Grant Agreement N° 872592

PLATOON

Digital platform and analytic tools for energy

Deliverable D1.2

Report on requirements for open, secure and flexible communication and coordination in energy value chain

Contractual delivery date:
M06

Actual delivery date:
30 June 2020

Responsible partner:
P17: INDRA-Minsait, Spain

Project Title	PLATOON – Digital platform and analytic tools for energy
Deliverable number	D1.2
Deliverable title	Report on requirements for open, secure and flexible communication and coordination in energy value chain
Author(s):	ENGIE, TECN, UBO, IAIS, ENG, CEPV, GIR, SAM, TIB, ROM, CS, UDGA, IND
Responsible Partner:	P17 – INDRA-Minsait
Date:	30.06.2020
Nature	R
Distribution level (CO, PU):	PU

Work package number	WP1 – Energy Management System Challenges
Work package leader	ENGIE, France
Abstract:	This document aims to identify potential issues that need to be addressed with the aim to make data exchange as easy as possible, for the energy sector and across sectors as a way to facilitate the development of new data-driven services. To meet this objective, this document covers an extensive number of topics and themes. The work was divided into two phases, a first phase of analysis, where the different partners that collaborated in the writing of the document worked individually preparing the different contents and a second phase of conception and elaboration of joint user stories, to give shape to the requirements expressed in this deliverable
Keyword List:	Data Governance and Sovereignty, Data Exchange, Energy management, Software Architectures, Big data, Real time data analytics, Energy systems, smart energy, smart grids, wireless energy transfer, Electricity Transmission/Distribution

The research leading to these results has received funding from the European Community's Horizon 2020 Work Programme (H2020) under grant agreement no 872592.

Editor(s):	P17 – INDRA-Minsait
Contributor(s):	ENGIE, TECN, UBO, IAIS, ENG, CEPV, GIR, SAM, TIB, ROM, CS, UDGA, IND
Reviewer(s):	Philippe Calvez (ENGIE) – Platoon Coordinator Erik Maqueda (TECN) – Technical Coordinator
Approved by:	Philippe Calvez (ENGIE) – Platoon Coordinator Erik Maqueda (TECN) – Technical Coordinator Eduardo Jimenez (IND) – Exploitation Coordinator
Recommended/mandatory readers:	Mandatory WP2-WP6 leaders and task leaders. Recommended the rest of the partners

Document Description

Document Revision History

Version	Date	Modifications Introduced	
		Modification Reason	Modified by
0.10	15-2-2020	Index proposal	IND
0.50	30-4-2020	Analysis contributions	IND
0.75	31-5-2020	User stories	IND
1	14-6-2020	Ready for internal review	IND
2	24-6-2020	Internal review by TECN	Erik Maqueda
3	29-6-2020	Review ENGIE	Philippe Calvez
4	30-6-2020	Final review	TECN-ENGIE

Table of Contents

List of Figures

Terms and abbreviations

CA	Consortium Agreement
CO	Confidential
DM	Dissemination Manager
DoA	Declaration of Action
EC	European Commission
EM	Exploitation Manager
GA	Grant Agreement
GAM	General Assembly Meeting
PM	Project Manager
PU	Public
QA	Quality Assurance
RE	Restricted
SC	Steering Committee
TM	Technical Manager
WP	Work package
WPL	Work package Leader
AE	Alarm and Event management
AI	Artificial Intelligence
AMI	Advanced Metering Infrastructure
API	Application Programming Interface
ASHRAE	American Society of Heating, Refrigerating and Air-Conditioning Engineers

BMS	Battery Management System
CAPEX	CAPital Expenditure
CBM	Condition Based Maintenance
CEM	Customer Energy Manager
CEN	Comité Européen Normalisation
CENELEC	European Committee for Electrotechnical sTandardization
CEPRI	Chinese Electrical Power Research Institute
CIM	Common Information Model
COSMAG	Comprehensive Architecture for Smart Grid
CSP	Concentrated Solar Power
DAPP	Decentralized APPlications
DCMI	Dublin Core Metadata Initiative
DER	Distributed Energy Resources
DM	Device and network Management
DS	Data Sharing
DSO	Distribution System Operator
EC	European Commission
EMCS	Energy Management and Control Systems
ESCO	Energy Service Companies
ETSI	European Telecommunications Standard Instittute
EV	Electric Vehicle
FAIR	Findable Accessible Interoperable, Re-usable
GDPR	General Data Protection Regulation
HEM	Home Energy Management
HVAC	Heating Ventilating and Air-conditioning Control
ICT	Information and Communication Technologies
IDS	Industrial Data Space
IEA	International Energy Agency
IEC	International Electrotechnical Commission
IEEE	Institute of Electrical and Electronic Engineers
IIC	Industrial Internet Consortium
IoT	Internet of Things
IRPWind	Integrated Research Programme on Wind Energy
ISGAN	International Smart Grid Action Network
ISO	International Organization for Standardization
KSGA	Korean Smart Grid Association
LCOE	Leverage Cost Of Energy
LV	Low Voltage
MCM	Market-based Coordination Mechanism
MIM	Minimal Interoperability Mechanism
ML	Machine Learning
NASA	National Aeronautical Space Administration
NIST	US National Institute of Standards and Technology
O&M	Operation & Maintenance
OA	Open Access

OD	Open Data
OEM	Original Equipment Manufacturer
OPEX	OPErating Expenditure
OWL	Web Ontology Language
PAS	Publicly Available Standards
PET	Privacy Enhancing Technologies
PHEV	Plug-in Hybrid Electric Vehicle
PKI	Public Key Infrastructure
PV	Photo Voltaic
R&D	Research & Development
RDF	Resource Description Framework
RES	Renewable Energy Source
SaaS	Software as a Service
SAREF	Smart Appliances REFerence ontology
SCHED	Schedule
SEC	Smart Energy Collective
SGAM	Smart Grid Architecture Model
SGD	Stochastic Gradient Descent
SME	Small and Medium Enterprise
SMPC	Secure Multi Party Computation
TCP	Transmission Control Protocol
TSO	Transmission System Operator
USEF	Universal Smart Energy Framework
V2G	Vehicle To Grid
W3C	World Wide Web Consortium
WEF	World Economic Forum
WT	Wind Turbine
XMPP	XML Messaging and Presence Protocol

Executive Summary

This document aims to identify potential issues and constraints that need to be addressed with the aim to make data exchange as easy as possible, for the energy sector and across sectors as a way to facilitate the development of new data-driven services.

To meet this objective, this document covers an extensive number of topics and themes, structured in several parts and explained below.

The work was divided into two phases, a first phase of analysis, where the different partners that collaborated in the writing of the document worked individually preparing the different contents and a second phase of conception and elaboration of joint user stories, to give shape to the requirements expressed in this deliverable.

In the analysis phase, the development of the contents was also divided into several topics.

The first one deals with the exchange of data within the electricity market in its current configuration, where new players have an increasing presence. The first analysis is carried out precisely to identify these actors, both the traditional ones and those of new incorporation.

The identification of these actors will be very useful later on, also for the construction of user stories, which are part of the table of requirements.

In general, in this first topic we deal with the Analysis of possible data exchanges, differential characteristics, semantics and protocols of communication.

We will repeat the following remark later, but note that this document is mainly concerned with "horizontal" data exchange, i.e. between different electricity market platforms and between these and platforms belonging to other sectors. Hence, the next section of the document deals with the identification of these other actors and platforms, with whom it makes more sense to talk about data exchange.

When carrying out the research corresponding to several of these topics, we find numerous European projects and even meta-projects such as Bridge (Cooperation group of Smart Grid, Energy Storage, Islands and Digitalisation H2020 projects)[i] where the vertical analysis of data exchange, reference architectures, standards and applicable legislation for the central actors of the TSO-DSO electricity market are treated with great profusion. Hence, we consider an orientation of Platoon and this particular deliverable for horizontal data exchange to be the right choice.

However, although we deal mainly with "horizontal" exchange, we also review the types of data exchange of a "vertical" nature and the Architectures, legacy formats, interfaces, and operating systems of the energy system, due to their possible implications in this exchange between platforms and between economic and productive sectors and also due to the fact that some of these issues are referred to on numerous occasions in Platoon's Grant Agreement[ii].

We also devote a section to the main regulatory standards to be met.

The second block of topics discussed, was the analysis of the most relevant European and International initiatives, related to the interoperability and standardization: Smart Grid Architecture Model (SGAM), standardization, Universal Smart Energy Framework (USEF)...(which are also references and topics reflected in Platoon's Grant Agreement) together with examples of current and past relevant projects in the field of energy and IoT, in which several of the consortium partners have had direct participation and which provide us with conclusions and lessons learned.

In the third block of topics discussed in the analysis phase, once the actors have been identified and how they currently exchange data and covered the references in terms of standardization and interoperability, we moved on to analyze the state of the art in a broad sense.

We started by addressing key concepts in the digital data value chain and its applications, Platform and API economy as data exchange makes sense also as an intangible asset, which can be used as a fundamental ingredient or basis for new business models

In the following sections, the key concepts that allow to overcome the barriers for data exchanges, data sovereignty, the mechanisms of establishment of trust, in their different variants.

We would also like to highlight other important contents and concepts such as those referring to data privacy, data security, data access rights and access to heterogeneous data and data Governance, which will lead to the extraction of most of the requirements of this deliverable.

In the fourth and last block of topics, we grouped the analysis of techniques and paradigms that can be part of the technological stack of data exchanges, related again to the field of IoT, Blockchain and advanced techniques that allow to overcome the problems in terms of generating large volumes of data for use with artificial intelligence techniques, but respecting the necessary privacy.

In the second phase we worked together on the development of requirements, using the technique of user stories, due to in Platoon has been established, for technical WPs, that they will follow an agile methodology.

Therefore, this deliverable contains the mandatory requirements for data exchange and has been divided into five groups of requirements, which do refer to specific parts of the analysis performed.[1]

These five groups of requirements are those relating to:

- Data governance applied to an ecosystem
- Data privacy, data security, data access rights and access to heterogeneous data
- Data exchanges analysis within electricity market and with other actors and platforms
- Data needs for Artificial Intelligence applications
- Technological stack of data exchanges

In the final structure of the document, in order to facilitate its reading and understanding, it was decided to move to annexes, those topics, for which NO user stories had been chosen that entailed mandatory requirements and that could be considered specific solutions/technologies that can be used in the following WPs to meet these requirements.

So, we expect than different solutions/technologies will be used according to the analyses to be carried out in the different WPs, following the general rule, that the user stories prescribe what you want to obtain and the development teams decide the "how and with which technology”.

[1] Only in the final revision, one of the requirements has been kept, but as "optional".

1 Introduction

The concept of data-driven innovation considers the availability of data for the fields of Artificial Intelligence and the Internet of Things to be of critical importance and have made data access and sharing more crucial than ever. According to the report of the OECD "Enhancing access to and sharing of data"[iii], despite a growing need for data and evidence of the economic and social benefits, data access and sharing has not achieved its potential. Individuals, businesses, and governments often face barriers to data access, which may be compounded by a reluctance to share.

To facilitate, encourage and enhance data access and sharing for the benefit of all, the following three major challenges need to be addressed, according to the OECD report:

1. Balancing the benefits of enhancing data "openness" with the risks, while considering legitimate private, national and public interests.
2. Reinforcing trust and empowering users through pro-active stakeholder engagement and community building.
3. Encouraging the provision of data through coherent incentive mechanisms and sustainable business models while acknowledging the limitations of (data) markets.

Europe has an excellent opportunity to define standards and to create platforms with global reach in the B2B area as it combines excellent industrial processes and IT know-how. The "Digitising European Industry Initiative" of the European Commission is a key element of the Digital Single Market strategy and heavily supports and guides this approach.[iv] Equally the European Commission has recently published the EU Data Strategy[v] which aims to make the EU a leader in a data-driven society creating a single market for data will allow it to flow freely within the EU and across sectors for the benefit of businesses, researchers and public administrations.

PLATOON project should acknowledge these challenges as a first set of high-level requirements and take them into account when defining the specifications and architectures of the platforms to be developed.

2 Part 1: Electricity market data exchange analysis

In this part we deal with the analysis of possible data exchanges, starting with the identification of the different actors in a modern energy market and adjacent sectors.

2.1 Structure of the market and actor

2.1.1 Actors in a modern electricity market

The energy sector is on the brink of a large-scale disruption of business models and supply chains. The traditional roles in energy production, distribution and consumption are changing and new technologies and market entrants are rapidly creating new dynamics. While the traditional business model was based on a utility-centric approach supported by a network of electric equipment manufacturers and associated service providers, newcomers are appearing on stage with growing strength.

The World Economic Forum ("The Future of Electricity: New Technologies Transforming the Grid Edge", WEF) highlights three trends in particular that are converging to produce game-changing disruptions:

- Electrification of large sectors of the economy such as transport and heating, as a key factor for long-term carbon reduction goals, through an increasingly relevant share of renewable energy.
- Decentralization, boosted by the sharp decrease in costs of distributed energy resources like Renewable Energy Sources (RES), distributed storage, demand flexibility and energy efficiency.
- Digitalization of both the grid (smart metering, sensors, automation, etc.), and beyond the meter, with the advent of the Internet of Things (IoT) and a surge of power-consuming connected devices.

EU policies, such as the Clean Energy for All Europeans package ("Winter Package") released in November 2016, set some initial policy responses to these changes, by means of encouraging the development of decentralized electricity generation, in which integration of renewables, energy storage, electric vehicles and flexible demand response are expected to play a significant role.

Key elements and stakeholders of the new energy scenario, most of them represented by partners in PLATOON, are described below:

- **Renewable power plants:** Renewable energy plants are very relevant sources of data from the energy transition point of view. Access to data related to energy production, availability of the plants and efficiency rates of the equipment (wind turbines, solar panels, balance of plant) can be very useful to analyse the performance of the systems under different operating conditions and to obtain relevant conclusions that could improve the sustainable generation of renewable energy. The developers of the renewable power plants are the owners of this information, so they constitute one of the key stakeholders group to be considered in terms of the data exchanges they require and they might offer: windfarm developers, investors in PV solar plants, operators of CSP (Concentrated Solar Power) plants.
- **Distribution System Operators (DSO)**: the increase of RES combined with changing consumption patterns places significant challenges on the traditional business model of DSOs. Providing a set of services adapted to the needs of the different agents connected and/or involved in the grids seems to be their best choice for the future. New industry partnerships are being formed, as large incumbent organizations recognize that they need access to more digital skills in their workforce.
- **Aggregators and energy service providers:** As a result of all these previous mentioned changes, the way of doing business and getting revenues in the electric market is changing dramatically. The power and decision-making capacity, which relied so far on the DSOs, is now

starting to split into other actors, thus forcing the market to evolve from a single-buyer model to a multi-agent model, with new stakeholders such as aggregators and energy service providers taking increasing roles.

- **Energy "prosumers"**: energy consumers are increasingly taking an active role in the energy system as "energy prosumers". Prosumers are households, SMEs or communities that, in addition to having the choice of their electricity and gas suppliers in fully liberalised markets, are also producing energy themselves and could, if energy market design is adapted accordingly, eventually become important participants in the energy via, for example, collective self-consumption of solar PV.
- **Building-related energy consumption**: changes are strongly driven by the European building legislation, including the "nearly zero buildings" approach (buildings that combine high levels of energy efficiency with high shares of on-site renewable energy). Also, building-level electricity storage technologies start to approach economic viability, allowing the share of self-supply to increase even further.
- **ICT (Information and Communication technologies) companies**: Increasingly, players from the ICT sector are also entering the sustainable energy arena in Europe. As in other business fields, the next years will be decisive to set which platforms will be key interfaces for supply and demand of energy-related data and services, as well as to confirm how the new business models will work. It is crucial that European SMEs are well equipped for this new competition, since it will allow them not only to defend their current position in national and local supply chains, but also to profit from new international markets.

A shift is coming in the energy industry from a focus on hardware to the increased importance of software to make systems more efficient, resilient, and digital. Digital technologies are including, among others, big data, analytics and machine learning, blockchain and cloud computing. These technologies can help overcome some of the key challenges in the energy sector: intermittency, aging grids, balancing distribution-connected generation, managing consumer self-generation, and coping with increasing system complexity. Besides, they are supporting the new panorama of the service-oriented energy responding to new expectations of customers, new players and to the new roles for existing players.

Technology and innovation are transforming traditional business models and creating opportunities for new products and services over an increasingly decentralized and digitized electricity grid at all levels, from generation to "beyond the meter". Therefore, new business models around the digitalization of the electricity industry will emerge for European companies if efficient and trustworthy data platforms and data exchange mechanisms are available.

2.1.2 Other actors and platforms

The cross-fertilization and integration of data and know-how from stakeholders which are active on different value chains, not represented or actively engaged in PLATOON, can enable the production of innovative solutions, beyond the current state-of-the-art.

Besides the specific actors of the electricity market and the platforms that will be defined and developed in PLATOON, other actors and market segments should be considered for a cross-sectoral approach. The interconnection of the electricity market data with data sources and platforms from other value chains and applications will increase business opportunities and enhance the exchange of data that is transforming the electricity system and creating a new and broader "Smart Grids" value chain. Some of them are especially highlighted:

- **Offshore wind turbine components condition monitoring:** The cost of wind power, and specifically of the power generated in offshore windfarms, has dropped dramatically, to a great

extent as a result of new auction systems which have forced developers and investors in windfarms to reduce costs both at investment (CAPEX) and operation (OPEX) levels. At the same time, sites for offshore windfarms are being built in more complex and "harsh" environments (deeper waters, longer distances from the coast), with highly demanding requirements in terms of quality, integrity, efficiency and reliability. In this market context, digitalization of offshore wind turbine components and systems has been identified as a main source of relevant competitive advantages. But most European components suppliers and engineering firms cannot access, manage, analyze and learn from the performance data produced by the wind turbines in real-life operation. To overcome this "data access gap" a digital platform with a sustainable business case would be required, as a marketplace where data owners (windfarm owners and wind turbine OEMs) would share relevant data from wind turbines and their systems and components and make them accessible for data users along the wind energy value chain (component suppliers, engineering firms, ICT companies). The digital platform should be designed to mitigate the "openness" risks, generate trust, encourage the provision of data through coherent revenue mechanisms and empower data users, while reducing uncertainties about "data ownership" by acknowledging and ensuring data privacy and security rights.

- **Electromobility**: mass deployment of Electric Vehicles (EVs) as new electricity consumers will have a significant impact on energy production and consumption patterns and is expected to increase interest in self-supply from own renewable electricity ("vehicle-to-building" as well as "vehicle-to-grid"). Battery-based energy storage from EVs is a powerful resource capable of optimising grid management and operation, increasing grid resilience and supporting the deployment of intermittent renewable energy sources. It is clear that smarter cars will have an impact in the energy ecosystem. But at the same time, cars are becoming data platforms with self-diagnostics, satellite navigation, entertainment systems, and links to critical infrastructure and traffic systems to reduce city congestion. As a consequence, the interconnection of data platforms from different actors, such as charging services managers (EV battery state of charge, grid load, charging timings and patterns), automotive industry (Battery Management System supervision, status and performance of power electronics devices, information about other related systems) and Smart cities applications (traffic supervision and management, mobility patterns, electricity flows) should be considered as a key source of synergies and new business opportunities.
- **Decentralised Energy systems**: generating electricity from multiple small energy sources, instead of centralized production present larger possibilities of RES integration makes energy available on local bases (Saving transportation and transformation losses). This power production model facilitate business for local ESCOs and entitle them for new business models as Demand Response, using storage (active and passive storage) as a resource for profits

2.2 Types of exchanges

2.2.1 Edge vs cloud exchanges

Cloud computing has introduced "unlimited computing power", able to respond to large data volumes ingested and process it with scalable computing power on demand.

Edge Computing potentially enables faster processing time/lower delays (with real-time guarantees), higher privacy control (confidential information is processed locally and not sent), reduced network use costs and impacts on energy consumption.

Fog computing is regarded as an evolution of Cloud computing paradigm, where computing capabilities are distributed along the edge-cloud continuum, and where the advantages of cloud technologies (e.g., containerization, virtualization, orchestration, manageability, availability) are retained.

Fog computing is a novel computing paradigm appeared recently that provides computation power in between the system and the Cloud. It allows the means for managing efficiently the increasing number of complex and dispersed data sources associated to IoT devices, which were initially handled with cloud resources only. Fog computing aims to selectively move computation from the cloud to the edge, based on the data availability, latency and response times, application requirements and edge features.

This will in turn reduce the workload on cloud data centres while providing a quicker response time whenever needed. The principle behind this paradigm is that both computational and storage resources must be distributed in a smart, dynamic and elastic way, so that decision making happens as close to the data origin (regardless if the edge or the cloud) as possible.

The fog paradigm fulfils the demands of Big Data analytics and Pilot projects such as those to be deployed by PLATOON project in the energy domain, in which the underlying systems can be designed following two different priorities:

- **First priority**: to provide quick and reactive information. This typically implies processing data streams, focusing on their most relevant aspects (operation oriented).
- **Second priority**: to provide more accurate and consistent responses, which typically implies aggregating as much information as possible, in larger and better detailed models (maintenance oriented).

2.2.2 Device management

Next generation energy digital platforms shall be able to sense and process large volumes of sparse and heterogeneous data sources: electrical grid elements (substations, HV lines, etc.), smart meters and submetering devices, weather forecasts, renewable power plants production, sensors in wind turbines and PV plant components, etc.

This large scale of data will be ingested by systems that enable added-value services, such as the automation of sensible decisions and actuation whenever needed. Internet of Things (IoT) has become the key technology for interconnecting the above-mentioned systems and devices, thereby digitalizing the physical world into a mesh representation of information systems (e.g. big data models, digital twins).

In the energy domain, IoT systems are in the dozens of thousands of devices that sense, (pre-)process, transmit and actuate. Moreover, the amount and capacities of IoT devices has been growing steadily over the years. As a result, traditional manual controlling functions have been replaced by remote functions that make the processing and analysis of relevant physical variables possible (e.g. associating with the deterioration of assets or specific elements). The provision of secure and lightweight IoT data protocols (e.g. MQTT over TLS, CoAP over DTLS) has alleviated the overhead of communications.

2.2.3 Operations

Currently, the main field of application of Data exchanges in the energy sector for Operations purposes is the management and control of the electricity transmission and distribution grids. As more devices and elements in these assets become digitalized and interconnected (building the so-called “Smart grids”), there is an increasing flow of data and, accordingly, the need of all the stakeholders involved to properly collect, manage, analyze, share and exploit this huge amount of data generated. In fact 6 out of the 7 Pilots in PLATOON project are dealing with data for operational applications in electricity grids or consumption sites:

- Pilots #2a, #2b and #4a: improvements for the operational of the grid (microgrids in 4a).
- Pilots #3a, #3b and #3c: energy management and operations performance in buildings and smart cities.

Some of the most relevant business opportunities linked to data exchanges in the operation of the electricity networks are listed in the following paragraphs, mentioning the specific devices and data requirements in each of them:

- **Low Voltage (LV) monitoring and automation advanced algorithms:** These applications are based on the availability of consumption data from the installed Advanced Metering Infrastructure (AMI). AMI includes the Smart Meters and Data Concentrators, together with the PLC communication infrastructure and the integration with the data collecting system. Smart metering rollouts have been largely deployed in most European countries.
- **Network management (substation automation, grid monitoring and control):** Although the tools developed so far to observe the network are reliable, improvements are required in: standardization (being main reference IEC 61850); interoperability (defined as the ability of two or more networks, systems, devices, applications, or components to interwork, to exchange and use information in order to perform required functions); cybersecurity (grid infrastructure is very critical, as many processes in health, banking, telecom and industry are dependent on a secure and reliable electricity supply). At transmission networks level, emphasis is also placed on the development of tools for the coordinated operation of pan-European networks.
- **Integration of Distributed Renewable Energy Sources (RES):** New business models should focus on offering new control schemes and new hardware/software solutions based on grid data processing for integrating small/distributed RESs, while assuring system reliability and security. The management of the flow of electricity and data in real time, including revised roles of network operators and proper network technology, are yet to be fully developed.
- **Storage solutions**: Another breakthrough development will be the implementation of storage solutions in substations and in RES power plants to optimize the dispatchability of energy. This kind of solutions will require advanced communications and data exchange (energy demand, generation forecast and pricing, status of batteries, . . .) between the generation facilities, the DSOs and the demand side, as well as efficient Battery Management Systems (BMS) to process the information and optimize its technical and economic performance.
- **“Prosumers” and smart buildings (“smart customers”):** This is one of the areas with greater breakthrough potential, so that the grid can provide the new proactive consumers (“prosumers”) with the information and services they demand, as the best way to implement energy efficiency measures. Demand response and flexible consumption should be integrated in the market as soon as possible, by working out appropriate market products. Technologies to collect and exploit data from smart appliances, domestic storage devices and home automation should boost new tariff schemes and new aggregation mechanisms, like virtual power plants. On the other hand, consumption data contain highly sensitive information that can relate to individual persons, and hence need to be protected at all times.
- **Electric Vehicle integration in the grid**: The smart integration of electric vehicles (EVs) and Plug-in Hybrid Vehicles (PHEV) in the electricity network has already been demonstrated in specific areas (pioneering cities and regions) and demo projects, but the real challenge will be faced when EV chargers are deployed at large scale with increasing consumption of electricity and potential demand peaks along the day. Availability of real-time data will then be key to offer a wide range of services, both in the EV and the grid sides: advanced services to the EV owners (variable pricing, charging points availability, interoperability of charging managers, etc.); battery monitoring; integration of information from the low voltage monitoring systems and electricity prices to regulate the grid capacity and program the slow chargers during the night (or low demand “time-windows”); management and optimization of quick and “ultra-fast” charging points; implementation of protocols and conditions for vehicle-to-grid (V2G) operations , … The EV paradigm is the “meeting point” for 2 traditional value chains (electricity distribution and car manufacturing), that have scarcely interacted before but are now facing

common challenges and interests. Data exchange may become the way to facilitate their effective integration, in order to create new business opportunities not even foreseen so far.

2.2.4 Maintenance

As the International Energy Agency (IEA) has stated[vi], *"Digital data and analytics can reduce power system costs in at least four ways: by reducing O&M costs; improving power plant and network efficiency; reducing unplanned outages and downtime; and extending the operational lifetime of assets. The IEA estimates that the overall savings from these digitally enabled measures could be in the order of USD 80 billion per year over 2016-40, or about 5% of total annual power generation costs based on the enhanced global deployment of available digital technologies to all power plants and network infrastructure."* ("Digitalization & Energy", International Energy Agency (IEA), 2017). As a matter of fact, the four sources of cost reductions mentioned by the IEA can be classified under the "Maintenance" concept, as they have to do with activities and costs usually accounted under this chapter.

It can be stated that renewable energy plants are the main field of application of data exchanges for maintenance purposes in the power system. Access to data related to energy production, availability of the plants and efficiency rates of the equipment (wind turbines, solar panels, balance of plant) are very useful to monitor the status and performance of the systems under different operating conditions and to obtain relevant information about the maintenance tasks and activities to be prioritized and executed.

The developers of the renewable power plants are the owners of this information ("Data owners"): windfarm developers, investors in PV solar plants, operators of CSP (Concentrated Solar Power) plants, etc. These companies constitute the key segment to be addressed in order to collect their conditions and requirements regarding data sharing with other stakeholders ("Data users") engaged in maintenance activities: plant operators, maintenance companies, OEMs, components suppliers, etc.

Probably wind energy is the renewable energy source (RES) that offers a more rewarding "business case" but, at the same time, demands more challenging requirements in terms of quantity and quality of data to be exchanged for maintenance applications. Precisely Pilot #1a in PLATOON will deal with "Predictive Maintenance of Wind farms".

The cost of wind power (and especially of the power generated in offshore windfarms), has dropped dramatically, to a great extent as a result of new auction systems which have forced developers and investors in windfarms to reduce costs both at investment (CAPEX) and operation (OPEX) levels. At the same time, sites for offshore windfarms are being built in more complex and "harsh" environments (deeper waters, longer distances from the coast), with highly demanding requirements in terms of quality, integrity, efficiency and reliability.

The relevance of operational expenditure (OPEX) in the wind sector is reflected in the percentage of expenditure that it involves, between 20% (average for onshore) and 35% (or even higher in offshore windfarms) of the life cycle of a wind turbine. This fact has shifted the focus of reducing costs from CAPEX to OPEX in order to reduce the cost of energy (Leverage Cost of Energy, LCOE).

The massive implementation of sensors in wind turbines and the consolidation of data processing technologies (Big Data, Data analytics, IA) is allowing the adoption of predictive maintenance techniques, also known as condition-based maintenance (CBM), as a key element in designing an effective O&M strategy. Predictive maintenance evaluates the actual condition (health) of the asset continuously over time to detect symptoms of degradation in the functioning of the systems and components of the asset before any significant deterioration of the physical components occurs. The benefits of a good predictive program can mean savings of 8% to 12% compared to using preventive maintenance based on periodic actions or based on cycles/hours of operation. Assessment of the

performance of components and their remaining lifetime is key also to improve operational programs and implement optimized maintenance strategies.

Some of the subsystems in wind turbines are critical in terms of the severity of the potential failure during operation: rotors and blades, generators, gearboxes and bearings, pitch control systems, yaw systems, power electronics or electric controls. The status and performance of the subsystems can be evaluated gathering data through a collection of techniques, such as:

- Vibration analysis
- Acoustic emissions
- Oil analysis
- Strain measurement
- Electrical parameters (intensity and voltage),
- Shock pulse method
- Physical condition of materials
- Thermography.

Data have to be sampled at regular time intervals using sensors and measurement systems. The frequency ranges of data and quality control should be defined depending on the sensors implemented and the level of accuracy of the analysis to be performed (from 15 minutes to 1 second, or higher for some applications). Using data processing and analyses, CBM systems can determine the status of the critical components. By processing the data history, faults can be detected (diagnosis) or predicted (prognostic) and the appropriate maintenance strategy can be chosen.

From the "Data Users" perspective, there are some relevant requirements they will demand when it comes to collecting data from different sensors for different purposes in different wind farms:

- **Metadata**: To accurately locate specific datasets, they should be tagged with a series of information, metadata, using so-called metadata cards. Besides preserving the information on data for a future re-use, metadata are used for indexing datasets to refine their findability. Metadata are classified into three categories: descriptive, administrative and structural. Descriptive metadata provide information on e.g. what (associated topic, type of variables, etc.), where data were collected (external conditions or geographical location, etc.) or how data were collected (instruments, activity type). Administrative metadata provide information on e.g. who collected the data (data owner), access rights, links to data, etc. Structural metadata provide information on e.g. data format. For the architecture and toolbox to be developed in PLATOON, a standard metadata element set has to be specified and used, e.g. the Dublin Core Metadata Initiative (DCMI)[vii], probably the most extended and up-to-date for scientific purposes.
- **Data taxonomy**: Taxonomy is the descriptive type of metadata containing terms that assign textual information to the data. In a broad sense, it is any means of organizing concepts of knowledge. In a narrow sense, taxonomy is a hierarchical classification or categorization system as we know from e.g. the classification of species. For data exchange purposes, taxonomy is used to put data into the correct context by defining and hierarchically classifying the research area topics and organizing data within topics. A good taxonomy should enable Data users to immediately grasp the overall structure of the knowledge domain and the associated data.

  Furthermore, the taxonomy is a basic requirement to ensure Interoperability between different platforms, data sources or applications, and to comply with the Open Access (OA) policy of the European Commission. The Open Access (OA) to knowledge is a principle established by the EC, underlying the H2020 EU Framework Programme for Research and Innovation, that aims at optimizing the impact of publicly funded projects, by making information openly available and

reusable to everyone in Europe. The Open Data (OD) policy is part of the OA strategy and is widely acknowledged as a fundamental step to support a fast track from research to innovation. To boost OD, the EC declared that data must be at the same time "Findable, Accessible, Interoperable and Re- usable (FAIR)".

Specifically, in a wind energy context, taxonomy terms can be used as a controlled wind energy vocabulary by data owners for tagging data in the metadata card and by data users as "facets" to filter content progressively via a "faceted search". This means that the taxonomy structure should include the topics distinctive of the Wind energy sector and the data type relevant to different topics and taxonomies of other facets.
Consequently, data exchanges in the framework of PLATOON project will have to adopt any of the Wind energy taxonomy structures proposed through different standardization bodies or projects. Some of the most relevant or used so far are included as references[viii] , [ix],[x],[xi]

- **Data quality control criteria.** Relevant IEC standards should be followed for data quality control in the PLATOON pilots and specifically in #1a. In order to be consistent with the comparisons of the same quantities measured by different methods or between different wind turbines, the data should be collected and produced in a consistent way. As an example, comparison of the wind measurements obtained from anemometer measurements, nacelle lidars or ground based lidars needs to be thought in advance, so that the companies who deliver the data make sure that these criteria are followed. Another example is the comparison of forces measured by strain gauges or fiber optics in different locations. In this particular case, the experience shows that comparing forces obtained from different sources requires significant effort to make sure at least the axis system is consistent. Although it is common to use "GL coordinate system" during the design of components, other coordinate systems are also used to represent the load distribution on components, which could make the comparisons useless, even for the same wind turbine. It is important to set a data standard and quality criteria already from the specifications of the data to be shared, in order to reduce this kind of inconsistencies and uncertainties,

2.2.5 Related business models: existing and new ones

Data sharing platforms in the energy sector should boost the definition of new **business models** connected to the **collection, management and secure and effective exploitation of the huge amounts of data** that the electricity power plants, the smart grids and the consumption sites are increasingly generating. Very few companies are presently using data as a source of value and business opportunities. Therefore, not much can be said about existing business models in the use of data in these fields. Instead, potential new business models for the exploitation of key results of PLATOON project must be identified and explored, as a first step in the definition of their specific Exploitation Plans.

In general, a business model can be defined as a unit of analysis to describe how the business of a firm works **("The St. Gallen Business Model Navigator"[xii]**). A business model is often depicted as an overarching concept that takes notice of the different components a business is constituted of and puts them together as a whole[xiii].

The paper of the University of St. Gallen referenced above identifies 55 patterns of business models. Some of them can be selected as the most likely business models for the products and services that will be developed in PLATOON. A brief description of each of them and their connection to the exploitable results of PLATOON are listed below:

- **“Leverage Customer Data”**: New value is created by collecting customer data and preparing it in beneficial ways for internal usage or interested third-parties. This business model is especially suitable for 2 kinds of applications and pilots to be developed in PLATOON:
    - The exploitation of energy consumption data from consumers (buildings, cities, . . .). External companies can collect and process this data in order to provide additional services to the energy consumers: efficiency programs and recommendations, detection of technical and non-technical losses, failures or misfunctioning of equipment and facilities, . . . This information could also be of interest to third-parties like utilities and DSOs, in order to optimize the energy dispatching and the operation and management of the grid.
    - The exploitation of data coming from digital devices in energy assets, such as the electricity grids and renewable power plants. The collection and analysis of this data (using for instance the platforms and “Data Analytics Toolbox” to be developed in PLATOON) can be a very relevant value proposition for the owners of these assets (utilities, DSO, windfarm developers, solar PV plant owners, . . .). Asset owners can perform this data analytics with their own internal resources (“software as a service”). But, if data privacy and security is ensured, these platforms could also provide data access to third-parties, so that they can offer services to the assets owners based on data analysis: operational efficiency, predictive maintenance, remaining lifetime estimation, etc. The enhancement of this value proposition would lead to the “Data Trustee” business model described below.

- **“Make More of it”**: Know-how existing in the companies as an intangible asset is not only used to build own products, but also offered to other companies. Most of the companies that are developing architectures and applications in PLATOON will be able to add value to their customers offering their know-how regarding knowledge models for energy facilities, reference architectures and standards for data sharing.

- **“Peer to Peer”** : This model is based on a cooperation that specializes in mediating between individuals belonging to a homogeneous group. For instance, a platform can offer an online database and communication service as a meeting point for energy consumers, where best practices, know-how and energy consumption recommendations could be exchanged between consumers. Consumers would be classified in categories (buildings, public services, industrial processes, . . .) and segments (consumption rates, timelines, patterns, . . .), so that individuals accessing the platform could easily search the most interesting peers or use cases.

- **“Performance based Contracting”**: A product's price is not based upon the physical value, but on the performance or valuable outcome it delivers in the form of a service. This is the basic existing business model for the Energy Service Companies (ESCOs), that may design, purchase, install and/or operate (some or of all of them) the energy facilities of their customers, and get their revenues not selling the physical assets but the energy the customer demands as a service. This business model can be reinforced if data is available and used to optimize the operation of the assets and to be more efficient on both sides: in the energy consumption and in the purchasing and supply. This new (or reinforced) business model is referenced in some papers as “Value adding services in operation”.

- **“Two-sided Market”**: A two-sided market facilitates interactions between multiple interdependent groups of customers. The value of the platform increases as more groups or as more individual members of each group are using it. In the case of PLATOON, the platform could facilitate interaction between energy retailers and consumers. Retailers could make offers more adapted to the energy consumption patterns and demands of consumers, and

consumers could compare the value propositions from different bidders. Different kinds of agents (aggregators, ESCOs, prosumers, . . .) could join the platform if they find the data exchanged relevant or interesting for their business model.

- **"Data Trustee".** The Working Group **"Digital Business models" of "Platform Industrie 4.0",** has produced a paper under the title "Digital business models for Industrie 4.0.", published by the Federal Ministry for Economic Affairs and Energy of Germany. In that paper, a specific new business model is proposed with regards to data trading. The data trustee provides a neutral platform that integrates data from different companies, combined with additional data across supply chains, continents and existing business relations. The value proposition includes the assessment of data quality, taking care of IT security, and ensuring that the terms regarding data use are complied with. As it is defined, this "Data Trustee" business model can be considered as an enhanced evolution of the "Leverage Customer Data" pattern and is probably one of the most ambitious in terms of the value added to data. In the case of PLATOON, this business model is ideally the one that would perfectly fit the exploitation of an overall platform integrating most of the connectors, applications and toolboxes to be developed in the project. The compliance of all these developments with the interoperability, data sovereignty and scalability specifications of the project, would ensure that the resulting integrated platform offers a unique value proposition as a "Data Trustee" platform.

2.2.6 Regulations in data exchanges

The management and exchange consumer data (metering and consumption data, data required for switching date, DR, etc.) is essential for a well-functioning retail market. The Third Energy Package and the General Data Provision Regulation to the Clean Energy Package (CEP)[xiv] allows consumers to access and share their own energy data.

Integrating national retail markets is more difficult than integrating wholesale market due to differences in market models, legislation, market processes and data exchange procedures across Member States (MS). The differences in data management was listed as a possible market entry barrier for new actors by the EC and having common criteria and principles was seen as the way to overcome this barrier.

Clean Energy Package

TSOs and DSOs share the opinion that one data management model that applies to every case is not applicable in all European countries and that each consumer data management has to be assessed nationally. However, it is also agreed that a lack of standardisation and interoperability can pose barriers, and this is shown in CEP.

- Data management
  - MS shall designate a competent authority to specify rules on the access to final customer data by eligible parties.
  - MS shall organise data management to ensure efficient and secure data access and exchange, on top of data protection and security.
  - Data access and storage rules shall comply with relevant Union law (Regulation EU 2016/679).
  - MS or designated competent authorities shall authorize/certify or supervise parties responsible for data management.
  - No additional costs shall be charged to final customers and MS shall be responsible for setting relevant charges, ensuring charges imposed are reasonable and duly justified.

- Interoperability requirements
    - MS shall facilitate full interoperability of energy services within the Union.
    - The Commission shall adopt interoperability requirements and non-discriminatory and transparent procedures through act implementation (Article 23(1)).
    - MS shall ensure electricity undertakings apply interoperability requirements and procedures (based on existing national practices).

Network code and guideline areas

The CEP confers consumer data management organization and rule specification to MS, however European level guidelines may be necessary in order to facilitate full interoperability. Two network code areas are described in Art. 59:

- Rules on DR, aggregation, energy storage and demand curtailment.
- Data exchange, settlement and transparency rules (demand-side flexibility). The responsibility for cybersecurity and data protection is a shared task of TSOs, DSOs and regulatory authorities.

3 Part 2: State of the art

3.1 Key concepts in the digital data value chain and its applications

3.1.1 Platform economy

A platform refers to a business model that creates value by facilitating exchanges between two or more interdependent groups. There are different players:

- Producers: who feed the platform
- Consumers: who purchase or use the products/services offered on the platform
- Platform owner

Platforms are born out of digitization, any social economic activity that utilizes a platform is referred to as platform economy. The term platform has several informal interpretations, such as:

- Online matchmakers: utilizing the internet infrastructure to match between demand and offers (e.g., Amazon, Uber, AirBnB, etc.)
- Matchmakers in general, without necessity of being online (e.g., business parks)
- A channel to connect without having means for production.

On a more formal terms, platform refers to a collection of core, low-variety components that forms a system supported by high variety collection of peripheral components. The algorithmic revolution and cloud computing are the foundations of the platform economy, where computing power is only the beginning.

Platforms are the core organizational form of the emerging informational economy that is set to replace markets. Platforms are able to shape the flow of information, where it does not just act as a network but shapes the exchange.

The platform provisions an infrastructure for producers and consumers of value to use. The platform benefits as the total value of the platform is increased as more producers create value, attracting more consumers, which in turn attracts more producers. Platforms control the interactions, so the kind of power of the platform is important as it may interact and lead to market control.

The strength of the platform economy relates to the elimination of barriers, increased information sharing between the different players through data flow, which leads to a greater participation.

The importance of platforms in businesses has been highlighted in well-known corporate success and failures, as often illustrated by the fall and rise of business in the past two decades. For example, the fall of Blockbuster against the rise of Netflix and other on-demand video streaming platforms, the fall of Nokia and Blackberry against the rise of Google's Android and the Apple's iOS ecosystem, and the profit decline of local offline retailers against e-commerce giants. However, despite the business tendency, platform economy is not always used in a commercial context.

Creating digital platforms requires technical expertise and the understanding of technical requirements. With the availability of open technologies stack that can be reused along with community supports, developing digital platforms becomes cheaper. Minimum viable products, however, still need more custom development such as a platform with an intuitive usability, and may drastically rise platform development costs. In terms of PLATOON platform, some functionalities can be developed by utilizing available technology stack, but still need additional functional requirements.

Since the platform economy refers to platform matchmaking, the PLATOON platform may be needed to perform this function on different aspects regarding data exchanges:

1. It needs to have easily deployable data exchange tools and components to be used by energy-related stakeholders as a software on their own platforms.

2. The PLATOON platform can offer an algorithm-specific analysis over the datasets provided by energy-related stakeholders. This enables custom requests for certain analysis with the option of keeping the data private between the data analyst and data owner.

3. Since developing the data requires a technical background, the technical team can provide training for stakeholders that are interested to use PLATOON internally in their institutions.

4. The PLATOON platform can also ideally be utilized by the public parties (e.g. think tank organizations, civil organizations, journalists) to analyze energy-related open data so that the knowledge acquired from this analysis can be used to build a renewable energy-related policy/regulation for respective government or public administration.

3.1.2 API economy

An **Application Programming Interface (API)** is a set of functions which are provided from a software component to be usable for further development of foreign technological stacks. The API specifications describe how to perform the usage of the underlying software component. The parameters and data types are given. Also, the output format and type are specified. The Data-In/and -Output is often aligned to technological or industrial standards to make foreign developers the usage easier and enhance interoperability. The specifications ideally should not change when its internal function construction changes leading to similar output, because other programs rely on the stability of the API specifications.

The **API Economy** is a business term that describes the opportunity regarding how to provide and use APIs to create value[xv]. API users don't need to develop every tool they need on their own to build up desired software projects. They can use APIs to stack their digital system by using the developed expertise of specific API providers. The API provider does not need to provide an entire program with graphical user interface but can focus on the desired functionalities. APIs in the API economy can be understood as Software as A Service (SAAS). The provided API can utilize open source to make the API development faster and more transparent. However, the usage of underlying open source tools and platforms is not mandatory. APIs can be implemented to provide functions as a web service. One of the most common web API over HTTP is REST[xvi]. The Usage of APIs can be implemented as a paid usage (by call, …) and with authentication needed.

A simplified **example** of stacking some APIs is a use case to make the reporting of electricity meters easier. The user only should make a photo from the current electric meter. Then we can use an API likes the one in OpenCV[xvii] to read the photo from file system and improve its quality. A Second API e.g. Google Visions (paid) derives the displayed power consumption numbers and serial numbers with Object Character Recognition (a machine learning functionality). The information of power consumption numbers and serial numbers can be identified from the electric meters photo. The Web API of power provider makes it easy to simply hand over this reported number. Also, Payment APIs can be used to make it a paid service to cover the costs for used APIs along with the further gained value of increasing user convenience. One can further stack this approach by APIs which can detect anomalies or fraud detection. The gained value could be increased by making use of this gained data by providing the user recommendations for better fitting energy providers to save money in future. This example shows how much development expertise and effort can be outsourced to easier generate a desired digital product.

The **PLATOON project** should showcase **APIs** from various consortiums partners being able to be stacked and to interact. Data and fitting algorithms can be provided over APIs. The API descriptions

need to be transparent, easily accessible, using state of the art methods and reliable. It would be important to clarify as early as possible how the desired use cases can be translated into the needed provided APIs. The API should be published in a standardized documentation format like OpenAPI. If not all algorithms and data are available, it is desirable to have at least some sample data and some representative algorithms available over the APIs. Therefore, current implementations can use those as placeholders until upcoming features are released. Additionally, a reporting tool is needed for suggesting future enhancement, API extensions as well as reporting API-related issues.

3.1.3 Data sovereignty

The energy-related data offered by different partners/stakeholders in PLATOON, serve as a strategic resource that can be used to create innovative value offerings. Key to success is to share and jointly maintain data within the ecosystem, as end-to-end process support can only be achieved if the stakeholders team up and jointly utilize their data resource. However, it is important to protect their data more than ever before, since the importance of data has grown.

Data sovereignty is about finding a balance between the need for protecting one's data and the need for sharing one's data with others. To find that balance, it is important to take a close look at the data itself, as not all data requires the same level of protection, as the value contribution of data varies. Public data, for example, which can be accessed by anyone, requires a lower level of protection than private data.

3.1.3.1 Data exchange and Data Sharing

Data exchange is a mere transfer of data from one participant to another. Data sharing includes data exchange that takes place between participants to achieve a common goal, for example, to enable a new business model by generating additional value out of data (Data markets). Data sharing implies a mode of collaboration between participants in the hope of mutually beneficial results.

As it is the first time in the document that we will quote it, we include a brief description of the International Data Spaces Association (IDS Association or IDSA). IDS was founded in February 2016. The constituent meeting was held at Fraunhofer Forum in Berlin attended by all founding members and its mission statement is devoted to digital transformation as a key factor for the success of companies worldwide to ensure that the special economic interests of business are specifically integrated into the research work of International Data Space and "to foster the general conditions and governance of a reference architecture for International Data Spaces and interfaces with the aim of achieving an international standard"[xviii]

To enforce data sovereignty, IDS specifications could be invoked, in which the participants exchange, share and process digital content by means of a dedicated software component: the Connector. It is the logical building block that ensures data sovereignty of the Data Owner is always guaranteed. Here, participants refer to the partners/stakeholders in the PLATOON ecosystem.

The Connector is the communication interface to all other participants. The common language of all Connectors is the Information Model. The information model facilitates compatibility and interoperability, thereby all connectors can exchange data. Each Connector that participates in the IDS ecosystem must provide a self-description for other IDS participants to read. The Connector self-description must contain information about the respective organization, about who maintains the Connector (i.e. the Service Provider), and about the content and type of the data offered or requested. The Information Model therefore supports the description, publication and identification of Digital Resources.

The Data Provider may attach Metadata to the data transferred using the IDS Vocabulary. In doing so, the terms and conditions to ensure data sovereignty can be defined unambiguously (e.g., data usage, pricing information, payment entitlement, or time of validity).

Data is always stored decentralized and is only transferred on demand. To exchange data, the Data consumer may directly contact the data provider, if it knows a suitable data provider. Otherwise, the Data consumer could use a Broker Service Provider, to find out about the data providers capable of providing the desired data.

3.1.4 Data needs for AI applications

For Artificial Intelligence applications, more analysis can be done if the datasets are provided in a semantic format. When the data is provided in a semantic format, more information can be derived and inferred from the relationship between data items.

A common language used is Resource Description Framework (RDF[xix]). RDF is a specification by the World Wide Web Consortium (W3C) to model information using specific syntax notations and supported serialization formats. RDFSchema[xx] add a layer on top that allows to create classes and hierarchical links and constitute a vocabulary that allows to describe simple ontologies. OWL[xxi] is another layer on top that allows to describe more complex ontologies. These specifications are languages and vocabularies to describe ontologies and they are independent of all domains.

Since the concepts covered in RDFSchema are rather conceptual, a more advanced ontology needs to be developed to represent a more domain-specific information, such as energy-related information. This ontology ideally provides the list of terms, classes, axioms, interrelationship between classes and so on. If there have been ontologies developed for the represented domain, ideally the ontology should be reused. In case some needed classes/terms/axioms are not yet presented in the ontology, the ontology should be extended instead of recreating a new domain-ontology from scratch.

If it is desired to support Supervised Machine Learning as one possible Implementation for AI Applications, a well-documented labelling should be provided. In addition, it is always valuable, if it is also documented, how the data was created, where missing values are expected and how they are treated with domain specific knowledge. Also, a documentation of reasons for outliers and noise helps to select and fit the ML pipeline (especially pre-processing and evaluation).

The need for greater access to data for AI applications is well understood across all sectors, and the energy sector is no different. Even if sufficient training data can be compiled from one stakeholder for specific objectives, the addition of datasets from key partners can result in stronger applications (more precise pattern recognition, confident predictions, etc.) and the identification of new applications that rely on multiple data sources. Apart from access issues in such distributed, multi-stakeholder settings (how data can be made discoverable, accessible, exchanged in a trusted and secure manner), data that is to be used for reliable AI applications has other core needs. The two primary concerns can be categorised under veracity and interoperability.

In order for trusted, high-quality data to produce useful and reliable results, veracity is the primary issue that needs addressing. Data needs to be of high-quality if the adage of ‘garbage in, garbage out’ has to be avoided. This is even more crucial when multiple data providers are involved. In addition, for insights resulting from AI applications to be considered seriously, the exact data used needs to be reproducible (together with the algorithms) and its source/s verifiable. The sections that cover Data

Quality and Data Provenance provide more in terms of the state-of-the-art addressing data veracity as a whole.

The need for data interoperability becomes apparent when multiple datasets or data sources are to be ingested for AI applications, regardless of whether they involve the same provider or are contributed by multiple stakeholders. Without sufficient interoperability readiness, significant manual effort will be required to aggregate and 'normalise' the different datasets prior to jointly processing or analysing them. Interoperability-by-design is therefore recommended as a best practice, and requires data to always be accompanied by metadata that facilitates its interpretation within the context of other relevant (and similarly 'meaningful') data, rather than in isolation. The consolidation and use of existing shared data models, in this case covering the concepts and descriptions characterising the energy sector, is thus a prerequisite for use-cases driven by mixed and heterogeneous data and their application.

3.1.5 Data quality

Data quality plays a crucial role in all data-intensive applications, hence, also in the energy domain. The term "data quality" is commonly defined as "fitness for use". As the definition indicates data quality is relative and depends on the context of use. The quality of data may depend on the type of characteristics or dimensions considered. Data that is sufficient quality for one use case might be insufficient for other scenarios. There are four data quality dimensions: intrinsic, contextual, representational and accessibility[xxii]. Intrinsic dimension defines a set of quality metrics such as accuracy, objectivity, believability and reputation. Contextual dimension metrics highlight the requirements that data quality must be considered within the context of the task at hand, i.e., timeliness, relevancy, completeness, value added, and appropriate amount of data.

Representational data quality dimensions emphasize the importance of a system, i.e., the system must present data in such a way that they are interpretable, understandable, concise and consistently represented. Similarly, the accessibility dimension focuses on the role of the system in terms of its accessibility and security. Representational data quality includes two aspects: format and meaning of data. To enhance data quality, one need to understand what data quality means to data consumers. Data quality issues need to be assessed and monitored early in the data value chain.

3.1.6 Data roles

To exchange data within the PLATOON ecosystem, IDS specifies core participants that are involved and required every time data is exchanged. Following are the roles:

- Data Owner
- Data Provider
- Data Consumer
- Data User
- Broker Service Provider
- Vocabulary Provider

3.1.6.1 Data owner

A data owner could be a legal entity or a natural person creating data and/or executing control over it. The data owner has the (technical) means and the responsibility to:

- Define Usage Contracts and Usage Policies, and to provide access to data.
- Define the Payment Model, including the model for reuse of data by third parties.

Usually, a participant acting as a Data Owner automatically assumes the role of Data Provider as well. In the unlikely case, the only activity of Data Owner would be to authorize a Data Provider to make its data available to be used by a Data Consumer

3.1.6.2 Data provider

The main activity of a Data Provider is to make data available for being exchanged between a Data Owner and a Data Consumer. If a Data Provider is the same as Data Owner, then it assumes all the responsibilities of a Data Owner. Other (Optional) activities of a Data Provider include:

- Register metadata of the data at a Broker Service Provider (optional IDS component), to facilitate a data request from Data Consumers. However, to establish a connection between Data Provider and a Data Consumer, Broker Service Provider is not necessarily needed.
- Log data transaction details in the Clearing House (optional IDS component) to facilitate billing or resolve a conflict.
- Enrich or transform data using IDS certified Data Apps.

3.1.6.3 Data consumer

The Data Consumer receives data from a Data Provider. If the information to connect with the Data Provider is already known to the Data Consumer then the Data Consumer may request the data directly from the Data Provider. Otherwise, the Data Consumer can search for existing datasets by making an inquiry at a Broker Service Provider that provides the required metadata for Data Consumer to connect to a Data provider.

A Data Consumer is a mirror entity of Data Provider and it could provide the same facilities as Data Provider such as logging details of a data exchange transaction, using Data Apps to enrich or transform data.

3.1.6.4 Data user

Data User is the legal entity that has the legal right to use the data of a Data Owner as specified by the usage policy. In most cases, the Data User is the same as a Data Consumer, unless of an exceptional scenario.

3.1.6.5 Broker service provider

Broker Service provider acts as a metadata registry that stores and manages information about the data sources available in the PLATOON ecosystem. It must provide an interface for Data Providers to send their metadata. The metadata should be stored in an internal repository for being queried by Data Consumers in a structured manner.

After the Broker Service Provider has provided the Data Consumer with the metadata about a certain Data Provider, its job is done and it is not involved in the subsequent data exchange process.

3.1.6.6 Vocabulary provider

A vocabulary provider manages and offers vocabularies (ontologies, reference data models, metadata elements) which can be used to annotate and describe datasets. They provide domain specific vocabularies and their reference to the IDS Information Model.

3.1.7 Types of trust

Establishing trust for data sharing and data exchange is a fundamental requirement. There are two types of trust:

- Static Trust: Static trust is based on the certification of participants and core technical components in the PLATOON ecosystem. Every participant and connectors in the PLATOON ecosystem must undergo certification in order to establish trust among all participants.

  Any participant that wants to operate a connector in order to exchange data in the International Data Spaces as a Data provider or Data Consumer needs to acquire a unique identity in the form of a certificate. This certificate enables them to establish secure and trusted connections to other IDS participants.
- Dynamic Trust: Dynamic trust is based on active monitoring of participants and core technical components in the PLATOON ecosystem by enforcing Data Usage Policy.

3.1.7.1 Certifications

The digital certificate is based on the certification of the participant and the certification of the connector. Certification aims at determining and formally stating compliance of a Participant or Connector with a predefined set of evaluation criteria.

A **Certification Body** needs to govern the aspects of certifying components and entities seeking admission to the PLATOON ecosystem based on the evaluation criteria. Depending on the decision from the certification body, a digital certificate (eg., X.509 certificates) needs to be issued by a **Certification Authority**.

At this time it is expected that no certification will be obtained during the project. For validation purposes a "fictitious" Certification Authority will be created that will act as the certification body. Formal certification will be seek after project completion by individual partners that are interested in exploiting IDS components in the PLATOON ecosystem.

3.1.7.2 Relational contracts

A Usage Contract formalizes the expectations regarding the behavior of Participants involved in a data exchange transaction in a declarative, technology-agnostic way. A data provider could share the data with a data consumer, provided they agree upon a **Data Usage Policy**.

Data Usage policy specifies a Contractual agreement, which defines rules and conditions on how to use the data from Data Owner, which needs to be established between a Data Provider and a Data Consumer to exchange data in the PLATOON ecosystem.

3.1.8 Data provenance

Data provenance is the description of the origins of data and processes by which data arrives and exists in a database[xxiii]. Provenance is important to 1) understand the conditions in which the data has been collected 2) trace changes and transformations;, 3) derive trustability, relevance, and confidence; 4) increase reproducibility and reusability; and describe complex transformations.
Two granularities of provenance can be defined, i.e., workflow provenance (coarse-grained) and data provenance (fine-grained). Workflow provenance maintains the metadata of processes and services that are executed. The processes and services metadata can be a software program or a hardware used during the experiments[xxiv]. For example, during the integration of datasets, users may trust the data if they know the algorithm used during data integration and the datasets integrated. Data provenance records the origin and history of the data transformed during process execution. The

particular features of the original datasets; combined to produce the integrated datasets feature, are preserved as provenance. For example, the users can use the provenance information to find out the original sources of the longitude and latitude values in geospatial data. Data provenance needs to be maintained in several application areas like, data integration, data-warehousing, grid-computing, workflow management, and curated databases[xxv].

3.2 Analysis of common patterns of data exchanges

3.2.1 Business process synchronization

Business process is a sequence of activities or tasks designed to create something of value. It begins with an objective and ends with achievement of the business objective of providing a result that provides customer value. Synchronizing changes in business or technology is significant in achieving such results. Delays in dealing with such changes makes the process model become out of sync with results. Moreover, processes that share similar characteristics and use similar resources, e.g., data, should be synchronized which leads to reduction of costs. Synchronization could be between similar business processes as well as between different categories.

Business processes can be categorized into three types: Operations process: constitutes the core business and creates the primary value stream, Management process: the process that oversees/manages operational processes and Supporting process: which supports the core operational processes. A complex business process may be decomposed into several subprocesses, which have their own attributes but also contribute to achieving the overall goal of the business. Care must be taken while doing so as there may be crossover.

3.2.2 Data driven new business models

Data is often described as the 'new oil'. Several businesses are developing new business models that are designed to create additional business value by extracting, refining and capitalizing on data.[xxvi] The competitive advantage associated with effective big data utilization is driving the desire for existing mainstream businesses to become data-driven. Creating new sources of data, developing services and technologies to organise and analyse as well as repackaging existing data sources all have the potential to base a successful business model.

A variety of sources and offering a range of services, include platform as service and analysis as service. Three distinct types of big data business models: data users, data suppliers, and data facilitators. Data users are organisations that use data either for informing business decisions, or as an input into other products and services such as credit reports or targeted advertising campaigns. The second class of business model, i.e., data suppliers, are organisations that either generate data that is of intrinsic value and therefore marketable, or else serve a kind of brokerage role by providing access to an aggregation of first- and third-party data. The third class of business model, i.e., data facilitators, encompasses the range of activities that support third parties that are lacking in infrastructure or expertise including advice on how to capitalise on big data, the provision of physical infrastructure, and the provision of outsourced analytics services[xxvii].

3.2.3 Regulation Compliance

Data compliance refers to any regulations that a business must follow to ensure that any sensitive digital assets they possess are organized and managed, i.e., guarded against loss, theft, and misuse, to meet business rules along with legal and governmental regulationsxxviii. The goal of data security compliance regulations is to provide a set of rules and guidelines that help organizations protect their systems and data from security risk and help companies achieve integrity, security and availability of information systems and sensitive data. These rules can be industry standards, state or federal-level laws or national regulations that specify what types of data need to be protected, what processes

should be considered acceptable under the legislation, and what the penalties will be if not followed by organizations.

A number of laws and regulations have been put in place by governments and industry that are focused on data protection, including the General Data Protection Regulation (GDPR) , Health Insurance Portability and Accountability Act of 1996 (HIPAA), Payment Card Industry Data Security Standard (PCI DSS)[xxix]. The General Data Protection Regulation (GDPR) was in effect by the European Union to deepen and harmonize personal data protection regulations. GDPR lays out a range of rules regarding people's right to know what data businesses have on them, how companies should go about processing this data, and tighter rules on the reporting of breaches. It applies to all companies who collect and process personal data on EU residents, regardless of where they are based.

3.2.4 Data Discovery Mechanisms

Data discovery is the process of finding data that is valuable for a given use case. The keys to successful data discovery among organizations include effective handling of metadata and provenance[xxx]. Searching for data in a distributed environment makes use of vocabularies, metadata and provenance. Vocabulary defines meaning of data items in a controlled manner and organizes data in a structured hierarchy, metadata associates data elements with vocabulary terms, and provenance provides the lineage of data as when, how, where, and by whom the data is produced. Effective data discovery includes three basic elements, a) a controlled vocabulary or an ontology, b) standard encoding for metadata and mapping of data to ontology terms, and c) efficient search mechanism that makes use of the semantics, metadata and provenance of data sources.

3.2.5 Standarized data exchange mechanisms (IDS type) vs ad-hoc data exchange mechanisms

A central issue of ad-hoc data exchange mechanisms is that each different organization defines its own format and specification for the data that is being provided. The ad-hoc nature of provided formats makes it harder and slower to exchange data between different sources.
Lack of standard data representation and exchange mechanisms leads to the necessity of development of customized data integration and exchange technologies/applications for every source and every instance of business process.

Data exchange between businesses could be carried out in a secure, flexible, and reliable manner if the involved parties use standard exchange mechanisms (either proprietary or open standards) such as Industrial Data Space (IDS) or Industrial Internet Consortium (IIC). In order for data to unfold its potential for value creation, it must be described and traded according to a global and interoperable standard. For instance, IDS Association provides a reference architecture that enables the sovereign exchange of data with clearly defined usage rights. It defines a technical infrastructure and includes contractual regulations at the semantic level, data linking, or analysis can be allowed or prohibited[xxxi].

3.2.6 Metadata

Metadata provides information about other data, and makes finding and working with particular data instances easier. In the energy domain more efficient processes can be built by creating and consuming rich and powerful metadata. During a data integration process, metadata about the sources enhance the relationships and business definitions assigned to entities and their attributes and relationships. Furthermore, metadata about an original data storage provides the information about the different data layouts and models. Metadata describing the streaming data supports collecting the details about the sources from where the data originated, as well as the transformations applied on it. Such data will enable the ingestion of big data, as well as the resolution of the interoperability conflicts across all the relevant data sources and their representation in the integrated knowledge base. We consider two types of metadata:

- Active metadata: Active metadata is a collection of metadata and annotations that is used actively within the project or process that generates it and is capable of being reused within that project or by another project.
- Passive metadata: Passive metadata is not connected to the data and is stored separately, and is consulted only by humans

3.2.7 Pre-data exchange (on-boarding pattern processes)

In the case of Platoon, where the data will be used by the Data Analytics Toolbox, before data can be processed, it may be necessary or helpful to perform some sort of data pre-processing or normalization. There are two types of pre-processing tasks depending on whether these can be performed locally without information from other end users, or globally which would include information about other users (with privacy).

- Data merging and scaling algorithms: "Ad-hoc" local pre-processing algorithms result feature vectors where after a pre-processing algorithm has been used the raw data is transformed into useful numerical vectors. The algorithm would need to be shared so that every end user is able to transform its own raw data into a common representation. It is important that an output vector with the expected content and format is produced. Once the input data has been transformed into a manageable feature vector, a second level of pre-processing/normalization is necessary taking into account all the data from all the users. The global mean depends on all the patterns from all the users, which is where the privacy preserving mechanisms come into place to avoid revealing individual values when estimating said global mean.

- Data alignment: aims to detect if all contributing users are providing data that serves to the same task. It is a preliminary check.

3.2.8 Post-data exchange pattern processes

Once data is transformed, data needs to pass through different processes before being used by target applications. Data needs to be validated of the constraints and checked whether it is up to the data quality standards. If data is valid and passes the quality test, then it can be integrated into the existing knowledge base, so that it can be explored and exploited. Different exploitation mechanisms can be applied on integrated data, such as query processing and analytics. Given the evolving nature of data, to ensure consistency and freshness, updates need to be propagated regularly and maintained its validity through time. Furthermore, data could have usage restrictions put by the data owner. Such restrictions should be respected and maintained accordingly.

3.2.9 Types of data usage restrictions/ data owner's usage policy

Data owners put different types of usage restrictions. These restrictions on usage controls collection of data, transfer/sharing of data, and the inferences that can be made over the data. Such restrictions may be determined by a contract, by statue, by custom, or by common decency[xxxii]. Sharing data and information needs to be balanced and controlled to maximize its effect, as this will facilitate organizations in establishing close connections and harmonization with their business partners.

3.2.9.1 Data policies

Data policies are norms that regulate management and publication of data range from recommendations to enforcements. The scope and content of such policies varies across organizations, countries, and disciplines. In order to promote sharing of data, mechanisms need to be put into place to provide the data owners a means to control how their data is being processed and used[xxxiii].

3.2.9.2 Data pricing model

Data marketplaces enable completely new business models where data can be considered as a commodity to be bought and sold in the market[xxxiv]. There are different data pricing strategies utilized by organizations/data owners for maximizing profit. Muschalle et al.[xxxv] present four main categories of data pricing models: 1) free data usage model, 2) usage-based pricing model, 3) package pricing model, and 4) flat fee tariff model. In free data usage model data is available in public storage and marketplace for free, in hope that it attracts customers/consumers into buying or paying for the complete data or other premium services on it. Usage based pricing model considers the measurement of each single data commodity and time counts for pricing, e.g., fee varies on the number of API calls at peak time to during normal hours. Package pricing model refers to a pricing model that offers a customer a certain amount of data or API calls for a fixed fee. The flat fee tariff is one of the simplest pricing models with minimal transaction costs based on time as the only parameter. Hence, it provides safety in planning future activities. On the other hand, it lacks flexibility for data consumers.
Other pricing models can be used by combining the four main pricing models. For instance, a two-part tariff pricing model is a combination of package pricing with a flat fee tariff model. In this scenario consumers pay a fixed basic fee and on top of that an additional 'fee per unit consumed'. Another pricing model is known as Freemium which provides access to basic or limited services for free but charges for premium content and services. The payment model for additional content and services can be according to one of the pricing models.

3.2.10 Barriers to data exchanges

Participation in data sharing and exchange remains difficult as data producers feel their ownership rights cannot be guaranteed. Different strategies and governance models have been proposed to overcome this challenge, including different ownership models and data rights management. Interoperability issues[xxxvi] are another barriers to data exchange between applications that process different data models. Interoperability is one of the key aspects in every complex system that deals with heterogeneous data and protocols. Different platforms based on different standards have to interact in terms of communication protocols, but also in terms of data structures, semantics, software and hardware.

In order to realise a cross-border data space with platforms capable of processing mixed proprietary, personal and open public data, there are technical and legal compliance challenges that we have compiled and are detailed in the table below:

Barrier	Description
Handling and distribute sensitive information in an appropriate way (GDPR)	All actors must handle (e.g. within the smart grid) and distribute sensitive information in an appropriate way, following GDPR principles, which impose restrictions on the capture, storage and distribution processes that should be carefully analyzed by the partners. • Purpose limitation principle: only allows the use of the data for the purpose it was collected. • Data minimization principle: data collected and processed should not be held or further used other than for the original purpose. The techniques employed to ensure privacy could be: • Physical: system creates a logical boundary that does not allow data flow.

	• Logical: information is sanitized before exchange (i.e. anonymization).
Data life-cycle management that is not designed around sharing in the electric sector	Existing data life-cycle management models need to improve how they incorporate all relevant processes, including preparing data for sharing, and finding the right data. The maturity of data services (e.g. cleaning, aggregation) in data sharing ecosystems is as crucial as the availability of the data itself.
Managing and respecting data ownership	Marketplaces rely on transferable ownership of data, so different ownership models or suitable data rights management frameworks have to be explored.
Decentralized data sharing and processing architectures	Standard data exchange protocols in decentralized architectures are required.
Weak verification and provenance support	Data veracity and traceability is crucial, so advanced provenance is required.
Secure data access, storage and restrictions	Secure control access and standardized security solutions and exchange protocols are required to enable a trusted network. Strict access right policies should be defined with user-based classification and complex authentication systems. GDPR defines storage minimization principle, where sensitive data is to be kept in a way that allows identification of data subjects for no longer that necessary for the specific purpose. Data store must be tailored so that the request of the end users rights is possible and do not break the system. In real energy projects, the mechanisms to provide end users a way to request for the application of their rights should be implemented and made publicly available.
Maturity of privacy-preserving technologies for big data	Current technical solutions for secure and trustworthy data sharing are in place and in continuous development. However, the uptake is lagging, so a more flexible uptake needs to be explored.
Legal blockers to free-flowing data	Free data flow across Europe is not in place yet, and legal matters surrounding data ownership, access, portability, retention, etc. need to be explored
Uncertainty around data policies and regulations	Inadequate regulation holds back development while preventing progress from happening. Questions on how to incorporate and adjust for the effects of the regulatory landscape within the Digital Single Market and specific to the energy sector need to be explored.
There is no common strategy for data management model.	There are two types: • Message-based: the raw physical data repository is only accessed by some data services that allows other processes and services to query for pieces of information. • Share database model: the data repository is unique and both ends exchange information to read, write the same resource.

Table 1: Data exchange barriers

3.3 Data privacy and security requirements

3.3.1 Data security

Article 32 of GDPR[xxxvii] states that the data controller shall take into account the "state of the art" about data processing security from a technical and organizational point of view. These technical and organizational measures refer to: pseudonymisation and encryption of personal data; the ability to ensure confidentiality, integrity, availability and resilience of systems and services; the ability to restore the availability and access to personal data; and a process for regularly evaluating the measures for ensuring the security of processing.

TeleTrusT in collaboration with ENISA (European Union Agency for Network and Information Security) provides a detailed analysis of the "state of the art" about processing security[xxxviii]. The analysis focuses on the following objectives:

- **Availability:** this concept refers to the ability of the user to access an information asset.
- **Integrity:** integrity is guaranteed when the data sent reaches its recipient complete and unchanged.
- **Confidentiality:** this principle states that sensitive data is only made available to authorised person.
- **Authenticity:** it ensures the unique identity of the communication partners.

Security solutions include broad issues on security in network, cloud, platform, application and IoT devices, and, consequently, differ in each case. Therefore, it is important to guarantee the detection of attacks according to the state of the art. ENISA in its "Threat Landscape report 2018"[xxxix] lists the main threat that may be used by an attacker to exploit vulnerabilities and to access persona data. Among the threats reported: **Web Based Attacks** use web systems and services for comprising the target and examples of this threat are man-in-the-browser[xl], an alternative of the man-in-the-middle[xli], and watering-hole[xlii] attacks; **Web Application Attacks** that exploit a vulnerability in the service an application on the web by using their APIs or services and examples of this threat are SQL injection[xliii] and cross-site-scripting attacks[xliv]; **Distributed Denial of Service** (DDoS) focus on making a resource unavailable for the purpose it was designed and is one of the highly impactful threats that takes advantage to the increasing dependency of IoT devices and APIs.

The basis of many IT security measures relies on cryptographic procedures, for instance authentication and authorization, access control and anonymization procedures. In the TeleTrusT report are recommended the following cryptographic algorithms:

- Symmetric encryption: AES-128, AES-192, AES-256. Is recommended to use with GCM mode[xlv] or EAX[xlvi] mode.
- Asymmetric encryption: ECIES-250 (384 bits or more), DLIES-2000 (3072 bits or more), RSA 2000 (3072 bits or more), curve25519, curve448 or ECC Brainpool.
- Hash functions: SHA-256, SHA-384, SHA-512, SHA3-256, SHA3-384 and SHA3-512.
- Key derivation functions (KDF) and password hashes: Argon2, PBKDF2, scrypt and bcrypt.
- Transport Layer Security: TLS 1.3 with forward secrecy.

Cryptographic procedures may suffer of side-channel attacks that take advantage of physical parameter rather than trying to exploit the weakness of the algorithm itself. Moreover, with the advancing of quantum computers someone might succeed in performing brute-force attacks in a much shorter time. Therefore, is recommended to check cryptographic measure at least once a year.

Following sections describes techniques for Secure communication, Data access control and Data usage control.

3.3.1.1 Secure communication

Secure communication among parties it is crucial to avoid data leakage. Unsecure communication may result in the capture of plain-text credential by an attacker. Several techniques can be adopted to ensure secure communication among parties. In order to guarantee, for instance, the identity of the communication parties and the authenticity of the transmitted contents it is recommended to rely on a Public Key Infrastructure (PKI). Applications of PKI are, among the others, digital signature that are used to ensure authenticity and integrity for sharing documents and HTTPS that is used to exchange encrypted data and assure confidentiality, authenticity and integrity of the data. Other approaches that can guarantee secure communication are VPNs by which the data transported is encrypted and the endpoints that belongs to the VPN are authenticate and authorised among each other.

Blockchain can be used to secure communication among parties. Khacef et al. propose solution using blockchain as a distributed ledger of identity and their associated public keys[xlvii]. The proposed approach uses the blockchain as a PKI removing the central authorities and taking advantage of smart-contracts. The authors use the blockchain to store public keys, digital signature and peer information. Moreover, they state that this approach will guarantee confidentiality, message integrity and authentication, and reliability. The limitation of this approach are the performance of the system and the scalability of the system with the growing number of smart-contract.

Following this approach, a blockchain-PKI based solution for IoT devices taking advantage of edge nodes is proposed[xlviii] where IoT devices are connected to edge nodes within a location. In this solution, edge nodes are used for verifying and validating the transaction. IoT devices are grouped by location and registered to an edge node providing its device ID, a public key and a timestamp. Further, the Edge node create a transaction that is shared with the neighbouring nodes which validate the transaction and add the same to the blockchain. When an IoT device needs to establish a communication with a device in other location, it contacts its edge node that verifies the device's public key based on the latest transaction.

3.3.1.2 Data access control

Authentication and authorization capabilities are critical aspect to support services and applications. An access control policy is defined as sets of conditions that define whether users have access granted to a protected resource. The authorization function can support different mechanisms, such as Access Control List (ACL), Role Based Access Control (RBAC), Attribute Based Access Control (ABAC), etc. Among the authentication and authorization standard solutions we can mention Security Assertion Markup Language (SAML), eXtensible Access Control Markup Language (XACML), Open Digital Rights Language (ODRL), JSON Web Token (JWT), OAuth2 and OpenID.

XACML architecture[xlix] comprises the following set of components:

- Policy Enforcement Point (PEP): which intercepts the requests to enforce access control on resources. This component forwards the request to the PDP and wait for its result.
- Policy Decision Point (PDP): which evaluates the policies and returns the authorization decision.
- Policy Information Point (PIP): which finds any missing attributes coming from the PEP and provide such attributes to the PDP to evaluate the policy.
- Policy Administration Point (PAP): which is used to create and manage the policy.

Most of the solutions take advantage of a centralized trusted entity that is in charge of managing access control logics. A decentralized access control solution for IoT devices using blockchain is proposed in the FairAccess framework[l]. In the workflow, depicted in Outchakoucht, Aissam & ES-SAMAALI, Hamza & Philippe, Jean. (2017)[li], a subject who wants to perform an action on a protected resource submit the request to the so-called authorization management point (AMP) acting as a PEP. This component

creates a so-called GetAccess transaction that is broadcasted to the network of nodes with the aim of reaching miners, that act as a distributed PDP, which accept or reject the transaction. After evaluating the request, the PDP executes a SmartContract already deployed in the blockchain thanks to previous transaction called GrantAccess. The execution of the SmartContract defines if the request should be permitted or not, and if it is permitted it provides the sender with an access token through an AllowAccess transaction.

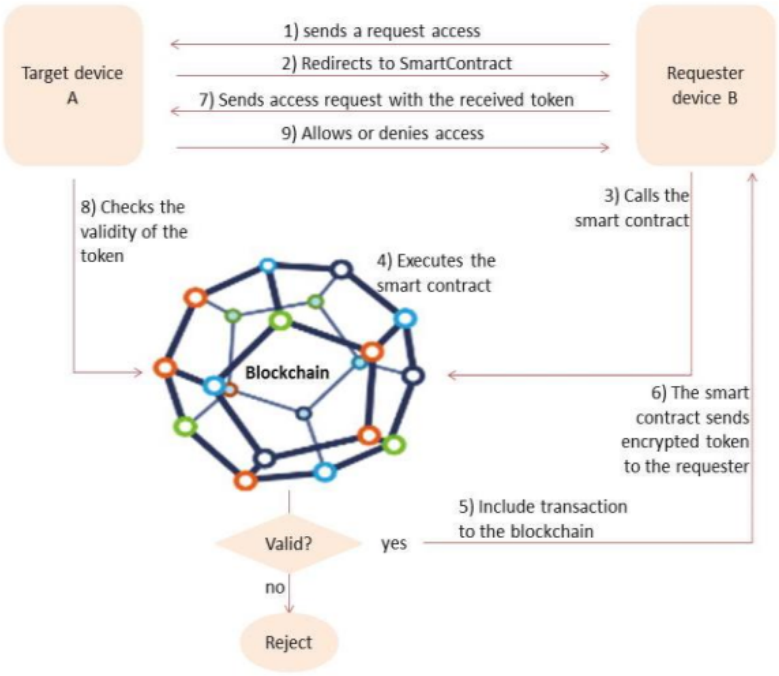

Figure 1: FairAccess Framework Workflow

3.3.1.3 Data usage control

Data usage control refers to the policies that must be applied in the data processing phase rather than data access. Hence the object is to control how data is treated and to assure that third-party entities use data in the way they are allowed to. UCON model[lii] is considered as the starting point for the development of data usage control.

In IDS the usage control is considered as an extension of the access control[liii]. Moreover, the usage control goal is to enforce the execute policies to data after the access has been granted controlling how data is processed, aggregated or forwarded. The control is demanded to a PEP component that intercepts the data flow and a PDP that is in charge of evaluate the policy and provide to the PEP the corresponding authorization decision. How the data should be treated is represented as event with attributes linked to data, but the decision may also depend on additional information that, as for the access control case, is provided by a PIP. Finally, a Policy Execution Point (PXP) is added to perform action based on the policy rules (e.g. send an email). The definition of the usage restrictions is demanded to a PAP. The management of such restrictions is demanded to a Policy Management Point (PMP) which manages, among the other actions, the instantiation and revocation of the policies. IDS list two possible ways of attaching policies to data in the first one the policies stick to data when it is exchanged (sticky policies) and in the second one the policies are stored independently from data in the PMP. The next figure illustrates IDS Data usage control flow.

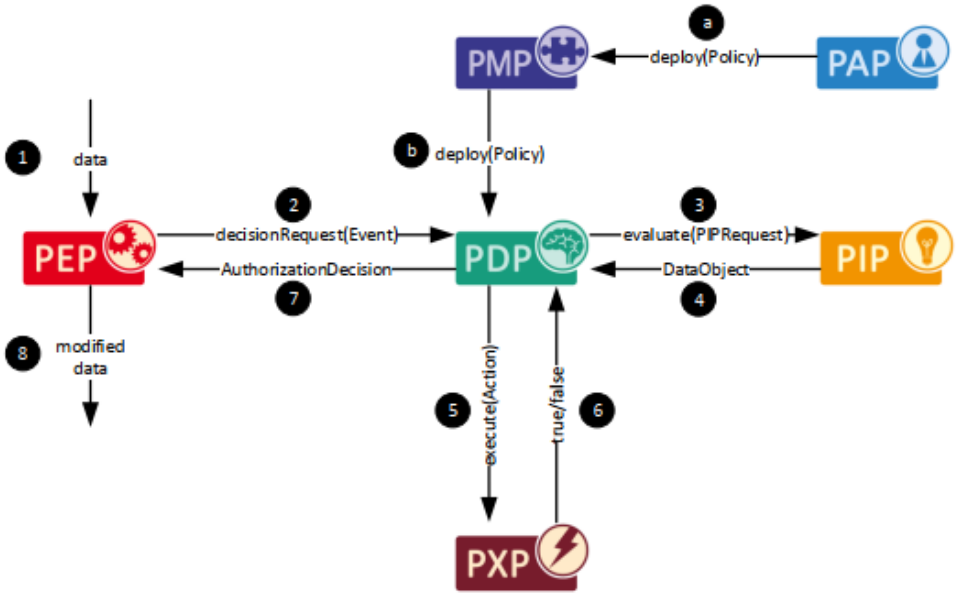

Figure 2: IDS Data Usage Control Flow

3.3.2 Data privacy

The General Data Protection Regulation[liv] (GDPR), Regulation (EU) 2016/679, introduced in May 2018, provides regulation to help people to have more control over their personal data, including right to access, erasure and portability. Personal data is defined as *"any information that relates to an identified or identifiable living individual"*[lv] meaning that any information that can be used to identify a person falls in the context of personal data, even if this has been de-identified, encrypted or pseudonymised. Personal data anonymised in such a way that the individual is not identifiable is no longer considered personal data.

Article 25[lvi] provides obligations for the data controller[lvii] about data protection by design e by default both in the form of appropriate technical and organizational measures and the necessary safeguards into the processing of the personal data following the data protection principles defined in Article 5[lviii] of GDPR Regulation. These principles are analysed by the European Data Protection Board[lix]:

- **Transparency:** the data subject must be clearly informed from the beginning of the process about how the data controller will collect, use and share her/his personal data.
- **Lawfulness:** in order to process the personal data, the controller shall identify a valid legal basis.
- **Fairness:** personal data shall not be processed in a way that is discriminatory or misleading to the data subject.
- **Purpose limitation:** the data controller must specify the purpose for the collection of the data and any other processing activities that are incompatible with the purpose shall not be execute.
- **Data minimisation:** the data controller must identify the minimum set of information about user that is adequate, relevant and limited to what is necessary in the process**.** Furthermore, as soon as the identification of a user is no longer needed (e.g. in statistics) the data shall be

anonymized, on the contrary, if the identification is needed for other processing activities, personal data should be pseudonymized.

- **Accuracy:** in order to avoid possible risk to the data subjects, the personal data shall be accurate and keep up to date.
- **Storage limitation:** data controller must ensure that identifiable data should not be stored for no longer than is necessary for the process.
- **Integrity and confidentiality (security):** the data controller must ensure that the process follows the current "state of the art" about data security in order to avoid possible data breach or vulnerabilities.
- **Accountability:** this means that data controller (e.g. companies/organizations) subject to GDPR are accountable for their handling of people's personal information.

Furthermore, Chapter 3 (art. 12-23)[lx] of the GDPR defines the rights of the data subject, that can be summarized as follow[lxi]:

- The right to be informed
- The right of access
- The right to rectification
- The right to erasure
- The right to restrict processing
- The right to data portability
- The right to object
- Rights in relation to automated decision making and profiling

Starting from the legal basis of data privacy, Hoepman[lxii] defines eight privacy design strategies that, further, are distinguishes into data-oriented strategies and process-oriented strategies[lxiii].

The following data-oriented strategies address the necessity[lxiv] and data minimisation principles:

- **MINIMISE**: this strategy strictly refers to the concept of data minimisation, therefore states that the minimum amount of personal data should be processed.
- **HIDE**: this strategy states that any personal data should be hidden from plain view meaning that the personal data shall be hidden from anybody not allowed to see or process the data.
- **SEPARATE**: this strategy states that personal data should be processed in a distributed fashion in order to limit the chance of a complete profile disclosure.
- **AGGREGATE**: this strategy suggests processing personal data at the possible highest level of aggregation in order to let data become less sensitive.

About process-oriented strategies the following are defined:

- **INFORM:** strictly related to the concept of transparency, the subject must be informed about which information is processed, for what purpose, and by which means.
- **CONTOL:** this strategy states that the data subject must be able to control her/his personal data by means of view, update and delete the data. Furthermore, it states that data subject should be able to decide about using a certain system and control the information processed by the system.
- **ENFORCE:** this strategy states that a privacy policy compatible with legal requirements should be in place and should be enforced. This implies that technical protection mechanisms that prevent violation of the policies are implemented.

- **DEMONSTRATE:** this strategy implies that the data controller is able to demonstrate compliance with the privacy policy and any applicable legal requirements.

Data privacy requirements are defined in more detail in the deliverable D1.5.

From a technological point of view, Privacy enhancing technologies (PET) are a set of technologies and approaches that can contribute to enhance privacy and data protection. Cryptographic techniques are often used to provide data protection, for instance:

- Encryption only: offers confidentiality protection. Payload is protected using symmetric keys.
- Signature only: which offers source authentication, integrity protection and (when asymmetric digital signatures are used) non-repudiation. This uses either symmetric keys-based MIC or asymmetric digital signatures verified using source end-point certificates.
- Nested Sign-then-Encrypt: This is used in cases where encryption is required in addition to source authentication and/or non-repudiation using a source end-point certificate. A digital signature on the payload is signed first, and then encryptions is applied to combination of the payload and digital signature.

Furthermore, the Royal Society in its report "Protecting privacy in practice"[lxv] describes a set of five PETs identified as promising in to enable privacy-preserving computation. The first of the PETs analysed is **Homomorphic Encryption** that allows to execute operations on encrypted data whose result, when decrypted, matches the result of the operations performed on the pre-encrypted data. This technique provides confidentiality and can be used when there is no trust among parties and sensitive data should not be accessible. Existing variation of the Homomorphic Encryption are the Fully homomorphic encryption (FHE), the Somewhat homomorphic encryption (SHE) and Partially homomorphic encryption (PHE). Their differences rely on the number of operations that can be performed over the data. The second PET reported is the **Trusted Execution Environment** (TEE) that is a secure area in a main processor. This kind of PET addresses the problem of insecurity and exposure and is a hardware-based way to ensure that data cannot be read by an external component (e.g. a cloud server) since TEEs are designed to be isolated from the rest of the system. **Secure Multi-party Computation** (MPC) is the third PET described in the report, this protocol enables distributed computation guaranteeing the privacy among the involved parties. As for TEEs, MPC addresses the problems of insecurity and exposure, moreover, it addresses the risk of revealing sensitive information of person or organization. MPC solution can be used to implement Private Set Intersection (PSI) or Private Information Retrieval (PIR), that, respectively, allows to two or more parties to compare datasets without revealing them in an unencrypted form (PIS), and allows a user to query a database whilst hiding the identity of the data retrieved (PIR). The fourth PET is **Differential Privacy** which states that when a computation result is released, it should not provide more information about an individual than if that individual had not been included in the dataset. This PET addresses the privacy in disclosure concept. The last PET reported are the **Personal Data Stores** (PDS) these are systems that provide access and control over personal data to the relative individual. Individuals, in PDS, are able to decide the information they want to share and with whom. These systems provide transparency and agency to individuals and address the problem of aggregation, exclusion, disclosure and the risk of undesirably sharing information.

An important initiative which aims at transforming the current organization centric system to a human centric system where personal data is a resource that the individual can access and control is MyData[lxvilxvii]. It is a progressive approach to personal data management that combines digital human rights and industry need to have access to data. MyData Principles can be summarized in the following:

- Human centric control and privacy: Individuals are empowered actors, not passive targets, in the management of their personal lives both online and offline – they have the right and practical means to manage their data and privacy.

- Usable data: It is essential that personal data is technically easy to access and use – it is accessible in machine readable open formats via secure, standardized APIs (Application Programming Interfaces).
- Open business environment: Shared MyData infrastructure enables decentralized management of personal data, improves interoperability, makes it easier for companies to comply with tightening data protection regulations, and allows individuals to change service providers without proprietary data lock-ins.

MyData framework focus on the consent management, that refers to the Article 6 of GDPR[lxviii] which states about "Lawfulness of processing" that, among other concepts, "*the data subject has given consent to the processing of his or her personal data for one or more specific purposes*". MyData. Consent management is the primary mechanism for permitting and enforcing the legal use of data. Via MyData accounts individuals can instruct the services to fetch and process data in accordance with consents that the individual has granted to data services. In technical and legal terms, consent is equivalent to authorization.

3.4 Data governance requirements applied to an ecosystem of platforms

Data is a valuable resource in any digital, data-driven business and is necessary to enable participants to leverage the potential of their data within a secure and trusted business ecosystem. In a nutshell, Data Governance is concerned with data lifecycle management decisions that ensure the safe, fair and secure (determined by pre-defined rules, legislative requirements, etc.) handling of data within and across a network of nodes over which data is passed on to fulfil pre-identified data value chains. In a data ecosystem that involves data exchange between distinct entities to fulfil such data value chains, as in the case of the PLATOON use-cases and the envisioned energy marketplace, Data Governance shape the foundations of the architecture (e.g. decentralised, centralised or hybrids) and therefore requirements need to be defined and agreed on early on in the project.

Data Governance models define a framework of decision-making rights and processes with regard to the definition, creation, processing, and use of data. As PLATOON has identified the IDS as the of-choice 'de facto' standard for data sharing ecosystems, the project will adhere to the IDS governance model, which governs and determines usage rights of data exchanged within IDS-compliant ecosystems.

To ensure Governance and compliance, IDS defines roles, functions, and processes that need to be met by the business ecosystem to achieve secure and reliable interoperability. Next diagram, shows the basic interaction taking place in the IDS ecosystem. The decision rights are distributed among different roles in the ecosystem, to limit the influence of authority by an individual actor.

In IDS, the fundamental mode of communication between a Data provider and Data Consumer happens via an **IDS Connector**, a dedicated software component, which acts as a communication interface for both data providers/consumers and app providers. They use a common language, **Information Model**, to communicate with each other. The information model facilitates compatibility and interoperability, thereby all connectors can exchange data. The management of **metadata** specifies data about data and comprises both syntactical, semantic and pragmatic information. This is of particular importance in a distributed system environment.

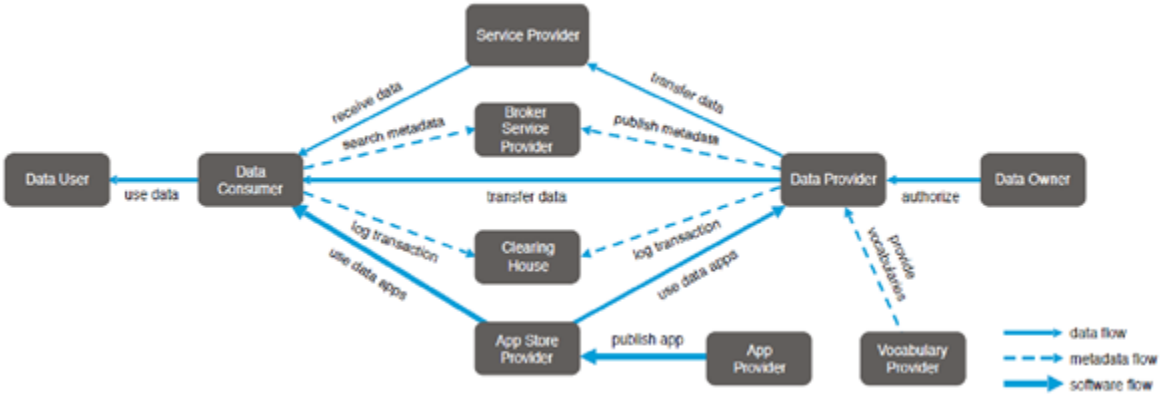

Figure 3: Basic interactions in IDS ecosystem

In order to realise an IDS-compliant governance model, the following list of requirements needs to be met by the various roles defined in the diagram:

- Data Owner/Data Provider
  - needs to define the usage constraints for data resources
  - publish metadata including usage constraints to the Broker
  - transfer data with usage constraints linked to data
  - receive information about data transaction from Clearing House
  - bill data, if required
  - monitor policy enforcement
  - manage data quality
  - describe data source
  - authorize data provider, if data provider is not the data owner
- Data Consumer
  - use data in compliance with usage constraints
  - search for existing datasets by making inquiry at Broker Service Provider
  - receive information about data transaction from Clearing House
  - monitor policy enforcement
- Broker Service Provider
  - metadata registry, to match demand and supply of data
  - provider registration interface for data provider
  - provide query interface for data consumer
- Clearing House
  - monitor and log data transactions and data value chains
  - monitor policy enforcement
  - provide data accounting platform

- App Store Provider
  - hosts all data apps in the data analytics platform
  - provide metadata and a contract based on the metadata for app user
  - provide interfaces for publishing and retrieving data apps
- App provider
  - register data apps at the App store using IDS connectors
  - the metadata and the contract bound to the data apps is registered at the App store
  - provides data apps that offer services such as data visualization, data quality, data transformation, etc.

As per the IDS governance model, in order to ensure trust in the PLATOON ecosystem, all Data Providers/Consumers, App Providers, Data Apps should undergo certification from a Certification body. It is to be decided who can fulfil this role in the envisaged PLATOON Energy Data Marketplace, both during (e.g., Steering Committee fulfils the role) the project and after the end of the project (e.g. appropriate independent entities that will by then be available).[2]

Next figure gives an overview of how the data governance fits into the PLATOON ecosystem.

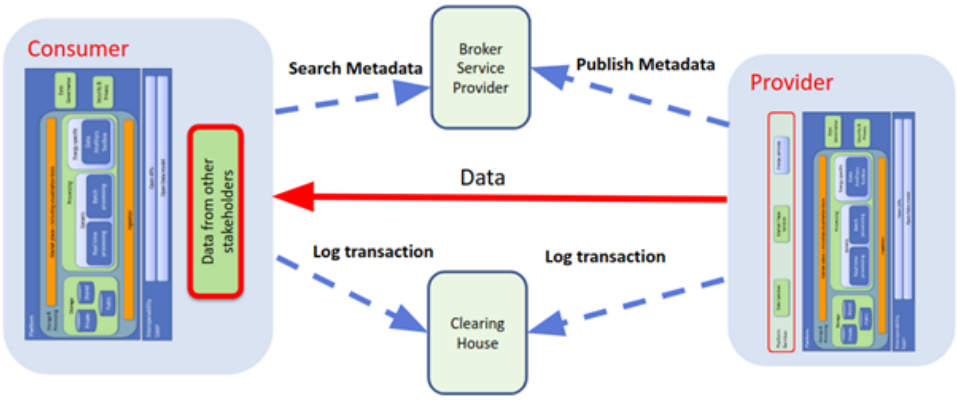

Figure 4: Data governance in PLATOON ecosystem

Similar interactions take place between a Data Provider/Consumer in need of a data app and app provider. The Data Provider/Consumer looks up at the App store for a specific data app and retrieves information about the respective App Provider. Afterwards, the data provider will contact the app provider, where they negotiate the contract bound to the data app. Once the contract negotiation is successful, the app provider shares the data app with the data provider, who is bound to use the data app as per the negotiated contract.

2 The IDS Association is working on providing Certification services for IDS-compliant data spaces. However, it is not possible to foresee the state or maturity of these services by the project's end or whether they are appropriate for the PLATOON data marketplace.

4 Requirements

The development of requirements has been carried out using the technique of user stories, due to in Platoon has been established, for technical WPs, that they will follow an agile methodology.

As we have mentioned, this phase was undertaken jointly by all the partners once the analysis phase was completed and the final list has been divided into five groups, which do refer to specific parts of the analysis performed.

For the elaboration of the user stories, a template was used that listed the different actors identified in the analysis phase(sections 2.1 and 3.1.6) and a series of objectives identified in the analysis phase (section 3.4 for example) complemented with what was expressed directly in the Grant Agreement.

To facilitate the understanding and review of the user stories, small simplifications have been introduced. For example, while IDS compliance is assured as much as possible, it is not needed to implement/realise all IDS roles and functionality for PLATOON (inc. the marketplace). Therefore, the IDS roles are reduced to a few less Platoon roles. In particular, as in PLATOON we do not foresee stories where data providers act on behalf of data owners - we do not see the need to distinguish between provider and owner in many of the user stories. We make the assumption that all providers are the owners (with one exception and initial entry that stresses the need for retaining control/sovereignty/rights management purely as a data owner). Similarly, roles like broker, clearing house are subsumed under a generic 'PLATOON Marketplace Operator' in several user stories, which provides all the intermediary functions necessary.

Some of these user stories will actually correspond to epics, which the different WP's will have to divide into user stories that they can undertake according to the duration of their sprints.

Also, in order to verify the fulfilment of the requirements, the user stories should be completed in the different WP's with acceptance criteria, which ideally should be based on the results that can be obtained in the different pilot projects.

The requirements tables contain the following fields:

- **Requirement ID:** 5 digit number starting with Deliverable number and then following consecutive numbers
- **Description:** requirement description in user story format (i.e. As [Actor Name] I want [Requirement] so that [Reason]).
- **Requirement Group :** Applicable deliverable: Pilot/Business, Data Exchange/Security, Platform, Legal/Ethics.
- **Requirement Type:** Functional / Non-functional
- **Applicable WP(s):** Downstream WPs that need to consider the defined requirement.
- **Mandatory/Optional**

- **Pilot Specific/General**
- **Specified in the DoA:** Yes/No.

4.1 Group 1: Data governance applied to an ecosystem

Requirement ID	Description	Requirement Group	Requirement Type	Applicable WPs	Mandatory / Optional	Pilot Specific / General	Specified in the DoA
12001	As a Platoon Vocabulary Provider, I want to create Platoon-energy related vocabularies so that I can contribute to open communication and integration.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12002	As a Platoon Data Consumer, I want to select a specific vocabulary from a Platoon Vocabulary Hub so that I can correctly interpret the data.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12003	As a Platoon Data Provider, I want to describe the dataset properties (e.g. data format, date and time of creation, dataset owner, metadata, etc.) and register the metadata at the Broker Service Provider, so that my dataset can be found through searches on its characteristics/metadata.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12004	As a Platoon Data Provider, I want to define a comprehensive pricing model for my datasets, so that I can establish different price types (e.g. pay per transfer, pay for access per day/month/year, etc.) so I can generate extra revenues.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12005	As a Platoon MarketPlace Operator, I want to be able to account usage of transferred and received data, so that I can perform clearing and settlement service duties for all financial and data exchange transactions.	Data Exchange / Security	Functional	All	Mandatory	General	Yes

12006	As a Platoon Data Provider/Consumer, I want to log data transaction details in the Clearing house and receive reports and statistics regarding transferred/received data usage, so that I can receive information about billing, correct use of datasets, demand and supply studies, pricing, be able to resolve conflicts, etc.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12007	As a Platoon Data Owner, I want to define a Data Usage Policy so that I can retain management rights, define rules and conditions on how data must be used by Data Consumers (e.g. who can see my data and which parts, prohibit forwarding to 3rd parties and other participants, merging data, the use that can be given to my data, etc.)	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12008	As a Data User/Administrator, I need a data management framework so that I can easily integrate and process data with different time resolutions.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12009	As a Platoon Data Provider/Consumer, I want to have an IDS-compliant Connector so that I can manage data and participate in the data exchange process as per IDS standards	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12010	As a Platoon App Provider, I want different data providers to use a common data model and APIs, so that I can easily use data from different sources to train my models without extra integration required and to be able to offer my models to different app consumers.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12011	As a Platoon App Consumer, I want different app providers to use a common data models and APIs, so that I can easily use models from different providers.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12012	As a Platoon Data Provider, I need to have a metadata based on IDS vocabulary so that I can explicitly define terms and conditions that guarantee data sovereignty.	Data Exchange / Security	Functional	All	Mandatory	General	Yes

12013	As a Platoon App Provider, I want data providers to convert data in a semantic format (RDF) by reusing/extending domain ontology or creating a new domain ontology from scratch, so that I can be able to derive and infer more information using data relationships.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12014	As a Platoon Data Provider, I want to publish different versions of a data source and mark versions as deprecated.so that my dataset is always up to date	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12015	As a Platoon Data Consumer, I want to search/query for a relevant dataset in the Broker Service Provider, so that I can find the relevant dataset useful for my business.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12016	As a Platoon Data Owner, I want to update my dataset´s metadata at the Broker Service Provider so that the dataset is up to date.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12017	As a Platoon Vocabulary Provider, I want to manage, edit, update, extend and publish different versions and mark versions as "deprecated" of Platoon-energy related vocabularies so that any modification results in a new version of the vocabulary in order to stay consistent with its users.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12018	As a Platoon MarketPlace Operator, I want to provide an registration interface for data providers, so that they can register their dataset´s metadata.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12019	AS a Platoon MarketPlace Operator, I want to store data provider´s metadata so that it will be visible to all participants within the Platoon ecosystem.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12020	As a Platoon MarketPlace Operator, I want to provide a query interface (optional: GUI) for data consumers, so that they can search for a specific dataset.	Data Exchange / Security	Functional	All	Mandatory	General	Yes

12021	As a Data Provider/Consumer, App Provider/Consumer, I want to have a unique identity in the Platoon ecosystem in the form of a certificate, so that secure and trusted connections to other participants can be established during the data exchange.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12022	As an App Provider, I want to be sure that companies I am going to share my models with meet minimum legal and IPR requirements so that they don´t make fraudulent use and exploitation of provided analytics tools.	Data Exchange / Security	Non-Functional	All	Mandatory	General	Yes
12023	As a Data Owner, I want clear definitions of possible licenses for the data I provide as open data, so that the usage limitation is clear.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12024	As a Data User, I want the API to be described in a standard manner (e.g. OpenAPI specification), so that it will be more transparent (e.g. data access, type of authorizations, possible responses, etc.)	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12025	As a Data Owner, I need to limit data access via API only for authorized entities, so that data is only available for intended parties.	Data Exchange / Security	Functional	All	Mandatory	General	Yes

4.2 Group 2: Data privacy, data security, data access rights and access to heterogeneous data

Requirement ID	Description	Requirement Group	Requirement Type	Applicable WPs	Mandatory / Optional	Pilot Specific / General	Specified in the DoA
12026	As a Platoon Data Provider, I want personal data to be anonymized so that I comply with GDPR.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12027	As a Platoon Data provider, I want personal data aggregated so that I keep confidential critical business keys.	Data Exchange / Security	Functional	All	Mandatory	General	Yes

12028	As Data Provider, I want companies that are going to have access to data, to meet certain minimum security requirements, so that I can be sure that my data is going to be safe outside my system and that GDPR is going to be complied with	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12029	As Data Consumer, I want the companies that are going to provide my data to meet security requirements so that they do not create cybersecurity threats to my system.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12030	As Data Consumer, I want to know the specific data privacy and usage requirements so that I can be sure to comply with GDPR and only use the data sticking to them.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12031	As a Data Consumer, I want to be sure that the companies that provide the data I am going to use meet ownership and sovereignty requirements, so that I have the permission to use data (e.g. data has been retrieved legally, not stolen).	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12032	As a Data Consumer, I want to be sure that the companies that provide the data meet quality and provenance requirements (original source and all the subsequent transformations), so that I can be sure that the models I develop based on the data are going to perform well when applied to other datasets.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12033	As Data Provider/Consumer and App Provider/Consumer, I want to be part of an ecosystem where all the stakeholders meet data quality, security and privacy requirements, so that I don´t have to check them every time I want to create a new data connection.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12034	As a Data Provider, I want my data to be encrypted every time it is sent to a Data Consumer, so that were there to be a malicious attack and the data is intercepted, they cannot extract any valuable information from my data.	Data Exchange / Security	Functional	All	Mandatory	General	Yes

12035	As a Platoon Data Provider/Consumer, I want to manage consent over personal data, so that I will be able to manage who can access and process my personal data.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12036	As a Platoon Data Provider/App Provider/Service Provider, I want to access control functionalities so that only authenticated and authorized people can access the data/services.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12037	As a Platoon Data Provider/Consumer, I want to comply with state of the art solutions regarding security, so that the data exchange process is secure.	Data Exchange / Security	Functional	All	Mandatory	General	Yes

4.3 Group3: Data exchanges analysis within electricity market and with other actors and platforms

Requirement ID	Description	Requirement Group	Requirement Type	Applicable WPs	Mandatory / Optional	Pilot Specific / General	Specified in the DoA
12038	As a Data Consumer / Energy Company (i.e. Energy production company, TSO, DSO, ESCO, etc.), I want to be able to access data from different companies within the same/different sectors, so that I can do benchmarking with similar scenarios in order to be able to assess competitiveness level.	Data Exchange / Security	Functional	All	Mandatory	General	No
12039	As a Data Consumer/ Equipment Manufacturer (OEMs, Tier 1 suppliers, etc.), I want to be able to access asset operational data, so that I can understand equipment behavior in real scenarios and be able to improve the design.	Data Exchange / Security	Functional	All	Mandatory	General	No

12040	As an App Consumer(energy company such as production companies, TSO, DSO, ESCO, etc.), I want to be able to use data analytics tools that have been already developed and validated by other companies so that I can get value from my data at low cost and low time to market and be able to use my time to focus on the critical parts of my business while offering a better service.	Data Exchange / Security	Functional	All	Mandatory	General	No
12041	As Data Provider, I want to manage/post-process/enhance the data produced at my facilities, so that I can assess the availability and performance of the plant (subsystems and assets) and its remaining lifetime, in order to improve operation programs and implement optimized maintenance strategies that will reduce operation and maintenance costs (OPEX) of the plants.	Data Exchange / Security	Functional	All	Mandatory	General	No
12042	As Data Provider/User, I want to have real-time access to the Renewable Energy production data and consumption demands and patterns, so that I can have a reliable overview on the grid capacity and provide services adapted to the needs of the different agents connected and/or involved in the grids (DR, frequency balance, etc.)	Data Exchange / Security	Functional	All	Mandatory	General	No
12043	As Platoon MarketPlace Operator, I want to have real-time access to energy consumption needs of customers I aggregate and represent with online information of the electricity wholesale market so that I can offer them an optimized service (energy quality and availability at best prices) for their specific quantities and timeframes.	Data Exchange / Security	Functional	All	Mandatory	General	No
12044	As Data Consumer, I want to have access to databases and analytical tools that can analyze my generation/consumption patterns so that I can better match my energy needs with the electricity I generate and demand from external suppliers to be able to optimize the return of my investment. (prosumers)	Data Exchange / Security	Functional	All	Mandatory	General	No

12045	As Data Consumer, I want to have easy access to big volumes of data (from renewable power plants and smart grids) at low cost, so that I can develop applications and can offer added value products or services (SaaS and PaaS).	Data Exchange / Security	Functional	All	Mandatory	General	No
12046	As Data Provider/Consumer (wind turbine OEM), I want to obtain data from different subsystems and components in the operational wind turbines I manufacture to be able to extract knowledge of how the performance of each of them is affected by others, so that I can acquire deeper knowledge of the technical features and performance of the critical systems and components and improve new model design.	Data Exchange / Security	Functional	All	Mandatory	General	No
12047	As Data Provider/Consumer (wind turbine OEM), I want to have legal access to competitor data so that I can offer an increased portfolio to customers (benchmarking, predictive maintenance, etc.)	Data Exchange / Security	Functional	All	Optional	General	No
12048	As a Data User (energy community), I want to access open data with open access and license, so that I can use the data to engage the renewable community (e.g. hackathon, journalism, policy making, etc.)	Data Exchange / Security	Functional	All	Mandatory	General	No
12049	As a Data User (energy community), I want to see a demo and tutorial of possible analysis (sample data provided), so that I can learn to use the analytic tool to perform analysis.	Data Exchange / Security	Functional	All	Mandatory	General	No

4.4 Group 4: Data needs for Artificial Intelligence applications

Requirement ID	Description	Requirement Group	Requirement Type	Applicable WPs	Mandatory / Optional	Pilot Specific / General	Specified in the DoA

12050	As a Platoon App Provider, I want to be able to receive raw data from different data providers, so that I can train my models with a wide range of scenarios and achieve a strong generalization capacity.	Data Exchange / Security	Functional	All	Mandatory	General	No
12051	As a Platoon Data Consumer, I want to use well-documented data (creation, conditions, missing values, domain specific knowledge treatment, etc.), so that I can select and fit the ML pipeline (pre-processing models and evaluation)	Data Exchange / Security	Functional	All	Mandatory	General	No
12052	As a Platoon Data Consumer (AI app), I need to have consistent and high-quality data so that I can produce useful and reliable results	Data Exchange / Security	Functional	All	Mandatory	General	No
12053	As Data User (energy community), I want to see data analysis show case and examples so that I know what type of specific analysis can be performed and what data type/structure/format is required for this analysis.	Data Exchange / Security	Functional	All	Mandatory	General	No
12054	As a Data User (ML Engineer/Data Scientist), I need the data to be properly labelled so that it can be used for development and testing of AI-based analytics.	Data Exchange / Security	Functional	All	Mandatory	General	No
12055	As a Data user, I want to be able to see if API stacking scenario can be done so that I can see what type of data mash up is possible.	Data Exchange / Security	Functional	All	Mandatory	General	No
12056	As a Data User (ML Engineer), I want an option of pseudonymization and data sampling, so that I can use the data for scientific disseminations which showcase opportunities and quality of the provided ML approaches of the generic Big Data Analytics toolbox.	Data Exchange / Security	Functional	All	Mandatory	General	No
12057	As a Data User (Data scientist/ML API developer), I want an easy set up (e.g. pipelines), so that API stacking can be performed correctly.	Data Exchange / Security	Functional	All	Mandatory	General	No

12058	As a Data User (Data Scientist) and Data Owner, I need an assessment method to see if the data is provided as a high-quality data (i.e.. data veracity ensured), so that we can use it to assess the data and avoid an undesired result cause by “garbage-in, garbage-out” concern.	Data Exchange / Security	Functional	All	Mandatory	General	No
12059	As a Data User (Data Scientist) and Data Owner, I need verifiable date, so that the provenance is clear.	Data Exchange / Security	Functional	All	Mandatory	General	No
12060	As a Data User (Data Scientist) and Data Owner, I need the data to be accompanied by metadata, so that it has a higher level of interoperability due to a more facilitated interpretation.	Data Exchange / Security	Functional	All	Mandatory	General	No

4.5 Group 5: Technological stack of data exchanges

Requirement ID	Description	Requirement Group	Requirement Type	Applicable WPs	Mandatory / Optional	Pilot Specific / General	Specified in the DoA
12061	As a Platoon Data Provider/Consumer, I want to be able to pre-process some of the data at the edge using IDS certified Data Apps, so that I can enrich/transform data making the whole process faster and reducing cloud service costs.	Data Exchange / Security	Functional	All	Mandatory	General	No
12062	As a Platoon Data Provider, I want to load data on the broker through punctual or batch operation so that I can take advantage of both of the solutions	Data Exchange / Security	Functional	All	Mandatory	General	No
12063	As a Platoon Data Provider/Consumer, App Provider/Consumer and Service Provider, I want to follow MIM´s paradigm (context information management API, shared data model and Marketplace API), so that interactions and interoperability among different systems is allowed.	Data Exchange / Security	Functional	All	Mandatory	General	No

12064	As a Platoon Data Consumer/Provider, App Provider and Service Provider, I want to be able to access/send the date from/to distinct data sources/cloud providers, so that vendor lock-in is avoided an interoperability among heterogeneous solutions is guaranteed.	Data Exchange / Security	Functional	All	Mandatory	General	Yes
12065	As a Platoon Data consumer I want to be able to aggregate, query messages based on existing widespread protocols such as Modbus, BacNet so that I am able to link that information together (relying on existing or developed ontologies) and be able extract or generate new knowledge.	Data Exchange / Security	Functional	All	Mandatory	General	No

5 Annex 1

As mention before, in the different annexes we have collected the specific solutions/technologies that can be used in the following WPs to meet the resulting requirements.

5.1 Architectures, legacy formats, interfaces and operating system of the energy system

5.1.1 Comprehensive Architecture for Smart Grid (COSMAG)

The main goal of COSMAG[lxix] is to understand how the energy system can transition to a Data Economy becoming part of the Digital Single Market. It is an ongoing work where new ideas and results can be added up to reach the right level of maturity.

The definition of COSMAG is based on a set of fundamental requirements, being two of them of special relevance for PLATOON:

- The architecture is built in such a way to offer "open gates", i.e. data interaction points that can be used for future expansions and novel use cases
- COSMAG does not introduce any new standards but rather exploits and collects results of previous projects or standardization activities.

COSMAG presents a preliminary overview of data exchange and data format in the Smart Energy sector.

5.1.1.1 Analysis of possible data exchanges

As a **Comprehensive Architecture for Smart Grid** (COSMAG) refers to the analysis and the collection of specifications, which are able to define possible **data exchange process among various possible actors**.

This exercise is intended to check if current standards offer the proper roles interfaces to enable business processes, including new ones and to identify where new standards may be needed.

The starting point of the analysis is the structure of the market and actor interactions diagram depicted in Figure1.

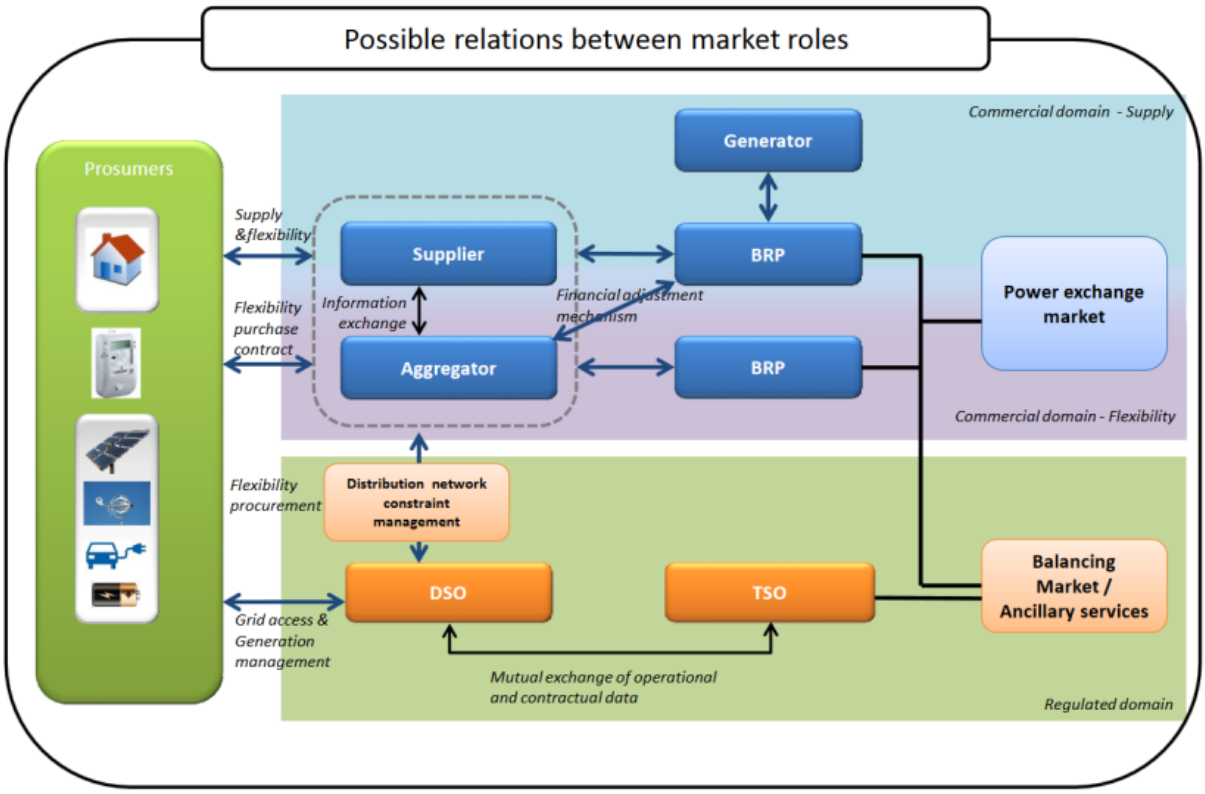

Figure 5: Market Roles

This figure shows the structure of the market and actor interactions in the context of the Deployment of Flexibility.

To further structure the analysis, COSMAG considers each actor separately and consequently all the interfaces for each of the actors. The following actors have been included in the analysis:

- Prosumer
- DSO
- TSO
- Supplier
- Aggregator
- Wholesale Market.

This list of actors is, in any case, not complete and **other actors could be identified**, such as community manager or data manager and others. A complete assessment of all the possible roles could be useful follow-up work, for example through a working group to be set up.

Furthermore, the increasing number of connected objects and the availability of data may require an assessment of new roles and actors in a digitalized energy market.

5.1.1.2 Semantics and protocol of communication

COSMAG provides information about the protocols and data modes available for the different data exchanges, including the following:

- **SAREF**: Initially for smart appliances (to be extended to other energy domains.)
- **Saref Extension for Energy**: SAREF4ENER is an extension of SAREF for the Energy domain. SAREF4ENER focuses on demand response scenarios, in which customers can offer flexibility

to the Smart Grid to manage their smart home devices by means of a Customer Energy Manager (CEM). The CEM is a logical function for optimizing energy consumption and/or production that can reside either in the home gateway or in the cloud. SAREF4ENER is published as an ETSI tecnical specification (ETSI TS 103 410-1

- **FIWARE Smart data models**.
    - Energy
    - Device
- **IEC61850** both as an automation protocol and as a data model for substations.
- **Common Information Model (CIM**)is a complete data model for power system used to exchange also data among grid operators both at TSO and DSO level (IEC61970-301 and IEC61968-11)
- **OpenADR**. IEC has approved **OpenADR** as Publically Available Standard (PAS) (IEC/PAS 62746-10-1) [14]. As part of this process OpenADR data model has been also mapped to CIM. This process is part of the wider IEC work PC118 (Smart Grid User Interface). Recent work from TNO has shown the possibility of integrating the **OpenADR** approach with **SAREF**.
- **SmartMeter**: Three protocols emerged as standard in this area: M-BUS, DLMS/COSEM and SML.

One of the interesting aspects offered by the availability of data platforms is the possibility to create open **interfaces** that could be exploited by third party providers which can bring innovative services to the energy domain. **Open API** offers the perfect bridge between private infrastructure spaces.

COSMAG includes **ETSI CIM standards** as a common language to exchange context information. The Application Programming Interface Specification/API (named NGSI-LD with OMA authorization) aims to enable applications to discover, access, update and manage data and context information from many different sources as well as to publish it through interoperable data publication platforms like Open Data platforms.

5.1.1.3 Differential characteristics

According to COSMAG, the goal of reference architectures is to contribute to the creation **of secure, trust and controlled collaboration spaces** in which existing and emerging technologies could better exploit in safe and trustable ways the data provided by energy actors and the insights derived by data innovators. Therefore, COSMAG introduces the concept of **Federated Data Solution Space** composed by:

- **Data valorisation Platform**: Data collection, integration and analysis.
- **Data Governance Platform**: Security and Privacy, Data Marketplace and Toolbox management.
- **Knowledge Warehouse**: Open, shared and private.

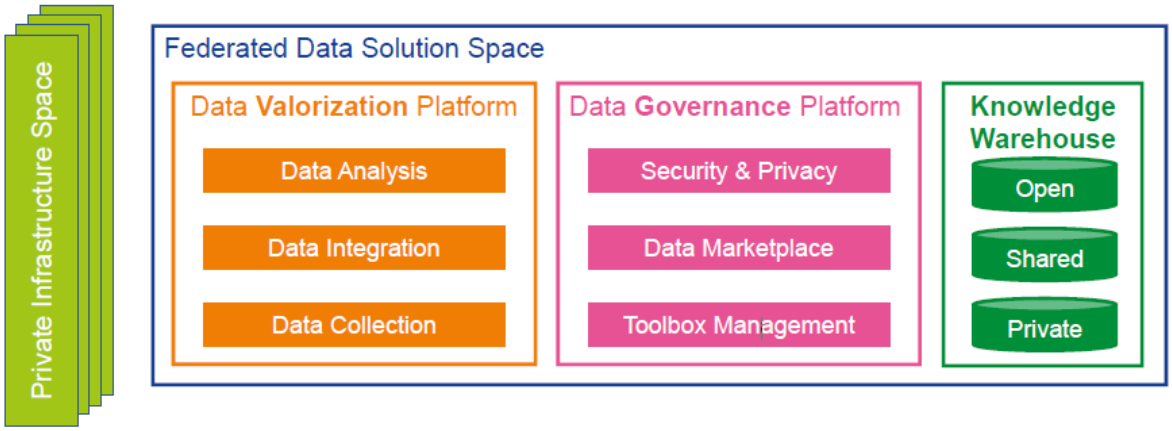

Figure 6: Federated data solution space

The goal of the Federation Data solution Space is to define a data collaboration framework that provides data and advance data governance and valorisation functionalities to every company, from big companies with their own big data analytics infrastructure to SMEs willing to provide new services to the energy domain.

COSMAG identifies several key elements in data management in support of a data economy, including:

- **Data business model**: definition of an appropriate data model beyond a single sector is a key ingredient for interoperability.
- **Data Commodity Monetisation**: how to create business models out of data in sectors such energy without violating privacy.
- **Semantic**: creating interoperability through clear and recognized data modelling.
- **Context:** the definition of the context is a key ingredient for bridging through different verticals .
- **Sovereignty**: to unlock the market, it is of key importance to offer the possibility to define who can manage the data and for which purpose to protect privacy and customer interest.
- **Open API**: close solutions will not create a real open and competitive market. Open API offers the perfect bridge between private infrastructure spaces.

5.1.1.4 Conclusions and further work

COSMAG presents a preliminary overview of data exchange and data format in the Smart Energy sector. Some conclusions of the analysis are very interesting from the PLATOON perspective:

- One important element to keep into account is the emerging role of sector coupling, making it critical to avoid data silos, not just between electricity, heat, gas etcetera within the energy sector, but also coupling of services with other sectors such as health, security, etcetera.
- It is important that data platforms will be based on open standards to support open competition.
- Data models are also a critical aspect. In this respect, SAREF extended to cover the whole energy value chain is a very valuable candidate.

Furthermore, according to COSMAG, the analysis included should be considered as an open draft to be continuously updated. Therefore, there is a good opportunity for Platoon to improve and complement COSMAG in the following key topics:

- assessment of possible new roles and identification of promising new cross-sector business models and services (e.g. energy and health in smart homes);
- Definition of a strategy for data platform management and integration deployment of those platforms in the energy architecture
- Integration of the IoT world in the energy context and vision

5.1.2 SAREF data modelling energy extension

SAREF[lxx] "Smart Appliances REFerence ontology" is an ontology developed in order to address the issue of multiple overlapping and competing standards within the smart home industry. SAREF ontology enables semantic interoperability between smart appliances. Thanks to SAREF different smart appliances from different manufactures can talk to each other through their common semantics while still keep using their own terminology and data models for its internal execution.

SAREF has been built and on a solid ontological foundation, based on DUL. Moreover, SAREF has mappings to the W3C SSN10 ontology, which is in turn related to DUL.

SAREF uses the concept of device, which is defined as “a tangible object designed to accomplish a particular task in households, common public buildings or offices. In order to accomplish this task, the device performs one or more functions”. For example, a washing machine is designed to wash (task) and to accomplish this task it performs the start and stop function.

A device must have some properties that uniquely defines it such as its model and manufacturer. In addition, it can have other properties such as, a description and location of the device within the building. Moreover, a building space contains several devices, which might interact with each other.

Buildings have also specific SAREF properties such as the type of space (e.g. living room). Building properties in SAREF are linked with the FIEMSER[lxxi] data model, which defines building related concepts, and takes into account other building-related approaches such as IFC.[lxxii]

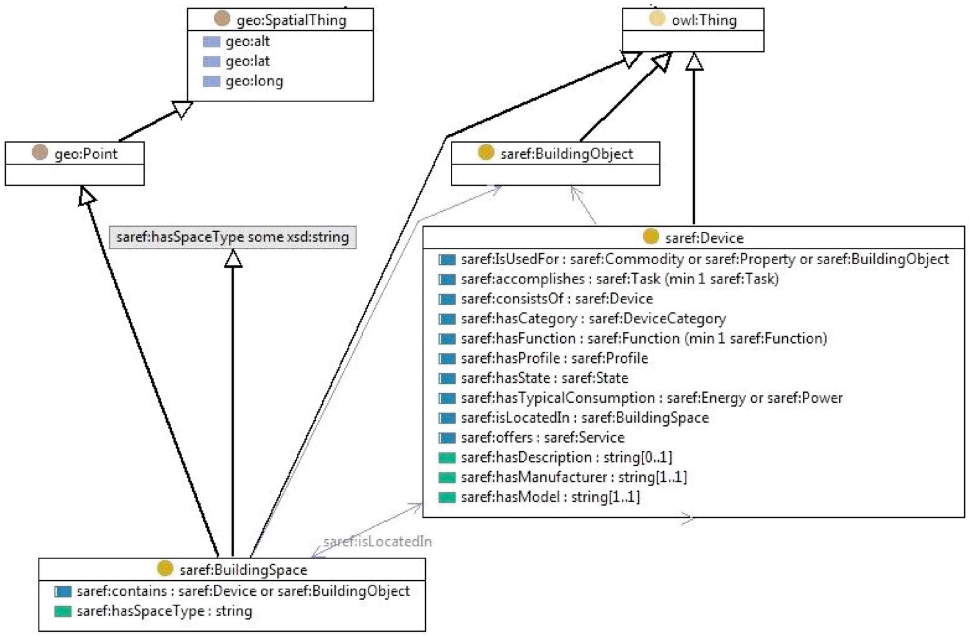

Figure 7: Device, building space and building object properties

On the other hand, devices have a category property, which is used to classify them into the following three different semantic groups:

1)Devices and sensors, and their specification in terms of functions, states, and services.

2)Energy consumption information and profiles to optimize energy efficiency.

3)Concepts coming from building related data models.

Devices must accomplish at least one function. A function is defined as “the functionality necessary to accomplish the task for which a device is designed”. For example, the Actuating Function allows to transmit data to actuators, such as level settings (e.g., temperature) or binary switching (e.g., open/close, on/off)”. Functions can be used for one of the following purposes:

- Offering a commodity, such as Water or Gas.
- Sensing, measuring and notifying a property, such as Temperature.
- Controlling a building object, such as a door or a window.

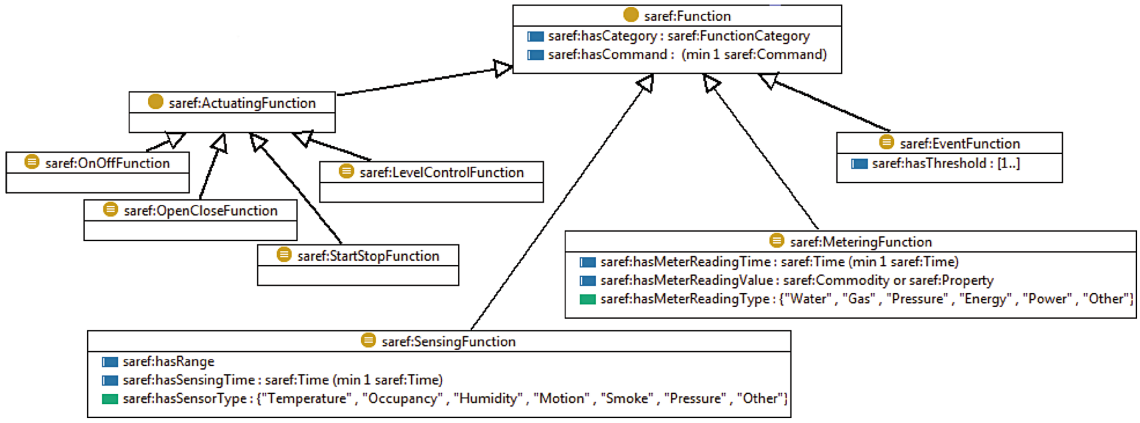

Figure 8: SAREF functions

Depending on the function, a device can have different states. For example, a switch can be found in the on or off state.

On the other hand, a device may consist of other devices. For example, a smoke sensor is a device that consists of a sensor which performs the Sensing function and Event function, and is used for the purpose of sensing a property of type Smoke and notifying that a certain threshold has been exceeded.

Besides, a device can offer a service. A service is a representation of a function to a network that makes this function discoverable, registerable and remotely controllable by other devices in the network. For example, a light switch can offer the service of remotely switching the lights in a home through mobile phone devices that are connected to the local network. This remote switching service represents the OnOffFunction, it must have an on/off state as input parameter, and it must have an off/on state has output parameter (“off” if the input state was “on” and vice versa).

In addition, devices can have a different load profile regarding its energy production and consumption (energy/power). Eventually, the production and consumption can be calculated over a time span and can be linked to a price. The time can be specified in terms of instants or intervals according to the W3C[lxxiii] time ontology.

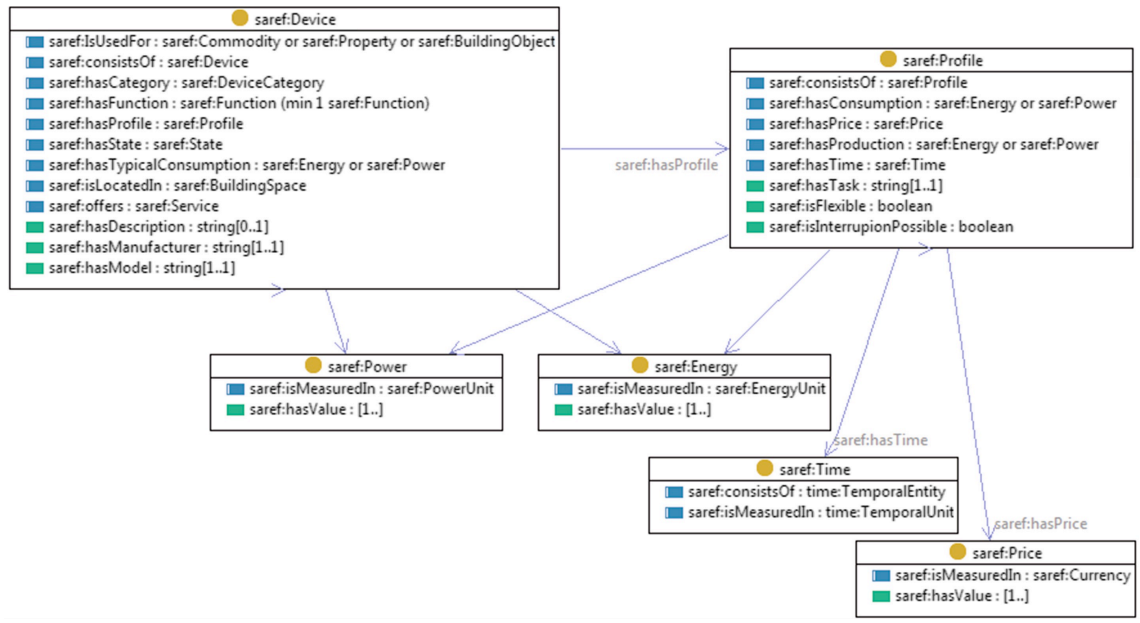

Figure 9: Device Profile, Power, Energy, Time and Price

Different ontologies for different semantics assets are publicly available in the SAREF google site[lxxiv]. Each ontology has a subpage that contains its description and a URL that points to the file in which the ontology is specified. The process of translating and matching the internal device ontology to SAREF ontology can be time consuming process. Therefore, it is recommended the use of tools for automatic translation and matching. This is definitely beneficial, but still may need a significant amount of human intervention.

In the context of PLATOON, the scope of SAREF was initially defined for smart appliances for smart home domain which applies to pilots 3a, 3b and 3c. However, during the last years SAREF has been extended for other domains such as **SAREF4ENER**[lxxv] and **SAREF4EE**[lxxvi] are an extension of SAREF for the Energy domain that was created in collaboration with energy@home[lxxvii] and EEBus[lxxviii], the major Italy- and Germany-based industry associations, to enable the interconnection of their data models. SAREF4ENER focuses on demand response scenarios, in which customers can offer flexibility to the Smart Grid to manage their smart home devices by means of a Customer Energy Manager (CEM). SAREF4EE focuses on the white goods industry (refrigerators, freezers, washing machines…). Additionally, **oneM2M Base ontology** is the extension of SAREF with the oneM2M initiative including standards for M2M and the Internet of Things.

5.1.3 ETSI Context Information Management (CIM) and NGSI-LD API

A “smart” service can be depicted as an interconnection of context providing services and context consuming applications. These work together to ensure that each application has the information it requires to deliver knowledge and insight, and to exercise control. The context of an application is composed by all the relevant aspects of its operating environment that are required for it to work as intended. Each application needs a different mix of data (context) from one or more sources. A context producer may be a sensor, a gauge, a database an open data repository, etc.

Context Information Management API allows users to provide, consume and subscribe to context information in multiple scenarios and involving multiple stakeholders.

ETSI NGSI-LD API defines a standard API for Context Information Management enabling close to real-time access to information coming from many different sources (not only IoT data sources). NGSI-LD API enables applications to perform updates on context, register context providers which can be

queried to get updates on context, query information on current and historic context information and subscribe to receive notifications of context changes.

NGSI-LD leverages on the former OMA NGSI 9 and 10 interfaces and FIWARE NGSIv2 to incorporate the latest advances from Linked Data. Latest version of NGSI-LCD specification can be found at https://www.etsi.org/deliver/etsi_gs/CIM/001_099/009/01.02.02_60/gs_cim009v010202p.pdf

NGSI-LD specification defines the so-called Context Information Management framework which includes:

- **The NGSI-LD Information Model**: Composed by the Core Meta Model, the Cross-Domain Ontology and Domain-Specific Ontologies:
- **A set of NGSI-LD Architectural options**: Centralized architecture, Distributed architecture and Federated architecture:
- **The NGSI APIs**: Context Information Provision, Consumption and Subscription and Context Source Registration, Discovery and Registration Subscription

In the ETSI ISG CIM framework, context information considers any relevant information about entities, their properties (temperature, location, or any other such parameter), and their relationships with other entities. Entities may be representations of real-word objects but may also be more abstract notions such as a legal entity, corporation, nation state, or groups of entities.

For example, a smart electric meter may be modelled as an entity of a defined type, installed in a house at a given location, measuring a dynamically changing power consumption, and connected to a particular distribution transformer. A Smart Energy application may require all this information as context.

The main considerations shaping the design of the NGSI-LD API are the following:

- The **information model and API should model context informatio**n such as entities, their properties, and relations. Communication protocols and other IoT system parameters are not explicitly modelled.
- **Context producers** should be able to register and update the broad categories of information they can offer. **Context consumers** should be able to discover relevant context information and receive notifications of updates.
- **Flexible query options** should be supported.
- Commonly needed cross-domain constructs such as time and location should be explicitly defined in the information model, to prevent minor variations leading to system incompatibilities. The **information model should be extensible** so as to allow the creation of domain/application **specific definitions and semantics**.
- A range of situations and architectures should be supported from the simple to the very complex systems with huge numbers of entities. Deployed architecture should be able to evolve, from **centralized** to **distributed** to **federated**, without needing to reinstall software implementations.

5.1.3.1 The NGSI-LD Information Model

The NGSI-LD Information Model prescribes the structure of context information that shall be supported by an NGSI-LD system. It specifies the data representation mechanisms that shall be used by the NGSI-LD API itself. In addition, it specifies the structure of the Context Information Management vocabularies to be used in conjunction with the API.

The NGSI-LD Information Model is defined at two levels: the foundation classes which correspond to the **Core Meta-model** and the **Cross-Domain Ontology**. The former amounts to a formal specification of the "property graph" model. The latter is a set of generic, transversal classes which are aimed at

avoiding conflicting or redundant definitions of the same classes in each of the domain-specific ontologies.

Below these two levels, **domain specific ontologies or vocabularies** can be devised. For instance, the SAREF Ontology ETSI TS 103 264 [i.4] can be mapped to the NGSI-LD Information Model, so that smart home applications will benefit from this Context Information Management API specification.

Next figure shows an example of the kind of concepts included in each level:

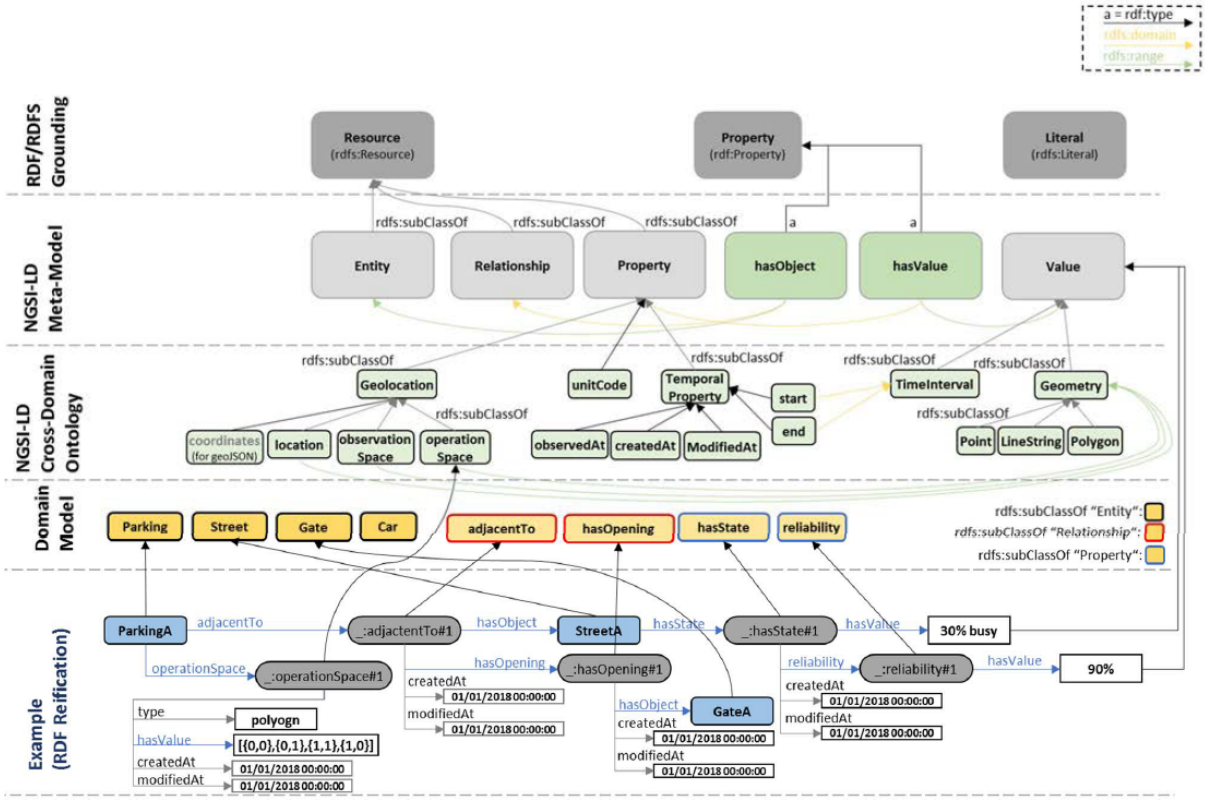

Figure 10:: NGSI-LD concepts

5.1.3.2 NGSI-LD Architectural options

The NGSI-LD API does not define a specific architecture. It is envisioned that the NGSI-LD API can be used in different architectural settings. Three prototypical architectures are the following:

Centralized architecture: This is a common architecture, to provide a Central Broker which acts as the central point of context management and storage for many context producers.

Distributed architecture: The actual information would be stored and provided by Context Sources, but applications can still access all information through the Distribution Broker. To make this work, Context Sources must register the information they can provide with the Context Registry

Federated architecture: This is a model for aggregation of NGSI-LD infrastructure in order to extend access to context information across multiple NGSI-LD systems.

5.1.3.3 The NGSI APIs

The NGSI-LD API supports a number of operations, with messages expressed using JSON-LD. It allows context consumers and context producers to interact with context information systems. The API

operations allow applications to discover, query and explore the graph-based data by specifying any combination of entities, types, relationships and/or properties as criteria for data queries.

One group of NGSI-LD operations allow Context Producers to create NGSI-LD Entities i.e. insert an object with a defined URI into the system, and to allow Context Consumers to retrieve and subscribe to Entities:

- **Context Information Provision**: a set of operations through which a Context Producer can create, modify, and delete an NGSI-LD Entity.
- **Context Information Consumption**: operations through which a Context Consumer can retrieve or query for NGSI-LD Entities. Queries can filter out Entities by Attribute Values (target value of a Property or the target value of a Relationship).
- **Context Information Subscription**: operations through which regular or event-driven update notifications of the context of one or more Entities can be created, updated, retrieved, queried for.

Another group of NGSI-LD operations allow Context Sources to be registered as potential sources of information meeting certain conditions. A Distribution Broker can query a Registry to ascertain which Context Sources may be able to provide the information requested.

- **Context Source Registration**: a set of operations through which a Context Source (i.e., the entire collection of information which it could provide) can be registered, updated, and deleted (removed from the registry). The registration information includes the types of Entities, Properties, and Relationships about which the Context Source can provide information, as well as geographic and temporal constraints on the information (e.g., “only in the region Germany”, “only for years 2017 and later”). For example, a particular Context Source could register that it can provide the indoor temperature for Building A and Building B or that it can provide the speed of cars in a geographic region covering the centre of a particular city.
- **Context Source Discovery**: operations through which a Context Consumer or Producer can retrieve or query Context Source registrations.
- **Context Source Registration Subscription**: a set of operations through which a Context Consumer can create, update, retrieve, query for, or be notified regarding Context Source registration subscriptions. In other words, subscribers may be notified about new Context Source Registrations that can potentially provide the requested information.

NGSI-LD OpenAPI specification can be found here:

https://forge.etsi.org/swagger/ui/?url=https://forge.etsi.org/gitlab/NGSI-LD/NGSI-LD/raw/master/spec/updated/full_api.json

5.1.3.4 NGSI-LD for PLATOON

According to ETSI, Context information exchange using NGSI-LD has three major advantages which can be very useful in the context of PLATOON:

- Firstly, within the NGSI-LD framework, **applications can flexibly discover and query relevant information**. The data discovery is dynamic, and the built-in query patterns support the most common questions that are practical in unbounded federated information systems.
- Secondly, **NGSI-LD helps to precisely communicate the nature of the context information for a given service**, such as its period of validity, its geographic constraints, and other semantically important information, by enabling direct inclusion of pointers to the relevant parameters and definitions. To **ensure interoperability**, the NGSI-LD API defines the meaning of the most commonly needed terms and provides the tools to create domain-specific extensions to model any other type of information.

- Thirdly, NGSI-LD provides a **scalable solution to connect, publish and federate diverse data sources** using a developer-friendly interface for data sharing and usage.

5.1.4 Modbus and BACNet open protocols for the End Use of Energy

5.1.4.1 Modbus protocol

5.1.4.1.1 Background

The Modbus protocol was first introduced in 1979 as a result of the development of Modicon (acquired in 1977 by Gould Electronics and then AEG 1989). The Modbus protocol is now, following the merger of AEG and Schneider, maintained by Schneider Electric. In view of the success and to support continuous development and a quality ecosystem, Schneider Electric has established the Modbus Organization (and transferred rights to it in 2004).

5.1.4.1.2 Definition

The Modbus protocol is primarily a message protocol between master-slave/client server that supports communication interactions between devices. The protocol is mainly designed to enable simple, reliable and communication between automation and field devices.

There are actually two main versions:

- one for the serial interface (RS-232 and RS-485);
- one for ETHERNET.

5.1.4.1.3 Scope of application

As mentioned previously there are three main different ways how the Modbus can be operated.

- Serial communication
  - Modbus RTU
  - Modbus ASCII
- Ethernet Communication
  - Modbus TCP

Some other versions of the Modbus Protocol exist as well such as Modbus Plus, Modbus over UDP or some version created by specific company et which subsequently became shared forms of the Modbus protocol, particularly in the oil & gas industries(Pemex, Enron).

The basic principle of the Modbus protocol is the deployment of a master and several slaves (for instance for measuring and controlling a system). Those two entities and sub entities can be connected via MODBUS. [lxxix]

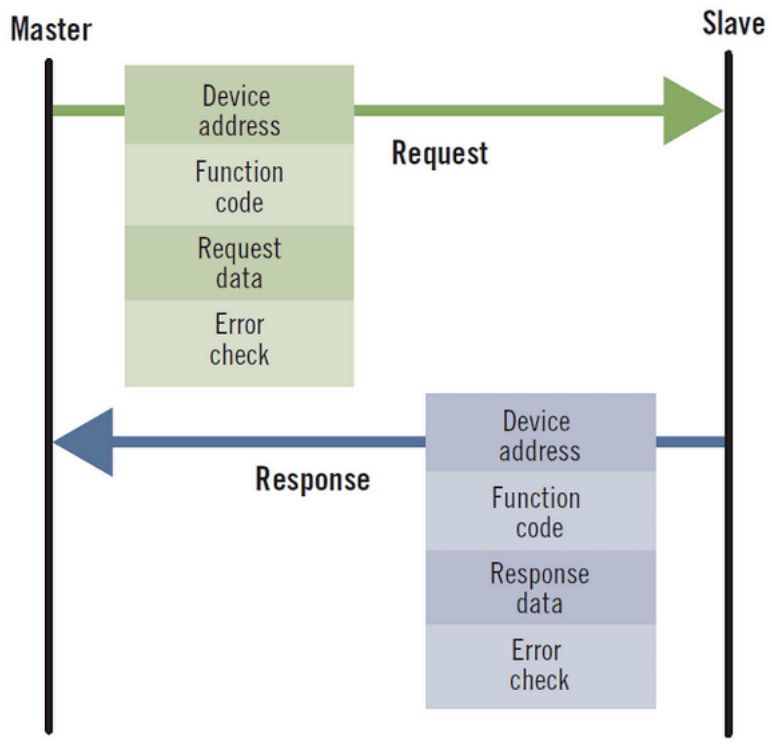

Figure 11: Example of Modbus TCP transaction

5.1.4.1.4 Main technical specifications

One of the most important aspects of the Modbus protocol and its message structure. Indeed, Modbus is above all a message communication protocol, the keystone of this protocol considers that Modbus messages are independent of the physical interfaces considered. This applies by definition to both RS-XXX (232-485) and Modbus TCP approaches. This could be considered as a form of interoperability defined at another time, by defining an abstraction at the level of the information transmitted and the structure of the messages. This allows hardware upgrades with a limited impact on the software brick.

Field	Description
Device address	Address of the receiver
Function code	Code defining message type
Data	Data block with additional information
Error check	Numeric check value to test for communication errors

Figure 12: Modbus Message Structure

If we consider the approaches with RS (232-485) type interface the messages are sent in plain text on the considered network, on the other hand they are integrated to packets in Modbus TCP/IP over Ethernet.

Since the global Modbus approach involves a Master to one or more Slaves, we find ourselves with two main architecture for Modbus , the Serial one in figure 2 and the TCP one in figure 3.[lxxx]

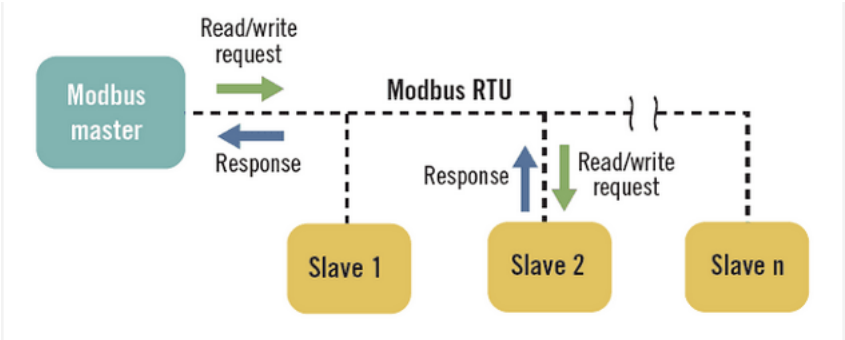

Figure 13:Modbus serial network architecture (1 Master up to 247 Slaves)

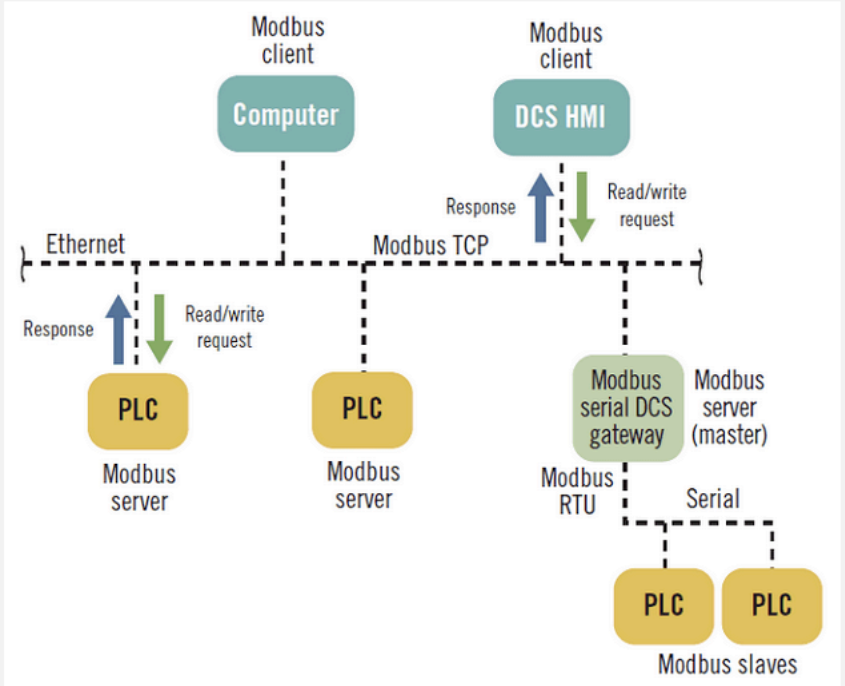

Figure 14:Modbus TCP network architecture (Client/server Approach with IP address)

5.1.4.1.5 Advantages and disadvantages

If you refer to Schneider Electric Frequently Asked Question Section[lxxxi], The advantages between Modbus RTU vs. Modbus approaches are following. First, the simplicity of implementing Modbus instructions in a TCP/IP environment. Second, the fact that this approach is based on a standard such as Ethernet, the associated hardware for the devices that will be brought to the Modbus TCP/IP protocol, are de facto based on known standards and references. Third, The Modbus protocol is open since its transfer to the Modbus Organization and finally all the documentation, code and others allow to reinforce a very large compatibility between the vendors using the Modbus protocol in multiple type of devices. One could therefore imagine that integrating interoperability approaches in the sense of those considered in PLATOON, if we relied on a communication protocol such as Modbus, would benefit from a robust base especially in Use Cases related to Building (even if it is necessary to consider BACNET[lxxxii], or KNX[lxxxiii] for automation in smart buildings).

Modbus remains a reference in process control systems, especially in the industrial field. However, one of the main weaknesses, more related to its architecture and initial bias, is that when using the

Modbus protocol, particular attention must be paid to the cyber security dimension. Indeed, there is no particular protection against unauthorized commands and the ability to intercept data.[lxxxiv]

Depending on the type of Modbus Protocol (RTU - TCP/IP) some limitations are more or less important (number of devices addressable by a master, or bandwidth issues, these two limitations being for the RTU approach).

5.1.4.1.6 Interoperability

If you consider the Building sector as an example of the deployment of communication protocol besides BACnet, KNX, Modbus is indeed a form of limitation in terms of data integration. Initiatives exist to develop Interoperability support for communication protocol for automation purpose. Some of them support indirectly the Modbus protocol. (e.g.DogOnt[lxxxv]).

Nevertheless, although the Modbus protocol presents a rudimentary form of "interoperability" described in point 4 of this document, the understanding and definition of abstract concepts concerning interacting devices is not possible within Modbus protocol itself and by definition need to be done at another system level for instance SCADA and Cyber Security[lxxxvi],[lxxxvii].

5.1.4.2 BACnet Protocol: A Data Communication Protocol for Building Automation and Control Networks

5.1.4.2.1 Background

BACnet was and is developed under the supervision of the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE).

Its development began in 1987 and to this day it is still being promoted by ASHRAEto develop a common communication platform that allows perfect interoperability between devices and control elements for BMS systems from different manufacturers. Since October 2003 BACnet is the worldwide ISO standard 164845 for open communication in building automation.

5.1.4.2.2 Definition

BACnet is data communication protocol which aims to provide standardised communication rules that enable automation equipment within a building to communicate in a standardised manner.

BACnet therefore relies on these rules to enable different equipment, components of different providers to communicate with each other through a standardized sharing of methods that address the aspects of presenting, requesting, interpreting and transporting information.

5.1.4.2.3 Scope of application

BACnet was designed to allow communication of building automation and control systems for applications such as heating, ventilating, and air-conditioning control (HVAC), lighting control, access control, and fire detection systems and their associated equipment[lxxxviii].

5.1.4.2.4 Main technical specifications

The exchange of data using BACnet protocol follows the client-server approach. The BACnet client requests a service from the BACnet server. The BACnet server executes the service.

Three main concepts[lxxxix] are considered and constitute the central bricks of the communication architecture of the BACnet Protocol.

- **Objects:** BACnet instead defines a standard set of "*Objects*", each of which has a standard set of "*Properties*"; that describe the Object and its current status to other devices on the BACnet

internetwork. It is through these properties that the Object may be controlled by other BACnet devices. BACnet uses an object-oriented model for abstracting and representing information.

- **Services**: Services describe procedures that are available to the participants for reading and writing object propertie2.
- **Properties :** Properties refer to defined object-specific data sets, the fields of which contain infor- mation required for the object's functionality[xc].

The BACnet standard identifies 123 different Properties of Objects. A different subset of these Properties is specified for each type of Object. The BACnet specification requires that certain Properties must be present for each Object. [xci]

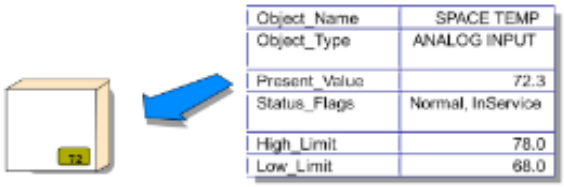

Figure 15:BACnet Objects with properties (capabilities, operation, related data)

5.1.4.2.5 Advantages and disadvantages

BACnet is designed specifically for building automation and control networks and it is strongly scalable. Indeed, BACnet BACnet has no limit on the number of BACnet devices that can be internetworked or the number of "points" that a given device can potentially contain[xcii]. System expansion was the major guiding force when the BACnet protocol was developed. As a result, BACnet is very open-ended. It allows you to choose from a large range of devices[xciii]
BACnet has many advantages and in particular if compared with the Modbus protocol, BACNET has a very strong network security layer and finally BACnet provides Web Services capabilities.

One of the disadvantages may emerge on the interoperability proposed by BACnet, which is mainly based on the fact of obtaining BTL certifications and which in an Internet of Things environment could, in front of the multiple and heterogeneous nature of the data coming from connected objects, face limitations in its capacity to address this volume of “new” equipment. Nevertheless, we can note the addition to BACnet WEB SERVICE, the support of RESTFUL API.

5.1.4.2.6 Interoperability

BACnet’s approach to interoperability between different devices allows each equipment to communicate with other equipment if that equipment strictly implements the BACnet protocol, regardless of the type of vendor. It is also possible for different device providers to make these devices available to submit each model to the BACnet Testing Laboratory and obtain a certification of conformity with the BACnet protocol.

The BACnet Standard identifies five interoperability areas[xciv]:

- Data sharing - **DS**
- Alarm and event management - **AE**
- Schedule - **SCHED**
- Trending - **T**
- Device and network management - **DM**

In terms of interoperability, it is important to note the existence of an ontology for BACnet - The BACnet Ontology (BACowl) is a description of the contents of the BACnet standard. [xcv]

6 Annex 2: European initiatives references

In this part we include the analysis of the most relevant European and International initiatives, related to the interoperability and standardization along with examples of current and past relevant IoT projects, and a final section on the new EU-Data Strategy

6.1 Initiatives related to interoperability and standardization

Standardization is a prerequisite to achieve interoperability. Many standardization initiatives currently exist in the smart grid arena, ranging from standardization of specific components to initiatives focusing on standardizing market roles. This section provides brief description of the most relevant European and International Initiatives, related to interoperability and standardisation.

6.1.1 Smart Grid Architecture Model (SGAM), standardization

The Smart Grid Coordination Group published the Smart Grid Architecture Model (SGAM)[xcvi] as a product of the standardization process in the EU Mandate M/490. It especially considers interoperability and aggregate interoperability categories, resulting in five Interoperability layers as shown in the next figure: Business Layer, Function Layer, Information Layer, Communication Layer, and Component Layer.

The business layer represents the business view on the information exchange related to smart grids, the function layer describes functions and services including their relationships from an architectural viewpoint, the information layer describes the information that is being used and exchanged between functions, services and components. The communication layer is to describe protocols and mechanisms for the interoperable exchange of information between components in the context of the underlying use case, function or service and related information objects or data models, and the component layer is the physical distribution of all participating components in the smart grid context.

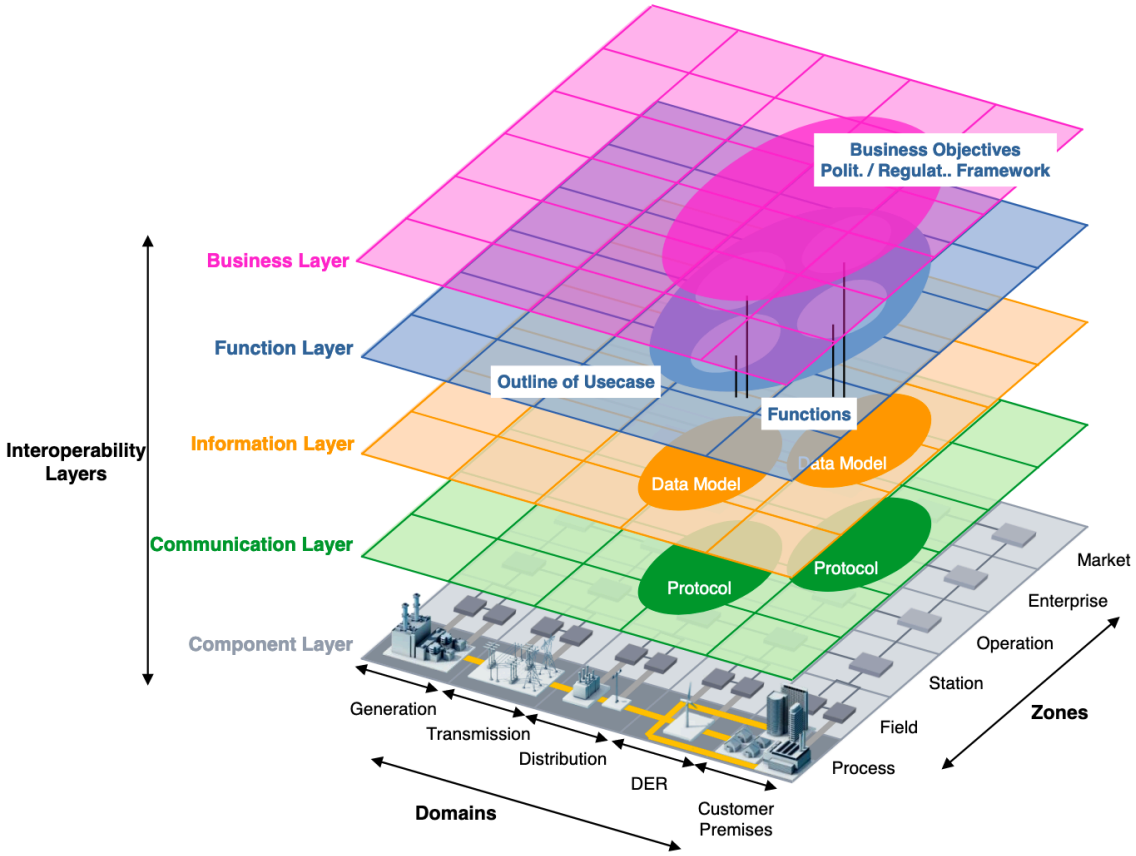

Figure 16: SGAM Framework by CEN-CENELEC-ETSI Smart Grid Coordination Group

The axis of Domains in the figure indicates five different domains:

- Generation representing large-scale power plants, ranging from nuclear to renewable hydropower.
- Transmission identifying transporting the generated energy over great distances.
- Distribution corresponding to distribute the transported energy to and from resources.
- Distributed Energy Resources (DER) indicating energy re-sources can be producing, storing or consuming energy.
- Customer Premises for customers ranging from industry to companies to private households.

The other axis of Zones shows the hierarchical levels of power system management:

- The Process includes the physical, chemical or spatial transformations of energy and the physical equipment directly involved.
- The Field includes equipment to protect, control and monitor the process of the power system.
- The Station represents the areal aggregation level for field level.
- The Operation Hosts power system control operation in the respective domain.
- The Enterprise Includes commercial and organizational processes, services and infrastructures for enterprises (utilities, service providers, energy traders ...)
- The Market Reflects the market operations possible along the energy conversion chain.

The document also detailed main elements of the different architectural viewpoints and shows the applicability. The SGAM provides a holistic view on the most important existing standards and architecture in Smart Grid standards and guides the other standard bodies and technical groups on dealing with standardization on the smart grid architecture.

6.1.2 Standardization committees

Smart Grids have received considerable attention worldwide in recent years. The concepts differ greatly in the main regions and this is also reflected in the respective roadmaps and studies. The number of technologies and systems, including hardware and software, coexisting in the Smart Grid ecosystem is very wide and so it is the number of organizations, committees and standardization groups working worldwide to agree on a common framework and roadmap. As countries were moving forward with their individual initiatives, it is very important that the efforts are coordinated and harmonized internationally.

NIST (US National Institute of Standards and Technology) is devoting considerable resources and attention to bilateral and multilateral engagement with other countries to cooperate in the development of international standards for the Smart Grids, to ensure interoperability addressing a common vision of devices, interfaces, communication, cyber security and systems integrity, system model and architecture, network and system management, business transactions and industry and market rules. NIST has led the establishment of ISGAN (International Smart Grid Action Network), a multinational collaboration of 23 countries and the European Union. Among the countries that have or will begin investing in substantial Smart Grid infrastructure are Canada, Mexico, Brazil, many of the member states of the EU, Japan, South Korea, Australia, India, and China.

Harmonization efforts are underway with (but not limited to) the following groups:

- The Institute of Electrical and Electronic Engineers (IEEE)

- The European Telecommunications Standard Institute (ETSI) together with the European Committee of Standardization (Comité Européen Normalisation-CEN) and the European Committee for Electrotechnical Standardization (CENELEC)
- The International Electrotechnical Commission (IEC)
- The Chinese Electrical Power Research Institue (CEPRI)
- The Korean Smart Grid Association (KSGA)
- The Japanese Federal Government

6.1.3 OpenADR Alliance

The OpenADR Alliance is a global alliance with over 150 members[xcvii]. Alliance main goals is to accelerate development, adoption and compliance of open Automated Demand Response (OpenADR) standards through the energy industry. OpenADR is open, secure, and a bidirectional information exchange model and Smart Grid standard. Demand Response (DR) and Distributed Energy Resources (DER) automation is addressed by connecting homes and businesses with the utilities to make it easier to power down during peak, demand, manage fluctuations, or to avoid electricity emergencies[xcviii]. OpenADR standardizes the message format used for Auto-DR and DER management so that dynamic price and reliability signals can be exchanged in a uniform and interoperable fashion among utilities, ISOs, and energy management and control systems, next figure [xcix].

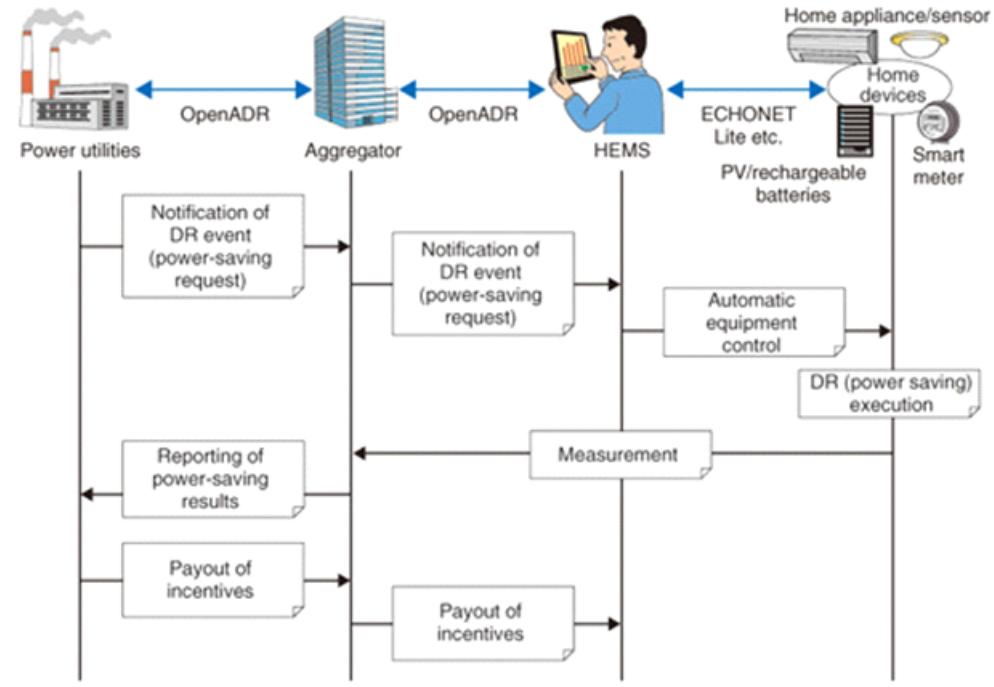

Figure 17: Example of an OpenADR sequence

OpenADR provides the following benefits[c]:

- Open Specification–Provides a standardized DR communication and signaling infrastructure using open, non-proprietary, industry-approved data models that can be implemented for both dynamic prices and DR emergency or reliability events.
- Flexibility–Provides open communications interfaces and protocols that are flexible, platform-independent, interoperable, and transparent to end-to-end technologies and software systems.

- Innovation and Interoperability–Encourages open innovation and interoperability, and allows controls and communications within a facility or enterprise to build on existing strategies to reduce technology operation and maintenance costs, stranded assets, and obsolesce in technology.
- Ease of Integration–Facilitates integration of common Energy Management and Control Systems (EMCS), centralized lighting, and other end-use devices that can receive Internet signals (such as XML).
- Supports Wide Range of Information Complexity – Can express the information in the DR signals in a variety of ways to allow for systems ranging from simple end devices (e.g., thermostats) to sophisticated intermediaries (e.g., aggregators) to receive the DR information that is best suited for its operations.
- Remote Access– Facilitates opt-out or override functions for participants to manage standardized DR-related operation modes to DR strategies and control systems.

OpenADR 2.0 Profile Specification contains following elements used to develop test and certificate framework for SG and customer system interoperability.

- A set of data models derived from the OASIS Energy Interoperation 1.0 standard.
- A set of services for performing various functions and operations for the exchange of the data models, also derived from the OASIS Energy Interoperation 1.0 standard.
- A set transport mechanisms for implementing the services. The transport mechanisms rely upon standard-based IP communications such as HTTP and XML Messaging and Presence Protocol (XMPP).
- A set of security mechanisms for securing each of the transport mechanisms.
- OpenADR 2.0 Schemas

OpenADR 2.0b ProfileSpecifcation was approved by the IEC as a Publicly Available Specification (PAS) IEC/PAS 62746-10-1 as a basis for a new commission standard to be developed. Additionally, the alliance is collaborating with different IEC committees – TC 57, PC118 and TC 65.

6.1.4 Universal Smart Energy Framework (USEF)

Universal Smart Energy Framework (USEF)[ci] is a non-profit partnership of ABB, Alliander, DNV GL, Essent, IBM, ICT Automation and Stedin that is developed by the Smart Energy Collective (SEC), an alliance of national and international companies including energy suppliers, network operators, electrical equipment manufacturers, consultancy and ICT companies.

One of its publication, “USEF: framework explained”[cii] introduces flexible value chains focusing on new market design with USEF framework for the various stakeholders with various roles active in the smart energy system.

It also develops open specifications and guidelines that accelerate the development of smart energy products, services and solutions for the large-scale rollout of smart grids in Europe. USEF provides a modular design for smart energy systems that can be customized to the needs of smart energy projects and a minimal set of specifications to secure the essential interoperability between all the components in a smart energy system.

The USEF specification is designed to drives Market-based Coordination Mechanism (MCM), and its operation scheme has four phases as shown in the next figure and describes detailed process flows per each step and their modifications to the wholesale processes. The specification also details grid operation including the specific grid processes such as active monitoring of the grid and graceful degradation to handle extraordinary grid conditions, as well as use cases and message exchange flows and description for each operation.

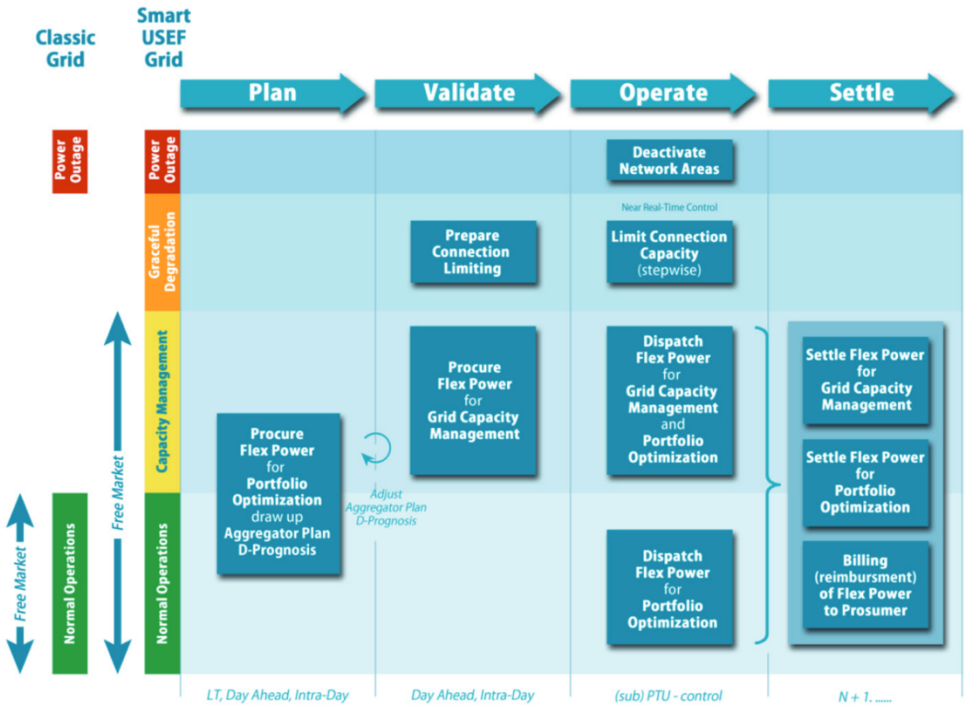

Figure 18: USEF Operation Scheme

It is said that Standardization is a prerequisite to development of USEF reference implementation, as the components of the smart energy system built upon the participation of multiple vendors and stakeholders need to be independent and easily interchangeable in the framework. The next figure explains the criteria used to rank and select standards for USEF. USEF puts a strong focus on interoperability and is aligned with the smart grid standardization developments (e.g. CEN-CENELEC, NIST and other relevant initiatives) for interoperability. It delivers one common standard on which to build all smart energy products and services.[ciii]

Criterion	Description criteria
International	The standard is valid for all regions of the world, without conflicting standards for specific regions.
Open	No royalties to be paid for applying the standard. The source codes are available for making extensions.
Scalability	The standard is applicable to real-life implementations without limitations to the number of implementations.
Future-proof	The standard can be used for technologies foreseen for the coming 10 years, and can be extended afterwards without breaking backwards compatibility.
Plug & play	The standard can be implemented with self-configuration for the intended applications in the SEC framework.
Critical Mass	The parties driving the standard together have a dominant share in the market.
Installed base	The standard is already applied for > 3 years in >10 real-life implementations with a scale > 2500 assets.
Aggregation	The standard has been successfully used in conjunction with other standards relevant for the smart energy systems.

Figure 19: The criteria used to rank and select standards

The list of the USEF specification is as followings and it can be downloaded from its website[civ]:

- USEF: The framework explained, Outlines the vision and approach to the flexibility market design, with a description of the structure, market roles, tools and rules.
- USEF: The framework specifications, technical guidelines for implementation of an optimised market-based energy system.
- USEF Flexibility Trading Protocol Specifications 1.01, a subset of the USEF Framework. Focused specifically on the exchange of flexibility between Aggregators and DSO's.
- USEF .XSD files USEF Flexibility Trading Protocol, a subset of the USEF framework. UFTP can be used as a stand-alone protocol for flexibility forecasting, offering, ordering, and settlement processes.
- USEF: The privacy and security guideline

6.1.5 IoT Big Data Harmonised Data Models developed by GSMA

The IoT Big Data harmonised Data Models from GSMA[cv] specifies harmonised data models for use by all the participants of the IoT Big Data Ecosystem Project. It uses JSON (JavaScript Object Notation) type specification[cvi], the NGSIv2[cvii] type specification and the schema.org type specification[cviii] for describing attributes of the data models.

The document includes data entities from diverse verticals such as agriculture, automotive, environment, industry, smart city and smart home and introduces a list of data models defined. During the data model definitions, the inputs from FIWARE[cix] and OASC[cx] have been included. The full set of data model entities and examples are in https://github.com/GSMADeveloper/NGSI-LD-Entities.

There are also accompanying documents of IoT Big Data Framework Architecture[cxi] and IoT Big Data NGSIv2 Profile[cxii], aim to define a framework of how mobile operators can approach the delivery of IoT Big Data services. While 'IoT Big Data Harmonised Data Models' introduces data model entities, the 'Framework Architecture' describes the role of mobile operators in IoT big data with challenges and area of services. It depicts general architecture for IoT big data as shown in the next figure, which well indicates common specification for interoperability.

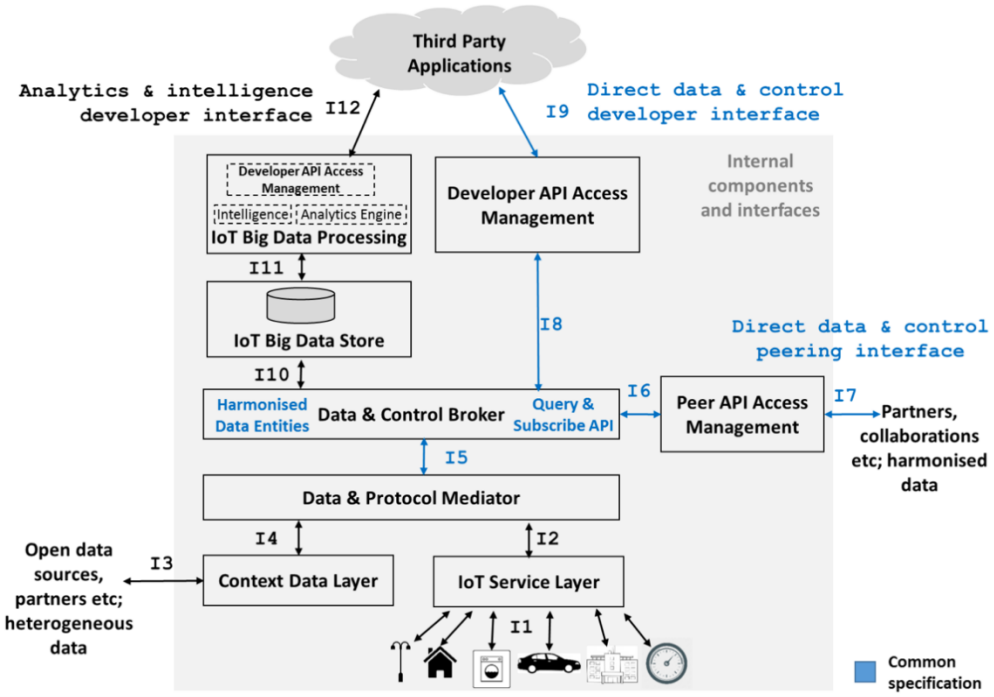

Figure 20: General Architecture for IoT Big Data from GSMA

As the figure indicates, harmonised data model entities are the key for data interoperability. There is no direct related data models for energy measurement or energy grid system in the current IoT Big Data Models from GSMA, but continuous efforts in the harmonised data models are ongoing and expect to collaborative work on open standard data models for the energy sector.

6.1.6 Big data domain European initiatives aimed to achieve interoperability through standardization: BDVA, European standardisation organisation ETSI

Interoperability is a crucial factor in the success of modern technologies, and market demand has ensured that interoperability holds a prominent position in standardization. Services and systems are often based on multiple standards from several standards-making organizations or on requirements published by industrial fora. Standardisation is a fundamental pillar in the construction of a digital single market and data economy. It is only through the use of standards that the requirements of interconnectivity and interoperability can be assured in an ICT-centric economy. As the development of standards is mainly initiated by market needs, industry plays an important role. European standards are then developed through one of the three European Standards Organisations: the European Committee for Standardisation (CEN), the European Committee for Electrotechnical Standardisation (CENELEC), and the European Telecommunications Standards Institute (ETSI). Collaboration between standards groups is therefore vital to achieve interoperability.

In this context, the **Big Data Value Association** (BDVA)[cxiii] aims to lead the way in the development of technology and interoperability standards by supporting Standards Development Organisations (SDOs), leveraging existing common standards, aligning the goals and integrating national and international efforts as the basis for an open and successful market. The BDVA is an industry-driven international not–for-profit organisation with 200 members all over Europe and a well-balanced composition of large, small, and medium-sized industries as well as research and user organizations[cxiv].

BDVA is the precursor of the **Big Data Value PPP program (BDV PPP),** that aims to create a functional data market and data economy in Europe, contributing to the European data strategy[cxv] and developing an interoperable data-driven ecosystem as a source for new businesses and innovations using big data. Besides energy, there are several relevant areas like mobility, manufacturing, economy, smart cities, infrastructures, food, finance… that will benefit from the work developed within the program.

The portfolio[cxvi] of BDV PPP projects[cxvii] includes a set of tools and methods for data management and processing, data privacy and protection, data sharing, integration of data analytics with other technologies, large scale deployments, etc, that could be used as the basis for the development of new initiatives and standards. More specifically, projects funded under DT-ICT-05-2020 are expected to provide data sharing tools and platforms, data governance frameworks, and availability, quality and interoperability, which are supposed to be the pillars of the new European data spaces.

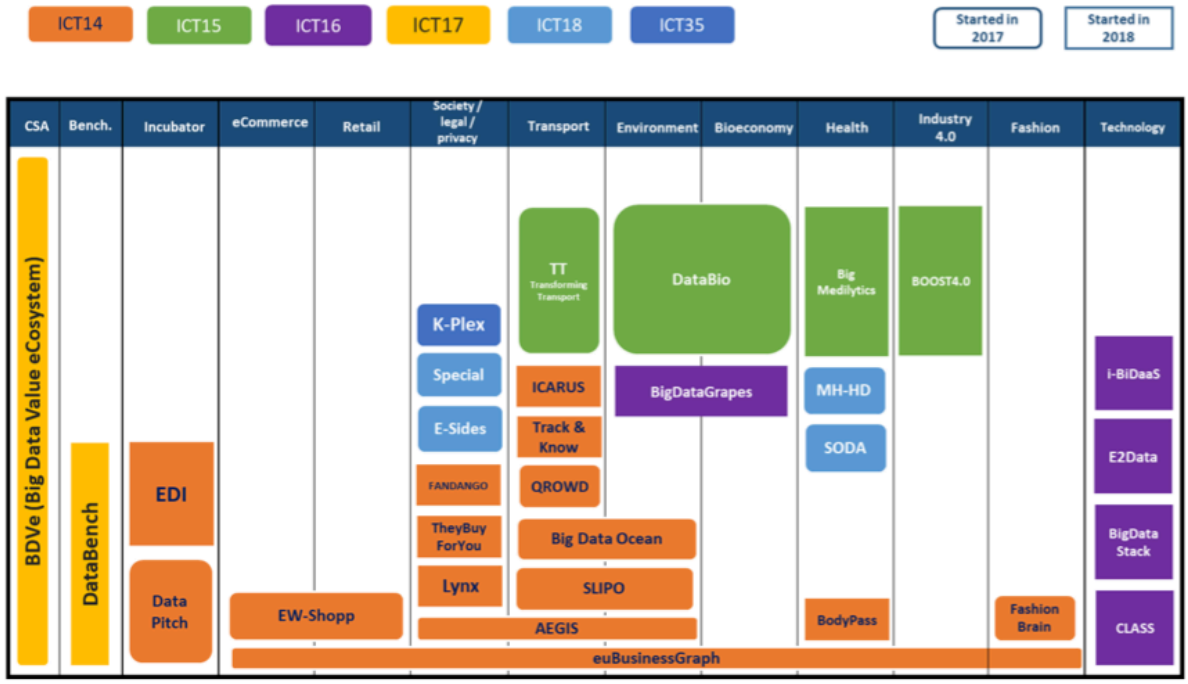

Figure 21: BDV PPP portfolio

6.2 European IoT and Energy pilots

6.2.1 TABEDE

1/11/2017 – 31/10/2022

To fully realize the European demand response potential, buildings must enter demand response schemes and expose all available flexibilities, including HVAC (Heating, Ventilation, and Air Conditioning) and thermal inertia, to the aggregator. However, several existing limitations must be overcome, such as interoperability and communication. Furthermore, it is often cheaper to install a new BMS (Building Management System) than to spend time and money on adapting an existing BMS.

TABEDE aims to allow buildings to integrate energy demand response schemes through a low-cost extender for BMS systems or as an autonomous system, which is independent of communication standards and integrates innovative flexibility algorithms. The proposed solution will reduce energy costs without compromising the comfort of buildings. The energy supplier will be able to exploit the flexibility of the building to maximize the use of renewable energy and ensure the quality of the energy.

TABEDE will allow:

- Indirect control of existing systems through set-points provided by TABEDE for equipment connected to the existing BMS, such as AHU, heating, FCU, boiler and chiller.
- Direct control of new installed equipment, which will be directly connected to the TABEDE system, such as lighting, smart sockets and EV battery chargers.

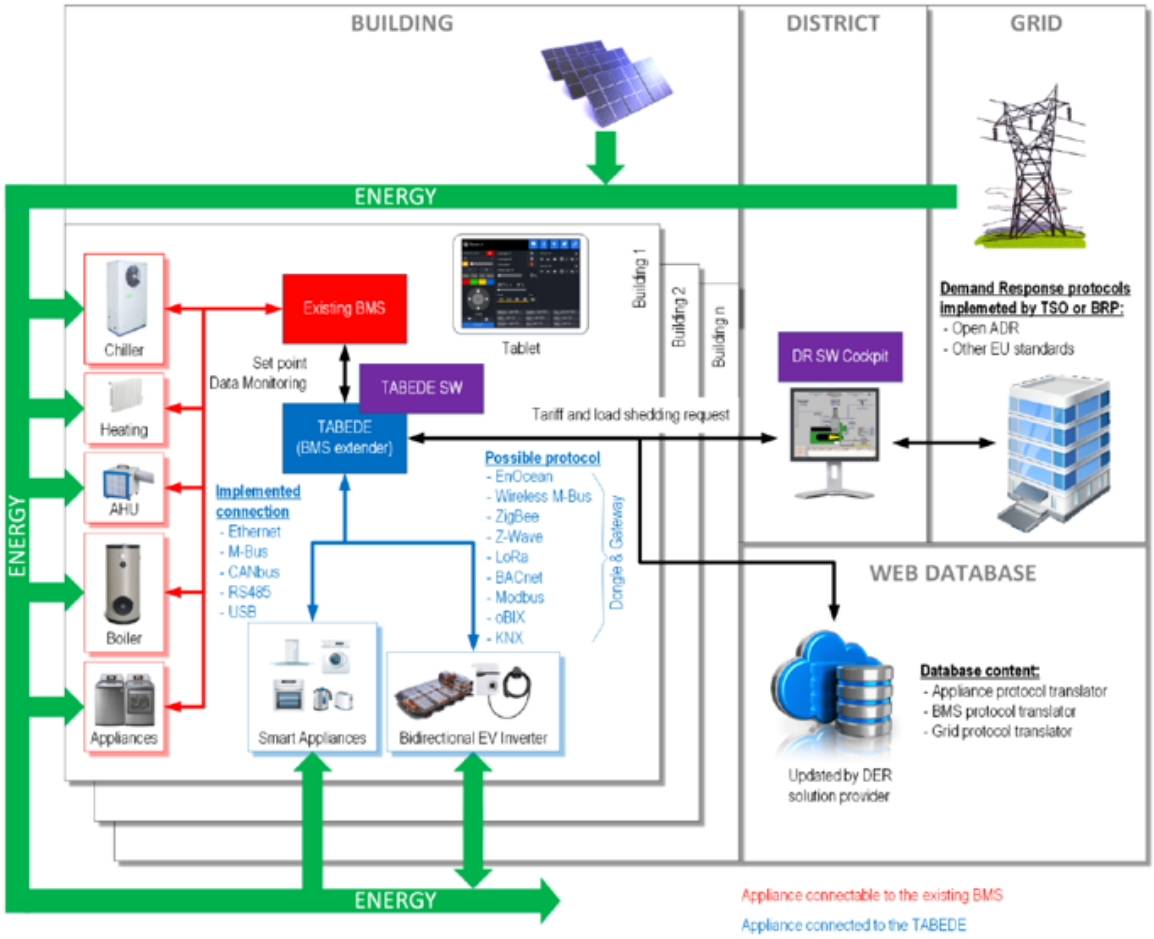

Figure 22: TABEDE

6.2.2 DRIMPAC

1/09/2018 – 31/08/2021

The European energy system is currently losing a vast source of demand flexibility which could offer multifaceted benefits to the system, energy generation and building demand flexibility. This is mainly caused by the lack of communication between the network / market and buildings, as well as the lack of interoperable intelligent building management systems capable of responding to network or market signals.

DRIMPAC aims to develop a solution to allow consumers to become active participants in the energy market. The project aims to bridge the communication gap between the network / market and buildings by providing a unique and universal technological framework that facilitates the end-to-end communication of the information necessary for the discovery and delivery of demand flexibility.

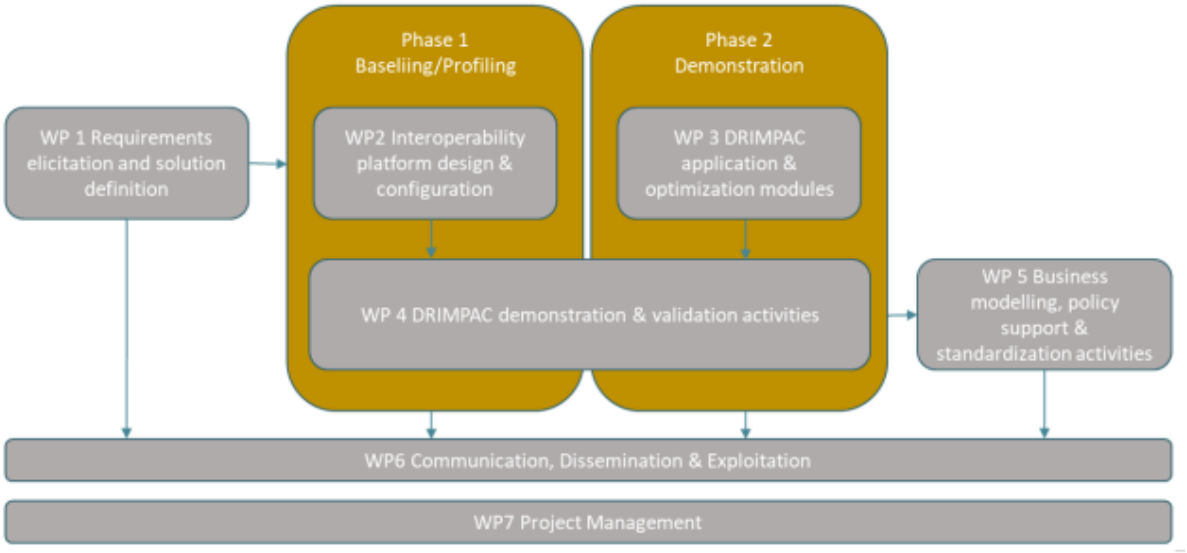

Figure 23: DRIMPAC

DRIMPAC finds the right balance between comfort and energy saving through environmental monitoring and preferences and intelligent algorithms. The project targets multiple building topologies that cover over 90% of the building stock and 40% of the final energy consumption in Europe and can support various energy mixes (electricity, gas, district heating) for a holistic optimization framework. Furthermore, it exploits the local generation and storage infrastructure (self-consumption, load shifting, thermal inertia), in order to provide an economic solution using autonomous components for the network and the sensory infrastructure. The system offers intuitive user interfaces taking into account end user profiles and their (lack of) technical expertise. Finally, DRIMPAC addresses the interoperability gaps and the fragmentation of standards by developing an end-to-end solution that covers the main standards.

6.2.3 RESPOND

01/10/2017 – 31/12/2020

Demand response (DR) refers to short-term targeted reductions in energy consumption during peak demand periods in exchange for financial incentives. DR is most commonly implemented in the industrial sector, where energy consumption is high and the peak of energy demand has a significant cost. Considering that distributed energy resources and renewable energies become mainstream, DR programs will have an increasingly important role in balancing energy demand also in the residential energy sector.

RESPOND will implement an interoperable energy automation, monitoring and control solution that will provide the answer to the question at the building, building and district level. Using an intelligent energy monitoring infrastructure, RESPOND will be able to detect energy saving opportunities and adapt to internal and external conditions and comfort levels in real time through optimal energy dispatching, taking into account both the offer and the of the question.

6.2.4 Synchronicity

01/01/2017 - 31/12/2019

SynchroniCity project[cxviii] (Delivering an IoT enabled Digital Single market for Europe and Beyond) represented the first attempt to deliver a Single Digital City Market for Europe by piloting its

foundations at scale in 11 reference zones - 8 European cities & 3 more worldwide cities - connecting 34 partners from 11 countries over 4 continents.

SynchroniCity delivered a harmonized ecosystem for IoT-enabled smart city solutions where IoT device manufacturers, system integrators and solution providers can innovate and openly compete. With an already emerging foundation, SynchroniCity established a reference architecture for the envisioned IoT-enabled city marketplace with identified interoperability points and interfaces and data models for different verticals.

SynchroniCity is built around the simple idea of building the minimal common technical ground needed in a global market for IoT-enabled services for cities and communities

Since cities and communities are quite different, but with many structural commonalities and common needs, SynchroniCity builds on a broad and inclusive baseline of inputs and requirements. This has led to the concept of minimal interoperability, meaning that the implementation can be different, as long as some pivotal points in any given architecture use the same interoperability mechanisms.

SynchroniCity defined an architecture that includes a set of logical components and functionalities that can enable different cities to be actively part of IoT Smart City digital single market.

SynchroniCity architecture is based on the concept of Interoperability points (red blocks in the picture).

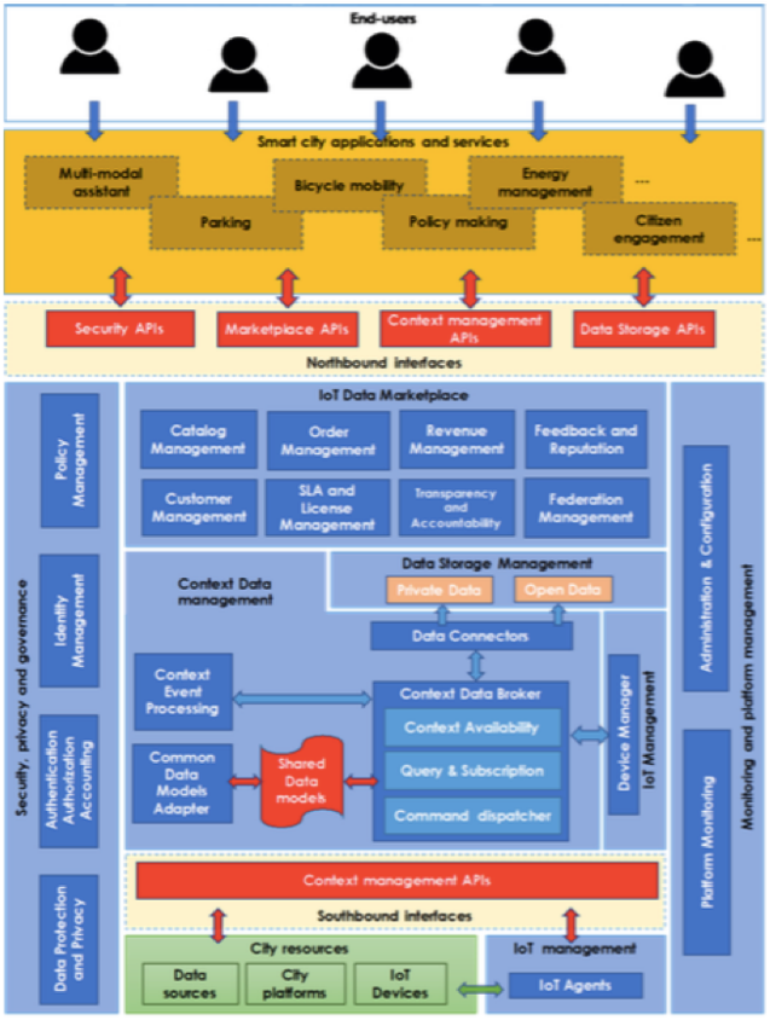

Figure 24: SynchroniCity architecture

Interoperability Points represent the main interfaces that allow a city and applications to interact with SynchroniCity platform: Interoperability points are independent from the specifications and software components, the mechanisms, that realise them and can be implemented by cities and communities in different steps to reach different levels of compliance.

The SynchroniCity architectural framework model, is the realisation of the Open & Agile Smart Cities "Minimal Interoperability Mechanisms", also known as MIMs (see specific section). The MIMs are vendor-neutral and technology-agnostic, meaning that anybody can use them and integrate them in existing systems and offerings.

From technical point of view SynchroniCity produced the following outcomes:

1. Synchronicity logical architecture and specifications[cxix]
2. SynchroniCity interoperability points and API definition: based on existing standards (FIWARE-NGSIv2, ETSI NGSI-LD OAUTH, DCAT-AP)
3. SynchroniCity common data models[cxx]: a set of NGSI compliant data models based on existing FIWARE ones and further extended by project partners
4. Reference implementation: a public repository[cxxi] that includes open source components (mostly based on FIWARE GE) to deploy a basic SynchroniCity framework.
5. 49 pilots deployments in 18 cities in Europe and beyond to validate SynchroniCity framework and specifications

The cities involved in SynchroniCity project have adopted the OASC principles to build IoT ecosystems based on open standards and existing datasets to build integrated services. The different deployments in the cities (that includes core cities and new ones added through the open calls) demonstrated the added value of the specifications and technical components provided by SynchroniCity, in relation to two main aspects: first of all, the open specification based on the concept of OASC MIMs represented a common interoperable baseline for the development of cross-border services based on different technologies and frameworks. Moreover, the reference implementation of the SynchroniCity architecture, based on open source and stable solutions, mostly developed by the FIWARE community, gave the opportunity to simplify the adoption of the open specification reducing costs of deployments and integration with existing systems. The cities, anyways, were not obliged to use reference implementation components but they were free to adopt alternatives compliant with standards defined by the project.

Another important achievement of Synchronicity was the provisioning of a set components to enable advanced Data Marketplace capabilities which would incentivize users to join the ecosystem and exchange data. For instance, the Data Marketplace provides tools to manage License, SLA, and pricing model that form the terms and conditions of the data offerings, which can be stored in a trustless and decentralized way so that both data providers and data consumers can rely on tamper- proof agreements enforceable in a court of law.

6.2.5 Roma capitale projects

6.2.5.1 The photovoltaic systems of Roma Capitale

2018 - 2019

Ensuring transparency and a clear picture of energy production from renewable sources, this is the function of mapping Rome's photovoltaic plants. The plants, in fact, are about 157 many of which are installed on the roofs of school buildings in Rome. This map indicates, through a geotag, the exact

location of each plant and its main technical data, accessible to a simple click of the user. From the map it is possible to access the main information relating to each photovoltaic system, including the power (kWp), the type of property to which it is subject and the status of the system and whether it is under maintenance or is properly functioning. In the system table there is also the link to the dynamic street-view picture. The map is available on the Roma Capitale website on the SIMU Department page (Public Works).

The plants built and surveyed to date have a total power of almost 2 MegaWatts, therefore a maximum annual production capacity of approximately 2.5 GWh, or the average equivalent consumption of over 900 homes, contributing to the reduction of CO2 emissions in the City of Rome.

This data will be accessible to everyone forever and thanks to the progressive updating of the same, the map will be a useful work tool. This will also help to understand the correct functioning of the installed systems and to allow for maintenance and improvements to be expected. Mapping is a useful and effective way to collect information and make it easily accessible, just a click away for citizens, as well as to keep an eye on the programming of interventions.

GIS Map
https://www.google.com/maps/d/viewer?mid=1w471f1gXgmCCTf02XFiMd_ui0qGCSHbm&ll=41.94130553633198%2C12.491152720123182&z=12

6.2.5.2 Metropolitan City of Rome - INTERREG ENERJ project

1/11/2016 - 31/10/2019

The public buildings sector represents an important potential for energy savings. However, in recent years the rate of energy renovations of the building stock has been far too low and public buildings are no exception. ENERJ is aimed at activating existing potentials and bridging this gap.

ENERJ focuses on the legal, technical and economic conditions for effective joint actions, increasing awareness of their added value and proposing procedures to simplify the decision-making process, plan and implement pilot actions.

The actions will be promoted in the various territorial realities by qualified Joint Action Coordinators (energy managers), who will provide consultancy and technical support services. The ENERJ platform will contain the actions of the municipalities involved and more generally it will also offer resources for the implementation of local energy plans, in particular sustainable energy plans. It will highlight best practices and act as a "meeting place" for local actors and other stakeholders.

The ENERJ web platform http://www.enerj-platform.eu/enerj/ and on a GIS basis http://www.enerj-platform.eu/enerjwm/ is an important tool to achieve the objectives of the project. To be fully operational, it must be powered by data on local public buildings and SEAPs from the largest number of municipalities, not just from the project partners. A database containing this information can be very useful in planning wider interventions of individual municipalities, creating the conditions for activating collaborations between multiple local authorities and attracting the attention of private operators who will be able to find useful information to evaluate proposals and offers.

Currently, hundreds of buildings have already been included in the platform by the partners. Given that this web tool will remain active for at least five years after the end of the project, the goal is to help municipalities to enter their data on buildings and SEAPs, facilitating the creation of joint actions between municipalities and with private investors, contributing to the financing and implementation of energy efficiency measures in the public sector.

6.3 EU Data Strategy

The European Commission released a data strategy in February 2020, part of "Shaping Europe´s Digital Future", where it lays a framework with the aim to increase data access and help European businesses monetize on the data generated within the EU. The idea is to make Europe a global hub, whose main elements include:[cxxii]

- A cross-sectoral governance framework for data access and use.
- Make more high-quality public sector data sets available for reuse.
- A Data Act to provide incentives for horizontal data sharing across sectors.
- Data access legislation.
- A centralized European cloud market place.
- Data portability individual rights.

The EU Data Strategy is structured around four building blocks:

1 **Cross sectoral governance framework for data access and use**: to facilitate and clarify cross-border data use, address interoperability requirements and standards from businesses across and within different sectors. The key actions proposed include:
   a. A legislative framework for common European data space governance. In the case of PLATOON, it is being covered by the IDS framework.
   b. Data Act and implementing act on high-value datasets
   c. review of data intellectual property framework

2 **Enablers**: Investment in data and strengthen Europe´s capabilities and infrastructures for hosting, processes and data usage and interoperability: to drive strategic investments, support data-driven innovation and strengthen Europe´s technological sovereignty. The key actions proposed include:
   a. Invest in a High Impact Project on European Data Spaces
   b. Sign Memoranda of Understanding (MoUs) on Cloud Federation and creat an EU self-regulatory cloud rulebook
   c. Launch a European Cloud Services Marketplace

3 **Competences**: empowering individuals, investing in skills and SMEs. The key action proposed is enhancing individual´s right to data portability under Art.20 of the GDPR (2021).

4 **Common European data spaces in strategic sector and domains of public interest**: to make large pools of data available. The strategy lists nice different data spaces such as industrial manufacturing, European Green Deal, mobility, health, financial, energy (the sector PLATOON is aimed for), agriculture, public administration and skills.

7 Annex 3: Analysis of techniques and paradigms that may be part of the overall technological stack of data exchanges

7.1 IoT paradigms

7.1.1 Fog computing paradigm

Fog computing is an extension of the cloud computing paradigm that aims to overcome the expected limitation of cloud computing itself due to the growth of smart IoT devices. One of the known drawbacks of Cloud computing is the latency included in the intercommunication between an IoT device and the Cloud. Therefore, the main principle of fog computing is to enable the computing at

the edge of the network closer as possible to IoT and end-user devices acting as an intermediary between the cloud and end devices bringing processing, storage and networking services.[cxxiii]

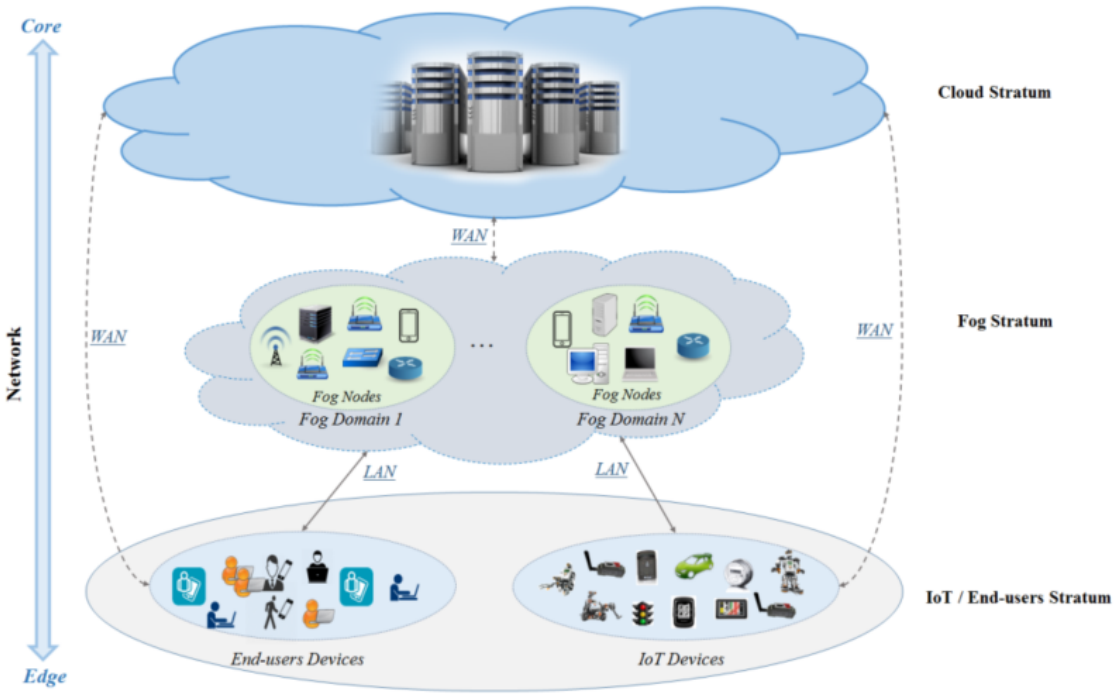

Figure 25: Fog system

The main features of a fog system should be the following:

- Location awareness and low latency: fog nodes might be deployed in different location and, as they are closer to edge devices, there is a low latency in communication among devices and fog nodes.
- Geographical distribution: the services provided by the fog nodes are distributed and can be deloyed anywhere.
- Scalability: the fog systems are expected to cover a large number of IoT or end user devices, thus the system is able to scale up or down.
- Support for mobility: IoT or end user devices and fog nodes can be mobile, therefore fog system supports mobility methods.
- Real-time interactions: fog applications provide real-time interaction between fog nodes due to the proximity of fog nodes and IoT or end user devices.
- Heterogeneity: the fog has the ability to work on different platforms.
- Interoperability: Fog components can interoperate and work with different domains and across different service providers.
- Support for on-line analytics and interplay with the cloud: The fog is placed between the cloud and end devices to play an important role in the absorption and processing of the data close to end devices.

The primary object of fog computing is to provide low and predictable latency for time-sensitive IoT applications. Moreover, the fog system should provide storage and processing capabilities to the devices that can virtually extends its resources or, since it acts as an intermediary between the device

and the cloud infrastructure, the fog system might pre-process data to be than sent to the cloud for further analysis.

Among the several fog compliant architectures proposed, next figure illustrates an architecture composed by six layers[cxxiv]. The physical and virtualization layer manages the several types of nodes in the fog system. The monitoring layer is used to monitor every aspect of the fog system, for instance which activity is performed by each node or the nodes' energy consumption. The pre-processing layer takes advantage of the temporary storage layer to store the results of real-time analysis that can be performed on the raw data coming from the edge devices. The security layer is in charge of managing the encryption/decryption of the data and, moreover, to check the data integrity. Finally, the transport layer uploads the pre-processed data stored into the temporary storage layer into the cloud.

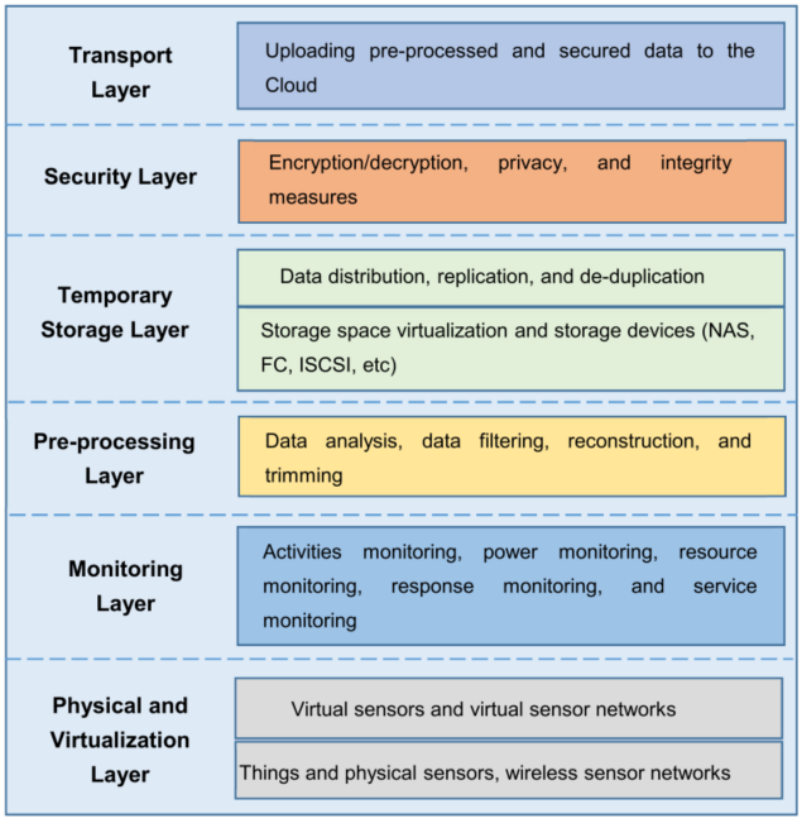

Figure 26: Layered architecture of fog computing

The following set of criteria to evaluate fog systems are defined:

- Heterogeneity: as the ability to decide which application component should be deployed and where taking into account the differences between cloud and fog capabilities.
- QoS Management: as the need to satisfy the QoS required for each application deployed in the fog system.
- Scalability: fog systems should be able to scale depending on the number of IoT/end user devices or fog nodes involved in the system.
- Mobility: the system should be able to handle the mobility of the several devices involved.
- Federation: since the fog system can be geographical distributed on large scale and the several fog nodes may be owned by different providers it is necessary to have a federation of these providers which might host the different modules that compose an application.

- Interoperability: the system must guarantee the interoperability among the several components of the system.

In the context of smart energy, several example of fog paradigm platforms as a service are proposed, of which two prototype implementations are described: [cxxv]

- Home Energy Management (HEM): where a HEM platform is provided to manage and monitor several devices deployed in a house.
- Microgrid-Level Energy Management: where a transformer powers several homes. The administrator can define certain power threshold for each home and the transformer will send signals to the home which have violated the threshold.

Fog computing approach can also be applied with smart grid architectures[cxxvi], where fog paradigm is used to manage and monitor a large-scale, geographically distributed micro grid system with millions of sensors and actuator. A distributed controller is in charge of the management of these sensors and actuators. The controller is built over the fog layer of the architecture to achieve very low latency between IoT sensors and the controller. Fog computing is also extensively used in RES and conventional generation in order to optimise data exchange, storage and processing capabilities. In fact, usually some of the data is pre-processed at device level (such as vibration analysis and certain types of aggregation operations…) and then pre-processed data from different asses is integrated in the cloud/on premise. Equally, some of the models are directly processed at the edge when the required response time is so small that cannot wait to send the data run it on the cloud/on premise and the get back the result.

7.1.2 Digital Twin paradigm

In the context of PLATOON, both data driven and physics-based digital twins will be developed and integrated as part of pilot 1a. Therefore, the following implications of digital twins needs to be considered.

The Digital Twin concept is not new and refers to the digital representation of a physical object. In 2012, the National Aeronautical Space Administration (NASA) defined the digital twin as an integrated multiphysics, multiscale, probabilistic simulation of an as-built vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its corresponding flying twin[cxxvii]. As depicted in W. Kritzinger, M. Karner, G. Traar, J. Henjes e W. Sihn [cxxviii], a digital twin exists when there is a fully integrated bidirectional data flow between a physical object and its digital counterpart, making possible to change the state of the physical object by changing the state of the digital one and vice versa.

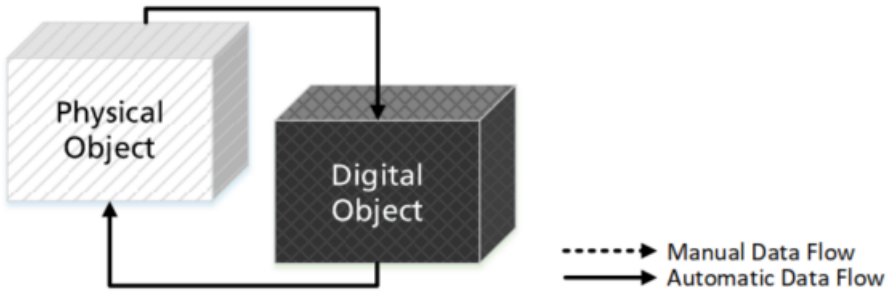

Figure 27: Digital Twin

The report "*Digital Twin: Enabling Technologies, Challenges and Open Research*"[cxxix] gives an overview of the enabling technologies, challenges and open researches for Digital Twin analysing the literature about this topic. Furthermore, it identifies some of the potential applications of the Digital Twin, namely: Smart Cities, Manufacturing and Healthcare. From the analysis of such potential applications the enabling technologies that can facilitate the growth of the Digital Twin paradigms are IoT, AI and Data Analytics. For instance, in the Smart City scenario, the Digital Twin approach can lead to the creation of a living testbed that can achieve two purposes: to test scenarios, and to allow for Digital Twin to learn from the environment by analysing changes in the data collected thanks to the increasing adoption of connected IoT sensors.

Moreover, the report focuses on the challenges about Digital Twin and it lists the following:

- **IT infrastructure**: to effective running a Digital Twin the IT infrastructure needs to guarantee high-performance in the form of up to date software, hardware and connection among the components.
- **Useful Data:** the data needs to have high-quality in the sense that should be noise free with a constant and interrupted data stream.
- **Privacy and Security:** the privacy and security must follow the best practice in the state of the art to avoid possible risks for sensitive data.
- **Trust:** it is mandatory to discuss and explain the Digital Twin technology to both users and organizations to ensure a good level of trust.
- **Expectations:** caution is needed to ensure that the expectations about Digital Twin technology will be met.
- **Standardised Modelling:** currently there is not a standardized approach for modelling Digital Twin.
- **Domain Modelling:** it is important to guarantee compatibility with domain such as IoT and/or Data Analytics for the successful use of Digital Twin.

The report discusses, also, about the open researches linked to the Digital Twin giving an overview about the researches in Smart Cities, Manufacturing and Healthcare sectors and, also, giving insights about the other open researches linked to the Digital Twin. About Healthcare, one of the most promising area of research describes the possibility to perform remote surgery taking advantage of the development of 5G technology and patient's Digital Twin. Furthermore, considering additional open researches, the topics are:

- **Data Models**: since no generic data model and/or architecture about digital twin is defined.
- **Heterogenous Systems**: to understand how to deal with the heterogenous digital twin systems, the researchers need to study and compare several approaches.
- **Artificial Intelligence**: researchers must evaluate the potential impact of AI on the advancements in Digital Twin technology.
- **Security**: since advancements in blockchain technologies could help in secure Digital Twins.
- **Data Exchange**: since Digital Twin solution needs to be scaled considering data exchange requirements.
- **IoT**: since sensors should be retrofit to ensure that data exchange guarantees high performances and accuracy.

7.1.3 Minimal Interoperability Mechanisms (MIMs) paradigm

Minimal Interoperability Mechanisms (MIMs)[cxxx] are universal tools for achieving interoperability of data, systems, and services between cities and suppliers around the world. As they are based on an inclusive list of baselines and references, MIMs take into account the different backgrounds of cities and communities and allow cities to achieve interoperability based on a minimal common ground.

Implementation can be different, as long as crucial interoperability points in any given technical architecture use the same interoperability mechanisms.

The MIMs can be considered vendor-neutral and technology-agnostic interfaces, that can be used to achieve interoperability integrating them in existing systems and offerings.

The following picture shows how the MIMs can be adopted to achieve interoperability starting from an initial status in which are present legacy systems with no standard or proprietary solutions to the final goal of being part of a common ecosystem of interoperable data and solutions (i.e. Marketplace).

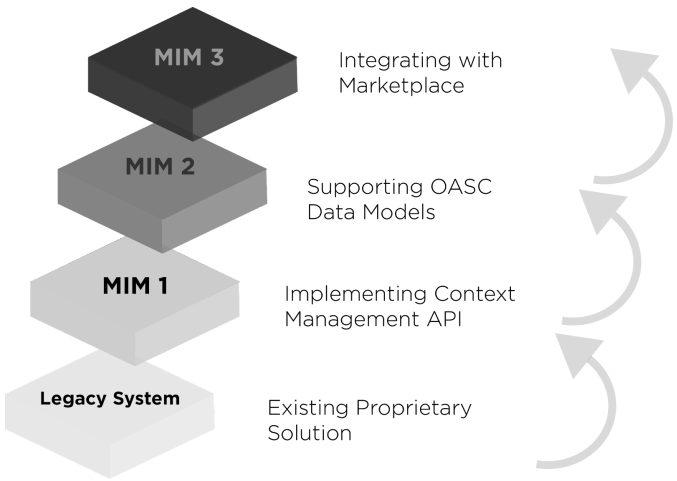

Figure 28: MIMs interoperability

The MIMs can be considered the implementation of logical interfaces (interoperability points) that allow interaction and interoperability among different systems.
In the following table are described the first three MIMs that have been officially adopted by the OASC Council of Cities in January 2019.

MIM	MIM Name	Interoperability Point	Description
1	OASC Context Information Management MIM	Context Information Management API	This API allows to access to real-time context information from different cities.
2	OASC Data Models MIM	Shared Data Models	Guidelines and catalogue of common data models in different verticals to enable interoperability for applications and systems among different cities.
3	OASC Ecosystem Transactions Management MIM	Marketplace API	The Marketplace API exposes functionalities such as catalogue management, ordering management, revenue management, Service Level Agreements (SLA), license management, etc. Complemented by marketplaces for hardware and services.

Figure 29: First three official MIMs

MIMS are concretely implemented through the adoption of baselines and standards (indicated in the next table) that have been identified, selected and further improved in the SynchroniCity project (section 6.2.4).

MIM	Name	Standards & [Baselines]	Reference
1	OASC Context Information Management MIM	ETSI NGSI-LD API[1], OMA NGSI, ITU- T SG20/FG-DPM [FIWARE NGSI]	Reference Architecture for IoT-Enabled Smart Cities (SC-D2.10)
2	OASC Data Models MIM	[SAREF, FIWARE, GSMA, schema.org, SynchroniCity RZ + partner data models]	Guidelines for the definition of OASC Shared Data Models (SC-D2.2) Catalogue of OASC Shared Data Models for Smart City domains (SC-D2.3; to be released)
3	OASC Ecosystem Transaction Management MIM	[TM Forum Business Ecosystem API, FIWARE Business Ecosystem and Marketplace Enabler API, SynchroniCity API]	Basic Data Marketplace Enablers (SC-D2.4) Guidelines for the integration of IoT devices in OASC compliant platforms (SC-D2.6)

Figure 30: SynchroniCity MIMs

7.1.4 Hybrid cloud and intercloud environments

The common deployment model of the cloud environment is based on the following definitions[cxxxi] [cxxxii]:

- **Public Cloud**: the cloud infrastructure is owned by the cloud service provider and application or services are provided to public users or organizations on a form of service agreement.
- **Private Cloud**: the cloud infrastructure is operated solely for an organization. It can be managed by the organization itself or by a third party.
- **Hybrid Cloud**: the cloud infrastructure is a composition of two or more clouds linked together by a standardized or proprietary technology.
- Community cloud: the cloud infrastructure is shared by several organizations which have shared concerns.

Furthermore, intercloud is defined as an interconnected cloud of clouds focus on interoperability and integration, and its main types are:

- **Federation Clouds**: in a federation of clouds, the cloud providers share their resources among each other.
- **Multi-Cloud**: this approach does not imply the sharing of the specific resources among the providers, but a client or service uses multiple independent clouds.

All hybrid clouds are multi-clouds but not all multi-clouds are hybrid clouds.[cxxxiii] The reverse relationship is valid when multiple clouds are connected by some form of integration or orchestration.

Three design patterns, depicted in next Figure[cxxxiv], are used to build hybrid cloud management solution:

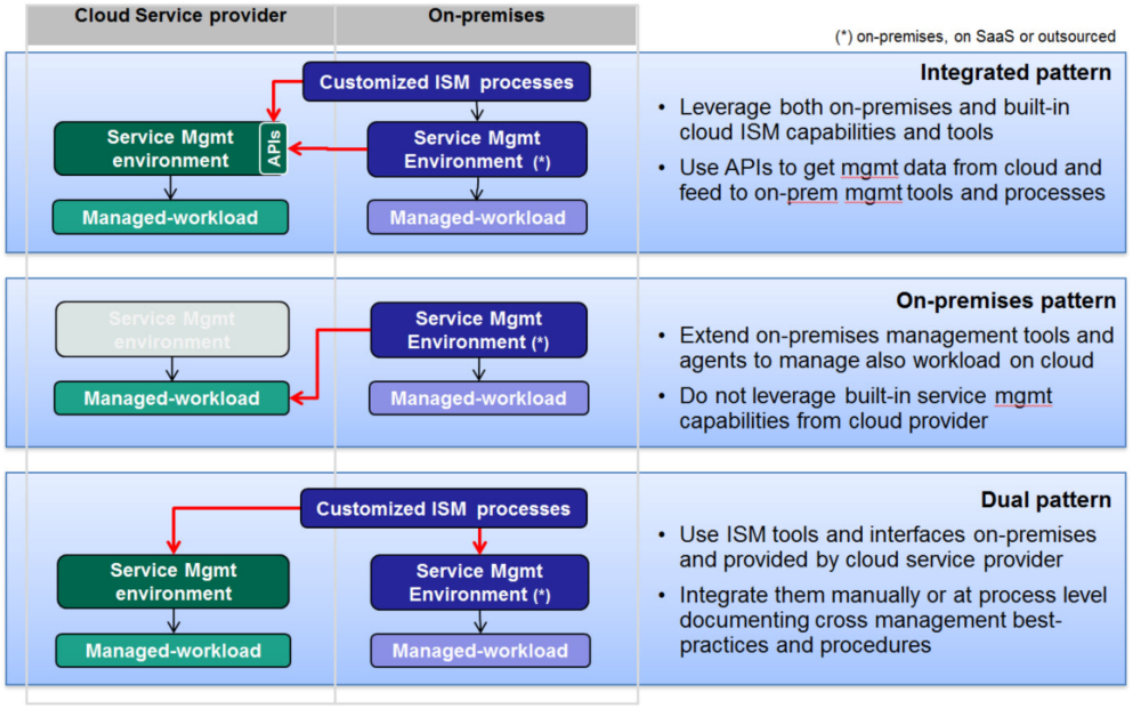

Figure 31: Patterns for Hybrid cloud management solutions

The "Dual Pattern" is useful if the final objective of the integration is the management of capabilities across on-premises and cloud workloads and for scenarios that cannot be implemented with other methods. The "On-premises pattern" is recommended if cloud customers rely on the cloud service provider to manage the cloud resources/services below the service responsibility line and they want to manage a limited amount of workloads. Finally, the "Integrated Pattern" leverages the on-premises and cloud-provider management services using existing APIs to automate the gathering, transfer and consolidation of management data from both environments.

In the scenario where multiple cloud environments cooperate, customers need avoid possible risk of vendor lock-in. Interoperability and portability among the cloud providers are key enablers for customers to make the best use of heterogenous cloud services[cxxxv]. SaaS applications are the biggest challenges for cloud interoperability and two possible approaches that can be used to handle these challenges are described. The first solution consists in the development of an isolation or mapping layer between the specific application and systems and the cloud service interfaces; the other approach is to use the services offered by an inter-cloud provider or a cloud service broker who is responsible of mapping an interface offered to the customer to a set of interfaces offered by a number of different cloud service providers. Portability's biggest challenge is at PaaS level since different cloud providers usually provides different PaaS platform and describes two alternative scenarios. The first is to increase the adoption of open source PaaS platform and the second relies on the usage of containerization technologies.

Furthermore, the following aspects need to be considered[cxxxvi]:

- Scalability and Elasticity: as the possibility of the cloud system to adapt to the growing of the number of devices managed, and the ability to provision or de-provision computing resources on demand.
- Data Bandwidth: in order to optimize the cloud system for big data.
- Data Sovereignty: as the compliancy with the regulation of the location in which the data is stored.

- Data Volume: as the ability to meet the performance requirements although the data volume increase.
- Resilience: the system should not depend on one single component and therefore should be made resilient through multiple instances of programs and cloud services together with data replication and redundancy procedures.
- Security and Governance: as the proper management of identity and data privacy of devices and individual.

7.2 Blockchain and Decentralised Ledger Technologies

"The creation of digital business-ecosystems and data marketplaces can only be successful by providing a mechanism for the documentation of transactions for billing, clearing, provenance tracking, etc. Blockchains and Distributed Ledger Technologies are suitable, mature, and accepted technologies for the application in digital-driven business ecosystems".[cxxxvii]

7.2.1 Decentralized vs distributed processes in digital business and Daaps

Digital business ecosystems are dependent on decentralized processes to different degrees, where processes span multiple entities and participants.

Decentralized processes differ from distributed processes in the way tasks are delegated, in the case of decentralization these are delegated to multiple parties while policies and controls remain centralized. So, participants interact in order to establish order and coordination for the achievement of goals and objectives.

Hence, decentralized processes require the following:

- Uniquely identified digital assets among the different participants.
- Authentic state tracking for digital assets in participants´, computing environments.
- Autonomous state change following contractual agreements.
- Identity compliance and access management.

7.3.1.1 Decentralized Applications (Dapps)

Decentralized systems require a new breed of applications that can run autonomously, with minimal logical or physical dependency on any single entity or technical component, in a peer-to-peer network, referred to as decentralized applications (or simply dapps).

Blockchain platforms establish the trust assurance foundation for dapps that inherently ensure correctness, integrity, privacy, provenance, performance and scalability of its operation and records. Therefore, blockchain platforms' differences depend on the degree of inherent trust assurance and decentralization in systems and their related decentralized applications (dapps).

Blockchain complete solutions, implement decentralized apps (dapps) in digital ecosystem process automation (digital asset or token) as peer-to-peer protocols. So it provides a decentralized computing environment to securely manage digital assets.

Dapps represent digital assets as tokens, controlling properties and behaviors autonomously by collective validation which is later confirmed by blockchain nodes. This infers capabilities such as autonomous control message response through immutable contracts, control messages leading to token state change and immutable storage of token states in distributed ledgers.

7.2.2 Blockchain Technology Core Features

7.2.2.1 Blockchain as a "peer-to-peer based distributed database"

Blockchain technology essentially integrates networks with databases resulting in a peer-to-peer based distributed database spread across multiple entities, with no single owner or single point of failure. It removes the need for trust in individual endpoints because immediate synchronization ("near real time") across entities participating in the blockchain takes place, meaning no single trusted third party is needed to guarantee that the transaction occurs. Also, a permanent record is guaranteed as data is appended, not deleted.

7.2.2.2 Blockchain as a distributed computer: 'Smart Contracts'

Smart contracts are object codes that are immutably stored on a blockchain platform and run autonomously to respond to internal or external events. Each object defines a set of state variables and related operations to establish the business logic and rules that must be followed in advance of changing the object state. Dapps use smart contracts as a trusted interface to interact with a blockchain platform ledger, embodying the business model and related rules. Smart contracts validate relevant interactions or transactions in the form of a technical protocol before any state change is stored on the platform ledger as an immutable fact.

Smart contracts are the on-chain codes, which allow a dapp to connect and interact with a blockchain platform. Smart contracts enable dapps or external data feeds to interface with the blockchain platform And augment the network-level validation protocols by defining/executing contextual business logic.

A dapp is similar to a traditional application. The key difference is that, instead of using a pure API call for connecting to a database, dapp uses a smart contract for connecting to a blockchain. That means smart contracts similarly play the role of the database connector to the blockchain while executing business logic and rules.

Smart contracts can read from and write only to the blockchain, so all the off-chain interactions have to be handled by trusted agents that map off-chain and on-chain assets.

7.2.2.3 Permissioned vs. Permissionless blockchains

A blockchain peer-to-peer network is composed of nodes that are generally implemented and deployed in two different models: public (permissionless) or enterprise (permissioned).

Early blockchain technologies applies the concept of 'permissionless' blockchains, meaning that anyone being able to physically access the blockchain could read and write data (e.g. bitcoin). When using blockchains in a more closed business community it becomes important to be able to govern the roles and rights of the various participants ('permissioned blockchain'). This is important in the context of International Data Spaces, given the principle of data sovereignty.[cxxxviii]

7.2.2.4 Blockchain Nodes

Nodes are participants within the blockchain network, that are grouped by their functionalities or properties. These functions and properties may vary in different platforms and their naming conventions. However, all nodes, regardless of their functionality, must be identified on the network as a prerequisite to use their roles as a contributor to the network or simply use the network services. Therefore, identity services are core to the operations of each platform.

Nodes are typically computing devices that use a standard messaging protocol to communicate in a public consortium or private network. Platforms may use different naming conventions for nodes or

how they break down workload among different types of nodes. There are at least two types of nodes:

- **Wallet nodes** (also known as users, clients): Client nodes with a digital identity (traditional PKI-based identity or a decentralized identity) that consume a decentralized application.
- **Full nodes** (also known as hosts, validators or miners): Server nodes with a digital identity (traditional PKI-based identity or a decentralized identity) that maintain a copy of the blockchain data structure, usually run the back-end piece of a dapp.

7.2.2.5 Tokenization

One of the most critical tasks in the solution development is the modeling of digital assets. The most common are tokens as digital asset representation, where the properties (state) and behavior (control message response methods) are captured.

Tokens are typically classified by five key characteristics:

- Token Type
  - Fungible (interchangeable tokens): Any quantity of fungible tokens in the same class has the same value as another equal quantity (e.g. physical cash or cryptocurrencies).
  - Non-nofungible (unique tokens): are not interchangeable with other tokens of the same type as they have a different value (e.g. a property title).
- Token Unit
  - Fractional: can be divided into smaller fractions.
  - Whole: cannot be divided into smaller fractions, only whole numbers.
  - Singleton: quantity that cannot be divided.
- Value Type
  - Intrinsic: the token itself is a valuable entity.
  - Reference: references a valuable entity elsewhere, like a property title.
- Representation Type
  - Common tokens: share a single set of properties, are not distinct from one another. These tokens are simply represented as a balance or quantity attributed to an owner's address where all the balances are recorded on the same balance sheet in a distributed ledger (e.g. back account balances). Only balances between accounts can be traced, not each individual token
  - Unique tokens: have their own unique identity and can be individually tracked. They can carry a unique state that cannot be changed in one place and cascade to all (e.g. paper bills are interchangeable but have unique properties such as a serial number).
- Key components
  - Template: defines token formula and defines capabilities and restrictions)
  - Class

7.2.2.6 Blockchain Coordination and Validation Protocols

Public and enterprise blockchains differ in how they validate and coordinate interactions or transactions before allowing the records to be written to the immutable ledger. The validation and coordination mechanisms are either consensus-based or endorsement-based.

Consensus-based protocols usually use a random mechanism, such as competition or voting, to select a proposed block as authoritative to extend the chain. Distributed consensus protocol (DCP) is a class of coordination protocol that is typically preferred in public blockchain platforms. DCP allows the participating nodes to agree on what should be written to the blockchain, relying on probabilistic algorithms that rely on randomness (e.g., Proof-of-Work or Proof-of-Stake) to decide which node proposes the next block.

Distributed consensus algorithms are generally complex and require many properties (inherently or with supplementary solutions) to ensure their correctness and security, such as being fair, fast, provable.

7.2.3 Blockchain Platforms Architecture

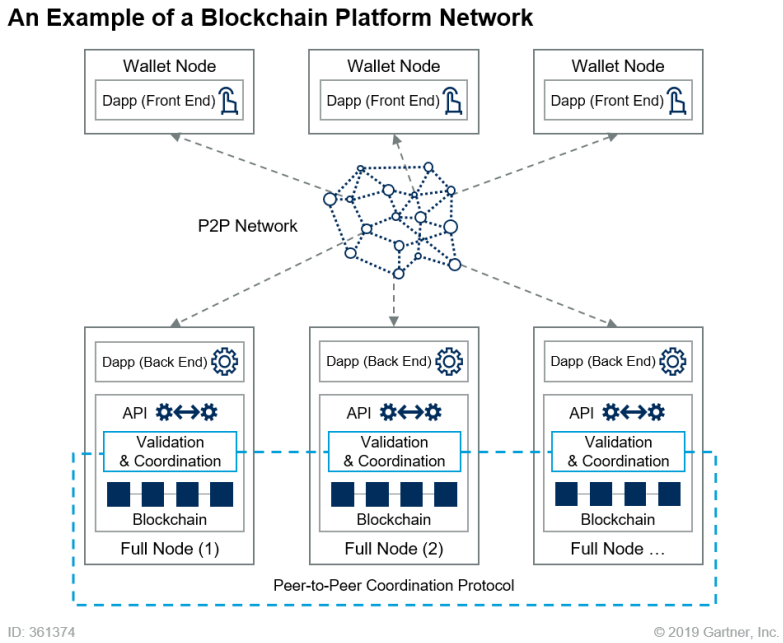

Figure 32: Blockchain platform network

A blockchain platform is a decentralized state transition machine that immutably stores and autonomously operates on the state of digital assets, which may be modeled by tokens and implemented as objects that encapsulate state (properties) and rules (behavior). It is built on a peer-to-peer network of distributed nodes that use a blockchain framework to implement a validation and coordination protocol to write records of interactions or transactions to an immutable data structure.

7.2.3.1 Platform interoperability

The interoperability options in the case of using blockchain technology in Platoon platforms, must be evaluated as this is the foundation of streamlining and automating complex multiparty processes. Interoperability must ensure safe cross-chain digital asset change. Interoperability can be achieved three ways:

- **Notary schemes** (exchange of arbitrary data): employ trusted federation of nodes to corroborate events on a chain in relation to another chain, checking if the event that took place in another chain is true. The notaries come to an agreement through a consensus

algorithm and then issue a signature that can be used to finalize the state change. This is called federated sidechain.

- **Relays** (exchange of arbitrary data); allow event verification of another chain. This can be used in cross-chain applications. Relay chains are distinct blockchains, whose architectures require flexible multisig capability and fast consensus finality.
- **Hashed time locks** (exchange of digital assets only): a hashed time lock contract (HTLC) is a class of blockchain-based payments, where hash and time locks are used to require the receiver to acknowledge receipt.

7.2.4 General business requirements

The business stream includes non-technical tasks to make a blockchain solution operational. A dapp is not an isolated application running on a server and depends on many components that are run on many nodes by operators in a decentralized model. This ecosystem needs a set of principles and policies to govern operational and business processes and must evolve in compliance with applicable laws.

The key requirements for decentralized processes include assets (tokens for identity and state), contracts (rules and trusted data), control (functions and events), interoperability (networks and mechanisms) and actors (roles and entitlements).

Blockchain networks can be boostrapped or joined, but a minimum number of network participants are needed to make a multiparty computing environment and dapp operational. In this case, the blockchain platform operators (responsible for running blockchain nodes), solution operators (responsible for dapp development), and suppliers and consumer participants make up the key participants.

A participation mechanism must be set up to define the power of blockchain operators, ensure proportional power, and enable and implement open, collaborative and transparent decision making. Open digital ecosystem governing is more challenging as it is made up of many unknown and loosely connected entities, unlike closed digital ecosystems (consortium or associations).

7.2.4.1 Blockchain adoption barriers

Blockchain solutions present two complex issues:

- The first one is related to the business and technical design convergence: A blockchain solution is an embodiment of business agreements and processes in technical protocols which are difficult, especially considering the extent of the infrastructure that supports technology and business operations of different organizations internally and externally.

- The second one is related to the evolution and maturity of blockchain technologies, a blockchain solution architecture must anticipate the upgrade and evolution of blockchain technologies while meeting operational requirements, including decentralized trust, security, safety and liveness in worse-case scenarios, especially when volume changes and/or when the network comes under malicious attacks.

7.2.5 IDS & Blockchain

"The International Data Spaces (IDS) reference architecture focuses on the concept of 'data sovereignty', allowing organizations to share datasets in a secure and controlled way using the International Data Spaces Connector concept"[cxxxix]. Some of the features of blockchain technology are consistent with features of the International Data Spaces architecture, such as the absence of a single trusted party (e.g. where all data is being stored) and the decentralized nature. Other features are

complementary, such as the permanence of the blockchain record. This makes it highly interesting to explore how IDS and blockchain technology fit together and how blockchain technology could be used in future versions of the International

The promising combination of Industrial Data Space and blockchain could lead to an audit-proof data transaction log.[3]

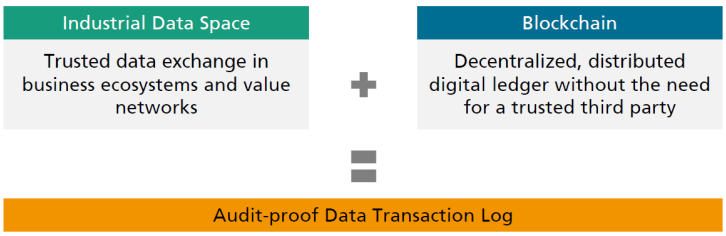

Figure 33: Blockchain and IDS

Blockchains play an important role when data is considered a 'shared asset', which needs to be stored in an immutable way between partners in the ecosystem and in several projects, IDS is used in conjunction with blockchain technology:

- Within the IDS architecture a 'Blockchain app' connects an IDS Connector with the API/SDK of a blockchain client, which in turn is connected to a node in the blockchain. Through this API, data can be shared with the blockchain, either acting as a data provider or data consumer.
- The IDS architecture contains the concepts of a Broker and Clearing house. Both concepts can be potentially implemented using blockchain technology.

Projects that are currently exploring the usage of blockchain technology in the context of IDS include:

- The AMAble project that aims to provide a data infrastructure for 3D printing.
- The BOOST 4.0 project that aims to enhance manufacturing through the use of big data. IDS is used as a cornerstone for this to share data in various use cases.
- The MARKET 4.0 project that aims to develop a marketplace for equipment manufacturers.
- The MIDIH project (Digital Innovation Hub for the Manufacturing Industry) experiments with open source industrial data platforms in the domains of Smart Factory, Smart Product and Smart Supply Chains.

7.3 Private and Secure AI

Private and secure AI consists of an ensemble of techniques that allow ML engineers to train models without having direct access to the data used for the training and also using cryptography in order to avoid getting any information about the data.

The overall justification of the need of these techniques are the concerns regarding data privacy in AI-machine learning training tasks. The issue of data privacy has come under the increasing attention of

[3] Figure by Prof. Jan Jürjens Director Research Projects, Fraunhofer Institute for Software & Systems Engineering ISST, Dortmund (Germany)

regulators and public due to several big scandals in several companies in recent years, so data privacy is a very important concern leading to enterprises having legal risks and competitive advantages.

With data availability essential to any machine learning model, creative ways must be devised to circumvent restrictions and enable model training to happen without the data actually having to leave its collection and storage location.

Organizations could implement encryption as part of their security baseline. However, this technique is not applicable when sharing the data with other organizations and it is not enough to avoid data breaches. This encryption creates a “crypto boundary”, where data is accessible in plain text, reducing is protection capability. Also, simple data anonymization can lead to privacy being breached using alternative auxiliary information to de-anonymize data.

Data masking can provide protection where the encryption has its limitations, by implementing a wide range of field-level data transformations, where the original data´s syntax and semantics may be preserved. However, this technique cannot provide the support for transactional or behavioural data such as location or banking transaction history.

Due to these limitations, new techniques have to be applied for data protection where the original data is not necessary outside the trusted boundaries. Data encryption is not enough for the cases the data has to be protected at all stages (“at rest”, “in transit” and “in use”). There are also cases where the data that needs to be protected cannot be masked using traditional masking techniques, as the real data needs to be used.

The scenarios that require privacy enhancement techniques normally include:

1. **Limited data collection** due the focus on privacy and the increasing regulations.
2. **Cloud and other third-party data processing** where the organisation has no control over the software and the computing environments.
3. **Internal and individual third party sharing** relating to how entities are granted authorisation to data sources.
4. **Multiple data sharing and processing** where an organisation may share data with multiple parties, without actually disclosing respective information.

Once the need is justified, the next step is to decide on the technique to be used. Private and Secure AI framework relies on three main techniques:

- Federated learning
- Differential Privacy
- Secured Multi-Party Computation

7.3.1 Federated Learning

Federated learning, also known as collaborative learning, is a ML technique that trains algorithms across multiple decentralized edge devices or servers holding local data samples, without the need to exchange these samples. Federated learning differs from traditional approaches in the fact that there is no need to upload all the data samples to one server (centralized ML techniques), or assume that local samples need to be identically distributed (classical decentralized approaches).

Federated learning enables multiple actors to build a common, robust machine learning model without sharing data, thus addressing critical issues such as data privacy, data security, data access rights and access to heterogeneous data. Its applications are spread over a number of industries including defence, telecommunications, IoT or pharmaceutics.

Since data never leaves its original location, federated learning opens up the possibility for different data owners at the organizational level to collaborate and share their data.

Federated learning can be divided into horizontal and vertical learning: [cxl]

An example of horizontal federated learning could be the case of two regional DSOs that have no overlapping clientele, but their data will have similar feature spaces since they have very similar business models. Here, they might come together to collaborate.

In vertical federated learning, two companies providing different services but having a large intersection of clientele might find room to collaborate on the different feature spaces they own, leading to better outcomes for both of them.

In both cases, the data owners are able to collaborate without having to sacrifice their respective clientele's privacy.

Another illustrative example is the healthcare sector, hospitals and other healthcare providers stand to gain if they are able to share patient data for model training in a privacy-preserving manner.

Federated learning can be applied with the following types of algorithms:

- Linear models and neural networks can be trained using Stochastic Gradient Descent (SGD) variants.
- Some clustering algorithms may be easily implemented such as K-means.
- Kernel methods such as SVMs, contain training data, so to keep privacy, semiparametric approximations will be used where a group of centroids are selected to obtain weights using the stochastic gradient descent procedure.

Federated learning is already being used by Google and Apple to train machine learning models on millions of phones. However, since the model is being sent to these devices, it can be stolen and analyzed to reveal private information, so this technique has to be complemented with the others in the framework.

7.3.2 Differential privacy

"Differential privacy describes a promise, made by a data holder, or curator, to a data subject: You will not be affected, adversely or otherwise, by allowing your data to be used in any study or analysis, no matter what other studies, data sets, or information sources, are available."[cxli]

The goal of differential privacy is to ensure that different kinds of statistical analysis do not compromise privacy. So, differential privacy proposes and develops tools and techniques to allow data holders and curators to maintain privacy. It uses data perturbation to transform data by adding noise and hide the individual data values.

Also, differential privacy allows us to measure how much private information is being leaked or revealed by our model. Noise can be added to model operations to effectively hide private information that would otherwise be leaked, and even adjust the noise to keep the leakage below some threshold. Adding noise comes at the cost of performance but we get the benefit that information about our users is kept secret.

Technically, differential privacy is a form of field-level data masking where data can be used for querying aggregate statistics while limiting exposure. It consists of an algorithm that hides the presence or absence of individual data.

There are two types of differential privacy: global and local. In global privacy, controlled noise is added to the result of the query, so that accuracy is not sacrificed. If the data owner can trust the curator, it is the recommended method.

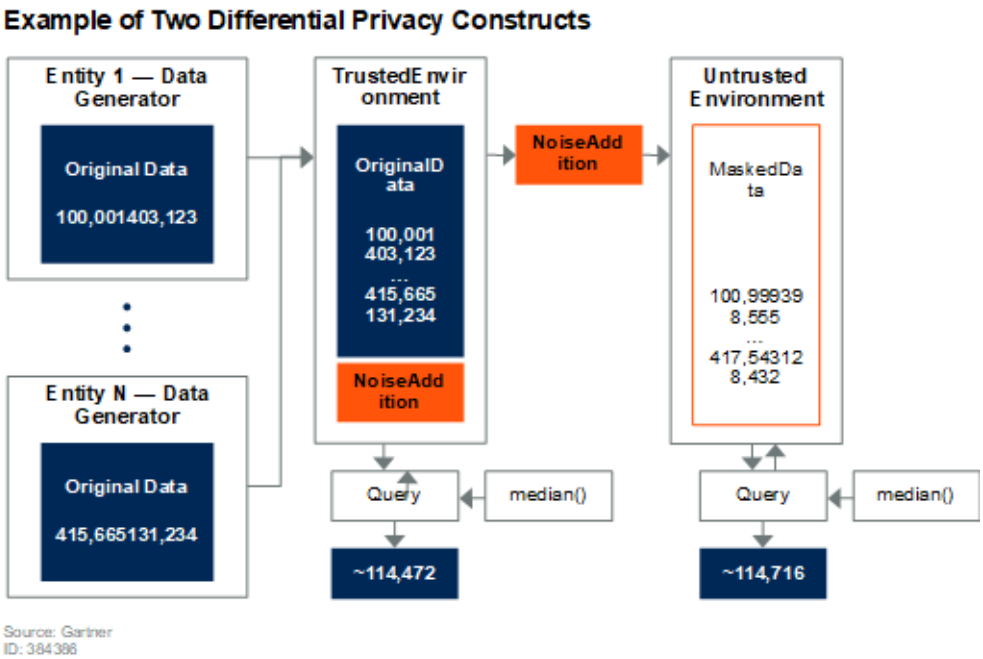

Figure 34: Global and local differential privacy

In local privacy, the noise is added before the data is stored in individual points for the whole database, sacrificing accuracy in order to achieve privacy. This technique has limitations, as the number of data types and use cases applicable is small, hence needing to be combined with additional data masking techniques for anonymisation, or the addition of deterministic noise. Also, in the context of deep learning would be the case when removing a data point or row from the dataset and training the neural model results in a neural model that is comparable to the model generated with the original dataset. The fact that neural models rarely converge to the same location even when trained on the same dataset poses a problem to demonstrate the privacy of neural models.

7.3.3 Secure Multiparty Computation

Secure multi-party computation, (SMPC) also known as secure computation, multi-party computation (MPC), or privacy-preserving computation, is a subfield of cryptography with the goal of creating methods for parties to jointly compute a function over their inputs while keeping those inputs private. [cxlii]

Unlike traditional cryptographic tasks, where cryptography assures the security and integrity of communication or storage and the adversary is outside the system of participants (an eavesdropper on the sender and receiver), the cryptography in this model protects participants' privacy from each other. Therefore, computations can be performed on data that is distributed among multiple parties,

and there need not be a trusted third party as each party can only see their contribution. Participants only have access to the results and not the specific data that was contributed by the other parties. Most SMPC protocols, make use of secret sharing where encrypted portions of data needed are divided among multiple participants.

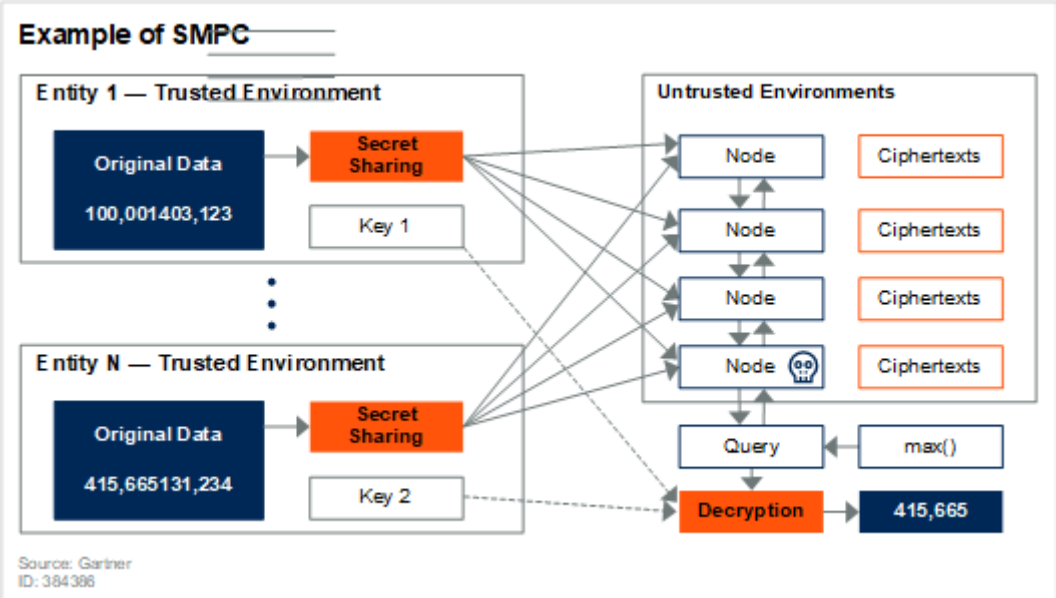

Figure 35: SMPC example

There is a new type of SMPC: Private Join and Compute, which combines two fundamental cryptographic techniques to protect individual data[cxliii]:

- Private set intersection allows two parties to privately join their sets and discover the common identifiers.
- Homomorphic encryption allows certain types of computation to be performed directly on encrypted data without previous decryption. This helps preserve raw data privacy.

7.4 Data preparation

In the case of Platoon, several partners have expressed the need for high quality data that allow meaningful analysis, so the development or introduction of data preparation tools with similar characteristics to those described below could be considered, as the first step in a data exchange process.

Data preparation is the act of manipulating (or pre-processing) raw data (which may come from disparate data sources) into a form that can readily and accurately be analysed. It is the first step into data analytics projects and it can include many discrete tasks such as data loading or ingestion, fusion, cleaning, augmentation and delivery. [cxliv]

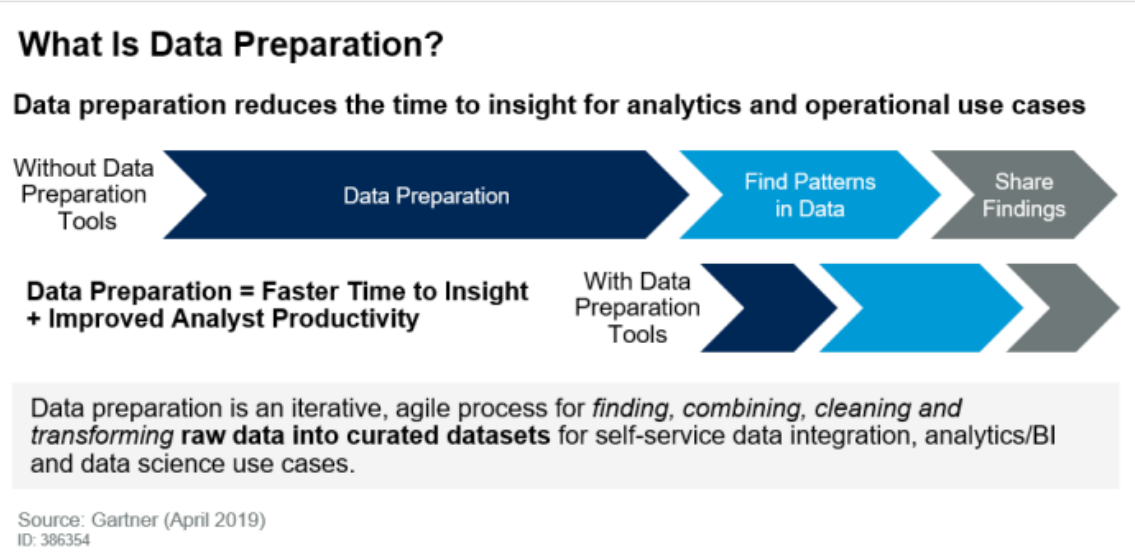

Figure 36: Data preparation advantages

The issues to be dealt with fall into two main categories:

- Systematic errors involving large numbers of data records, probably because they come from different sources.
- Individual errors affecting small numbers of data records, probably due to errors in the original data entry.

Data preparation saves time and allows analysts to generate insights from data.

7.4.1 Self-service data preparation

The objective in modern and state of the art Analytics &Business Intelligence platforms and especially in data preparation-oriented tools is to make this work simpler, visual and more intuitive.

Instead of involve data scientists / data engineers with a deep knowledge on how to find, combine, clean, and transform raw data into curated datasets for analysis, aspiring self-service data preparation tools help business users to understand data preparation from a conceptual and procedural standpoint.

Gartner[cxlv] for example, consider self-service Data Preparation as a vital component of the next generation of Business Analytics and Business Intelligence to make Advanced Data capabilities accessible to team members and business users no matter their skills or technical knowledge.

Regarding implementation, self-service Data Preparation consists of a set of tools, designed for ease-of-use and access by business users in a self-serve environment. Some of their characteristics are:

- Grant access to data from multiple sources.
- Interactive Exploration: Detailed visual representations allow for deeper data exploration, with an automatic understanding of the data at its most granular level provides business users with powerful capabilities to explore, manipulate and merge new data sources.
- Intuitive and flexible interface to allow a non-technical user to be able to compile and prepare data, test hypotheses, visualize and share data and prepare and execute analysis.

- Predictive Transformation: Every transformation leads to a prediction and the tool provides a ranked list recommendation for users to evaluate or edit.
- Collaborative Data Governance: support is provided for collaborative security, access, data lineage and metadata.

Again, these tools allow business users to transform, shape, reduce, combine, explore, clean, sample and aggregate data, without the need for SQL skills, ETL or other programming language.

We can also talk about augmented data preparation, when apart from intuitive interfaces, the tool uses machine learning automation to augment data profiling and data quality, harmonization, modeling, manipulation, enrichment, metadata development and cataloguing.

So the more advanced tools in this category of self-service data preparation uses state of the art visualization and UX/UI interfaces and machine learning to clean and transform data sets.

Figure 37: Self-service data Preparation Tools[4]

[4] Image from https://www.smarten.com/products/self-serve-data-preparation.html

8 Internal Review

8.1 Internal Review 1

Mark with X the corresponding column:

Y= yes	N= no	NA = not applicable

Name of reviewer: Philippe CALVEZ
Organisation: ENGIE
Date: 29.06.2020

ELEMENT TO REVIEW	Y	N	NA	Comments	Author
FORMAT: Does the document …?					
…include editors, deliverable name, version number, dissemination level, date, and status?	X				
… contain a license (in case of public deliverables)?		X			
… include the names of contributors and reviewers?	X				
… contain a version table?	X				
… contain an updated table of contents?	X				
… contain a list of figures?	X				
… contain a list of tables?	X				
… contain a list of terms and abbreviations?	X				
… contain an Executive Summary?	X				
… contain a Conclusions section?		X		Perhaps considering the Part 4 Requirements as a conclusion	P.Calvez
… contain a List of References (Bibliography) in the appropriate format?	X				
… use the fonts and sections defined in the official template?	X				
… use correct spelling and grammar?	X				
… conform to length guidelines (50 pages maximum (plus Executive Summary and annexes)	X				

ELEMENT TO REVIEW	Y	N	NA	Comments	Author
… conform to guidelines regarding Annexes (inclusion of complementary information)	X				
… present consistency along the whole document in terms of English quality/style? (to avoid accidental usage of copy & paste text)	X				
About the content…					
Is the deliverable content correctly written?	X				
Is the overall style of the deliverable correctly organized and presented in a logical order?	X				
Is the Executive Summary self-contained, following the guidelines and does it include the main conclusions of the document?	X				
Is the body of the deliverable (technique, methodology results, discussion) well enough explained?	X				
Are the contents of the document treated with the required depth?	X				
Does the document need additional sections to be considered complete?		X			
Are there any sections in the document that should be removed?		X			
Are all references in the document included in the references section?	X			Perhaps considering another way to number the references instead of Roman numerals.	P.Calvez
Have you noticed any text in the document not well referenced? (copy and paste of text/picture without including the reference in the reference list)		X			
TECHNICAL RESEARCH WPs (WP2-WP5)					
Is the deliverable sufficiently innovative?					

ELEMENT TO REVIEW	Y	N	NA	Comments	Author
Does the document present technical soundness and its methods are correctly explained?					
What do you think is the strongest aspect of the deliverable?					
What do you think is the weakest aspect of the deliverable?					
Please perform a brief evaluation and/or validation of the results, if applicable.					
VALIDATION WP (WP6)					
Does the document present technical soundness and the validation methods are correctly explained?					
What do you think is the strongest aspect of the deliverable?					
What do you think is the weakest aspect of the deliverable?					
Please perform a brief evaluation and/or validation of the results, if applicable.					
DISSEMINATION AND EXPLOITATION WPs (WP8 & WP9)					
Does the document present a consistent outreach and exploitation strategy?					
Are the methods and means correctly explained?					
What do you think is the strongest aspect of the deliverable?					
What do you think is the weakest aspect of the deliverable?					
Please perform a brief evaluation and/or validation of the results, if applicable.					

SUGGESTED IMPROVEMENTS

PAGE	SECTION	SUGGESTED IMPROVEMENT

CONCLUSION

Mark with X the corresponding line.

X	Document accepted; no changes required.
	Document accepted; changes required.
	Document not accepted; it must be reviewed after changes are implemented.

Please rank this document globally on a scale of 1-5.
(1-Poor; 2–Fair; 3–Average; 4–Good; 5–Excellent)
Using a half point scale.
Mark with X the corresponding grade.

Document grade	1	1.5	2	2.5	3	3.5	4	4.5	5
								X	

8.2 Internal Review 2

Mark with X the corresponding column:

Y= yes	N= no	NA = not applicable

Name of reviewer: Erik MAQUEDA
Organisation: TECNALIA
Date: 30.06.2020

ELEMENT TO REVIEW	Y	N	NA	Comments	Author
FORMAT: Does the document …?					
…include editors, deliverable name, version number, dissemination level, date, and status?	X				
… contain a license (in case of public deliverables)?		X			
… include the names of contributors and reviewers?	X				
… contain a version table?	X				
… contain an updated table of contents?	X				
… contain a list of figures?	X				
… contain a list of tables?		X		I would put a title for each of the requirements tables and add them to the list of tables.	
… contain a list of terms and abbreviations?	X				
… contain an Executive Summary?	X				
… contain a Conclusions section?		X			
… contain a List of References (Bibliography) in the appropriate format?	X			Some references need to be added at the end (see comments in the document).	
… use the fonts and sections defined in the official template?	X				
… use correct spelling and grammar?	X				

ELEMENT TO REVIEW	Y	N	NA	Comments	Author
… conform to length guidelines (50 pages maximum (plus Executive Summary and annexes)	X				
… conform to guidelines regarding Annexes (inclusion of complementary information)	X				
… present consistency along the whole document in terms of English quality/style? (to avoid accidental usage of copy & paste text)	X				
About the content…					
Is the deliverable content correctly written?	X				
Is the overall style of the deliverable correctly organized and presented in a logical order?	X				
Is the Executive Summary self-contained, following the guidelines and does it include the main conclusions of the document?	X				
Is the body of the deliverable (technique, methodology results, discussion) well enough explained?	X				
Are the contents of the document treated with the required depth?	X				
Does the document need additional sections to be considered complete?		X			
Are there any sections in the document that should be removed?		X			
Are all references in the document included in the references section?	X			Some references need to be added at the end (see comments in the document).	
Have you noticed any text in the document not well referenced? (copy and paste of text/picture without including the reference in the reference list)	X			Some references need to be added at the end (see comments in the document).	

ELEMENT TO REVIEW	Y	N	NA	Comments	Author
TECHNICAL RESEARCH WPs (WP2-WP5)					
Is the deliverable sufficiently innovative?					
Does the document present technical soundness and its methods are correctly explained?					
What do you think is the strongest aspect of the deliverable?					
What do you think is the weakest aspect of the deliverable?					
Please perform a brief evaluation and/or validation of the results, if applicable.					
VALIDATION WP (WP6)					
Does the document present technical soundness and the validation methods are correctly explained?					
What do you think is the strongest aspect of the deliverable?					
What do you think is the weakest aspect of the deliverable?					
Please perform a brief evaluation and/or validation of the results, if applicable.					
DISSEMINATION AND EXPLOITATION WPs (WP8 & WP9)					
Does the document present a consistent outreach and exploitation strategy?					
Are the methods and means correctly explained?					
What do you think is the strongest aspect of the deliverable?					
What do you think is the weakest aspect of the deliverable?					
Please perform a brief evaluation and/or validation of the results, if applicable.					

SUGGESTED IMPROVEMENTS

PAGE	SECTION	SUGGESTED IMPROVEMENT

CONCLUSION

Mark with X the corresponding line.

	Document accepted; no changes required.
X	Document accepted; changes required.

	Document not accepted; it must be reviewed after changes are implemented.

Please rank this document globally on a scale of 1-5.
(1-Poor; 2–Fair; 3–Average; 4–Good; 5–Excellent)
Using a half point scale.
Mark with X the corresponding grade.

Document grade	**1**	**1.5**	**2**	**2.5**	**3**	**3.5**	**4**	**4.5**	**5**
						X			

9 References

[i] https://www.h2020-bridge.eu/
[ii] Grant Agreement-872592-PLATOON
iii OECD (2019). Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies. OECD publishing. Paris https://doi.org/10.1787/276aaca8-en
iv European Commission, Digitising European Industry. https://ec.europa.eu/digital-single-market/en/policies/digitising-european-industry
[v] https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy
vi https://www.iea.org/reports/digitalisation-and-energy
[vii] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
[viii] • NERC-GADS, GADS Wind Turbine Generation: Data Reporting Instructions: Effective January 2010; NERC: Atlanta, GA, USA, 2010.
[ix] • ReliaWind system structure. Reder, M.D.; Gonzalez, E.; Melero, J.J. Wind turbine failures—Tackling current problems in failure data analysis. J. Phys. Conf. Ser. 2016. Gayo, J.B. Final Publishable Summary of Results of Project ReliaWind; Gamesa Innovation and Technology, Spain, 2011
[x] • RDS-PP. VGB PoweTech e.V. VGB-Standard RDS-PP: Application Guideline Part 32: Wind Power Plants: VGB-S823-32-2014-03-EN-DE; Verlag Technisch-Wissenschaftlicher Schriften: Essen, Germany, 2014. This seems to be the most comprehensive and up-to-date standard. Some European manufacturers and operators of WT already make use of it and were also involved in the development of the guideline.
[xi] • IRPWind (Integrated Research Programme on Wind Energy), Taxonomy and meta data for wind energy R&D. Project funded by the EC, led by DTU and developed by the Joint Programme on Wind Energy of the European Energy Research Alliance (EERA JPWind), December 2017.
[xii] Oliver Gassmann, Karolin Frankenberger, Michaela Csik, working paper of the University of St. Gallen
[xiii] Demil and Lecocq 2010; Osterwalder and Pigneur, 2010
[xiv] Full name of the e-Directive: DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on common rules for the internal market for electricity and amending Directive 2012/27/EU (recast)
[xv] https://www.gartner.com/smarterwithgartner/the-road-to-the-api-economy/
[xvi] https://de.wikipedia.org/wiki/Representational_State_Transfer
[xvii] https://opencv.org/
[xviii] https://www.internationaldataspaces.org/the-association/#mission
[xix] https://www.w3.org/RDF/
[xx] https://www.w3.org/TR/rdf-schema/
[xxi] https://www.w3.org/OWL/
[xxii] R. Y. Wang, D. M. Strong, 2013
[xxiii] Buneman, P., Khanna, S. and Wang-Chiew, T. (2001). Why and where: A characterization of data provenance. In Database Theory—ICDT, Springer, 316-330.
[xxiv] Omitola, T., Gibbins, N. and Shadbolt, N. (2010): Provenance in Linked Data Integration. In Future Internet Assembly, Ghent, Belgium.
[xxv] Glavic, B. and Alonso, G. (2009): Perm: Processing provenance and data on the same data model through query rewriting. Data Engineering, In ICDE, IEEE 25th International Conference on, IEEE, 174-185.
[xxvi] Brownlow, J., Zaki, M., Neely, A., & Urmetzer, F. (2015). Data and analytics-data-driven business models: A blueprint for innovation. Cambridge Service Alliance.
[xxvii] Schroeder, Ralph. "Big data business models: Challenges and opportunities." Cogent Social Sciences 2.1 (2016): 1166924.
[xxviii] https://www.insightsforprofessionals.com/it/leadership/data-compliance-standards
[xxix] https://blog.netwrix.com/2019/08/06/data-security-compliance-essentials-only/

[xxx] Rocha, Artur & Oliveira, Marco & Freire, Filipe & David, Gabriel & Vilar, Pedro & Taboada, Begoña & Iglesias, Isabel & Lazaro, Clara & Bastos, Maria & van Golde, Ilmer & Jorge da Silva, António. (2016). Data discovery mechanisms and metadata handling in RAIA Coastal Observatory

[xxxi] Capiello, C., Gal, A., Jarke, M., & Rehof, J. (2020). Data Ecosystems: Sovereign Data Exchange among Organizations (Dagstuhl Seminar 19391). In Dagstuhl Reports (Vol. 9, No. 9). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

[xxxii] Hanson, C., Berners-Lee, T., Kagal, L., Sussman, G. J., & Weitzner, D. (2007, June). Data-purpose algebra: Modeling data usage policies. In Eighth IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY'07) (pp. 173-177). IEEE

[xxxiii] Huu, Quyet Cao. Policy-based usage control for trustworthy data sharing in smart cities. Diss. 2017

[xxxiv] Liang, F., Yu, W., An, D., Yang, Q., Fu, X., & Zhao, W. (2018). A survey on big data market: Pricing, trading and protection. IEEE Access, 6, 15132-15154.

[xxxv] Muschalle, A., Stahl, F., Löser, A., & Vossen, G. (2012, August). Pricing approaches for data markets. In International workshop on business intelligence for the real-time enterprise (pp. 129-144). Springer, Berlin, Heidelberg

[xxxvi] Vidal, M. E., Endris, K. M., Jozashoori, S., Karim, F., & Palma, G. (2019). Semantic data integration of big biomedical data for supporting personalised medicine. In Current Trends in Semantic Web Technologies: Theory and Practice (pp. 25-56). Springer, Cham.

xxxvii https://gdpr-info.eu/art-32-gdpr/

xxxviii TeleTrust, «Guideline "State Of The Art" Technical and organisational measures,» 2020. [Online]. Available: https://www.teletrust.de/fileadmin/docs/fachgruppen/ag-stand-der-technik/2020-01-TeleTrusT_Guideline_State_of_the_art_in_IT_security_ENG.pdf.

xxxix Enisa, «ENISA Threat Landscape Report 2018,» January 2019. [Online]. Available: https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-2018.

xl https://owasp.org/www-community/attacks/Man-in-the-browser_attack

xli https://owasp.org/www-community/attacks/Man-in-the-middle_attack

xlii https://searchsecurity.techtarget.com/definition/watering-hole-attack#:~:text=A%20watering%20hole%20attack%20is,the%20target's%20place%20of%20employment.

xliii https://owasp.org/www-community/attacks/SQL_Injection

xliv https://owasp.org/www-community/attacks/xss/

xlv https://www.cryptopp.com/wiki/GCM_Mode

xlvi https://www.cryptopp.com/wiki/EAX_Mode

xlvii Kahina Khacef, Guy Pujolle. Secure Peer-to-Peer communication based on Blockchain. 33rd International Conference on Advanced Information Networking and Applications (AINA-2019), Mar 2019, Matsue, Japan. pp.662-672, ff10.1007/978-3-030-15035-8_64ff. ffhal-02180329ff

xlviii Pavithran, Deepa & Shaalan, Khaled. (2019). Towards Creating Public Key Authentication for IoT Blockchain. 110-114. 10.1109/ITT48889.2019.9075105.

xlix https://docs.wso2.com/display/IS530/XACML+Architecture

l Ouaddah, Aafaf & Elkalam, Anas & Ouahman, Abdellah. (2017). Towards a Novel Privacy-Preserving Access Control Model Based on Blockchain Technology in IoT. 10.1007/978-3-319-46568-5_53.

li Outchakoucht, Aissam & ES-SAMAALI, Hamza & Philippe, Jean. (2017). Dynamic Access Control Policy based on Blockchain and Machine Learning for the Internet of Things. International Journal of Advanced Computer Science and Applications. 8. 10.14569/IJACSA.2017.080757.

lii Park, Jaehong & Sandhu, Ravi. (2002). The UCON ABC usage control model. ACM Transactions on Information and System Security. 7. 128-174.

liii https://www.internationaldataspaces.org/wp-content/uploads/2019/11/Usage-Control-in-IDS-V2.0_final.pdf

liv https://gdpr-info.eu/

lv https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en
lvi https://gdpr-info.eu/art-25-gdpr/
lvii https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/controller-processor/what-data-controller-or-data-processor_en
lviii https://gdpr-info.eu/art-5-gdpr/
lix European Data Protection Board, «Guidelines 4/2019 on Article 25 Data Protection by Design and by Default,» [Online]. Available: https://edpb.europa.eu/sites/edpb/files/consultation/edpb_guidelines_201904_dataprotection_by_design_and_by_default.pdf
lx https://gdpr-info.eu/chapter-3/
lxi https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/individual-rights/
lxii J.-H. Hoepman, «Privacy Design Strategies,» in *ICT Systems Security and Privacy Protection: 29th IFIP TC 11 International Conference, SEC 2014, Marrakech, Morocco, June 2-4, 2014. Proceedings*, 2014, pp. 446-459.
lxiii *ENISA, «Privacy and Data Protection by Design - from policy to engineering,» December 2014. [Online]. Available: https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design.*
lxiv https://edps.europa.eu/data-protection/our-work/subjects/necessity-proportionality_en
lxv Royal Society, «Protecting privacy in practice,» [Online]. Available: https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/privacy-enhancing-technologies-report.pdf.
lxvi https://mydata.org/
lxvii https://github.com/mydata-sdk/mydata-docs
lxviii https://gdpr-info.eu/art-6-gdpr/
lxix https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=57632
lxx L. Daniele, F. den Hartog, J. Roes, "Created in Close Interaction with the Industry: The Smart Appliances REFerence (SAREF) Ontology", Formal Ontologies Meet Industry. Springer International Publishing, pp. 100-112, 2015.
lxxi http://www.fiemser.eu/?page_id=112
lxxii http://www.buildingsmart-tech.org/ifc/IFC4/final/html/
lxxiii http://www.w3.org/TR/owl-time/
lxxiv https://sites.google.com/site/smartappliancesproject/ontologies
lxxv https://ontology.tno.nl/saref4ener/
lxxvi https://ontology.tno.nl/saref4ee/
lxxvii http://www.energy-home.it/
lxxviii https://www.eebus.org
lxxix https://www.controlglobal.com/articles/2019/introduction-to-modbus/
lxxx https://www.controlglobal.com/articles/2019/introduction-to-modbus/
lxxxi https://www.se.com/uk/en/faqs/FA293162/
lxxxii http://www.bacnet.org/
lxxxiii https://www.knx.org/knx-en/for-professionals/index.php
lxxxiv Palmer; Shenoi, Sujeet, eds. (23–25 March 2009). Critical Infrastructure Protection III. Third IFIP WG 11. 10 International Conference. Hanover, New Hampshire: Springer. p. 87. ISBN 3-642-04797-1.
lxxxv https://iot-ontologies.github.io/dogont/
lxxxvi https://www.scitepress.org/Papers/2016/57455/pdf/index.html
lxxxvii https://hal.archives-ouvertes.fr/hal-00988944
lxxxviii https://en.wikipedia.org/wiki/BACnet
lxxxix http://www.bacnet.org/Bibliography/ES-7-96/ES-7-96.htm
xc KBOB_Empfehlung_BACnet_e_2017_v1.1.pdf
xci https://www.ccontrols.com/pdf/BACnetIntroduction.pdf

xcii http://www.bacnet.org/Bibliography/RSES-1-99.htm
xciii https://www.bacnetinternational.org/page/faq#Q5
xciv KBOB_Empfehlung_BACnet_e_2017_v1.1.pdf
xcv http://bacowl.sourceforge.net/intro.html
xcvi Smart Grid Reference Architecture by CEN-CENELEC-ETSI Smart Grid Coordination Group, Nov 2012,
https://ec.europa.eu/energy/sites/ener/files/documents/xpert_group1_reference_architecture.pdf

xcvii https://www.openadr.org/
xcviii
https://www.openadr.org/assets/OpenADR%20for%20Smart%20Inverter%20Control_final.pdf
xcix 'Trends Concerning Standardization of OpenADR | NTT Technical Review'. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr201312gls_s.html (accessed Jun. 12, 2020).
c OpenADR_2_0b_Profile_sepcification_v1_1.pdf
ci https://www.usef.energy/
cii USEF, The Framework explained, 2015, https://www.usef.energy/
ciii Smart Energy Collective, An introduction to the Universal Smart Energy Framework v1.0, 2013, https://ec.europa.eu/energy/sites/ener/files/documents/xpert_group3_summary.pdf
civ https://www.usef.energy
cv GSMA, 2018, https://www.gsma.com/iot/wp-content/uploads/2018/07/CLP.26-v5.0.pdf
cvi JSON, http://www.json.org/
cvii FIWARE-NGSIv2 Specification, http://fiware.github.io/specifications/ngsiv2/stable/
cviii Schema.org, http://schema.org/
cix FIWARE, http://www.fiware.org/
cx OASC, http://oascities.org/
cxi GSMA, 2018, https://www.gsma.com/iot/wp-content/uploads/2018/11/CLP.25-v2.0.pdf
cxii GSMA, 2016, https://www.gsma.com/iot/wp-content/uploads/2016/11/CLP.24-v1.0.pdf
[cxiii] http://www.bdva.eu
[cxiv] http://bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf
[cxv] http://bdva.eu/PositionDataStrategy
[cxvi] http://www.bdva.eu/PPP
[cxvii] https://ec.europa.eu/digital-single-market/en/news/rolling-plan-ict-standardisation
cxviii https://www.synchronicity-iot.eu
cxix SynchroniCity D2.10: https://synchronicity-iot.eu/wp-content/uploads/2018/09/SynchroniCity_D2.10.pdf
cxx https://gitlab.com/synchronicity-iot/synchronicity-data-models
cxxi https://gitlab.com/synchronicity-iot
[cxxii] https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf
cxxiii C. &. N. D. &. Y. S. &. G. R. &. M. M. &. P. P. Mouradian, «A Comprehensive Survey on Fog Computing: State-of-the-art and Research Challenges.,» IEEE Communications Surveys & Tutorials. , 2017.
cxxiv H. &. W. R. &. W. G. Atlam, «Fog Computing and the Internet of Things: A Review.,» Big Data and Cognitive Computing., 2018.
cxxv M. A. A. F. a. K. Vatanparvar, «Energy Management-as-a-Service Over Fog Computing Platform,» IEEE Internet of Things Journal, 2016.
cxxvi M. &. B. M. Hussain, «Fog Computing for Internet of Things (IoT)-Aided Smart Grid Architectures.,» Big Data and Cognitive Computing., 2019.
cxxvii E. H. Glaessgen e D. S. Stargel, «The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles,» in *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference - Special Session on the Digital Twin*, Honolulu, HI, 2012.

cxxviii W. Kritzinger, M. Karner, G. Traar, J. Henjes e W. Sihn, «Digital Twin in manufacturing: A categorical literature review and classification,» *IFAC PapersOnLine,* vol. 51, n. 11, pp. 1016-1022, 2018.
cxxix A. Fuller, Z. Fan, C. Day e C. Barlow, «Digital Twin: Enabling Technologies, Challenges and Open Research,» [Online]. Available: https://arxiv.org/ftp/arxiv/papers/1911/1911.01276.pdf
cxxx https://oascities.org/wp-content/uploads/2019/06/OASC-MIMs.pdf
cxxxi B. &. D. B. &. D. A. Rani, «Cloud Computing and Inter-Clouds – Types, Topologies and Research Issues.,» 2015.
cxxxii NIST, «The NIST Definition of Cloud,» [Online]. Available: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf.
cxxxiii Red Hat, «Types of cloud computing,» [Online]. Available: https://www.redhat.com/en/topics/cloud-computing/public-cloud-vs-private-cloud-and-hybrid-cloud.
cxxxiv Cloud Standards Customer Council, «Pratical Guide to Hybrid Cloud Computing,» [Online]. Available: https://www.omg.org/cloud/deliverables/CSCC-Practical-Guide-to-Hybrid-Cloud-Computing.pdf.
cxxxv Council Cloud Standards Customer, «Interoperability and Portability for Cloud Computing: A Guide,» 2017. [Online]. Available: https://www.omg.org/cloud/deliverables/CSCC-Interoperability-and-Portability-for-Cloud-Computing-A-Guide.pdf.
cxxxvi Cloud Standard Customer Council, «Cloud Customer Architecture for IoT,» 2016. [Online]. Available: https://www.omg.org/cloud/deliverables/CSCC-Cloud-Customer-Architecture-for-IoT.pdf.
[cxxxvii] BLOCKCHAIN TECHNOLOGY IN IDS Position Paper | Version 1.0 | March 2019
[cxxxviii] BLOCKCHAIN TECHNOLOGY IN IDS Position Paper | Version 1.0 | March 2019

[cxxxix] BLOCKCHAIN TECHNOLOGY IN IDS Position Paper | Version 1.0 | March 2019

cxl https://arxiv.org/abs/1902.04885 Federated Machine Learning: Concept and ApplicationsQiang Yang et al.

cxliThe algorithmic foundations of differential privacy. Cynthia Dwork. Microsoft Research, USA dwork@microsoft.com. Aaron Roth. University of Pennsylvania ...

cxlii https://en.wikipedia.org/wiki/Secure_multi-party_computation
[cxliii] https://security.googleblog.com/2019/06/helping-organizations-do-more-without-collecting-more-data.html
cxlivhttps://en.wikipedia.org/wiki/Data_preparation
[cxlv] Magic Quadrant for Data Integration Tools Published 1 August 2019 - ID G00369547