structure of data. This data can be stored in a relational database. Semi-structured data: It's semi-organized data. It doesn't conform to the formal structure of data. Unstructured data: It's not an organized data and doesn't fit into rows and columns structure of a relational database. Example: Text files, Emails, images, videos, voicemails, audio files etc.
3. Velocity, it refers to the speed at which the data is getting accumulated. This is mainly due to IoTs, mobile data, social media etc.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

4. Veracity, it refers to the assurance of quality/integrity/credibility/accuracy of the data. Since the data is collected from multiple sources, we need to check the data for accuracy before using it for business insights.
5. Value, it refers to how useful the data is in decision making. We need to extract the value of the Big Data using proper analytics.

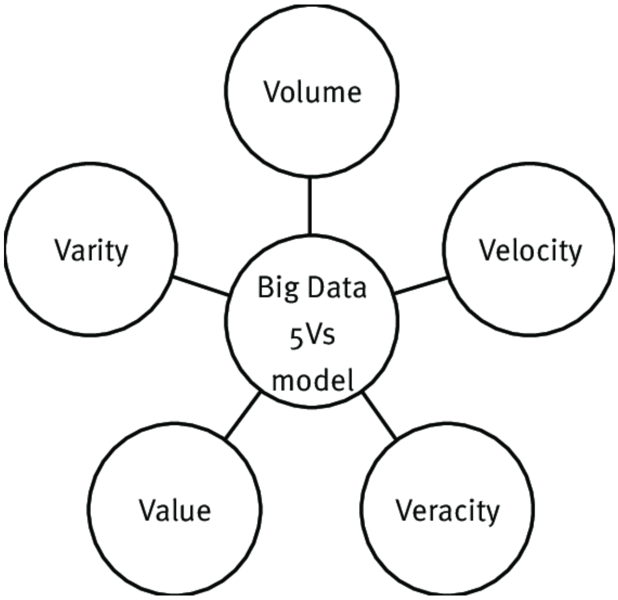

Figure 2: 5V's Big Data Model

2.1.2 Big Data in the Smart Grid

The Big data characteristics can be also represented in smart grid; in particular it becomes necessary to process large volume and varieties of both real-time and historical data to extract meaningful information in order to make data-driven decisions.

In fact, data from the smart grids are generated in real-time at a very high rate and volume [4]. The extraction of information from smart grid data is required for grid applications and calls for deep insight into the data sources. The data in smart grids can be classified into consumer, distribution, transmission, and generation data. These data are acquired from the sensors, smart meters, grid devices, detectors, Supervisory Control and Data Acquisition (SCADA), etc. The collected signals relate to power utilization habits of consumers, phasor measurement, energy consumption, energy pricing and bidding, operation, or financials for running the utility, etc.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

Also, in term of variety the Smart Grid data sources present a big amount of semi-structured, quasi-structured, and unstructured data in addition to structured data.

Table 1: Data Variety in Smart Grid Management

Type	Dataset
Structured Data	Meters' data, distribution management data, equipment parameters, load control data, marketing system data in relational format
Semi-Structured Data	Web service data, load monitoring, power quality data
Quasi-Structured Data	Web click-stream that contains erratic formats and data values, web scrapping data, search engine results
Unstructured Data	Meteorological information, customer service data, economy data of distribution regions

An example of smart grid data collected and processed is shown in Figure 3.

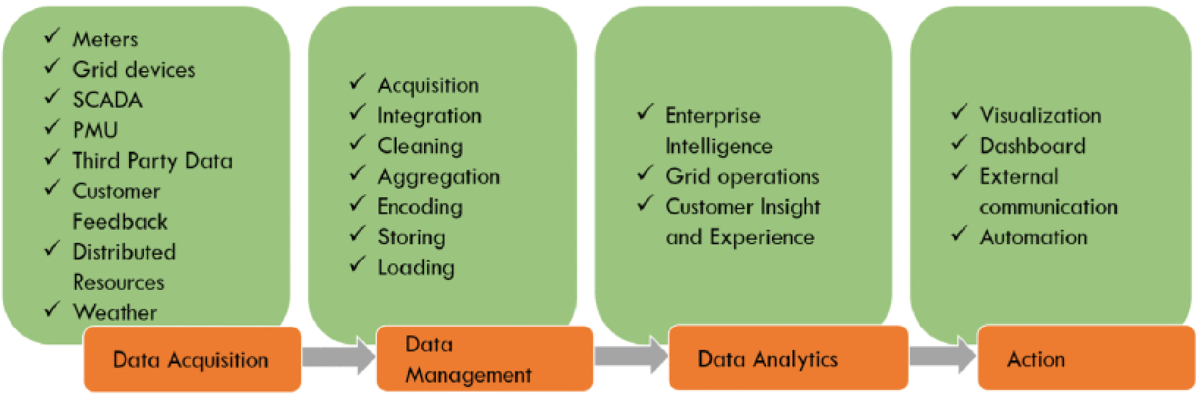

Figure 3: Data Process in Smart Grid Management

2.1.3 Big data architecture and platforms

Currently, there are no standard big data analytics architectures developed for power grid applications. Therefore, a clear understanding of big data architecture is required to identify how big data integrates with the existing power system control and operational architecture, what are the essential characteristics of big data environment, how they differ from traditional computational environments, and what scientific, technological and standardisation challenges are needed to deploy big data solutions [5].

For further information, an analysis of the most relevant open-source big data platforms is presented in the Appendix A – Big Data Architecture and Platforms.

2.2 Big Data Analytics

The most relevant application of the Big Data technologies for the Smart Grid Management is undoubtedly that of Big Data Analytics, including predictive and adaptive services based on Artificial Intelligence and Machine Learning approaches.

Benefits of Big Data Analytics

- Collect and analyse data to improve service quality
- Analysis of unstructured data
- Analytics comes in different formats
- Choose between on-premise and managed service
- Trends in the utility industry
- Trust is growing for smart grid data analytics
- Sensors are replacing MDM
- Integrating data is a core function
- Collecting the right data in the right place

2.2.1 Big Data Analytics Types

There are four major types of big data analytics [6]. These are described as follows:

Descriptive Analytics

Descriptive analytics illustrates what happened in the past using the historical data available and shows the data in an easily understandable form or visualization. In general, the data is illustrated using graphs, bar diagrams, pie diagrams, maps, scatter plots, etc. In short, descriptive analytics is performed to understand or illustrate the patterns in the data.

Diagnostic Analytics

It is a form of advanced analytics that examines data or content to answer the question, "Why did it happen?". It is characterized by techniques such as drill-down, data discovery, data mining, and correlations.

Predictive Analytics

It extrapolates from the data available to predict what can happen in the future. The tools that are used for predictive analytics are time-series analysis using statistical methods and other data mining algorithms. Predictive analytics is usually performed to predict which events can happen in the future.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

Prescriptive Analytics

It is used to discover the best outcome of past events when the features of the data and operating parameters of a system are given. It helps to develop strategies for future events under similar conditions. The techniques involve simulation tools, and these simulate the operating conditions or features to finally come up with the best outcome. The simulation techniques strategize how to plan for similar events in the future. Prescriptive analytics is basically performed to know how preferable events can be made to happen in the future. Example: power flow analysis, etc.

The Figure 4 compares the amount of added value to an organization versus the complexity required for a specific analytics type implementation.

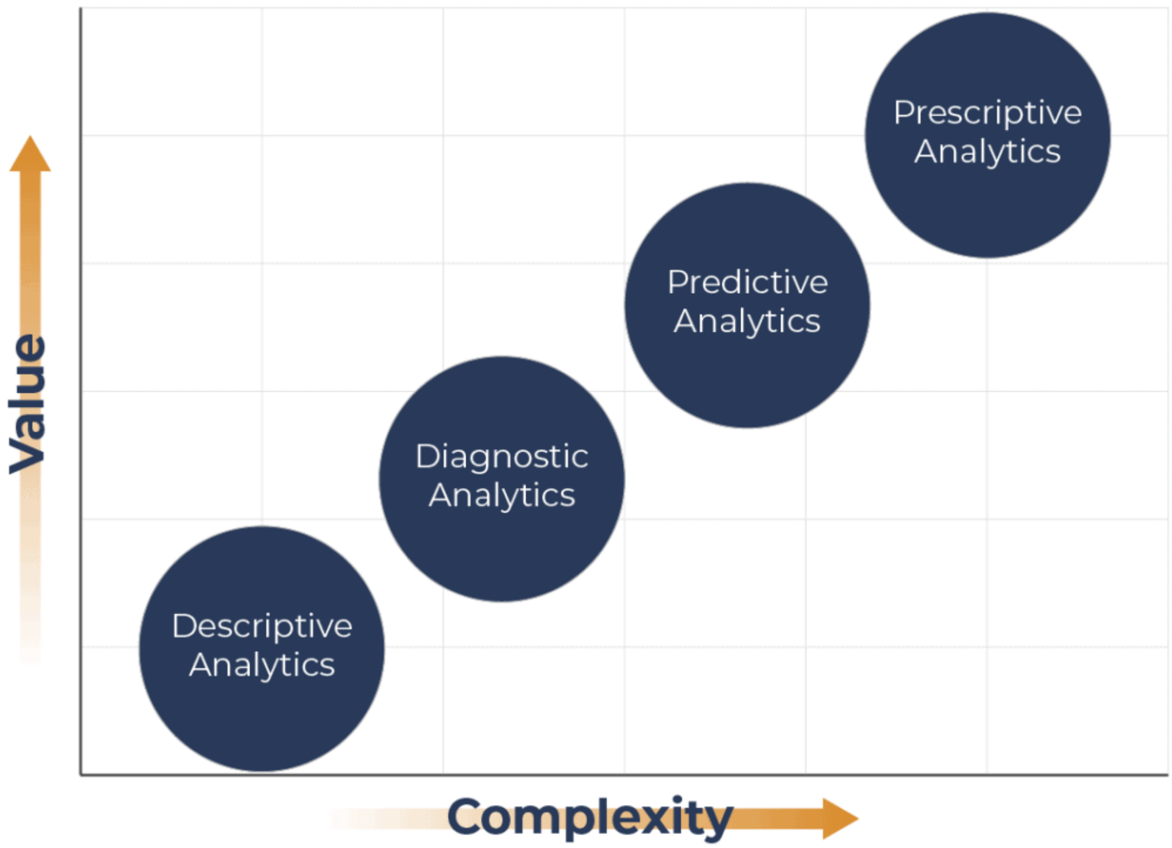

Figure 4: Big Data Analytics Type - Added-value/Complexity comparison [6]

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

2.2.2 Big Data Analytics Process

Big Data analytics requires pre-defined strategies because of the high volume of data. Also, as described in Chapter 2.1.1, the velocity and variety of data pose challenges in the data analytics process. It is very crucial that the data from the smart grid are processed in real-time because significant patterns can be recognized from the data to make better decisions.

The Big Data Analytics process typically requires three fundamental steps:

- Data Acquisition
- Data Processing
- Data Analytics Techniques Application

Data Acquisition

The first step in any of the data analytics process is the collection of data. The data in the smart grid are collected from various sources as mentioned in the earlier section. With the data collection already in place, the other sub-tasks in data acquisition are data communication and data pre-processing. The raw data need to be transmitted either to a real-time stream processing system or to a storage system from where the data can be sent to the offline batch processing system for further analysis. Since the data have been collected from diverse and multiple sources, the data aggregation and cleaning are the foremost and crucial steps. Data aggregation services should be in place to integrate the data from varied sources and furnish a unified view of the available data. In data pre-processing, the inconsistent and missing data are to be filled or one among the records, and the features are to be removed to improve the data quality. It is crucial to refine the features in the extracted data as there are noise and redundancy in the collected raw data. Refining the features involve either feature selection or feature extraction. If the data contain highly correlated features, then the machine learning algorithms, in general, perform poorly.

Data Processing

The data collected and transmitted should be stored in storage infrastructure for further processing. The stage, at which the data is processed, classifies data processing into the following types:

- Batch Processing
- Stream Processing
- Iterative Processing

Batch Processing

Batch analytics is fundamentally the analysis of data in batches. It involves the workflow on offline data where all the data are available, pre-extracted, and ingested using scripts and a huge group of data is analysed in a single execution. An example of batch processing in smart grids includes the training of data-driven models using offline data for applications of topology identification, predictive maintenance, energy forecasting, etc. These models would require re-training if new data became available and need to be included in the modelling performance. There is no specific time interval defined to term processing as batch analytics. However, it is usually considered that if the processing is scheduled to happen with an interval equal to or greater than 20 minutes, then it is batch processing.

Stream Processing

Stream processing is primarily the processing of each new data instance as soon as it is available instead of waiting for batches of offline data. The idea behind the stream processing is that the potential worth of information from data relies on the freshness of data. Hence, it is crucial that the stream processing model processes the data as soon as the data instance is available to obtain approximation results. If the data are continuously available in huge streams, a portion of the data can be stored in memory until it is processed.

Iterative Processing

There are a few big data problems that require the processing of data iteratively and demand a greater number of read and write operations than batch processing and stream processing. These involve a high number of Input-Output transfers and are time-consuming. In the context of the smart grid management, only batch and stream processing are relevant.

Data Analytics Techniques Application

Multiple machine learning algorithms are used as data analytics techniques. These techniques are used to map the relationship between the features in the data and the prediction label usually. If the labels exist, the techniques employed are named as supervised techniques. Whereas the data may not explicitly consist of labels and it is up to the algorithm to recognize the patterns in the data. These techniques that work on data without labels are termed as unsupervised techniques.

2.2.2.1 Cross Industry Standard Process for Data Mining (CRISP-DM)

Another standard methodology for the Big Data Analytics process and more generally for data mining is that of CRISP-DM, acronym for Cross-industry standard process for data mining which describes and codes the common approaches used by data mining experts. This methodology is one of the most used on an industrial level [7].

The CRISP-DM divides the data mining operation into 6 essential parts:

- Business Understanding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

• Data Understanding

• Data Preparation

• Modeling

• Evaluation

• Deployment

Unlike other conventional, typically linear and unidirectional processes, the phases that characterize CRISP-DM are designed to be repeated cyclically.

The Figure 5 symbolizes the cyclic nature of the CRISP-DM methodology.

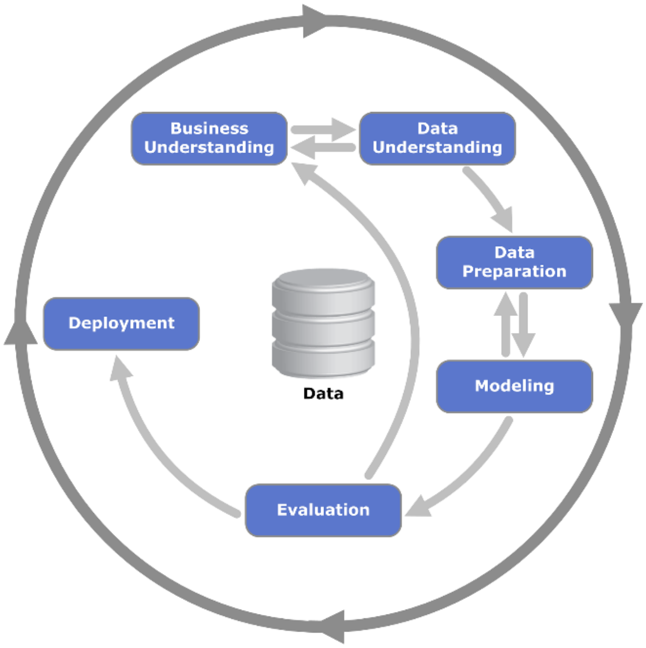

Figure 5: CRISP-DM Iterative Cycle [8]

2.2.3 Application Areas of Big Data Analytics in Smart Grid Management

There are many potential application areas which would leverage on the big data analytics in the smart grid.

- Fault Classification and Identification
- Preventive Maintenance
- Power Quality Monitoring

- Topology Identification
- Renewable Energy Forecasting

2.3 BDVA – Big Data Value Association - Reference Model

The Big Data Value Association (BDVA) [9] is an industry-driven international not–for-profit organisation which aims at creating a functional Data Market and Data Economy in Europe, to allow Europe to play a leading role in Big Data in the global market.

As already analysed in D5.2, the BDVA's vision is very aligned with the OneNet's one, including the necessity to develop and adopt a pan-European data sharing space. This is the primary driver, together with the opportunities presented by Artificial Intelligence (AI), for the evolution of the data economy.

Also, the main goals of the BDVA are aligned with OneNet perspective and in particular:

- To create the conditions for the development of a trusted European data sharing framework.
- To incorporate data sharing at the core of the data lifecycle for greater access to data.
- To provide supportive measures for European businesses to safely embrace new technologies practices and policies.
- To assemble a European-wide digital skills strategy to equip the workforce for the new data economy.

In the context of this document, is also important to report that the BDVA defined the BDV Reference Model [10] as common reference framework to locate Big Data technologies on the overall IT stack.

It addresses the main concerns and aspects to be considered for Big Data Value systems. The BDV Reference Model distinguishes between two different elements. On the one hand, it describes the elements that are at the core of the BDVA; on the other, it outlines the features that are developed in strong collaboration with related European activities.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

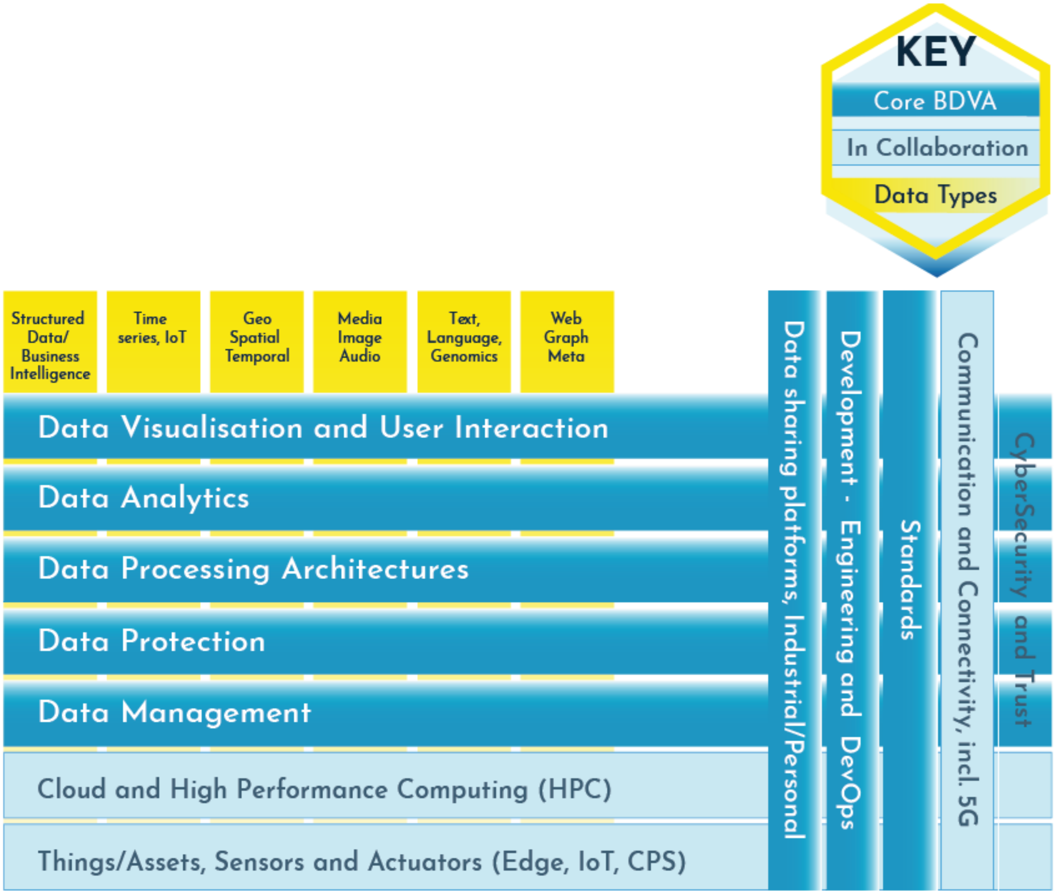

Figure 6: BDV Reference Model [10]

The BDV Reference Model is structured into horizontal and vertical concerns.

Horizontal concerns cover specific aspects along the data processing chain, starting with data collection and ingestion, and extending to data visualization. It should be noted that the horizontal concerns do not imply a layered architecture. As an example, data visualization may be applied directly to collected data (the data management aspect) without the need for data processing and analytics.

Vertical concerns address cross-cutting issues, which may affect all the horizontal concerns. In addition, vertical concerns may also involve non-technical aspects.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

2.4 High Level Architecture – AIOTI

The following proposal is a high-level IoT architecture referred to as the AIOTI High-level architecture (HLA). This proposal is the 5.0 release [11] and is the result of the work made within AIOTI WG03 and considered the works of SDOs, Consortia, and Alliances in the IoT space. The architecture was also made with the need to support instantiation in large scale pilot deployments in mind.

In summary, this proposal:

- Presents a Domain Model and discusses the “thing” in IoT
- Presents a Functional Model (Figure 7)
- 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

By recommendation of the AIOTI WG03 this architecture is partly described using the ISO/IEC/IEEE 42010 standard. This standard describes architectures in terms of their multiple viewpoints and models and further specifies minimal requirements for architecture descriptions, frameworks, description languages and viewpoints. In the following figure X, the architectural models that were described using ISO/IEC/IEEE 42010 can be seen.

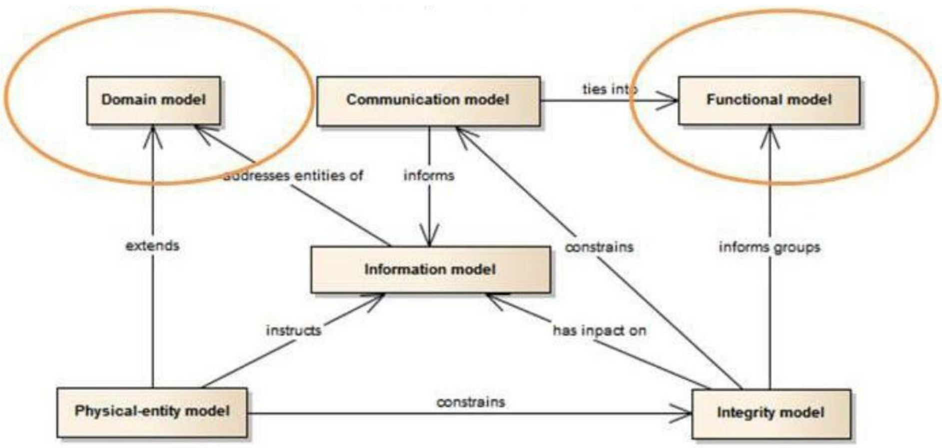

Figure 7: Architectural Models based on ISO/IEC/IEEE 42010

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 957739

The HLA focuses its recommendations on the Domain and Functional models (other models can be considered for future releases of the HLA):

1. The Domain Model describes entities in the IoT domain and the relationships between them.
2. The Functional Model describes functions and interfaces (interactions) within the IoT domain.

2.4.1 AIOTI HLA Domain Model

The main concepts and relationships of the architecture are described in the domain model (Figure 8). By naming and identifying these concepts and relationships a common lexicon is provided which is fundamental for this and other models.

In this model, a **User** (human or otherwise) interacts with a physical entity, a **Thing**. The interaction is mediated by an **IoT Service** which is associated with a **Virtual Entity**, a digital representation of the physical entity. The **IoT Service** then interacts with the **Thing** via an **IoT Device** which exposes the capabilities of the actual physical entity.

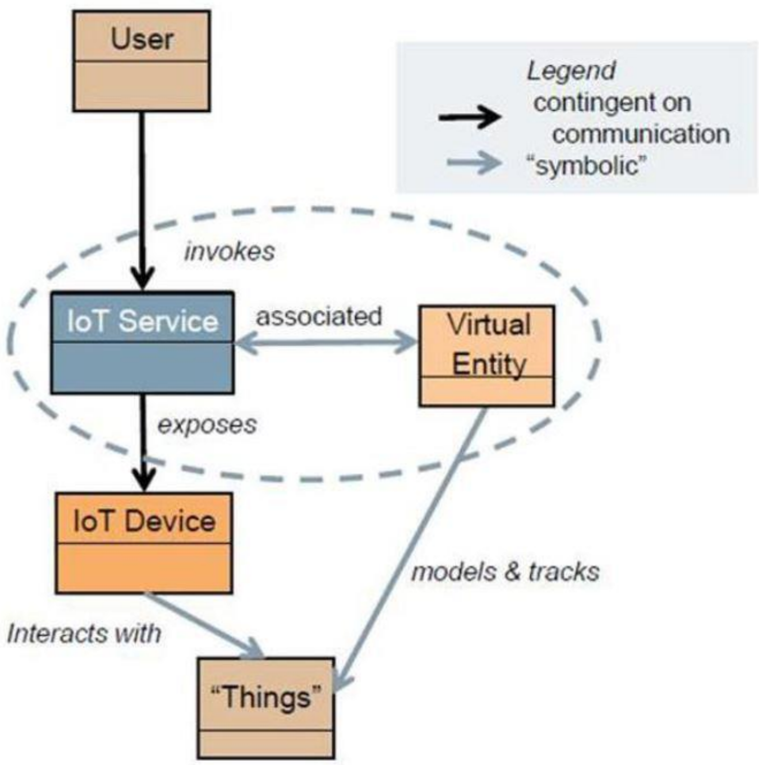

Figure 8: Domain Model

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

2.4.2 AIOTI HLA Functional Model

The AIOTI Functional Model describes interactions, functions, and interfaces within the domain but interactions outside of the domain are not excluded. The functional model is composed of three layers. Both the functional model and the three layers that compose it can be seen in Figure 9.

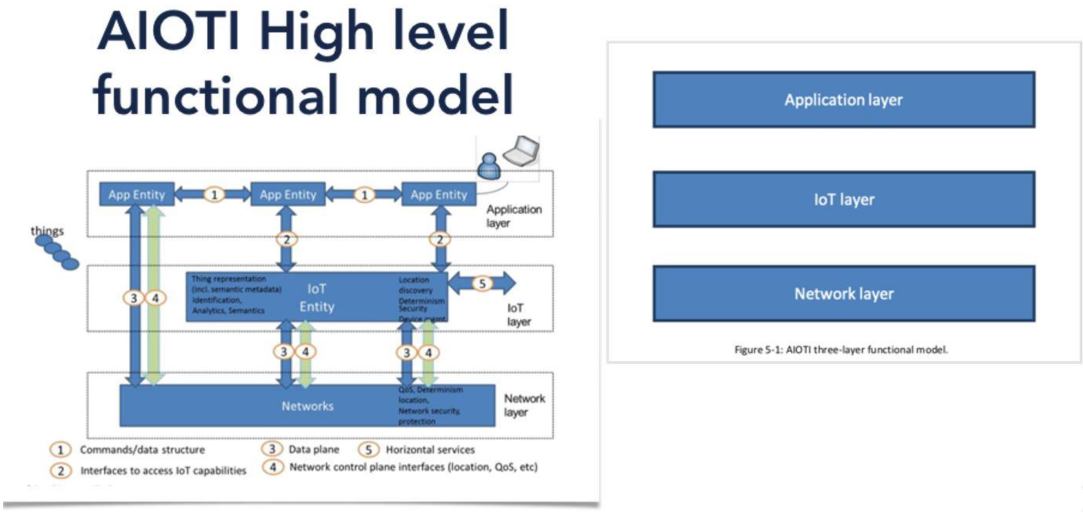

Figure 9: AIOTI High-level Functional Model

- The Application layer: contains the communications and interface methods used in process- to-process communications
- The IoT layer: groups IoT specific functions, such as data storage and sharing, and exposes those to the application layer via interfaces commonly referred to as Application Programming Interfaces (APIs). The IoT layer makes use of the Network layer's services.
- The Network layer: the services of the Network layer can be grouped into data plane services, providing short- and long-range connectivity and data forwarding between entities, and control plane services such as location, device triggering, QoS or determinism.

Functions do not mandate any specific implementation or deployment; therefore, it should not be assumed that a function must correspond to a physical entity in an operational deployment. Grouping of multiple functions in a physical equipment remains possible in the instantiations of the functional model.

2.4.3 HLA deployment considerations

In order to deploy the HLA the following technologies and concepts are identified as the major consideration to take into account:

- Cloud and Edge Computing: 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.

2.4.4 HLA 3D IoT Layered Architecture

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 10 depicts the three main views in the 3D Model (Layers, Cross-cutting functions, and Properties)

The layers' view in the 3D model, see Figure 10, 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.

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.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

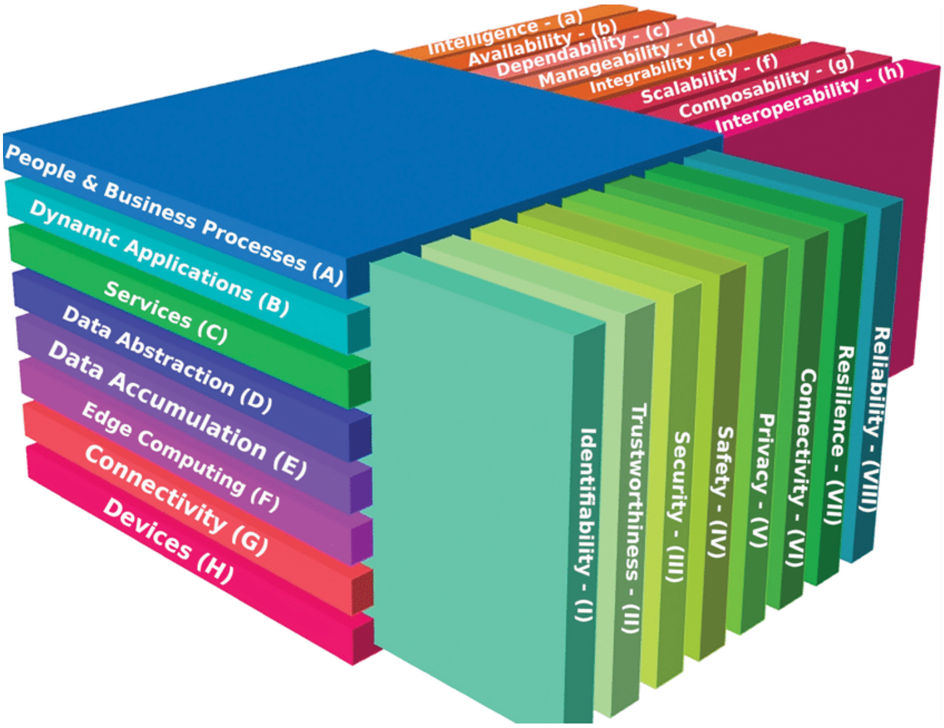

Figure 10: AIOTI 3D HLA [12]

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

3 Technology Alignment with OneNet Concept

3.1 Analysis of Use Cases and Requirements

OneNet Deliverable 5.1 [13] provides the overall concept of the OneNet System based on the collection of the Demo System Use Cases and General OneNet System Use Cases.

Starting from these SUCs, the D5.1 also provides the list of Functional and Non-Functional Requirements for the overall OneNet System. In this chapter, we analysed all the information related to the FIWARE Context Broker integration and the OneNet Orchestration Workbench and OneNet Analytics and Monitoring Dashboard implementation.

Starting from the analysis of the System Use Cases, it is evident that two of these are of fundamental importance for the purpose of this deliverable: "GSUC_02: AI, Big Data, IoT Data Orchestration for cross-platform services" and "GSUC_03: Integration of devices and other data sources to OneNet using FIWARE"

As described in the GSUC_02, two important objectives of the OneNet System and in particular of the OneNet Orchestration Workbench, are: Enable AI, Big Data and IoT data orchestration for cross-platform services; tracking the performance of the cross-platform services

The OneNet Orchestration Workbench allows to integrate data coming from the OneNet middleware and implement a data pipeline orchestration.

It also should include:

- Job Scheduling
- App/Service registry and discovery
- Error/Retries management
- SLAs tracking, alerting and notification

The GSUC_03 described the importance of the FIWARE Context Broker for the seamless integration of different data sources. The main objectives of this Use Case are: Connect different data sources using standardized FIWARE components; Adaption and evolution of the FIWARE context Broker for providing a data-model agnostic connector based on NSGI-LD.

This General OneNet SUC focuses on the OneNet System integration and adoption of the FIWARE Context Broker and FIWARE Smart Energy Architecture for enabling a standardized data exchange providing a context-based and data-model agnostic connector. This connector uses REST API named NGSI-LD, that introduces the Linked

Data (LD) concept in the already existing NGSI standard. The API operations allow applications to create entities, search the graph-based and subscribe to entities notifications.

The results of this analysis are also reflected in the Functional Requirements, and in which are reported in the Table 2.

Table 2: Functional Requirements relevant for FIWARE integration and OneNet Orchestration Workbench and Dashboard implementation

<table>
<tr><th>Requirement ID</th><th>Requirement Name</th><th>Description</th><th>Reference</th></tr>
<tr><td>OneNet_FUR_25</td><td>The OneNet system could use AI mechanism for empowering Data services</td><td>For improving the Data Services offered by the OneNet system, some AI mechanism could be implemented.</td><td>GSUC_02</td></tr>
<tr><td>OneNet_FUR_26</td><td>The OneNet system should be able to integrate any kind of data sources using Context Broker</td><td>The usage of the FIWARE context broker could facilitate the integration of any kind of data source, using a standard API based approach.</td><td>GSUC_03</td></tr>
<tr><td>OneNet_FUR_27</td><td>The OneNet Orchestration Workbench must be able to manage data and service orchestration</td><td>The OneNet Orchestration Workbench aims to support the data orchestration for the evaluation of the performance and scalability of the AI, IoT and Big Data cross-platform services for market and grid operations.</td><td rowspan="3">GSUC_02</td></tr>
<tr><td>OneNet_FUR_28</td><td>The OneNet Orchestration Workbench must be able to integrate data using the OneNet Middleware</td><td rowspan="2">The OneNet Orchestration Workbench allows to integrate data coming from the OneNet middleware and implement a data pipeline orchestration.
It also should include:
Job Scheduling
App/Service registry and discovery
Error/Retries management
SLAs tracking, alerting and notification</td></tr>
<tr><td>OneNet_FUR_29</td><td>The Service Provider must be able to register its service in the OneNet Orchestration Workbench</td></tr>
</table>

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

OneNet_FUR_30	The Service Provider must be able to create a data workflow using the Orchestration Workbench		
OneNet_FUR_31	The Service Provider must be able to evaluate the performance of its own service		
OneNet_FUR_32	The OneNet Orchestration Workbench should provide a service catalogue to the OneNet Participants		
OneNet_FUR_33	The OneNet system should offer a UI dashboard to OneNet Participants for monitoring and analytics	The OneNet system should implement a GUI for facilitating the OneNet Participants in the management, monitoring and analytics of the data transactions.	

3.2 Analysis of OneNet Architecture

One of the main outputs of the D5.2 [14] is a very detailed analysis of the most relevant initiatives, reference architectures and projects concerning to the OneNet main objectives and expected characteristics. The result of this analysis conducted to the identification of the IDS reference model and FIWARE interfaces, for implementing a hybrid solution using both the standard models for implementing the OneNet Decentralized middleware and the OneNet Connector. The usage of IDS Connector and FIWARE Context Broker was identified as the best solution to be adopted for ensuring a high level of standardization, interoperability, scalability, and reuse of OneNet solution.

The results of this analysis, together with the analysis of the OneNet main concept, Use Cases, and requirements, brought to the design and definition of OneNet Reference Architecture.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

The OneNet Reference Architecture (shown in the Figure 11) includes three core components relevant for this deliverable:

- the OneNet Decentralised Middleware and the OneNet Connector, which leverages in the FIWARE Context Broker and NGSI-LD API for the data exchange
- the OneNet Orchestration Workbench;
- the OneNet Monitoring and Analytics Dashboard

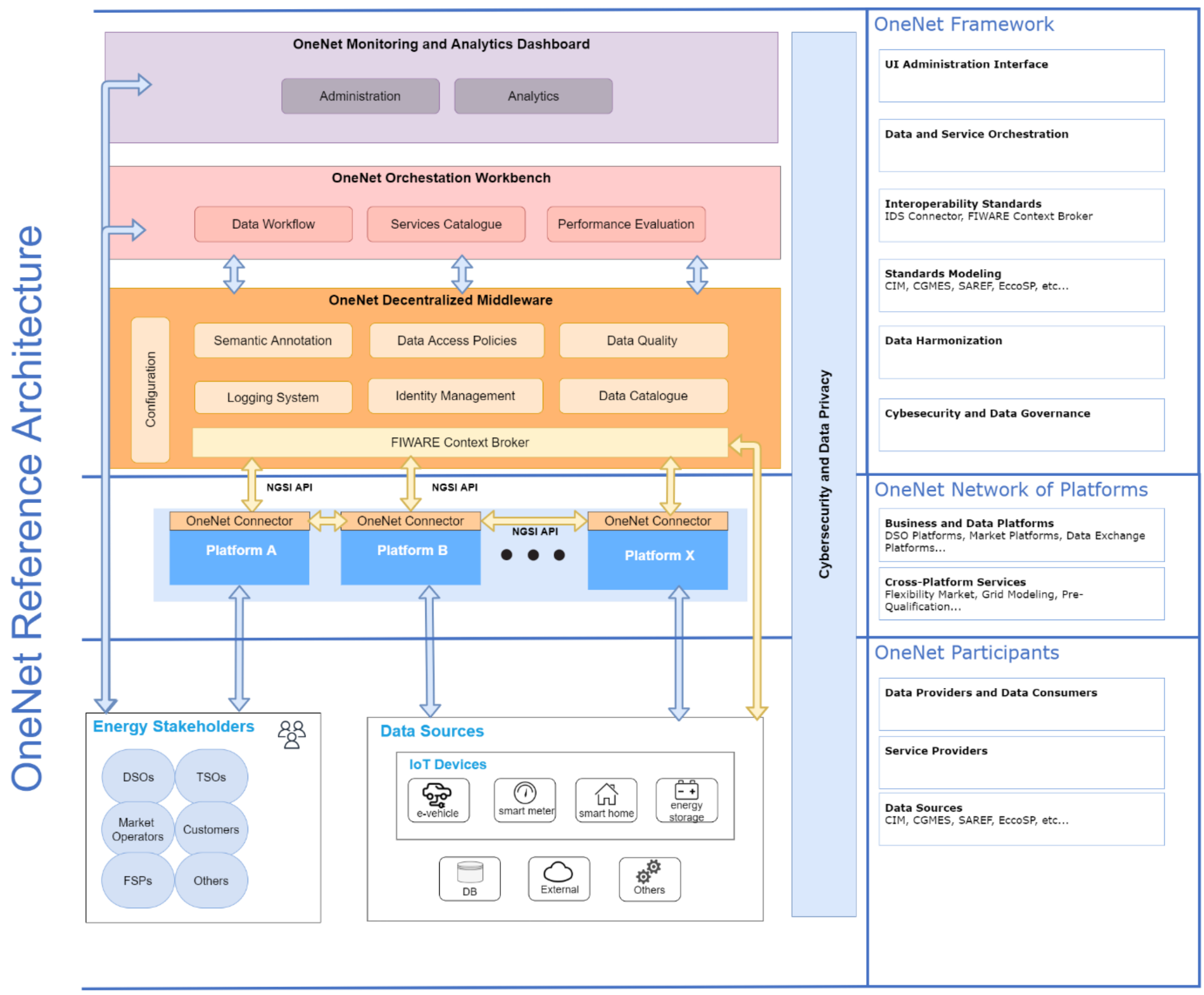

Figure 11: OneNet Refence Architecture [14]

OneNet Decentralized Middleware and OneNet Connector

As already described, the FIWARE ecosystem and the FIWARE Context Broker and the NGSI-LD API take on a fundamental importance in the implementation of the OneNet Decentralized Middleware and the OneNet Connector.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

The OneNet platform should make sure that data providers and data consumers can rely on the identity of the members of the data ecosystem between different security domains, leveraging in a complete decentralised approach and maintain a high level of interoperability and reusability.

The OneNet Connector, using both the design model of the IDS Reference and FIWARE ecosystem can ensure all these characteristics.

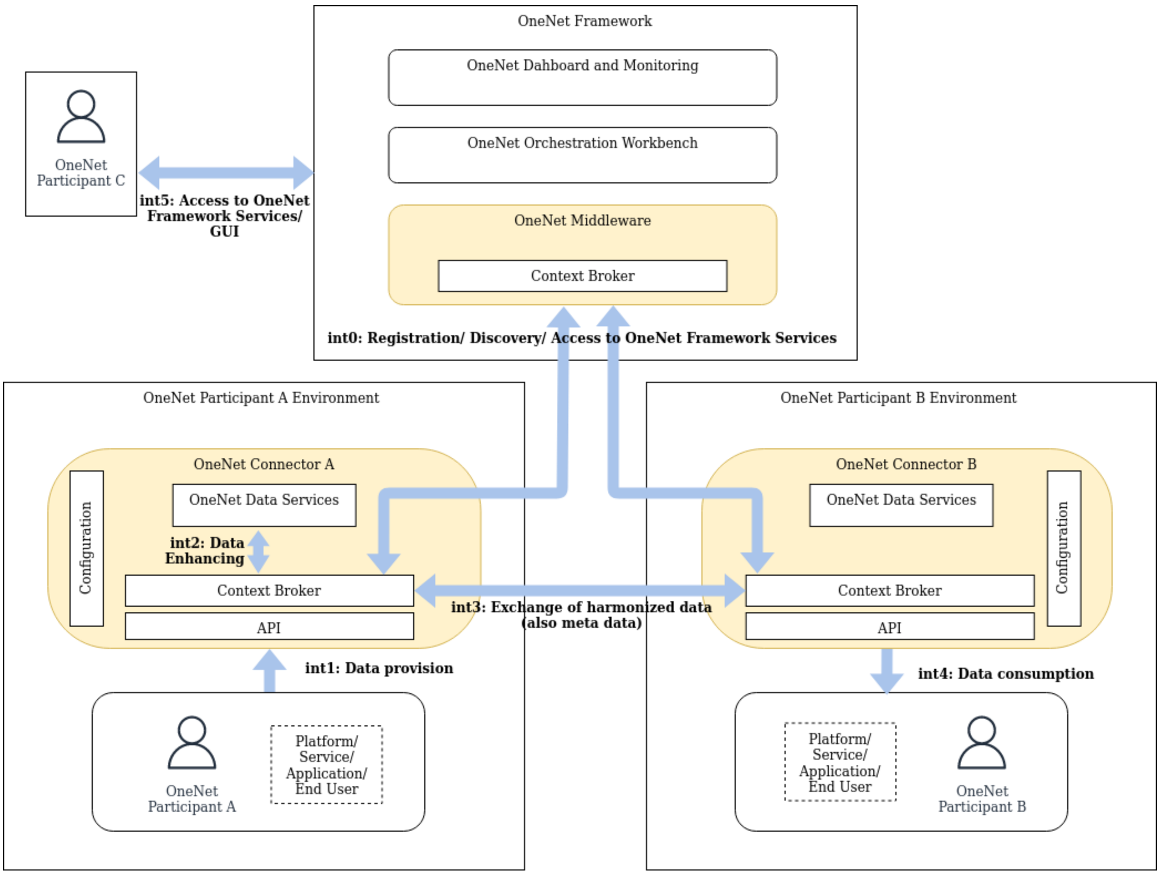

Figure 12: OneNet Decentralized Approach

OneNet Orchestration Workbench

OneNet Orchestration Workbench is described in the D5.2 such as the component able to orchestrate and evaluate the performance and scalability of the cross-platform services that will be integrated and implemented in the OneNet System.

Any OneNet Participant and in particular the Service Providers will be able to test and evaluate its own service exploiting the OneNet Orchestration Workbench, that allows to integrate data coming from the OneNet Middleware and to implement a data pipeline orchestration.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

From a functional perspective, following the functional requirements of D5.1 and listed on the Chapter 3.1, the OneNet Orchestration Workbench will support the integration and the evaluation of the performance and scalability of the AI, IoT and Big Data cross-platform services for market and grid operations.

The OneNet Orchestration Workbench will allow to integrate data coming from the OneNet middleware and will allows the interaction with the service providers which will be able to register and evaluate their own services, implementing services orchestration, data pipeline and evaluation schema.

The OneNet Orchestration Workbench should also include:

- Job Scheduling
- App/Service registry and discovery
- Error/Retries management
- SLAs tracking, alerting and notification

Specific interfaces of this component will be described in Chapter 5

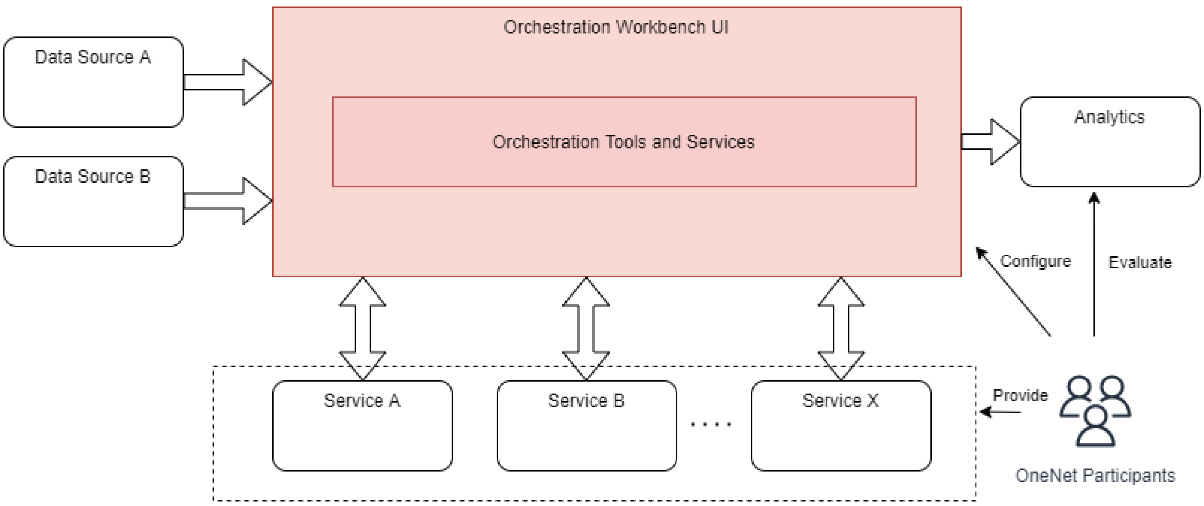

Figure 13: OneNet Orchestration Workbench Concept

OneNet Monitoring and Analytics Dashboard

The OneNet Monitoring and Analytics Dashboard is described such as the component that offers a GUI for facilitating the OneNet Participants in the management, monitoring and analytics of the data exchanges.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

The OneNet Monitoring and Analytics Dashboard will be the main User Interface that will allow the access to the OneNet Participants for monitoring and data analytics features, as well as for the OneNet Administrators for configuration and administration tools.

Most of all the functionalities expected from OneNet that could require a GUI will be implemented in this component.

Main features of the OneNet Monitoring and Analytics Dashboard should be:

- administrative and configuration tools;
- easy integration with the OneNet Orchestration Workbench and OneNet Middleware;
- data-analytics dashboard;
- monitoring and alerting dashboard for data processes and platform integrations;
- user-friendly selection of data sources and services from the catalogues;

3.3 Analysis of Cross-Platform Services and Functional Specifications

OneNet Deliverable 5.3 [15] provides the functional specifications and cross-platform services to be supported by the OneNet System. The OneNet Connector and Middleware are the key components to enable interoperability among stakeholders in the energy sector in a decentralized way. FIWARE -- more specifically the FIWARE context broker – is identified as the enabling technology that will be used in each OneNet Controller and the OneNet Middleware for data exchanges. The context broker of a OneNet Connector is the main gateway of each OneNet Participant to access services of the OneNet Framework through the OneNet Middleware as well as to exchange and provide data in the OneNet Network of Platforms in peer-to-peer connections.

A OneNet Participant will use the OneNet API to interact with its Connector and therefore a local instance of a context broker. Received data is forwarded by the local context broker to the OneNet Participant's data processing system. The cross-platform services specified in D5.3 exhibit either a one-at-a-time (e.g., sending a pre-qualification result or measurement value) or a publish/ subscribe (e.g., sharing a market result with all market participants) communication pattern. Both patterns must be implemented through the OneNet API and context broker. Furthermore, the context broker is responsible for interfacing OneNet data services inside the OneNet Connector with each other, for example data privacy service, semantic checking service, and data quality checking service.

The data quality checking service of the OneNet Connector will use big data analytics methods if required to check the data quality requirements that are identified as part of D5.3 for all cross-platform services. The more

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

data-intense a service (for example streaming of measurement data) the more we need big data analytics to check quality criteria such as for example consistency, efficiency, and precision.

Two other components of the OneNet System that will make use of big data analytics and artificial intelligence techniques are the OneNet Workbench and the OneNet Monitoring and Analytics Dashboard. Within these components, such techniques can be applied to facilitate a service discovery for OneNet Participants (Workbench), to test the correctness and performance of new data or cross-platform services prior to their practical use (Workbench), or to analyse and monitor the health and status of the OneNet Network of Platforms and all of its Participants and Components (Dashboard and Monitoring).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

4 FIWARE and Standardized Data Models

4.1 FIWARE Context Broker NGSI

FIWARE Orion Context Broker, which is the core component of the FIWARE platform and implements the Next Generation Service Interface NGSI standard to deploy the Publish/Subscribe architecture. In particular, the Orion Context Broker is one of the nine CEF building blocks [16], whose aim is to help to implement the most commonly needed digital capabilities among Europe. In addition to this, FIWARE is integrated onto the International Data Spaces [17] initiative whose purpose is to offer a common framework for IIoT (Industrial IoT). In view of all this, FIWARE is becoming one of the most relevant frameworks for the development of smart application in several domains.

In FIWARE framework, Orion Context Broker is a key service to ease the development and provisioning of smart and innovative applications that require context information and data stream management, processing and exploitation. The Orion Context Broker [18] is an HTTP Publish/Subscribe implementation—based on the NGSI standard—that enables management of the entire lifecycle of context information including updates, queries, registrations, and subscriptions. Orion Context Broker has been recognized as a CEF Building Block, which is one step forward on its path towards becoming a global standard for large-scale contextual information management. Orion allows defining a model of data (i.e., entity) to which publishers update values to be obtained by subscribers. Orion uses the NoSQL MongoDB database to store these entities and the last value recorded on them.

In NGSI information model, contextual elements are referred to as entities. Entities are physical objects (sensor, actuator, etc.), hardware, and software as presented in Figure 14. Based on this conceptual model, entities are uniquely identified by an Entityid. Each entity can have attributes that are related to an entity's characteristics. These attributes can have static or dynamic values and represent by the triplets <Name, Type, Value>.

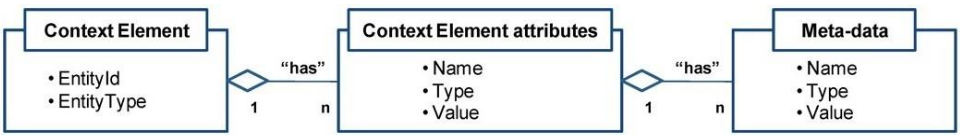

Figure 14: Context element conceptual model [19]

4.1.1 Evolution of FIWARE Context Broker NGSI into NGSI-LD

The FIWARE NGSI v2 information model has been extended to support Linked Data (LD) which is including entities relationships, property of graphs, and semantics. This revolution has been conducted under the ETSI

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

CIM initiative and named NGSI-LD. NGSI-LD targets semantic context information, its specification creates models of real-world entities, relationships, and properties and is expressive to connect and federate other existing information models, using JSON-LD as a lightweight linked data format. Linked data is structured data that is interlinked with other data so it becomes more useful through semantic queries. Furthermore, it is compatible with the Resource Description Framework (RDF) so that triple stores and application logic can be found in SPARQL.

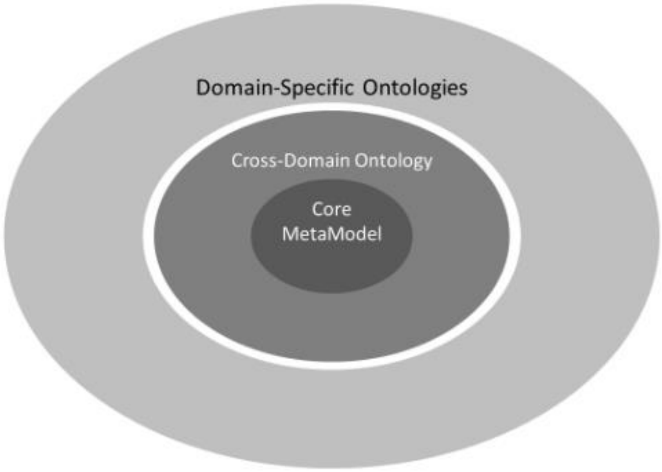

Figure 15: Overview of the NGSI-LD information model structure [20]

According to Figure 15, NGSI-LD is defined at two levels consisting of the Core Meta-Model, the Cross-Domain Ontology. The NGSI-LD Core-Meta Model represents Entities, their Relationships, and their properties with the value according to the “property graph” model [21]. The main constructs of NGSI-LD are Entity, Property, and Relationship. The letters in Cross-Domain are a set of generic and transversal classes to avoid redundant definitions in the Domain-Specific Ontologies. Indeed, NGSI-LD information model contains the core terms needed to uniquely represent the key concepts of the NGSI-LD information model as well as the terms that define the API-related data types. These are encoded using JSON-LD, which provides the advantage of being familiar and accessible to developers.

For example, an Entity, representing a device or other data source, is encoded using a JSON-LD object. The data in NGSI-LD are structured like a data graph model and linked with each other. The relationships between entities are easily handled in NGSI-LD in a similar way to graph databases. It also provides the update, query, and subscription support needed to allow applications to automatically access data from various data sources. Currently, there exist 4 implementations of NGSI-LD that are summarized in the table:

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 957739

Table 3: FIWARE Context Broker implementation

Name	NGSI v2	NGSI-LD	Description
Orion	Supported	Not Supported	The original Orion and supporting NGSI V2
Orion-LD	Supported	Supported	A fork from the original Orion repository and aims to be merged back at some point. It is the only context broker which can service both NGSI-v2 and NGSI-LD
Scorpio	Not Supported	Supported	pure NGSI-LD brokers which does not require the compromises of having to serve both syntaxes. Positions itself as the heavyweight broker, with a strong interest in federations.
Stellio	Not Supported	Supported	pure NGSI-LD brokers which does not require the compromises of having to serve both syntaxes. It is somewhere in the middle between Scorpio and Orion.

All NGSI-LD brokers are trying to align to the ETSI standard, however new features are still being added (the latest ETSI update, the 1.3.1 specification was a couple of months ago). Orion complies with 95% of the 1.2.1 specification (excluding the temporal API). Scorpio is probably higher but some minor differences can be found within payloads.

4.2 OneNet Connector – IDS and FIWARE ecosystem

As described in Chapter 3.2 the OneNet Decentralised Middleware and the OneNet Connector are the core systems for enabling a seamless platforms integration and cooperation for cross-platform market and network operation services and at the same time makes available and accessible data from different sources (the OneNet Participants) in a secure and trusted way ensuring data ownership and privacy.

As described in D5.2 [14], while the OneNet Decentralised Middleware offers central features to all the OneNet participants like identity management, sources discovery, semantic annotation, vocabularies and ontologies, the OneNet Connector is a decentralised instance of the OneNet Middleware itself and is responsible for the execution of the complete data exchange process.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

Starting from the analysis on the state of the art (also reported in D5.2), it was clear that IDS reference architecture model architecture (RAM) and its own connector, implement all the aspects and characteristics necessary to pursue the objectives of OneNet. Whit this is in mind, also the FIWARE initiative is following this path, incubated an open-source implementation of this technology that has already been tested how it can integrate with the rest of core FIWARE components [22].

The OneNet Connector will therefore consist of a hybrid integration of the IDS Reference Model (IDS connector) and FIWARE components (i.e., FIWARE Context Broker, using NGSI-API)

Each OneNet Participant will be able to deploy and configure its own connector that will include:

- UI Configuration tool
- Set of interoperable APIs for the connection with already existing Platform/Application/Services
- OneNet Data services (a detailed list of the Data Services is provided in D5.3)

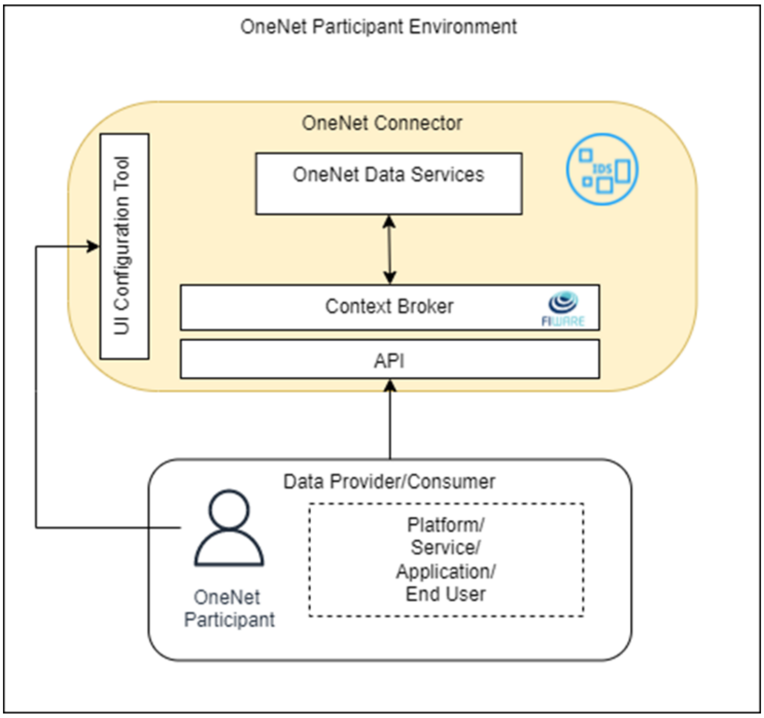

Figure 16: OneNet Connector

4.3 Standardized Data Models Adoption

A Standardized Data Model is an abstract model that organises elements of data and standardises how they relate to one another and to the properties of real-world entities. They play a crucial role because they define the harmonised representation formats and semantics that will be used by applications both to consume and to publish data. The adoption of Standardized Data Models is fundamental for facilitating interoperability within the community.

The FIWARE Foundation together with TMForum, and IUDX, have launched the Smart Data Models initiative [23] where data models are made available for the benefit of all. In it FIWARE Data Models have been harmonised to enable data portability for different application domains including, Smart Cities, Smart Agrifood, Smart Environment, Smart Sensoring, Smart Energy, Smart Water and others domains. The data models are intended to be used wherever you want, but specifically, they are designed to be compliant to FIWARE NGSI V2 and NGSI-LD.

A smart data model includes three elements [24]:

- **the schema**, or technical representation of the model defining the technical data types and structure;
- **the specification** of a written document for human readers;
- **the examples** of the payloads for NGSIv2 and NGSI-LD versions.

All data models are public and of royalty-free nature and grouped into subjects. The subjects can belong to one or several domains. The domains represent industrial sectors.

For the definition of a Standardized Data Models, a well-structured lifecycle wad defined. It includes three different statuses:

- **Official**: The data models are accepted and are fully documented and with schema. And with examples available but for specific exceptions.
- **Harmonization**: The data models are accepted but they are in progress to complete some of the mandatory elements.
- **Incubated**: The data models are elaborated by the users and the organization provides support to achieve an official data model.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

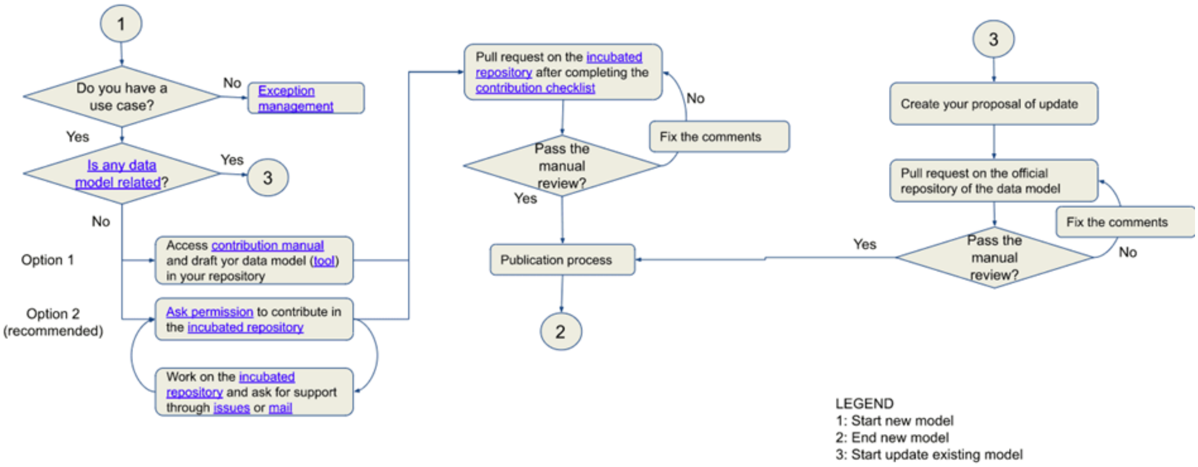

Figure 17: Standardized Data Model Lifecycle [24]

Following an open standard contribution approach, and to guarantee maximum adoption, FIWARE Data models are stored in Github dedicated repositories [25]. In the next chapter, we will see the list of the Smart Energy Data Models [26].

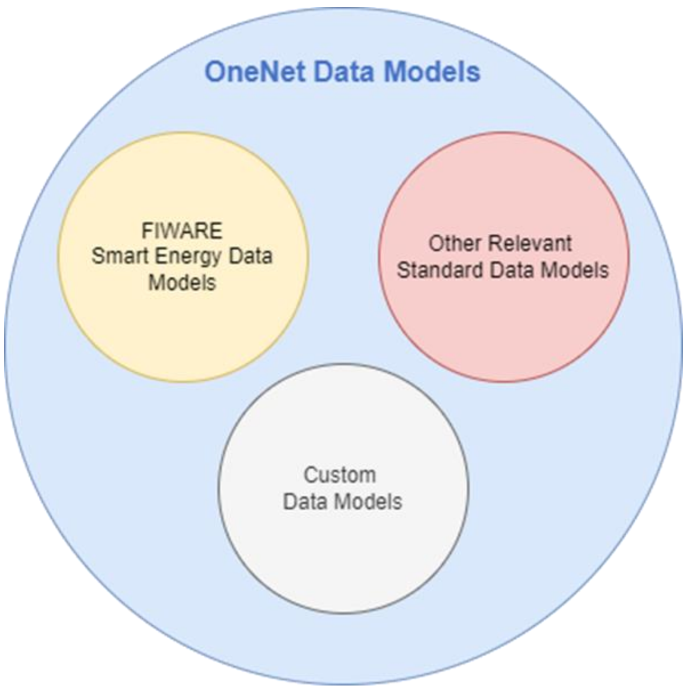

Figure 18: OneNet Data Models Adoption

4.3.1 FIWARE Smart Energy Data Models

As described in the previous paragraph, every domain is divided into subjects. In the Smart Energy Domain, 5 different subjects were defined so far:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

- Battery
- Energy
- EnergyCIM
- GreenEnergy
- Weather

Each subject includes many data models. All the FIWARE Smart Energy Data Models are detailed and described in the GitHub Open Repository [26].

The detailed list of the FIWARE Smart Energy Data Models to be addressed in OneNet is reported in the Appendix B – FIWARE Smart Energy Data Models.

4.3.2 Relevant Standard Data Models

One of the main objectives of the OneNet project is to support the cross-border and cross-domain data exchange. For this reason, an in-depth analysis was conducted to identify which are the most relevant standard data models, which should be considered and supported by OneNet. These standard data models, which extend the list of those already present and supported by FIWARE and the Smart Data Model initiative, will be identified and consolidated within Task T5.6.

The analysis that leads to the definition of these standard data models was made starting from the activities carried out in T5.3 for the definition and categorization of a list of cross-platform services, relevant both at the European level starting from the H2020 projects in progress or already concluded, but above all by the needs of the OneNet demos.

In Task T5.3 a first list of data exchanges between the various platforms were identified and this information was used in T5.6 to define a list of Business Objects and related standard data models to be integrated into the OneNet platform. So far, several data models from CIM were addressed for the cross-platform services (e.g., ESMP (62325), CGMES (61970), CDPSM (61968)) to address flexibility trading, activation and settlement as well as grid topology exchanges.

A first list of the extended standard data models will be published within the D5.6 in March 2022 (M18 of the project).

4.3.3 Creation of new Data Models

Although the use of standard data models facilitates the easier and faster data exchange and interoperability among different systems, it is often necessary to create and use customized data models or make changes to existing ones.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

For this reason, the OneNet architecture allows the OneNet participants to configure and enable new data models, customized for specific data exchanges.

OneNet Middleware, through the OneNet connector, will therefore give way to create new data models, as well as provide the Data Models catalogue of the existing ones, encouraging to follow the specific guidelines suggested by the Smart Data Models initiative [23], so that eventually the new data models can be reused within the OneNet ecosystem in an easy and accessible way.

Each Data Model is defined using a JSON Schema [24] that covers only the so-called key-value representation of NGSI v2 context data. Thus, the JSON Schema does not cover the normalised representation of context data.

In the NGSI v2 normalised format each attribute (key) of the Data Model is represented by a JSON object with the following syntax:

- The attribute value is specified by the value property, whose value may be any JSON datatype.
- The attribute NGSI type is specified by the "type" property, whose value is a string containing the NGSI type.
- The attribute metadata is specified by the metadata property. Its value is another JSON object which contains a property per metadata element defined (the name of the property is the name of the metadata element). Each metadata element, in turn, is represented by a JSON object containing the following properties:
    - value: Its value contains the metadata value, which may correspond to any JSON datatype.
    - type: Its value contains a string representation of the metadata NGSI type.

```
{
    "id": "entityId",
    "type": "entityType",
    "att1": {
        "value": "value1",
        "type": "Text",
        "metadata": {
            "metada1": {
                "value": "metavalue1"
            }
        }
    }
}
```

Figure 19: Normalised Data Model

```
{
    "id": "entityId",
    "type": "entityType",
    "att1": "value1"
}
```

Figure 20: Key-Value Data Model

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

As shown in the Figure 19 and Figure 20, the JSON schema of normalised format contains richer information, but on the other side it could be redundant and less efficient in term of transport.

The utilization of NGSI-LD also introduces the concept of the "*@context*" and the relationships between entities.

```
{
    "@context": "http://context/ngsi-context.jsonld",
    "id": "urn:ngsi-ld:TemperatureSensor:001",
    "type": "TemperatureSensor"
},
```

Figure 21: @Context Example

```
{
    "id": "urn:ngsi-ld:TemperatureSensor:001",
    "type": "TemperatureSensor",
    "controlledAsset": "urn:ngsi-ld:Building:farm001"
}
```

Figure 22: Relationship Example

The @context provides additional information allowing the systems to interpret the rest of the data with more clarity and depth. This enables interoperability of the solutions you could create, which is based on a standard data model with NGSI-LD. When you have a digital twin of your system, and you have for each entity a respective URL/URN that uniquely identifies it.

The relationship between entities allow to establish associations between instances using linked data. In practice, they are conveyed by means of a special NGSI v2 attribute with a special value (relationship's object), which happens to be a URI which points to another entity.

Previously the Smart data model defined the name of such attributes was normed to be ref followed by the name of the Entity Type referenced by the attribute, for example in the case of Entity Type Device, the full attribute would be "*refDevice*".

With the introduction of NGSI-LD such attributes are defined with the usage of a verb (plus optionally an object) such as *"hasStop", "operatedBy", "hasTrip"*, etc.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

4.4 Data Profiling – CGMES

4.4.1 Metadata and reference data to support TSO-DSO data exchange

This chapter provides an overview of the header, metadata, and reference data that was developed to satisfy new requirements for TSO-TSO data exchanges as well as to support enhanced TSO-DSO information exchange. The work presented here is based on the existing International standards, ENTSO-E specifications, and the work developed by international organisations (World Wide Web Consortium-W3C, European Commission) that are applicable for this topic.

4.4.2 Metadata introduction

Access to consistent, high-quality metadata is critical to finding, understanding, and reusing scientific data. General definitions of metadata can be found in many manuals. Most of them are very short and simple. The most used generic definition states that "Metadata are data about data" but more precise definition states: Metadata is data that defines and describes other data [27]

From the metadata perspective, the goal here is to use one single model for the header and metadata, covering the needs of the TSO and DSO community. However, considering the great variety of requirements and handling metadata, it is recommended to cover this in different places as presented in the Figure 23.

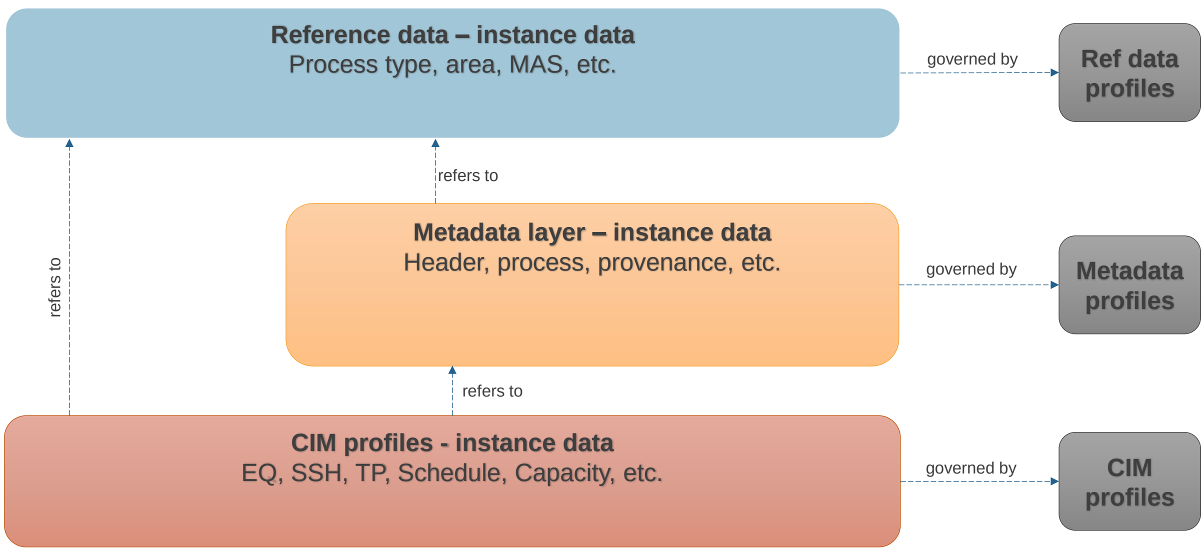

Figure 23: Relations between Reference data, Metadata, and CIM profiles.

In the data exchange there are three main layers: the instance data that is governed by CIM profiles. This data has header information; the metadata layer instance data which describes the processes and can describe the

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

provenance of the data; the reference data which describes common data to be referenced from either metadata layer or instance data of the core profiles that are exchanged for various profiles, e.g. EQ, SSH, etc.

4.4.3 Gap Analysis and Requirements

A gap analysis was conducted to check if the current header, metadata and reference data is covering the requirements of the TSO-DSO data exchange.

The gap assessment was performed for each component of the existing header and metadata developed by ENTSO-E. The gap analysis was based on the existing ENTSO-E documentations and on expert assessment from ENTSO-E and OneNet partners. The following sections describe some areas where special attention should be paid and extend the current model. Table 4 gives and overview of the requirements for TSO/DSO metadata and reference data.

Table 4: List of the requirements for TSO/DSO metadata and reference data.

Data present in CIM/CGMES	Header and metadata attributes	Reference data	Gap description
Meter	No need	Gap to cover	There is a need to define refence data for reading quality type
Meter	No need	Gap to cover	There is a need to define refence data for reading type
Meter	Gap to cover	No need	Information on Metering data operator (MDO) should be exchanged
Boundary Information Exchange	No need	No need	Boundary information exchange between TSO and DSO needs to be defined.
Exchange of Difference model	No need	No need	The exchange of difference model needs to be clarified and improved.

An exhaustive description on the TSO/DSO data exchange is provided in the Appendix C – CGMES TSO/DSO Data Exchange.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

4.4.4 Exchange of difference model

The exchange of difference model defined in IEC 61970-552:2016 is defined as a custom difference model exchange for CIM and in some definitions is not fully aligned with W3C and linked data principles. It is recommended to reply on newer methods to represent the different model exchange also known as incremental model exchange. For instance, there is a vocabulary JSON-LD patch that could be used and standardised in IEC.

4.4.5 Next steps

Based on the above recommendations it is proposed that sample data is prepared to demonstrate the approach. In order to prepare the data different demos in OneNet project will need to feedback which use cases are going to be part of the demos and the sample data can include some of these concrete cases.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

5 Interoperable interfaces for OneNet Orchestration Workbench and OneNet Monitoring and Analytics Dashboard

5.1 Integration with the OneNet Middleware

As described in D5.2 [14] and in Chapter 3.2 the OneNet Orchestration Workbench and the OneNet Monitoring and Analytics Dashboard will support and empower the data integration, orchestration and monitoring within the OneNet Framework.

To achieve this objective, it is necessary that the integration of the data and services to be orchestrated, monitored and evaluated complies with the specifications and requirements of the OneNet System and in particular that the two components are integrated with the OneNet Decentralized Middleware.

The integration of these two components with the OneNet Decentralized Middleware facilitates the management of the integration of data and services and allows to take advantage of all the features that the OneNet System provides, both in terms of data management and security.

The Figure 24 shows that for the integration of the OneNet Orchestration Workbench and OneNet Monitoring and Analytics Dashboard, two interfaces' layers are required. From one side the integration and communication between the components and the OneNet Decentralised Middleware for the data and services integration and the other side the interfaces for the interaction of the OneNet Participants with the components, exploiting both the User Interfaces and Services available.

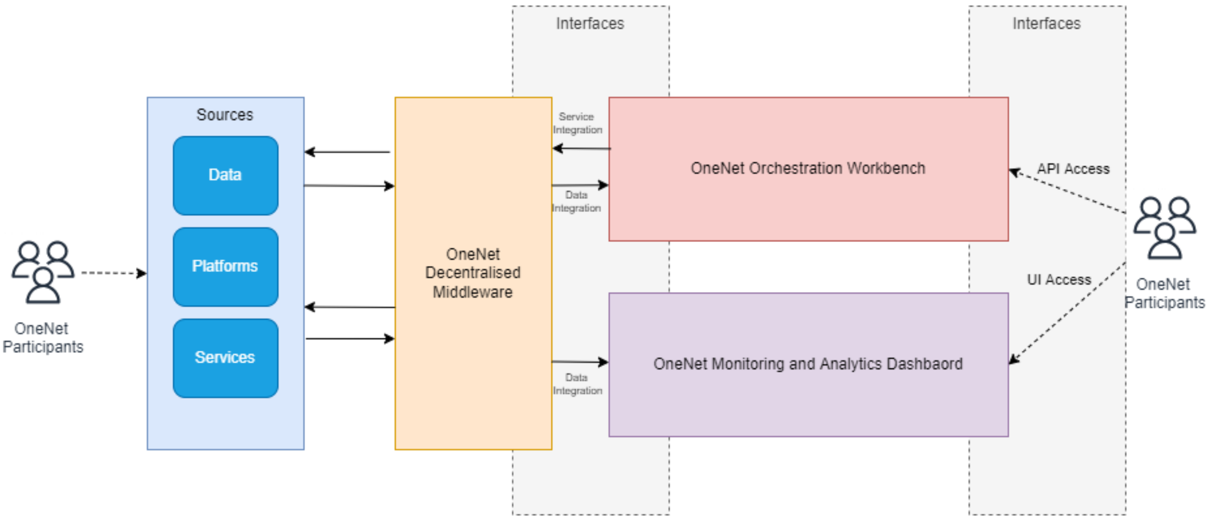

Figure 24: OneNet Orchestration Workbench and OneNet Monitoring and Analytics Dashboard Interface

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

5.2 Standard Interfaces – OpenAPIs and Graphical User Interfaces

The interfaces to be implemented between the OneNet Decentralised Middleware and OneNet Orchestration Workbench should follow standard approaches and exploit what already provided in terms of interoperability and communication with the OneNet Decentralised Middleware.

For this reason, the OneNet Workbench should exploit the FIWARE Context Broker standard interface (and standard NGSI API) for the integration of data sources through the OneNet Decentralised Middleware.

On the other hand, the functionalities offered by the OneNet Orchestration Workbench are not limited only to the integration of data sources but also to the registration, integration, and evaluation of services. Thus, it is also necessary that some additional interfaces based on REST services and standards OpenAPI are included.

The interfaces to be implemented for the OneNet Participants focus above all on the Graphical User Interfaces (GUI) of the OneNet Monitoring and Analytics Dashboard, even if some services offered by the two components could be used in automated way and therefore offered through REST interfaces based on OpenAPI standards.

The graphical user interfaces provided in the OneNet Monitoring and Analytics Dashboard should allow to manage all the activities envisaged within the OneNet System, including those of:

- Access and configure the environment as a OneNet Participant
- Exploration of data and service catalogues
- Analytics Dashboard
- Monitoring and Alerting Dashboard

The table below summarize all the interfaces expected to be implemented in WP6 for the OneNet Orchestration Workbench and OneNet Monitoring and Analytics Dashboard.

Interface ID	From	To	Type	Description
int0	OneNet Orchestration Workbench	OneNet Middleware (NGSI broker)	Broker/API	Interface between OneNet Middleware and the OneNet Orchestration Workbench. The OneNet Orchestration Workbench should exploit FIWARE Context Broker implementation for integrating data sources and services.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

int1	OneNet Middleware	OneNet Orchestration Workbench	Open API	Interface offered by OneNet Orchestration Workbench for including additional services in the OneNet Decentralized Middleware (e.g., register news services or retrieving Service Catalogue information).
Int2	OneNet Participant (User)	OneNet Monitoring and Analytics Dashboard	GUI	Graphical Interfaces for accessing, configure and utilize the OneNet Monitoring and Analytics Dashboard
Int3	OneNet Participant (User)	OneNet Monitoring and Analytics Dashboard	Open API	Standard OpenAPI interfaces for exploiting the services offered by the OneNet Monitoring and Analytics Dashboard in automated way
Int4	OneNet Participant (User)	OneNet Orchestration Workbench	Open API	Standard OpenAPI interfaces for exploiting the services offered by the OneNet Orchestration Workbench in automated way

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

6 Conclusions

The work conducted in Task T5.4 so far and documented in this deliverable gives a complete overview on how the usage and evolution of the FIWARE Context Broker together with the adoption of the standardised data models is necessary and crucial for achieving the main objectives that OneNet Solution foresees.

In addition, the application of Big Data Analytics Methods and Services is now widely used in the context of the Smart Grid Management and for the implementation and integration of the real time services.

The outcomes provided in the other activities of the WP5 clearly highlight how the FIWARE Context Broker, the OneNet Orchestration Workbench and the OneNet Analytics and Monitoring Dashboard assume a relevant role in the context of the OneNet Solution for empowering the seamless cross-platform, cross-country and cross-domain data exchange.

The result of this activity brought to the identification of the already existing Standard Data Models and the possibility to extend this data models, maintaining a standardised approach following the Smart Models Initiative.

Moreover, an exhaustive analysis and investigation on the CGMES for the TSO-DSO data exchange and cooperation was reported and will be integrated in the demo use cases for the evaluation phase.

Finally, the interfaces for the interoperable integration of the OneNet Orchestration Workbench and OneNet Monitoring and Analytics Dashboard were identified, in order to be implemented in the WP6 activities.

In the next steps, the Task T5.4 will continue the analysis on the FIWARE evolution and the interoperable interfaces, taking a look in the implementation and evaluation phase of the project, including any possible feedback and results in the later stage.

7 References

[1] D. Syed, A. Zainab, A. Ghrayeb, S. S. Refaat, H. Abu-Rub and O. Bouhali, "Smart Grid Big Data Analytics: Survey of Technologies, Techniques, and Applications," in IEEE Access, vol. 9, pp. 59564-59585, 2021, doi: 10.1109/ACCESS.2020.3041178.

[2] Zhang, Y., Huang, T. & Bompard, E.F. Big data analytics in smart grids: a review. Energy Inform 1, 8 (2018). https://doi.org/10.1186/s42162-018-0007-5

[3] A. S.Gillis, «5 V's of big data,» March 2021. [Online]. Available: https://searchdatamanagement.techtarget.com/definition/5-Vs-of-big-data.

[4] Y. Simmhan, S. Aman, A. Kumbhare, R. Liu, S. Stevens, Q. Zhou, et al., "Cloud-based software platform for big data analytics in smart grids", Comput. Sci. Eng., vol. 15, no. 4, pp. 38-47, Jul. 2013.

[5] Bhattarai et al. Big data analytics in smart grids: state-of-the-art, challenges, opportunities, and future directions. United States. https://doi.org/10.1049/iet-stg.2018.0261

[6] R. Kalsbeek, «Where to Start With The 4 Types of Analytics,» 21 December 2020. [Online]. Available: https://iterationinsights.com/article/where-to-start-with-the-4-types-of-analytics/.

[7] «CRISPy Data Mining in Marketing Organizations,» [Online]. Available: https://xperra.com/blog/crispy.html.

[8] «Wikipedia - Cross industry standard process for data mining,» [Online]. Available: https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining.

[9] «Big Data Value Association,» [Online]. Available: https://www.bdva.eu/.

[10] BDVA, «European Big Data Value Strategic Research and Innovation Agenda,» 2017.

[11] «High Level Architecture (HLA) - Release 5.0,» December 2020. [Online]. Available: https://aioti.eu/wp-content/uploads/2020/12/AIOTI_HLA_R5_201221_Published.pdf.

[12] Willner, Alexander & Gowtham, Varun. (2020). Towards a Reference Architecture Model for Industrial Edge Computing.

[13] OneNet Deliverable 5.1 "OneNet Concept and Requirements" https://onenet-project.eu/public-deliverables/

[14] OneNet Deliverable 5.2 "OneNet reference architecture" (Confidential deliverable)

[15] OneNet Deliverable 5.3 "Data and Platform Assets Functional Specs and Data Quality Compliance" (Confidential deliverable)

[16] «CEF Digital Connection Europe - Context Broker,» [Online]. Available: https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/Context+Broker.

[17] «International Data Space Association,» [Online]. Available: https://internationaldataspaces.org/.

[18] «FIWARE Orion Context Broker,» [Online]. Available: https://fiware-orion.readthedocs.io/en/master/.

[19] FIWARE-NGSIv2Specification,"Nextgenerationserviceinterfacesarchitectureapprovedversion [Online].Available:http://fiware.github.io/specifications/ngsiv2/stable

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

[20] https://www.etsi.org/deliver/etsi_gs/CIM/001_099/009/01.04.01_60/gs_cim009v010401p.pdf
[21] Graph Databases: "New Opportunities for Connected Data". O'Reilly 2nd Edition. Webber, Robinson, et al. ISBN:1491930896 9781491930892.
[22] «FIWARE TRUE Connector,» [Online]. Available: https://fiware-true-connector.readthedocs.io/en/latest/.
[23] «Smart Data Models,» [Online]. Available: https://smartdatamodels.org/.
[24] «FIWARE Smart Data Models,» [Online]. Available: https://www.fiware.org/smart-data-models/.
[25] «FIWARE Smart Data Models Repository,» [Online]. Available: https://github.com/smart-data-models/data-models/tree/master/specs.
[26] «FIWARE Smart Data Models - Smart Energy,» [Online]. Available: https://github.com/smart-data-models/SmartEnergy.
[27] «Handbook to set up a S-DWH: Roadmap for a Design phase. » 2017. [Online] https://ec.europa.eu/eurostat/cros/system/files/s-dwh_m_5.1_metadata1.1.pdf.
[28] «Apache Hadoop,» [Online]. Available: https://hadoop.apache.org/.
[29] «Apache Cassandra,» [Online]. Available: https://cassandra.apache.org/_/index.html.
[30] «Apache Kafka,» [Online]. Available: https://kafka.apache.org/.
[31] «Apache Spark,» [Online]. Available: https://spark.apache.org/.
[32] «Apache Storm,» [Online]. Available: https://storm.apache.org/.
[33] «Apache Zookeper,» [Online]. Available: http://zookeeper.apache.org/.
[34] «Apache Drill,» [Online]. Available: https://drill.apache.org/.
[35] «Metadata and Document Header Data Exchange Specification ENTSO-E. » 2021.
[36] «W3C SKOS Simple Knowledge Organization System» [Online] https://www.w3.org/TR/skos-reference/
[37] «W3C DCAT»: [Online] https://w3c.github.io/dxwg/dcat/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

8 Appendix A – Big Data Architecture and Platforms

Apache Hadoop

Apache Hadoop [28] is an open-source framework that provides a set of utilities that facilitate the distributed data storage and the big data processing. It consists of the core (for storage part) called the Hadoop Distributed File System (HDFS), the processing component that is the MapReduce programming model and resource scheduler called Hadoop YARN (Yet Another Resource Negotiator).

Hadoop core contains a pre-defined collection of utilities and libraries that can be used by other modules within the Hadoop ecosystem.

HDF is the default distributed storage system in Apache Hadoop. The huge datasets are dumped in the HDFS and when required, access to the data is provided to other Hadoop modules using utilities. HDFS component provides reliable and quick access to the data by creating several copies of the data block and these copies are distributed across multiple clusters.

Hadoop YARN is the dynamic resource management component that lets the user run multiple Hadoop applications without having to worry about the aggravating workloads. YARN provides for improved cluster utilization. Key components of YARN are Resource Manager, Application Master, Node Manager, and containers.

Hadoop MapReduce is the framework for parallel computations of massive data sets. MapReduce model is employed for the parallel computation and interpretation of massive-scale data and has three stages: map, shuffle, and reduce. All the jobs are written in a functional programming style to create map and reduce tasks. Dynamic systems for the MapReduce model are commonly clusters that perform tasks such as data partitioning, scheduling of jobs, and communication between the cluster nodes and hence, are more suitable when dealing with massive-scale data. In the map phase, the data are read from the DFS and partitioned into clustered systems where the input is processed to compute the intermediary results which are then stored on the local node of the cluster where the map phase has run and waits for all the map functions to generate output in key-value pairs. The output in key-value pairs is then given as input to the reduce function to generate the final result. The advantage of the MapReduce model is that it takes processing to where the data resides and hence, decreases the transmission of data and improves efficiency. Therefore, the MapReduce model is more apt for the distributed computing of massive-scale data.

Hadoop has provided for storing and analyzing data at massive scales. However, data analytics technology cannot be applied to real-time systems. The advent of the Internet-of-Things, smart meters, and devices has led to the possibility of real-time analysis of data for the benefits of business and many other advantages such as

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

smart grid stability, and management. The real-time handling of data falls under one of the categories: Stream processing or Iterative processing. The stream processing framework would work efficiently for big data analytics in the smart grid for real-time decisions about generation, control, etc.

Apache Cassandra

Apache Cassandra [29] is a NoSQL database ideal for high-speed and huge data transactions leveraging on a distributed and decentralized storage system. Cassandra is designed with scale, performance, and continuous availability as the foundation architecture principles. Cassandra operates using a masterless ring architecture—it does not rely on a master-slave relationship. In Cassandra, all nodes play an identical role; there is no concept of a master node, with all nodes communicating with each other via a distributed, scalable protocol. Cassandra stores data by dividing the data evenly around its cluster of nodes. Each node is responsible for part of the data. The act of distributing data across nodes is referred to as data partitioning.

Cassandra architecture is scalable by-design. It means that it is capable of handling large amounts of data and millions of concurrent users or operations per second—even across multiple data centers—as easily as it can manage much smaller amounts of data and user traffic. To add more capacity, you simply add new nodes to an existing cluster without having to take it down first. Unlike other master-slave or sharded systems, Cassandra has no single point of failure and therefore is able to offer continuous uptime and availability.

Apache Kafka

Apache Kafka [30] is a distributed stream processing engine for building real-time data pipelines and streaming applications. It allows you to receive data from different types of sources (called producers), processing them within its architecture and making them available to recipients (consumers). Within the architecture it is possible to build applications through the Kafka Streams library for filtering and data enrichment operations.

Apache Kafka can handle huge volumes of data and remains responsive; this makes Apache Kafka one of the preferred platforms when the volume of the data involved is big to huge. Kafka can also work seamlessly with most of other Big Data platforms and software like Spark for real-time ingesting and feed Data Warehouses/lakes and Hadoop for real-time analysis as well as to process real-time streams to collect Big Data.

Apache Spark

Apache Spark [31] is an open-source cluster computing framework for analyzing massive-scaled data. Spark has the capability for stream processing of big data and has many advantages over Hadoop MapReduce and Storm. In Apache Spark, data analytics is more stream processing than batch processing and hence, it avoids the reprocessing of the data. This provides the stream processing model of Apache Spark to be dynamic and it becomes more crucial during the real-time processing of huge volumes of data collected from different sources.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

Even for iterative processing, the leading framework currently is Apache Spark as it possesses the capability of processing and holding the data in the memory nodes across the cluster.

The main characteristics of the Spark Framework are:

- Speed: Spark extends the MapReduce model to execute computations of stream processing and interactive querying. In literature, it is proven to be 10 times faster than the Hadoop MapReduce model.
- Ease of Use: Applications written in any language such as Java, python, scala, etc. are compatible with Apache Spark.
- Advanced Analytics: Spark supports the MapReduce model of Hadoop, SQL-like querying, streaming data, machine learning algorithms, and graph algorithms as well.
- Iterative and Interactive Applications: Spark is designed to execute both in-memory and on-disk. It holds the intermediary results in memory rather than writing to disk to avoid reprocessing the data if required again. Spark operators perform external operations on the data if it does not fit into memory.
- In-memory Computation: The data is stored in memory rather than written on disk. Hence, Spark reduces the response time to a great extent when the data is queried.
- Directed Acyclic Graph (DAG): DAG in Apache Spark is a set of vertices and edges where the vertices are the representations of the Resilient Distributed Datasets (RDDs) while edges represent the operations to be performed on the RDDs. DAGs in Spark can contain any number of stages. Even the MapReduce model of Hadoop is a DAG of two stages - Map and Reduce. This allows for simple jobs to be completed in one stage and more complex jobs to be completed in one run of many stages unlike multiple jobs in the MapReduce model. Thus, jobs in Spark execute faster than they would in the MapReduce paradigm.

The Spark Framework is composed by multiple components.

- Spark Core is the base of all the Spark projects, and it allows basic input/output operations, distributed task dispatching, and scheduling through an Application Programming Interface (API) centered on RDD abstraction.
- Spark SQL: Spark SQL is the Apache Spark module that is commonly worked with structured data. It lies on top of Spark core and is used to execute SQL queries.
- Spark Streaming: It is the component of the spark that enables the processing of live streams of data. Spark streaming gives a programming interface for processing data streams. It resembles the Spark core's RDD API, pushes data in small chunks, and does RDD transformations on the batches of data.
- MLib: Apache Spark comprises a library with common machine learning functionality and this library is called MLib. It processes data faster when compared to Hadoop's disk-based machine learning library called Mahout.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

- GraphX: The GraphX API provides for users to view data in graphical format and to view RDDs without data movement or duplication. It uses the fundamental operators such as subgraphs, joinVertices, aggregateMessages, etc.

Apache Storm

Apache Storm [32] is a scalable and distributed framework for reliable computation and processing of streams of real-time data with processing latencies in the order of milliseconds. Apache Storm can ingest the data from multiple sources using Kafka or Kinesis. A storm cluster is very alike to the data cluster in Apache Hadoop. In Hadoop, MapReduce jobs are executed while topologies are executed in Apache Storm. Topologies are very similar to jobs, but topologies process messages or data forever until these are killed.

In a Storm cluster, there are two types of nodes, namely master node and worker nodes. A background process called Nimbus runs on the master node and this is analogous to Hadoop's job-tracker. Nimbus process distributes the code in the cluster i.e. assigns tasks to the machines and monitors for any failures. On the machines other than the master node, the process called Supervisor runs and it listens for the work assigned to its machine by the Nimbus daemon. It starts and stops the worker node process depending upon the task assigned to the machine. Every worker process runs a subset of topologies. That means the execution of topology requires multiple worker processes that are assigned to different machines across the cluster. It requires coordination between Nimbus daemon and Supervisor processes, and this is taken care of by Zookeeper which is the coordinating service in the distributed environment [33]. Zookeeper takes care of naming, configuration, synchronization, etc. The important point to note is that all daemons in the Apache Storm are stateless and fail-fast and these come back up even if these are killed by issuing manual commands. This provides for the stable and reliable real-time analysis of big data.

Apache Drill

Apache Drill [34] is an open-source software framework that provides for data-driven distributed applications requiring interactive processing of massive-scaled data. Apache Drill is the first and only distributed SQL engine that does not require schemas. Drill automatically understands the data when data are provided. This saves a lot of time and effort in defining schemas, transforming data, and maintaining those schemas. It is designed to handle Petabytes (PBs) of data spread across thousands of clusters and it responds to ad-hoc queries with high performance and low latency.

It is a query layer that functions even when multiple data sources are present. It primarily scans the full tables instead of maintaining indices. The workers in Apache Drill are named Drillbits and run on each of the data nodes in the cluster. The coordination between the drillbits, optimization, scheduling, and execution is performed in a distributed way.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

9 Appendix B – FIWARE Smart Energy Data Models

9.1 Battery

The Battery subject includes all the data models related with battery system to be used across different domains.

The following data models are available:

Battery. Represent a physical battery with its hardware specifications

BatteryStatus. Represent a status for a physical battery.

StorageBatteryDevice. The storage battery device data model is intended to describe the technical characteristics of the battery and the charging and discharging conditions of the energy.

StorageBatteryMeasurement. Storage Battery Observed Data Model is intended to measure the remaining energy capacity in a battery, which can be redistributed in the form of electrical energy. These functions apply from a source which depends on the type of battery (reference to the attribute 'batteryType' of the Data Model StorageBatteryDevice).

9.2 Energy

The Energy subject contains domain specific data models related to energy. The adaptation of IEC standards (CIM) is provided in the specific subject.

The following data models are available:

ACMeasurement. The Data Model intended to measure the electrical energies consumed by an electrical system which uses an Alternating Current (AC) for a three-phase (L1, L2, L3) or single-phase (L) and neutral (N). It integrates the initial version of the data Modem [THREEPHASEMEASUREMENT], extended to also perform single-phase measurements. It includes attributes for various electrical measurements such as power, frequency, current and voltage.

InverterDevice. The data model is intended to describe the mechanical, electrical characteristics of an Inverter according to DC - Direct Current Information supplied as input and AC - Alternating Current Information returned as output. Remark: This Data Model can be used directly as a main entity to describe the device [Inverter] or as a sub-entity of the Data Model {DEVICE} using a reference by the [refDevice] attribute.

TechnicalCabinetDevice. Technical Cabinet Device Data Model is intended to to describe the technical characteristics of the Device, designed to be placed in an urban or interurban environment. The main objective

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

of these cabinets for this Data Model is to protect the electrical equipment necessary for the control, surveillance, reading and management of urban lighting, signaling, video and electrical distribution. The scope of use of some of these cabinets can extend to an additional protection for installations of modular apparatuses of telephony, data processing, meteorological stations, photo-voltaic stations, wind turbines stations, telecommunications, networks, data, bre Optics , etc. Remark : This Data Model can be used directly as a main entity to describe the device Technical Cabinet or as a sub-entity of the Data Model DEVICE using a reference by the refDevice attribute. It can also refer to the list of all the components it contains, with the refDeviceList attribute, using the Data Model DEVICE

ThreePhaseAcMeasurement. An electrical measurement from a system that uses three phase alternating current.

9.3 EnergyCIM

The EnergyCIM subject try to map the IEC 61970 Standards (CIM) data models. Below are report a partial list of the data models that includes the most relevant for the OneNet context. A complete list of data models is available in the GitHub repo

ActivePower. Adapted from CIM data models. Product of RMS value of the voltage and the RMS value of the in-phase component of the current.

ActivePowerLimit. Adapted from CIM data models. Limit on active power flow.

ActivePowerPerCurrentFlow. Adapted from CIM data models.

Area. Adapted from CIM data models. Area.

BaseVoltage. Adapted from CIM data models. Defines a system base voltage which is referenced.

BasicIntervalSchedule. Adapted from CIM data models. Schedule of values at points in time.

Command. Adapted from CIM data models. A Command is a discrete control used for supervisory control.

ConformLoad. Adapted from CIM data models. ConformLoad represent loads that follow a daily load change pattern where the pattern can be used to scale the load with a system load.

ConformLoadGroup. Adapted from CIM data models. A group of loads conforming to an allocation pattern.

ConformLoadSchedule. Adapted from CIM data models. A curve of load versus time (X-axis) showing the active power values (Y1-axis) and reactive power (Y2-axis) for each unit of the period covered. This curve represents a typical pattern of load over the time period for a given day type and season.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

Control. Adapted from CIM data models. Control is used for supervisory/device control. It represents control outputs that are used to change the state in a process, e.g. close or open breaker, a set point value or a raise lower command.

ControlArea. Adapted from CIM data models. A control areais a grouping of generating units and/or loads and a cutset of tie lines (as terminals) which may be used for a variety of purposes including automatic generation control, powerflow solution area interchange control specification, and input to load forecasting. Note that any number of overlapping control area specifications can be superimposed on the physical model.

CoordinateSystem. Adapted from CIM data models. Coordinate reference system.

CurrentFlow. Adapted from CIM data models. Electrical current with sign convention: positive flow is out of the conducting equipment into the connectivity node. Can be both AC and DC.

CurrentLimit. Adapted from CIM data models. Operational limit on current.

Curve. Adapted from CIM data models. A multi-purpose curve or functional relationship between an independent variable (X-axis) and dependent (Y-axis) variables.

CurveData. Adapted from CIM data models. Multi-purpose data points for defining a curve. The use of this generic class is discouraged if a more specific class can be used to specify the x and y axis values along with their specific data types.

DayType. Adapted from CIM data models. Group of similar days. For example it could be used to represent weekdays, weekend, or holidays.

Discrete. Adapted from CIM data models. Discrete represents a discrete Measurement, i.e. a Measurement representing discrete values, e.g. a Breaker position.

DiscreteValue. Adapted from CIM data models. DiscreteValue represents a discrete MeasurementValue.

EnergyArea. Adapted from CIM data models. Describes an area having energy production or consumption. Specializations are intended to support the load allocation function as typically required in energy management systems or planning studies to allocate hypothesized load levels to individual load points for power flow analysis. Often the energy area can be linked to both measured and forecast load levels.

EnergyConsumer. Adapted from CIM data models. Generic user of energy - a point of consumption on the power system model.

EnergySchedulingType. Adapted from CIM data models. Used to define the type of generation for scheduling purposes.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

EnergySource. Adapted from CIM data models. A generic equivalent for an energy supplier on a transmission or distribution voltage level.

Equipment. Adapted from CIM data models. The parts of a power system that are physical devices, electronic or mechanical.

Frequency. Adapted from CIM data models. Cycles per second.

GeneratingUnit. Adapted from CIM data models. A single or set of synchronous machines for converting mechanical power into alternating-current power. For example, individual machines within a set may be defined for scheduling purposes while a single control signal is derived for the set. In this case there would be a GeneratingUnit for each member of the set and an additional GeneratingUnit corresponding to the set.

GeographicalRegion. Adapted from CIM data models. A geographical region of a power system network model.

IdentifiedObject. Adapted from CIM data models. This is a root class to provide common identification for all classes needing identification and naming attributes.

Length. Adapted from CIM data models. Unit of length. Never negative.

LimitSet. Adapted from CIM data models. Specifies a set of Limits that are associated with a Measurement. A Measurement may have several LimitSets corresponding to seasonal or other changing conditions. The condition is captured in the name and description attributes. The same LimitSet may be used for several Measurements. In particular percentage limits are used this way.

Line. Adapted from CIM data models. Contains equipment beyond a substation belonging to a power transmission line.

LoadAggregate. Adapted from CIM data models. Standard aggregate load model comprised of static and/or dynamic components. A static load model represents the sensitivity of the real and reactive power consumed by the load to the amplitude and frequency of the bus voltage. A dynamic load model can used to represent the aggregate response of the motor components of the load.

LoadArea. Adapted from CIM data models. The class is the root or first level in a hierarchical structure for grouping of loads for the purpose of load flow load scaling.

LoadComposite. Adapted from CIM data models. This model combines static load and induction motor load effects. The dynamics of the motor are simplified by linearizing the induction machine equations.

LoadDynamics. Adapted from CIM data models. Load whose behaviour is described by reference to a standard model A standard feature of dynamic load behaviour modelling is the ability to associate the same behaviour to multiple energy consumers by means of a single aggregate load definition. Aggregate loads are

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

used to represent all or part of the real and reactive load from one or more loads in the static (power flow) data. This load is usually the aggregation of many individual load devices and the load model is approximate representation of the aggregate response of the load devices to system disturbances. The load model is always applied to individual bus loads (energy consumers) but a single set of load model parameters can used for all loads in the grouping.

LoadGroup. Adapted from CIM data models. The class is the third level in a hierarchical structure for grouping of loads for the purpose of load flow load scaling.

LoadStatic. Adapted from CIM data models. General static load model representing the sensitivity of the real and reactive power consumed by the load to the amplitude and frequency of the bus voltage.

LoadUserDefined. Adapted from CIM data models. Load whose dynamic behaviour is described by a user-defined model.

Location. Adapted from CIM data models. The place, scene, or point of something where someone or something has been, is, and/or will be at a given moment in time. It can be defined with one or more position points (coordinates) in a given coordinate system.

Measurement. Adapted from CIM data models. A Measurement represents any measured, calculated or non-measured non-calculated quantity. Any piece of equipment may contain Measurements, e.g., a substation may have temperature measurements and door open indications, a transformer may have oil temperature and tank pressure measurements, a bay may contain a number of power flow measurements and a Breaker may contain a switch status measurement.

MeasurementValue. Adapted from CIM data models. The current state for a measurement. A state value is an instance of a measurement from a specific source. Measurements can be associated with many state values, each representing a different source for the measurement.

MeasurementValueQuality. Adapted from CIM data models. Measurement quality flags. Bits 0-10 are defined for substation automation in draft IEC 61850 part 7-3. Bits 11-15 are reserved for future expansion by that document. Bits 16-31 are reserved for EMS applications.

MeasurementValueSource. Adapted from CIM data models. MeasurementValueSource describes the alternative sources updating a MeasurementValue. User conventions for how to use the MeasurementValueSource attributes are described in the introduction to IEC 61970-301.

Money. Adapted from CIM data models. Amount of money.

NonConformLoad. Adapted from CIM data models. NonConformLoad represent loads that do not follow a daily load change pattern and changes are not correlated with the daily load change pattern.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

NonConformLoadGroup. Adapted from CIM data models. Loads that do not follow a daily and seasonal load variation pattern.

NonConformLoadSchedule. Adapted from CIM data models. An active power (Y1-axis) and reactive power (Y2-axis) schedule (curves) versus time (X-axis) for non-conforming loads, e.g., large industrial load or power station service (where modeled).

OperationalLimit. Adapted from CIM data models. A value associated with a specific kind of limit. The sub class value attribute shall be positive. The sub class value attribute is inversely proportional to OperationalLimitType.acceptableDuration (acceptableDuration for short). A pair of value_x and acceptableDuration_x are related to each other as follows: if value_1 > value_2 > value_3 >... then acceptableDuration_1 < acceptableDuration_2 < acceptableDuration_3 < ... A value_x with direction='high' shall be greater than a value_y with direction='low'.

OperationalLimitSet. Adapted from CIM data models. A set of limits associated with equipment. Sets of limits might apply to a specific temperature, or season for example. A set of limits may contain different severities of limit levels that would apply to the same equipment. The set may contain limits of different types such as apparent power and current limits or high and low voltage limits that are logically applied together as a set.

OperationalLimitType. Adapted from CIM data models. The operational meaning of a category of limits.

PositionPoint. Adapted from CIM data models. Set of spatial coordinates that determine a point, defined in the coordinate system specified in 'Location.CoordinateSystem'. Use a single position point instance to desribe a point-oriented location. Use a sequence of position points to describe a line-oriented object (physical location of non-point-oriented objects like cables or lines), or area of an object (like a substation or a geographical zone - in this case, have first and last position point with the same values).

PowerSystemResource. Adapted from CIM data models. A power system resource can be an item of equipment such as a switch, an equipment container containing many individual items of equipment such as a substation, or an organisational entity such as sub-control area. Power system resources can have measurements associated.

PU. Adapted from CIM data models. Per Unit - a positive or negative value referred to a defined base. Values typically range from -10 to +10.

Quality61850. Adapted from CIM data models. Quality flags in this class are as defined in IEC 61850, except for estimatorReplaced, which has been included in this class for convenience.

ReactivePower. Adapted from CIM data models. Product of RMS value of the voltage and the RMS value of the quadrature component of the current.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

RegularIntervalSchedule. Adapted from CIM data models. The schedule has time points where the time between them is constant.

RegularTimePoint. Adapted from CIM data models. Time point for a schedule where the time between the consecutive points is constant.

RegulatingControl. Adapted from CIM data models. Specifies a set of equipment that works together to control a power system quantity such as voltage or flow. Remote bus voltage control is possible by specifying the controlled terminal located at some place remote from the controlling equipment. In case multiple equipment, possibly of different types, control same terminal there must be only one RegulatingControl at that terminal. The most specific subtype of RegulatingControl shall be used in case such equipment participate in the control, e.g., TapChangerControl for tap changers. For flow control load sign convention is used, i.e., positive sign means flow out from a TopologicalNode (bus) into the conducting equipment.

RegulationSchedule. Adapted from CIM data models. A pre-established pattern over time for a controlled variable, e.g., busbar voltage.

RemoteInputSignal. Adapted from CIM data models. Supports connection to a terminal associated with a remote bus from which an input signal of a specific type is coming.

Season. Adapted from CIM data models. A specified time period of the year.

SeasonDayTypeSchedule. Adapted from CIM data models. A time schedule covering a 24-hour period, with curve data for a specific type of season and day.

Seconds. Adapted from CIM data models. Time, in seconds.

SetPoint. Adapted from CIM data models. An analog control that issue a set point value.

StringMeasurement. Adapted from CIM data models. StringMeasurement represents a measurement with values of type string.

StringMeasurementValue. Adapted from CIM data models. StringMeasurementValue represents a measurement value of type string.

SubGeographicalRegion. Adapted from CIM data models. A subset of a geographical region of a power system network model.

SubLoadArea. Adapted from CIM data models. The class is the second level in a hierarchical structure for grouping of loads for the purpose of load flow load scaling.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

Substation. Adapted from CIM data models. A collection of equipment for purposes other than generation or utilization, through which electric energy in bulk is passed for the purposes of switching or modifying its characteristics.

Temperature. Adapted from CIM data models. Value of temperature in degrees Celsius.

Voltage. Adapted from CIM data models. Electrical voltage, can be both AC and DC.

VoltageLevel. Adapted from CIM data models. A collection of equipment at one common system voltage forming a switchgear. The equipment typically consists of breakers, busbars, instrumentation, control, regulation and protection devices as well as assemblies of all these.

VoltageLimit. Adapted from CIM data models. Operational limit applied to voltage.

VoltagePerReactivePower. Adapted from CIM data models. Voltage variation with reactive power.

VolumeFlowRate. Adapted from CIM data models. Volume per time.

9.4 GreenEnergy

The GreenEnergy subject addresses the data models related to the generation of green energy

At the moment, the following data models are available:

GreenEnergyGenerator. A generic generator station which can generate energy from green energy

GreenEnergyMeasurement. An instantaneous measure of power generation using green energy sources

PhotovoltaicDevice. The Data Model is intended to describe the mechanical, electrical and thermal characteristics of photo-voltaic panels according to STC - Standard Test Condition and NOCT - Normal Operating Cell Temperature.

PhotovoltaicMeasurement. The Data Model is intended to measure the continuous power transferred by the photo-voltaic panel to an Inverter Device.

9.5 Weather

This subject includes data models useful for dealing with weather data. These data models are initially associated with the agriculture domain but is applicable to many different applications, such as the energy domain.

The following data models are available:

SeaConditions. This entity contains a harmonised geographic description of sea conditions

WeatherAlert. A weather alert generated by a user or device in a given location

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

WeatherForecast. A harmonised description of a Weather Forecast

WeatherObserved. An observation of weather conditions at a certain place and time. This data model has been developed in cooperation with mobile operators and the GSMA.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

10 Appendix C – CGMES TSO/DSO Data Exchange

10.1 Metering data

Metering data is generated by metering data operator (MDO) on behalf of the customer or generator. In many countries the MDO role is by default taken over by DSO, however, also cases exist where other actors take this role. Where DSO carry out their role as MDOs, they have a large experience in data collection, data validation and data storage.

An increased number of market parties will require access to metering data and the DSO/MSO is responsible to make these data available if the customer provide his/her consent.

When the meter class that is part of the CIM standard is exchanged, it is important for the DSOs to understand the metering data operator MDO.

10.1.1Reference data

TSO-DSO exchanges need to rely on common reference data such as lists of countries, companies, base voltages, and processes referred to EU legislation. The reference data should be expanded over time, but here related to metering the following figure gives an indication of needed reference data. Precise values to be defined as reference data will still need to be analysed and defined.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

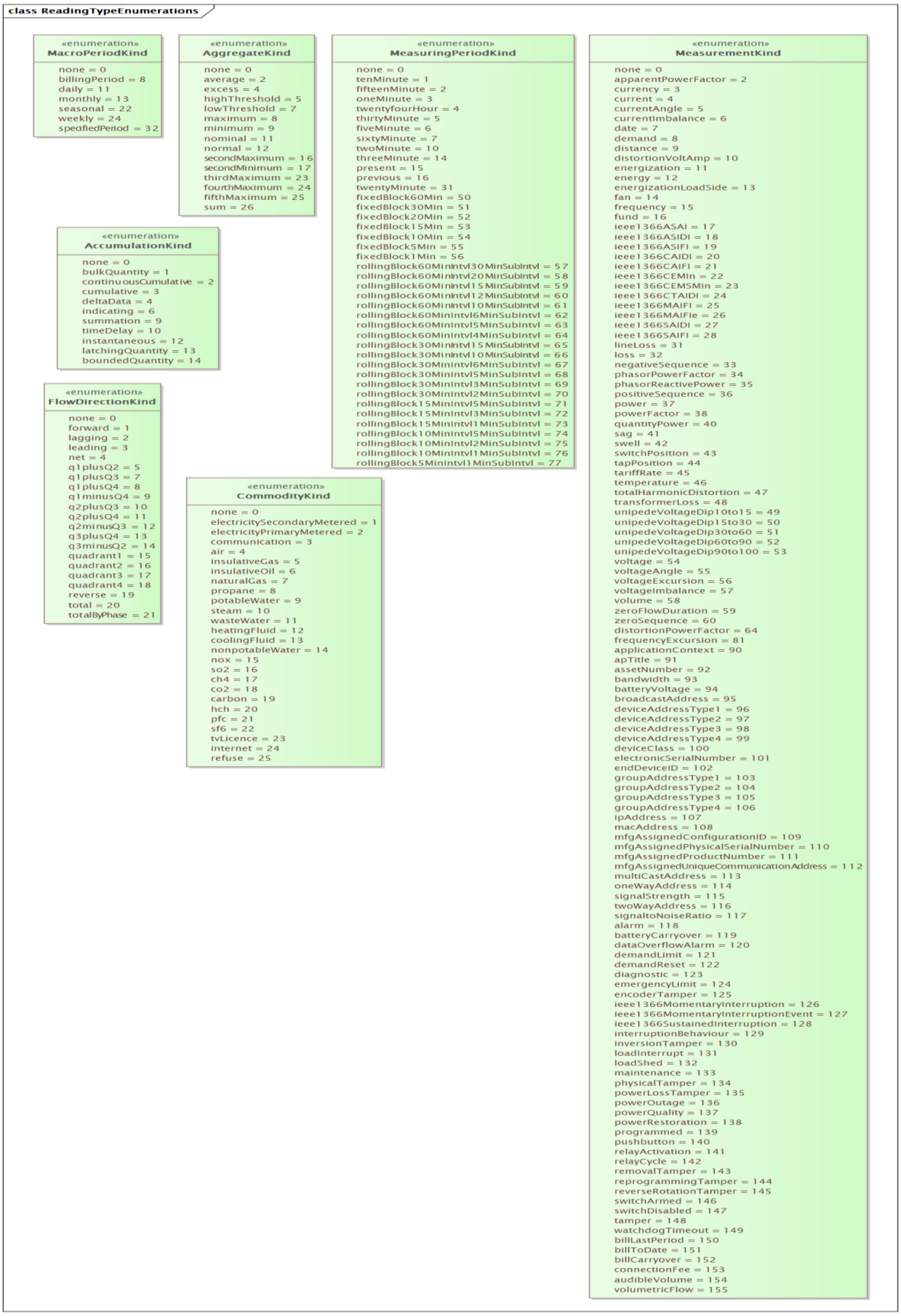

Figure 25: Reading type related enumerations from CIM.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

10.1.2 Boundary information exchange

TSOs and DSOs need to exchange network data and to do so they need to agree on connecting points, i.e., boundary points that define the TSO-DSO border. Related to these points there is a need to define commonly shared information which is agreed bilaterally. This common data is referred as boundary information/data exchange. This calls for a discussion between TSOs and DSOs to clarify this and to see what kind of system will be used to agree, maintain, and exchange such information.

10.2 Application recommendations

10.2.1 General

Considering the requirements in the above section the applications should use the recommendations provided in the following sections which give guidance on how you apply modified header and metadata solutions for TSO-DSO exchanges.

10.2.2 Header

At present stage the document header of CGMES related exchanges contains information about the metadata of the model which is serialised in an instance file. The data exchange between TSO-s use the header and metadata models developed by IEC TC57 working group 13 which is standardised in the IEC 61970-552:2016. However, based on the business need to implement the Common Grid Model building process and the Coordinated Security Analysis process, ENTSO-E concluded that the existing header and metadata described in IEC 61970-552:2016 are not sufficient to meet all the requirements of the TSO data exchange. ENTSO-E has developed a new header and metadata model that contains new attributes and descriptions that will serve the TSO needs [35].

Note that this header is containing old and new attributes to enable backwards compatibility with previous exchanges. For newly designed exchanges, attribute stereotypes with md shall not be used and the header shall rely on W3C DCAT and other used ontologies.

To clarify the specific usage of the attributes for the TSO-DSO data exchange processes, it is recommended to comply the following guidelines

- the information on the metering data operator is exchanged as dcat:creator in the metadata.
- prov:atLocation – it is necessary to be provided to exchange information on the region or on the domain.
- prov:generatedAtTime – it is considered a required attribute for the exchanges.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

- dcterms:conformsTo- it is considered a required attribute. There could be many values for this attribute to point to documents or profile URI or references to constraints to which the data conforms to.
- prov:specializationOf – it is considered a required attribute to reference to modelling authority set version sourcing the model. It is also in line with the DSO needs to identify the version of the modelling authority set which covers the area of responsibility
- in addition the following attributes are considered important for TSO-DSO exchanges: version related attributes, hasXSDDuration, inXSDDateTimeStamp, identifier.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

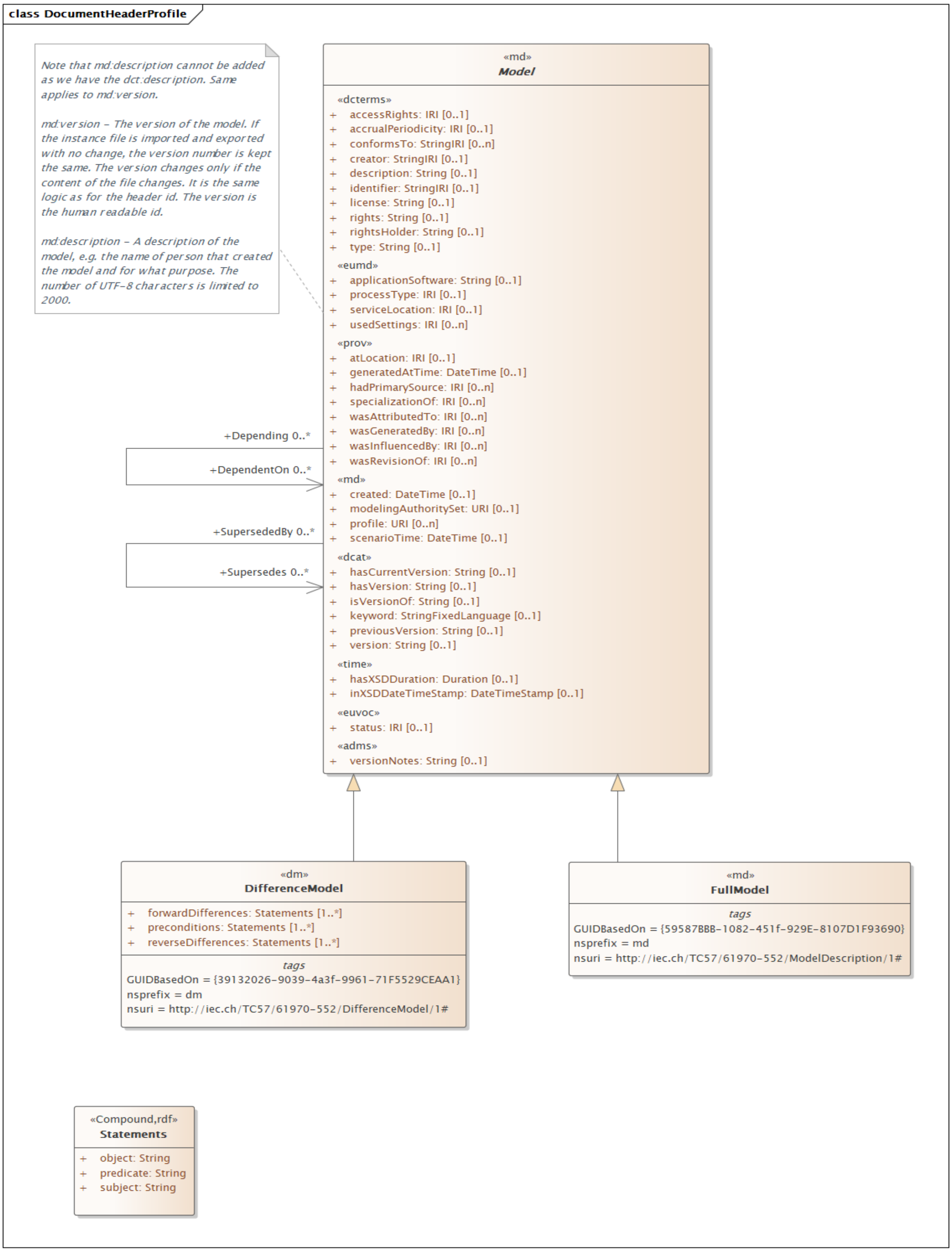

Figure 26: Class diagram DocumentHeaderProfile as part of the Metadata and Document Header Data Exchange Specification.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

Table 5: Attributes details for the TSO-DSO data exchange processes

name	mult	type	description
previousVersion	0..1	String	(dcat) The previous version of a resource in a lineage. This property is meant to be used to specify a version chain, consisting of snapshots of a resource. The notion of version used by this property is limited to versions resulting from revisions occurring to a resource as part of its life-cycle. One of the typical cases here is representing the history of the versions of a dataset that have been released over time.
created	0..1	DateTime	(md) The date and time when the model was created. It is the time of the serialization. The format is an extended format according to the ISO 8601-2005. European exchanges shall refer to UTC, e.g. <md:Model.created>2014-05-15T17:48:31.474Z</md:Model.created>.
generatedAtTime	0..1	DateTime	(prov) Generation is the completion of production of a new entity by an activity. This entity did not exist before generation and becomes available for usage after this generation. [CIM context: The date and time when the model was serialized in the document where the header is located. The format is an extended format according to the ISO 8601-2005. European exchanges shall refer to UTC.].
hasXSDDuration	0..1	Duration	(time) Extent of a temporal entity, expressed using xsd:duration. [CIM context: The duration of the validity period of the model that it is serialized in the document where the header is located. It is only used in relation to the inXSDDateTimeStamp property which indicates the beginning of the validity period of the model. The end of the validity period is derived from both inXSDDateTimeStamp and hasXSDDuration.].
inXSDDateTimeStamp	0..1	DateTimeStamp	(time) Position of an instant, expressed using xsd:dateTimeStamp, in which the time-zone field is mandatory. [CIM context: The date and time that this model represents, i.e. for which the model is (or was) valid. If used in relation with hasXSDDuration it indicates the beginning of the validity period. It is indicating either an instant (in cases where the model is only valid for a point in time) or the start time of a period. If not provided the model is considered valid for any time stamp. The format is an extended format according to the ISO 8601-2005. European exchanges shall refer to UTC.].

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

name	mult	type	description
status	0..1	IRI	(euvoc) Indicates the status of a skos:Concept or a skosxl:Label, or any resource related to controlled vocabulary management. [CIM context: The condition or position of an object with regard to its standing. (Validated, Primary, Backup etc.)].
applicationSoftware	0..1	String	(eumd) Identifies the application software which generated this instance file. The application software term is defined in ISO/IEC/IEEE 24765:2017. The application software can be identified either: - as a string which contains information on the software name and version, e.g. <tool_name>-<major_version>.<minor_version>.<patch>, or - as a reference to a software identification tag as defined by ISO/IEC 19770-2:2015 and ISO/IEC/IEEE 24765:2017.
description	0..1	String	(dcterms) A free-text account of the item.
modelingAuthoritySet	0..1	URI	(md) A URN/URI referring to the organisation role / model authority set reference. The organization role is the source of the model. It is the same for all profiles part of a model exchange.
atLocation	0..1	IRI	(prov) A location can be an identifiable geographic place (ISO 19112), but it can also be a non-geographic place such as a directory, row, or column. As such, there are numerous ways in which location can be expressed, such as by a coordinate, address, landmark, and so forth. [CIM context: Reference to a region or a domain for which this model is provided.].
keyword	0..1	StringFixedLanguage	(dcat) A keyword or tag describing a resource. [CIM context: The intended content type of the model, usually the profile keyword. Used to identify what profiles and content is expected in the document, e.g., Equipment, Boundary, SSH, AE, etc. The same keyword is used for different versions of same profile. It can be also used to identify different content based on the same profile. For instance, as the equipment profile can be used for both boundary data and equipment not related to boundary, the keyword is different to indicate that boundary data is exchanged. In order to avoid ambiguity the property is not exchanged in cases where the document contains multiple profiles referenced by dct:conformsTo.].
scenarioTime	0..1	DateTime	(md) The date and time that this model represents, i.e. for which the model is valid. The format is an extended format according to the ISO 8601-2005. European exchanges shall refer to UTC, e.g. <md:Model.scenarioTime>2030-01-15T17:00:00.000Z</md:Model.scenarioTime>.
accessRights	0..1	IRI	(dcterms) Information about who access the resource or an indication of its security

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

name	mult	type	description
			status.Access Rights may include information regarding access or restrictions based on privacy, security, or other policies. [CIM context: Reference to the confidentiality level that shall be applied when handling this model.].
profile	0..n	URI	(md) URN/URI describing the profiles that governs this model. It uniquely identifies the profiles and its version, e.g. http://iec.ch/TC57/61970-456/SteadyStateHypothesis/2/0.
wasInfluencedBy	0..n	IRI	(prov) Influence is the capacity of an entity, activity, or agent to have an effect on the character, development, or behavior of another by means of usage, start, end, generation, invalidation, communication, derivation, attribution, association, or delegation. [CIM context: A reference to the model on which the model serialised in this document depends on. The references are maintained by the producer of the model. Minimum requirements for the dependency are specified and can be restricted within a business process as long as they do not contradict requirements by standards. For instance, IEC 61970-600-1 defines minimum requirements for the profiles defined in that standard.].
conformsTo	0..n	StringIRI	(dcterms) An established standard to which the described resource conforms. [CIM context: An IRI describing the profile that governs this model. It uniquely identifies the profile and its version. Multiple instances of the property describe all standards or specifications to which the model and the document representing this model conform to. A document would normally conform to profile definitions, the constraints that relate to the profile and/or the set of business specific constrains. A reference to a machine- readable constraints or specification indicates that the document was tested against these constraints and it conforms to them.].
hadPrimarySource	0..n	IRI	(prov) A primary source for a topic refers to something produced by some agent with direct experience and knowledge about the topic, at the time of the topic's study, without benefit from hindsight. Because of the directness of primary sources, they 'speak for themselves' in ways that cannot be captured through the filter of secondary sources. As such, it is important for secondary sources to reference those primary sources from which they were derived, so that their reliability can be investigated. A primary source relation is a particular case of derivation of secondary materials from their primary sources. It is recognized that the determination of primary sources can be up to interpretation, and should be done according to

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

name	mult	type	description
			conventions accepted within the application's domain. [CIM context: Reference to a modelling authority set version sourcing the model. It is only used in cases where a model is modified by an agent which has different version of modelling authority set. The agent that makes a revision of a model indicates the primary source using this property and also refers to its own version of modelling authority set using prov:specializationOf.].
version	0..1	String	(dcat) The version number of a resource.
identifier	0..1	StringIRI	(dcterms) An unambiguous reference to the resource within a given context. Recommended practice is to identify the resource by means of a string conforming to an identification system. Examples include International Standard Book Number (ISBN), Digital Object Identifier (DOI), and Uniform Resource Name (URN). Persistent identifiers should be provided as HTTP URIs. [CIM context: A unique identifier of the model which is serialised in the document where the header is located. The identifier is persistent for a given version of the model and shall change when the model changes. If a model is serialized as complete (full) model or as difference model exchange the identifier shall be the same. The identifier shall not be used as an identifier of the document which can be different for a given version of a model.].
wasGeneratedBy	0..n	IRI	(prov) Generation is the completion of production of a new entity by an activity. This entity did not exist before generation and becomes available for usage after this generation. [CIM context: Reference to an activity or the exact business nature (process, configuration) which produced or uses the model.].
license	0..1	String	(dcterms) A legal document giving official permission to do something with the resource. Recommended practice is to identify the license document with a URI. If this is not possible or feasible, a literal value that identifies the license may be provided. [CIM context: Reference to the license under which the data is made available. If no license holder is defined, then the original data provider holds the license.].
wasAttributedTo	0..n	IRI	(prov) Attribution is the ascribing of an entity to an agent. [CIM context: Reference to the agent (or service provider) from which the model originates.].
usedSettings	0..n	IRI	(eumd) Reference to a set of parameters describing used settings (e.g. power flow settings, process

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

name	mult	type	description
			settings, etc.) applied to the model prior its serialisation.
wasRevisionOf	0..n	IRI	(prov) A revision is a derivation for which the resulting entity is a revised version of some original. The implication here is that the resulting entity contains substantial content from the original. Revision is a particular case of derivation. [CIM context: When a model is updated the resulting model supersedes the models that were used as basis for the update. Hence this is a reference to the model which are superseded by this model. A model can supersede 1 or more models, e.g. a difference model or a full model supersede multiple models (difference or full). In this case, multiple properties are included in the header. The referenced document(s) is (are) identified by the URN/MRID/UUID in the FullModel rdf:about attribute when full model(s) is (are) referenced and by the URN/MRID/UUID in the DifferenceModel rdf:about attribute when difference model(s) is (are) referenced.].
rights	0..1	String	(dcterms) A statement that concerns all rights not addressed with dct:license or dct:accessRights, such as copyright statements.
rightsHolder	0..1	String	(dcterms) Information about rights held in and over the resource. Typically, rights information includes a statement about various property rights associated with the resource, including intellectual property rights. Recommended practice is to refer to a rights statement with a URI. If this is not possible or feasible, a literal value (name, label, or short text) may be provided.
type	0..1	String	(dcterms) The nature or genre of the resource. Recommended practice is to use a controlled vocabulary such as the DCMI Type Vocabulary [DCMI-TYPE]. To describe the file format, physical medium, or dimensions of the resource, use the property Format.
processType	0..1	IRI	(eumd) The exact business nature. Reference to Business Process configurations.
accrualPeriodicity	0..1	IRI	(dcterms) The frequency with which items are added to a collection. [CIM context: Reference to the time frame.].
specializationOf	0..n	IRI	(prov) An entity that is a specialization of another shares all aspects of the latter, and additionally presents more specific aspects of the same thing as the latter. In particular, the lifetime of the entity being specialized contains that of any specialization. Examples of aspects include a time period, an abstraction, and a context associated with the entity. [CIM context: Reference to modelling authority set version sourcing the model. The agent that makes a revision

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

name	mult	type	description
			of a model indicates the primary source using prov:hadPrimarySource and refers to its own version of modelling authority set using this property.].
creator	0..1	StringIRI	(dcterms) An entity responsible for making the resource. Recommended practice is to identify the creator with a URI. If this is not possible or feasible, a literal value that identifies the creator may be provided. [CIM context: The name of the agent (Modeling Authority) from which the model originates].
serviceLocation	0..1	IRI	(eumd) Reference to a service location (region or a domain).
versionNotes	0..1	String	(adms) A description of changes between this version and the previous version of the resource.
hasVersion	0..1	String	(dcat) This resource has a more specific, versioned resource. This property is intended for relating a non-versioned or abstract resource to several versioned resources, e.g. snapshots [PAV]. The notion of version used by this property is limited to versions resulting from revisions occurring to a resource as part of its life-cycle. Therefore, its semantics is more specific than its super-property dcterms:hasVersion, which makes use of a broader notion of version, including editions and adaptations.
isVersionOf	0..1	String	(dcat) This resource is a version of a non-versioned or abstract resource. This property is intended for relating a versioned resource to a non-versioned or abstract resource. The notion of version used by this property is limited to versions resulting from revisions occurring to a resource as part of its life-cycle. Therefore, its semantics is more specific than its super-property dcterms:isVersionOf, which makes use of a broader notion of version, including editions and adaptations.
hasCurrentVersion	0..1	String	(dcat) This resource has a more specific, versioned resource with equivalent content. This property is intended for relating a non-versioned or abstract resource to a single snapshot that can be used as a permalink to indicate the current version of the content. The notion of version used by this property is limited to versions resulting from revisions occurring to a resource as part of its life-cycle.
description	0..1	String	(md) A description of the model, e.g. the name of person that created the model and for what purpose. The number of UTF-8 characters is limited to 2000.
version	0..1	String	(md) The version of the model. If the instance file is imported and exported with no change, the version number is kept the same. The version changes only if the content of the file changes. It is the same logic

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739

name	mult	type	description
			as for the header id. The version is the human readable id. [CIM context: It relates to the version of the document and not the version of the model which is serialized.].

10.3 Metadata layer

To link between reference data and regular CIM based model exchange, a metadata layer is introduced. In the metadata layer cim:IdentifiedObject will inherit from skos:Concept which will allow to instantiate a subclass of cim:IdentifiedObject with any of the skos concept related mapping properties. For instance cim:BaseVoltage with nominal voltage of 225 kV can have a property

<skos:exactMatch rdf:resource="http://data.europa.eu/energy/baseVoltage/225kV"/>

This technique will allow that current CIM based data exchanges are referring to objects serialised in the metadata layer which will ensure the completes of the RDF graph and in addition it enables validation of consistency between data serialised in the metadata layer and reference data published in a central location.
Another alternative which is more in line with linked data is that in the metadata layer objects can have multiple types i.e. have multiple inheritance and be identified with the same permanent URI which will create the linkage.

10.3.1Boundary information exchange

IEC 61970-600-1:2021 and IEC 61970-600-2:2021 (CGMES v3.0) include the profile “Equipment boundary” which is dedicated for an exchange of boundary information. However, it should be noted that this profile is included as deprecated profile, meaning that it still be used but it could be deleted in future standards. In order to support an exchange of boundary information CGMES v3.0 includes all necessary boundary information in the Core Equipment profile (EQ). It is recommended that new implementation of boundary information related profiles is based on Core Equipment profile and not on the Equipment Boundary profile. It is expected that new applications will be designed and build to support TSO-DSO data exchanges and it is important that these applications are as much as possible future proof.
In addition, it is expected that TSO-DSO data exchanges will often use a setup in which BoundaryPoint is defined in a Substation rather than in a Line. It was intended that CGMES v3.0 is supporting such setup. However, further gaps in the modelling were identified. The development of the solution is in progress and it is planned to release a fix soon.

10.3.2 Reference data

It is recommended that W3C [36] SKOS and DCAT 3 [37] are used to describe the reference data. For instance, each set of reference data contains skos:ConceptScheme which defines the group of reference data. Due to tooling limitations it is advised to use one skos:ConceptScheme per category of data. Then each value for the reference data is modelled as skos:Concept. An instance of both skos:ConceptScheme and skos:Concept are shown below as an example.

```
<skos:ConceptScheme rdf:about="http://data.europa.eu/energy/energyProductType">
 <owl:versionInfo>20210616-0</owl:versionInfo>
   <rdfs:label xml:lang="en">Energy Product Type</rdfs:label>
   <at:prefLabel xml:lang="en">Energy Product Type</at:prefLabel>
   <skos:prefLabel xml:lang="en">Energy Product Type</skos:prefLabel>
   <dcterms:description xml:lang="en">The identification of the nature of an energy product such as power, energy, reactive power, etc.</dcterms:description>
  </skos:ConceptScheme>
  <skos:Concept rdf:about="http://data.europa.eu/energy/baseVoltage/225kV">
   <eu:IdentifiedObject.shortName>AC-225</eu:IdentifiedObject.shortName>
   <cim:IdentifiedObject.mRID>63893f24-5b4e-407c-9a1e-4ff71121f33c</cim:IdentifiedObject.mRID>
   <skos:topConceptOf rdf:resource="http://data.europa.eu/energy/baseVoltage"/>
   <dcterms:identifier>AC-225</dcterms:identifier>
   <skos:prefLabel>AC-225</skos:prefLabel>
   <cim:IdentifiedObject.name>AC-225</cim:IdentifiedObject.name>
   <cim:IdentifiedObject.description>Base voltage defined by ENTSO-E 225 kV</cim:IdentifiedObject.description>
   <skos:inScheme rdf:resource="http://data.europa.eu/energy/baseVoltage"/>
   <cim:BaseVoltage.nominalVoltage rdf:datatype="http://www.w3.org/2001/XMLSchema#float"
   >225</cim:BaseVoltage.nominalVoltage>
  </skos:Concept>
```

This approach should be followed for “Reading quality type” and “Reading type” as these are widely used enumerated values and it is recommended to have them defined as reference data. An example for reading type

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 957739

with ID “0.0.7.6.1.1.12.0.0.0.0.0.0.0.0.3.72.0” is presented below. The data is not real one and there should be an effort to collect all IDs and related data for the reading types and reading quality types in order to populate/define the complete set of reference data. This will enable receiving system to access more detailed information and not only have access to the ID without knowing units, directions, etc. After the reference data is created there should be an instruction on how to refer to the readingType from the messages that are exchanged, e.g.

“http://data.europa.eu/energy/readingType/0.0.7.6.1.1.12.0.0.0.0.0.0.0.0.3.72.0” or other way.


```
<skos:ConceptScheme rdf:about="http://data.europa.eu/energy/readingType">
 <owl:versionInfo>20210616-0</owl:versionInfo>
   <rdfs:label xml:lang="en">Reading Type</rdfs:label>
   <at:prefLabel xml:lang="en">Reading Type</at:prefLabel>
   <skos:prefLabel xml:lang="en">Reading Type</skos:prefLabel>
   <dcterms:description xml:lang="en"> This is the concept scheme for the Reading Type.</dcterms:description>
  </skos:ConceptScheme>
  <skos:Concept rdf:about="http://data.europa.eu/energy/readingType/0.0.7.6.1.1.12.0.0.0.0.0.0.0.0.3.72.0">
   <eu:IdentifiedObject.shortName>XXX</eu:IdentifiedObject.shortName>
   <cim:IdentifiedObject.mRID>0.0.7.6.1.1.12.0.0.0.0.0.0.0.0.3.72.0</cim:IdentifiedObject.mRID>
   <skos:topConceptOf rdf:resource="http://data.europa.eu/energy/readingType"/>
   <dcterms:identifier>0.0.7.6.1.1.12.0.0.0.0.0.0.0.0.3.72.0</dcterms:identifier>
   <skos:prefLabel>XXX</skos:prefLabel>
   <cim:IdentifiedObject.name>XXX</cim:IdentifiedObject.name>
   <cim:IdentifiedObject.description>Reading type xxx</cim:IdentifiedObject.description>
   <skos:inScheme rdf:resource="http://data.europa.eu/energy/readingType"/>
  <cim:aggregate>none</cim:aggregate>
  <cim:measuringPeriod>PT1H</cim:measuringPeriod>
  <cim:accumulation>indicating</cim:accumulation>
  <cim:flowDirection>A02</cim:flowDirection>
```


This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 957739

```
  <cim:commodity>electricity</cim:commodity>

  <cim:measurementKind>energy</cim:measurementKind>

  <cim:multiplier>K</cim:multiplier>

  <cim:measurementUnit>WH</cim:measurementUnit>

  <cim:energyProduct>8716867000030</cim:energyProduct>

 </skos:Concept>
```

This paper reflects only the author's view and the Innovation and Networks Executive Agency (INEA) is not responsible for any use that may be made of the information it contains.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 957739