Orchestrating an interoperable sovereign federated Multi-vector Energy data space built on open standards and ready for GAia-X

D3.4 Data analytic services and requirements related to interoperability, security, privacy and data sovereignty

Document Identification	
Contractual delivery date:	31/10/2022
Actual delivery date:	31/10/2022
Responsible beneficiary:	UPC
Contributing beneficiaries:	ATOS IT, Tecnalia, ENGIE, EDP, EyPESA, Odit-e, RINA-C, REVOLT, PUPIN, MERCATOR, ASTEA, EW, ATOS SP, ELIA, IDSA, ICOM, AU, IMT, GIREVE, METEO, SENER ING
Dissemination level:	PU
Version:	1.0
Status:	Final

Keywords:

Data analytic services, digital twins, Data Space, architecture, energy

This document is issued within the frame and for the purpose of the OMEGA-X project. This project has received funding from the European Union's Horizon Europe Framework Programme under Grant Agreement No. 101069287. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the European Commission.

(*) Dissemination level: **(PU)** Public, fully open, e.g. web (Deliverables flagged as public will be automatically published in CORDIS project's page). **(SEN)** Sensitive, limited under the conditions of the Grant Agreement. **(Classified EU-R)** EU RESTRICTED under the Commission Decision No2015/444. **(Classified EU-C)** EU CONFIDENTIAL under the Commission Decision No2015/444. **(Classified EU-S)** EU SECRET under the Commission Decision No2015/444.

Document Information

Document Identification			
Related WP	WP3	**Related Deliverables(s):**	N/A
Document reference:	OMEGA-X_D3.4	**Total number of pages:**	176

List of Contributors	
Name	**Partner**
Marc Jené Vinuesa, Antonio Emmanuel Saldaña, Mónica Aragüés	UPC
Javier Gómez, Javier Valiño, Nataliia Diatlova	ATOS
Luis Pineda	IMT
Silvia Castellvi	IDSA
Maria Ines Marques, Tiago Coito	EDP
Tanguy Choné, Luc Richaud	Odit-e
Erik Maqueda, Mikel Fernandez	Tecnalia
Philippe Calvez, Lilia Bouchendouka	ENGIE
Isidoros Kokos	ICOM
Sara Frusone	RINA-C
Francesco Ferracuti	REVOLT
Ibon Salbidegoitia	METEO
Pablo González Perez, David Cuesta Ruiz, Segio Relloso Lassa	SENER ING

Document History			
0.1	30/07/2022	UPC, ATOS	ToC Proposal
0.2	15/09/2022	Tecnalia, SENER ING, UPC, METEO, EDF, ENGIE, REVOLT, Odit-e	Filling of Service Templates
0.3	20/09/2022	UPC	Introduction and Sections 3.1 and 3.2

Document History			
0.4	30/09/2022	EDP, Tecnalia, EDF, RINA-C (UC family leaders)	Added the introduction to UC families
0.6	07/10/2022	ATOS	Section 2 and 4.4
0.7	14/10/2022	IMT	Section 4.2
0.8	21/10/2022	IDSA	Section 4.1 and 4.3
0.9	21/10/2022	UPC, ATOS	Executive Summary and Conclusions
0.92	25/10/2022	UPC, ATOS	Apply corrections after review
0.95	27/10/2022	ATOS	Form Quality Review
1.0	31/10/2022	ATOS	Final version to be submitted

Quality Control		
Role	**Who (Partner short name)**	**Approval date**
Reviewer 2	Florian Mancel and Bruno Traverson - EDF	26/10/2022
Reviewer 1	Filipe Aranda de Sá and Pedro Pimenta - ISMAI	25/10/2022
Quality manager	María Guadalupe Rodríguez - ATOS	31/10/2022
Project Coordinator	Javier Valiño - ATOS	31/10/2022

Table of Contents

List of Tables

List of Figures

List of Acronyms

Abbreviation / acronym	Description
€	Currency sign used for the euro.
€/kWh	Euro per kilowatt-hour. Unit of energy price.
€/MWh	Euro per megawatt-hour. Unit of energy price.
3D-CAD	Three-Dimensional Computer Aided Design
A	Ampere. Unit of electric current.
ABAC	Attribute Based Access Control
AES	Advanced Encryption Standard
API	Application Programming Interface
ASP	Array State Parameter
BIM	Building Information Modelling
BIPV	Building Integrated Photovoltaics
BUC	Business Use Case
CAPEX	CAPital EXpenditure
CPO	Charging Point Operator
DER	Distributed Energy Resources
DHI	Diffuse Horizontal Irradiance
DHN	District Heating Networks
DHW	Domestic Hot Water
DID	Decentralized identifiers
DME	Detect Measurement Errors
DMP	Diode Model Parameters
DP	Data Provider
DRM	Digital Rights Management
DSBA	Data Spaces Business Alliance
DSBA-CTO	Data Spaces Business Alliance Chief Technology Officer
DSO	Distribution System Operator

Abbreviation / acronym	Description
DT	Digital Twin
EC	Energy Community
EMS	Energy Management System
EMSO	Electro Mobility Service Operator
EMSP	Electro Mobility Service Provider
ESCO	Energy service company
ETSI	European Telecommunications Standardization Institute
EV	Electric Vehicle
EVU	Electric Vehicle user
FBE	File-Based Encryption
FDD	Failure Detection and Diagnosis
FDE	Full-Disk encryption
FDI	False Data Injection
FF	Fill Factor
FSPs	Flexibility Service Providers
FTP	File Transfer Protocol
GCO	Granular Certificates of Origin
GDPR	General Data Protection Regulation
GHI	Global Horizontal Irradiance
GIS	Geographic Information System
HEMRM	Harmonised Electricity Market Role Model
HVAC	Heating, Ventilation, and Air Conditioning
IAM	Identity and Access Management
IDS	International Data Spaces
IDSA	International Data Spaces Association
IEC	International Electrotechnical Commission
IoT	Internet of Things
IQR	Interquartile Range

Abbreviation / acronym	Description
ISP	Interoperability service provider
KPI	Key Performance Indicator
kvar	Kilovolt-ampere reactive. Unit of reactive power.
kW	Kilowatt. Unit of power.
kWh	Kilowatt-hour. Unit of energy.
LCOE	Levelized Cost Of Energy
LEC	Local Energy Community
LV	Low voltage
M	M coordinate or measure.
m/s	Meter per second. Unit of speed.
m^3	Cubic meter. Unit of volume.
m^3/h	Cubic meter per second. Unit of volumetric flow rate.
MICE	Multiple Imputation by Chained Equations
min / '	Minute. Unit of time.
mm	Millimeters. Unit of precipitation.
MV	Medium voltage
MVF	Minimum Viable Framework
MW	Megawatt. Unit of power.
NA	Not Applicable
NIST	National Institute of Standards and Technology
NTL	Non-Technical Losses
nZEB	Nearly zero-emission building
º	Degree. Unit of angle.
O&M	Operation and Maintenance
ºC	Degree Celsius. Unit of temperature.
ODRL	Open Digital Rights Language
OEM	Original Equipment Manufacturer
OIDC4VP	OpenID Connect for Verifiable Presentations

Abbreviation / acronym	Description
OPEX	OPerational EXpenditure
p.u.	Per-unit system.
PAP	Policy Administration Point
PDP	Policy Decision Point
PEP	Policy Enforcement Point
PID	Potential Induced Degradation
PIP	Policy Information Point
PMP	Policy Management Point
PMU	Phasor Measurement Unit
PoA	Plane of Array
PR	Performance Ratio
PV	Solar Photovoltaic
RES	Renewable Energy Sources
SCADA	Supervisory Control And Data Acquisition
SCM	Secure Multiparty Computation
SD	Service Developer
SFTP	Secure File Transfer Protocol
SIOP	Self-Issued OpenID Provider
SM	Smart Meter
SSH	Secure SHell
STC	Standard Test Conditions
SUC	System Use Case
TMY	Typical Meteorological Year
TSO	Transmission System Operator
UCF	Use Case family
V	Volt. Unit of electric voltage.
W/m^2	Watt per square meter. Unit of irradiance.
WDN	Water Distribution Network

Abbreviation / acronym	Description
Wh	Watt-hour. Unit of energy.
Wh/m^2	Watt-hour per square meter. Unit of irradiation.
XBID	Cross-bid

Executive Summary

The OMEGA-X project aims at implementing a data space, including federated infrastructure, data marketplace and service marketplace, involving data sharing between different stakeholders and demonstrating its value for energy use cases. The project should also guarantee scalability and interoperability with other data space initiatives. The Energy Data Space will benefit from several European pilots grouped into four use case families (Renewables, Local Energy Communities, Electromobility and Flexibility).

This report captures the work done to define the different services to be developed within each family of use cases and the requirements related to the attributes of the Data Space, such as interoperability, security, privacy and data sovereignty.

A first approach to the architecture of the OMEGA-X platform is presented in this deliverable. State-of-the-art technologies regarding data spaces are listed, and a brief overview of federated infrastructures is included, detailing their main advantages and why they are becoming relevant in the last few years. Regarding data space technologies, the focus is on the two principal architecture proposals: GAIA-X and IDSA, studying and comparing both to better understand the benefits and drawbacks of each one. This analysis concludes with some preliminary components that might be part of the OMEGA-X platform to satisfy the architecture and service requirements.

Up to 33 services have been defined in the context of the project to benefit from the Energy Data Space, drawing on data from different pilots and other services. Table 1 indicates how many services have been defined within each use case family, and the report provides the first iteration of each of them.

Table 1. Data analytic services defined in each use case family.

Use case family	Number of services
Renewables	12
Local Energy Communities	10
Electromobility	2
Flexibility	9

Finally, the deliverable covers data sovereignty, security and privacy topics. A state-of-the-art of the existing data sovereignty solutions is presented, as well as a data security and privacy study of existing solutions for data encryption standards, secure transfer protocols or backup tools. These technologies are also linked with the current regulatory framework (GDPR). These analysis results are translated into the OMEGA-X approach for data sovereignty and privacy, which advanced the principal methodologies and guidelines to satisfy those requirements. Finally, some metrics are described to effectively quantify and measure the data quality and enable data certification mechanisms.

1 Introduction

1.1 Purpose of the document

This report provides an initial definition of the data analytic services developed within the OMEGA-X framework, benefitting from a European Energy Data Space. Also, this deliverable sets the grounds of the project architecture, focusing on essential requirements for Data Spaces such as interoperability, security, privacy and data sovereignty and governance. The impact of these requirements in the final architecture is also studied as well as some promising technologies that can satisfy the aforementioned requirements.

The report's primary audience is the partners involved in the four Use Case families of the project, from service developers to pilots. The document describes every service and presents the data needs of each of them, which should be matched by the pilots. Also, with a clear image of every service, Business Use Cases (BUCs) and System Use Cases (SUCs) will be outlined.

1.2 Relation to other project activities

This report is an outcome of tasks *T3.2 Definition of analytic services to be used as baseline for use case demonstration* and *T3.3 Analysis of requirements for interoperability, security, privacy and data sovereignty and certification*.

Services have been described within Use Case families and will feed *T3.1 Use case identification* to define SUCs and BUCs.

Requirements related to interoperability, security, privacy and data sovereignty that are being defined in this deliverable will provide the foundation for the activities of *T3.4 Full system architecture and building blocks*.

The services identified and defined in this deliverable will be developed within *WP5 – Data Space Marketplaces*, specifically by tasks *T5.1 Development of analytic services* and *T5.2 Digital Twins*.

Finally, this deliverable is also related to the *WP6 - Demonstration*. Services will be developed around Use Case families (UCF), which include selected pilots. The tasks elaborating on these Use Case families are *T6.2 Renewables Use case family*, *T6.3 Energy communities and sector integration Use case family*, *T6.4 Collaboration among Electromobility actors Use case family* and *T6.5 Flexibility Use case family*.

1.3 Structure of the document

The general structure of the deliverable is as follows:

Section 1 "Introduction" presents the purpose of the document, the relationship with other OMEGA-X tasks, the structure of the document and the glossary adopted in this document.

Section 2 "First approach to architecture" introduces federated infrastructures, focusing on Data Spaces, and provides a first analysis of the components, modules and methodology used to define the project architecture.

Section 3 "Data Analytic Services" gives a first characterisation of the different data analytic energy services included in the project, grouping them by use case family.

Section 4 "Data Sovereignty, Security and Privacy" analyses the state of the art of data sovereignty, security and privacy solutions and presents OMEGA-X approach to data sovereignty and privacy. It also introduces data quality and certification mechanisms that will be followed.

Section 5 "Conclusions" summarises the work presented and reports the conclusions.

Finally, **Section 6 "References"** lists the different sources used in the deliverable.

2 First approach to architecture

2.1 Introduction to federated infrastructures

The growing interest in data economy is something that goes beyond European boundaries. In order to be successful in such a competitive market Europe has to advance not only in the purely technical and technological aspects but also at the regulatory level. Several actions have already been taken by the European Commission in order to set the foundations for this new economy. In the General Data Protection Regulation (GDPR) [1] was approved by the European Parliament in 2016 and went into effect in May 2018. This regulation allows European citizens to control their personal data and establishes sanctions to the companies that operate in EU's internal market and not comply with such regulation.

More recently, the European Commission published the communication "A European strategy for data" [2] where they share their vision on the necessity of a European and cross-sector data space. This data space shall bring new opportunities and benefits not only for big companies, but also for SMEs and citizens, allowing all of them to have access to highly valuable data. As a result, new services shall take advantage of this data to build novel solutions that shall translate in a common benefit.

The ambition of having such an enormous data space faces several difficulties, from regulatory aspects to more technical ones. Focusing on the purely technical challenges, this section shall justify the rationality of using these federated infrastructures as the building blocks for the European data space.

Federation is understood as the ability to interconnect data or service platforms while still being able of maintaining their functionality when working standalone. Some examples of these federated services that are recently gaining relevance are Mastodon [3] (an open-source federated microblogging service, alternative to Twitter, which is already used by several EU institutions [4]), EU Video [5] (Official ActivityPub video platform of the EU institutions) and Matrix [6] (another open-source federated service, this time for instant messaging and voice communications). Although federation services are becoming more relevant, and they are considered as the foundation of the European data space, they are not new. Services that everybody use on a daily basis, such as email and SMS, have been the perfect example of federation, allowing their users not only to send emails to other users from the same federation (or mail/phone provider) but also to users from other federations.

To enable interoperability across different federations, those federations must establish standardised and compatible communication channels, information models and content catalogues at the very least. Keeping source code open and free, although it is not a compulsory requirement, facilitates the adoption, maintenance and extension of this type of solutions.

All things considered, federated infrastructures, despite not being a completely new solution, can leverage the huge advances in cloud and edge computing, the efforts in creating and standardising common information models (such as NGSI-LD [7]) and the current regulatory framework to implement the vision of creating a cross-sector data space at the European level.

2.2 Data Space infrastructures

As stated in the communication "A European strategy for data" [2] the European Commission acknowledges the efforts located during the last decade in order to promote the future data economy. Huge advances in digitalisation and data collecting, harmonising and sharing are recognised as well. However, it is also identified a lack of cross-sector data sharing. Although it is true that some synergies between close sectors have been already established, (an example of this could be projects relating energy and mobility sectors), there is not yet this common data marketplace that European Commission is ambitioning. Data spaces arise as a new paradigm that offers a sensible solution to face this problem.

Data spaces could be defined as federated and decentralised infrastructures that allow data sharing at the same time they enable and ensure data sovereignty, governance, privacy and security. The concept of federation was already introduced in the subsection 2.1. Nevertheless, to fully understand the flexibility and all the possibilities that data spaces offer, and why they are a completely new paradigm, it is important to see in more detail what this data sovereignty mean and how it could be achieved.

Data sovereignty and governance can be seen as the capability of data owners to retain the control over their data and, at the same time, establish a set of policies to openly share this data to whoever satisfy their requirements. These requirements or limitations might be in terms of price, permissions or any other. Even freemium schemes are feasible. A naive example would be a data owner that can offer their data for free under some limitations (in terms of data granularity or length) and unlock the whole dataset at a certain price.

2.2.1 Data Space technical requirements

From a technical point of view, there are some requirements and difficulties that data spaces must deal with before becoming an actual solution. This subsection aims at evaluating which of these requirements have been already satisfied and where main efforts should be allocated. In addition, following subsection 2.3 shall cover the main data spaces implementations that are currently under discussion, comparing them and analysing their components to understand their similarities and differences.

According to Design Principles for Data Spaces [8], there are three main technical challenges that data spaces must tackle: data interoperability, data sovereignty and trust and data value creation.

Data Interoperability: Covering data acquisition, harmonisation, exchange and traceability. As most of these requirements are the basis for any data service, there are already some well-known and mature solutions. Considering all these requirements, the most novel and challenging one might be traceability, although blockchain and distributed ledger technologies seem a really promising solution.

Data Sovereignty and Trust: Covering identity management, access and usage control and trusted exchange. Similarly to the previous requirements, there are also mature enough solutions to face these necessities. Some of them are Keycloak [9] as an open-source identity manager, or public key infrastructures for ensuring trusted data exchanges.

Data Value Creation: Covering metadata and discovery, accounting and publication and marketplace services. These requirements are more specific to data spaces, and consequently, technology readiness level of the existing solutions is lower and there is still room for researching and improving them.

Aside from these purely technical challenges, the European data space must be aligned with the regulatory framework, not only in terms of data (GDPR and other [10]), but also it must be aligned with European Green Deal [11] objectives, being carbon-neutral and energy-efficient. All these considerations have to be taken into account when selecting the technologies to build the data space.

2.3 First analysis of components and modules of the architecture

Aiming at maximizing the project's interoperability, the investigated architecture, heavily relies on the approaches taken by major EU initiative in the Data Spaces ecosystem - International Data Spaces Association (IDSA [12]) and Gaia-X [13]. That being said, it is important to offer an in-depth explanation for each approach.

2.3.1 IDSA

IDSA Reference Architecture Model [14] defines multiple different roles for the participants in the IDS to assume, as well as basic patterns of interaction, that may take place between for categories of roles.

- Category 1: Core Participant
- Category 2: Intermediary
- Category 3: Software/Service Provider
- Category 4: Governance Body

Interactions between said roles are briefly described on Figure 1.

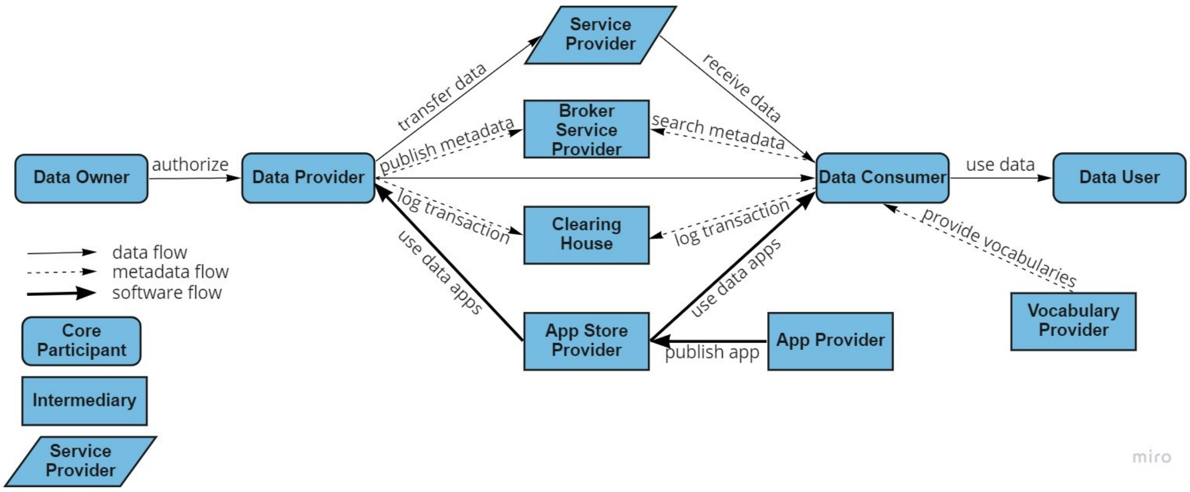

Figure 1. Roles and interactions in IDS

Category 1: Core Participant

Any data exchange in the IDS requires involvement of Core Participants. Roles reserved for this category, are Data Owner, Data Provider, Data Consumer, Data User, and App Provider

[14]. Any organization that owns, intends to provide and/or to consume data can assume the role of a Core Participant.

The **Data Owner** is a legal entity or a person, who is responsible for creating data and/or controlling it. A major concept of data ownership is having all technical means and responsibility to define:

- Usage Policies and Usage Contracts, and to provide data access; and
- the Payment Model (including a model covering the process of reusing of data by third parties).

The **Data Provider** is responsible for making the data available for an exchange between the Data Owner and the Data Consumer. In most cases, the Data Provider's role is identical to the Data Owner's. Aside from providing data, the Data Provider can use Data Apps to transform or improve data. Furthermore, the Data Provider can log the details of the transaction, after its completion (successful or unsuccessful), at the Clearing House.

The main task of the **Data Consumer** is to receive data from the Data Provider. The activities of the Data Consumer are somewhat mirroring the activities which Data Provider performs. There are two ways of establishing a connection between the Data Consumer and the Data Provider: a Data Consumer can either directly connect with a Data Provider and request the data; or make a request to a Broker Service Provider to receive metadata that is required to find and connect to a Data Provider. In a similar way to the Data Provider, the Data Consumer can log the data exchange result details at a Clearing House and use Data Apps when needed.

The **Data User** is a legal entity (person, group of people, institution, etc.), that has full legal rights to use the data which belongs to the Data Owner according to the usage policy. A Data User and a Data Consumer are identical in most cases.

The **App Provider** develops apps, which are intended for use in the International Data Spaces. Each app should be compliant with the architecture of IDS. It also can be certified by a Certification Body and must be published in the App Store Provider for Data Consumers and Data Providers to access it. It is also a standard practice to follow each app with metadata regarding its functionality.

Category 2: Intermediary

Intermediaries perform in IDS architecture as trusted entities. According to the resource [14], roles, that are assigned to this category are Broker Service Provider, Clearing House, Identity Provider, App Store Provider, and Vocabulary Provider.

The **Broker Service Provider** is defined as an intermediary, that both cumulates and controls information regarding data sources, that are available in the IDS. Multiple Broker Service Providers may coexist at the same time, since a Broker Service Provider plays a central, but not exclusive role in the architecture.

Providing and receiving metadata is the main activity of the Broker Service Provider. First, the Broker Service Provider has to offer an interface, so that Data Providers can publish their metadata. Then, said metadata has to be placed and kept in an internal repository, so that Data Consumers could request it when needed.

The **Clearing House** is defined as an intermediary, which job is to arrange clearing and settlement services for every data exchange transaction. It is possible for a Broker Service Provider to assume the role of a Clearing House, although in the International Data Spaces clearing activities are separated, as they are different from maintaining an internal repository with metadata.

All activities, that are performed at the time of data exchange, are logged in a Clearing House: once a data exchange is completed, the Data Provider and the Data Consumer each have to log the details of the transaction in order to confirm a data transfer. After that, the Clearing House can provide reports on the transactions that were logged for various purposes, such as billing or resolution of conflicts.

The main task of the **Identity Provider** is to offer a service for different operations with identity information of (and for) participants in the IDS, such as its creation, managing, maintenance, monitoring and validation. This service is crucial for secure operation of the IDS, and to protect the data from an unauthorized access.

The mission of the **App Store Provider** is to provide specific applications called Data Apps. Data Apps can be deployed inside of the core technical component required for every participant to join IDS – the Connector. The responsibility of an App Store Provider is to manage information about Data Apps that are offered by App Providers, along with putting up interfaces for publishing and retrieving these apps with their representing metadata.

The **Vocabulary Provider** manages and provides vocabularies (for example, reference data models, ontologies or metadata elements) for datasets description.

Category 3: Software/Service Provider

This category includes IT companies, that provide software and/or services to the participants in the IDS. Service Provider and Software Provider are roles comprised in this category.

If an IDS participant did not successfully deploy technical infrastructure (that is required for the participation in the IDS) there is an option to transfer the data to a Service Provider, which is responsible for hosting the correct infrastructure for some other organizations. Said data will eventually be available in the International Data Spaces.

The Service Providers also offer supplementary services to enhance the quality of data, that is exchanged inside IDS. Said services are used for data analysis, cleansing, integration or enrichment.

Providing software for IDS required functionality implementation is the core task of the **Software Provider**. Given software is not allocated in the App Store Provider. It is delivered over the chosen distribution channels of Software Providers. Rules of usage of the software are based on individual agreement between the Software Provider and an end user of this software – Data Provider, Broker Service Provider or Data Consumer.

Category 4: Governance Bodies

The Governance Bodies of the International Data Spaces, according to the resource [14], are the Certification Body, Evaluation Facilities, and the International Data Spaces Association.

The **Certification Body**, along with selected **Evaluation Facilities**, is responsible for the participants' certification and the core technical components in the IDS. These two Governance Bodies make sure, that the access to the business ecosystem is given only to the organizations that are truly compliant. The **International Data Spaces Association (IDSA)** is a non-profit organization, that supports Reference Architecture model's continuous development and certification process of the IDS participants.

2.3.2 GAIA-X

The GAIA-X ecosystem is formed by an Infrastructure Ecosystem and a Data Ecosystem, which are connected via Federation services. Detailed information on GAIA-X architecture is well explained in the report [15] and research [16].

General roles

In GAIA-X ecosystem participants (or entities) can assume the role of **Data Providers, Data Consumers** or both, if they simultaneously generate and consume data, respectively – the same or new one. A Data Provider can be seen as a supplier, a Data Consumer as a customer, and, similarly, the combined role of customer and supplier is both a Data Customer and a Data Provider. An asset may have a physical form, or it may be an immaterial service (such as software).

Main assets

The main assets of GAIA-X are Nodes, Services and Data Assets.

A **Node** is an element of data infrastructure, that meets all the requirements developed by the workstream, that deals with technical implementation of GAIA-X. Nodes can represent virtual machines, containers, data centres, etc.

In GAIA-X infrastructure, **Services** embody cloud offering, which implies all kinds of cloud services. A Service is called a Service Instance, if it is realized on a GAIA-X Node. Service Instances can run either on one, or on multiple nodes. Furthermore, it is possible to combine different Services with each other

A **Data Asset** is a dataset, provided to consumers by GAIA-X Services. Data Assets are available (or hosted) on GAIA-X Nodes. GAIA-X Participants (organizations, natural or legal persons participating in the ecosystem) can search for these Data Assets and consume them.

Federation services

Federation services interconnect the Infrastructure and Data Ecosystem, thus they are the core of GAIA-X. As it is stated in [15], [16], these services include Federated Catalogues, Identity and Trust Services, Compliance Services and Sovereign Data Exchange Services.

Federated Catalogues are considered as the main instrument for publishing and discovery of Self-Descriptions of both Data Assets and participants. GAIA-X Data Connectors are also a part of a Federated Catalogue.

The **Identity** & **Trust** category includes methods that ensure that the trust towards Participants' capabilities can be built and that their identities can be verified. Federated Identity model is the core method. Federated identity management connects several national or international identity providers, thus making identities able to operate in-between different domains.

In order to guarantee **Compliance**, legal relations have to be established between Service Provider and Service Consumer, along with rights and responsibilities for participants and procedures of onboarding and certification.

The **Sovereign Data Exchange** services are determined by GAIA-X in order to provide an ability for a natural or a legal person to exclusively determine specific rules for using their data. Usage Control is one of the key things that define data sovereignty. It can be explained as an enforcement of Usage Policies. Such Policies are meant to describe, under which terms and conditions customers should use data assets. Said Policies' enforcement allows to take control of the Provider's data.

Overall concurrence of IDSA and GAIA-X architectures was well analysed in research [16]. There are similarities between modules from both approaches, and all described services and functionalities are somewhat complementary or compatible.

In conclusion, it is vital to align OMEGA-X with already existing and well-established initiatives, such as IDSA and GAIA-X.

2.4 Architecture & service requirements for interoperability

As already seen, at the time of writing this document, there are two main architecture proposals for creating data spaces: IDSA and GAIA-X. Thus, the OMEGA-X data space must be aligned with them in order to be a valid and future-proof solution. Both GAIA-X and IDSA have some minimum requirements that must be satisfied to obtain their certification of a compatible solution. As a first preliminary analysis, this subsection aims at summarising those requirements referring to interoperability, bearing in mind that they are prone to be changed or updated in the future as Gaia-X and IDSA become more mature solutions. For the sake of flexibility and openness, if there is any requirement that is opposed between Gaia-X and IDSA, OMEGA-X data space shall take the least restrictive approach.

Despite their differences (studied in Section 2.3), Gaia-X and IDSA can be used together to build an even more complete data space solution. As studied in [16] Gaia-X can focus on the high-level services while the data related capabilities (data sovereignty, trust, etc.) relies on IDSA. In this sense IDSA connectors can be integrated inside a Gaia-X node. Similarly, Gaia-X Federated Catalogues can take advantage of IDSA Broker or other components.

Interoperability is commonly understood as the ability to ingest, collect, transform store or use data or services. On top of this general definition, when applying interoperability concept to data spaces, requirements related to trustworthiness, data governance, and sovereignty must be included. Consequently, the first requirement that OMEGA-X must satisfy is at architectural level.

OMEGA-X data space architecture must follow the microservice philosophy [17] rather than being a monolithic/centralised solution. Having an architecture based on loosely coupled and decentralised microservices, facilitates not only interoperability but also guarantees that the resulting platform is stable, scalable, and more secure.

Following the microservice philosophy also implies the necessity to define and standardise API communications between each component or microservice. In the same vein, some kind of message broker is probably needed, as well to enable asynchronous machine-to-machine communications. In the case of microservice, these API communications should be through RESTful API interfaces, while for message brokering there are some interesting options such as Apache Kafka, MQTT or AMQP.

Despite the many benefits that microservice architectures offer, they come with some difficulties mainly for shipping or deployment and orchestration. Although, it is also true that in the in the last years this labour has been enormously simplified by the arise of containerisation technologies, being Docker [18] the flagship, and orchestration technologies such as Kubernetes. The combination of these two technologies allows developers to seamlessly enhance the architecture by adding, updating or removing new services, in the form of containers, without impacting in the overall performance or user experience.

Regarding data, it must be meaningful both inside and outside the federation. Therefore, having a common information model that supports semantics, relationships, and linked data becomes extremely relevant. The NGSI-LD information model and API is an ETSI (European Telecommunications Standardization Institute) standard and might be used as reference as it natively supports all the requirements mentioned and is compatible with most novel FIWARE [19] components.

In addition to being meaningful, data or services must be accessible as well. Thus, OMEGA-X must offer some discovery and publish protocols. These discovery protocols shall take advantage of agreed and standard catalogues which must contain common description of the data or services published. This requirement easily aligns with Gaia-X Federated Catalogue component and is also related with IDSA broker service provider.

With relation to data governance and sovereignty, the principal requirement is to being able to uniquely and trustily identify the users or the actors in the data space. After identifying the actors OMEGA-X data space shall be able to control and limit their access to data and services based on some predetermined roles and policies. There are two main approaches to guarantee that this identification is trusted across federations: 1) on the one hand having a centralised and trusted entity in charge of verifying and certifying every actor, and 2) in the other hand, having a distributed self-sovereign identity approach and establish a web of trust that set a trustiness score to each participant according to the trustiness score of the peers that have acknowledge they. Although these two approaches might be opposed to each other, the ambition is to keep both coexisting in the ecosystem so actors might be able to identify themselves following the workflow they prefer. Some of the identity protocols that are proposed in [20] and shall be implemented in OMEGA-X platform are OAuth 2.0 [21], SAML [22] and OpenID Connect [23] (which actually is a framework built in top of OAuth 2.0 protocol), SIOPv2 [24] and OIDC4VP. A candidate technology to play the role of Identity Manager and, consequently, manage all these protocols is Keycloak which is an open-source identity and management solution.

3 Data Analytic Services

3.1 Energy services in the context of data spaces

Thermal energy, electrical energy, gas and even the water distribution grids, are part of the energy infrastructure system, which is understood as a multi-vector energy system. This system requires the participation of different stakeholders. For example, in the electricity market, entities such as Transmission System Operators (TSOs), Distribution System Operators (DSOs) or electric vehicle (EV) infrastructure owners require active data management to maintain system stability, improve grid operations and coordinate planning strategies among grid owners. Furthermore, the participation of end users with their flexible assets, such as batteries, electric cars, charging stations, solar panels, thermostats and private energy management systems, needs to be constantly coordinated with DSOs to avoid compromising the security of the energy system.

With an "all parties", multi-vector, energy data space, DSOs, TSOs, suppliers, producers and other involved parties from different countries are aimed to be connected. The benefits of this energy data space are that it ensures a common data platform among participants, makes large volumes of heterogeneous energy data discoverable and consumable, provides a distributed ecosystem environment and lowers barriers to data access.

In OMEGA-X, the analytic services aim to enable and facilitate interoperability of data within and across ecosystems to provide sovereignty and avoid vendor lock-ins. The services from different providers and countries will empower the European energy data space by leveraging data from different providers in multiple use case families.

3.2 Services characterisation

Task *T3.2 Definition of analytic services to be used as baseline for use case demonstration* aims to define the services to be used in the pilots within each Use Case Family. Thus, service developers and pilots must work together to define the services that will benefit from the energy data space.

Living documents have been created for each service, which is particularly important during the first months of the OMEGA-X project. These documents, known as *Service Templates*, aim to define the services, focusing on the objectives and data needed. Elaborated by the service developers, the service templates are used to facilitate and encourage cooperation with the pilots and Use Case Family leaders in defining the use cases. The document template, which has been conceived with the IEC 62559-1 [25] methodology in mind, includes the following sections to provide a clear description of the service:

- **Name**: Name of the service.
- **Use Case Family**: Use Case family to which the service belongs. This will define in which pilots it can be applied.
- **Pilots**: Pilots interested in applying for the service. At the end of month 5 (September 2022), some services are still undefined whether they will be applied in some pilots or not. If the service is going to be applied in more than one pilot, it is classified as interpilot, while if it is going to be applied within one pilot only, it is classified as intrapilot.

- **Category**: Within each Use Case family different categories are defined to group the services. These will also be used to define the BUCs and SUCs.
- **Description**: This section includes a brief introduction of the problem addressed by the service, a description of how it works and the main objectives. It also includes a paragraph arguing how the service will benefit from the energy data space.
- **Description of stakeholders involved**: Different stakeholders involved in the service are defined, such as the service user, the data provider and the algorithm developer.
- **Service to be offered in real-time/offline**: It defines whether the service is offered in real-time or offline and the execution frequency. If the service runs on demand, it will also be presented here.
- **Input data requirements**: It includes one or more tables with the necessary input data. The included columns are: data, granularity, description, units, volume, owner and optional/required. If the service needs training data, it is also presented here.
- **Expected outputs**: The output of the service is described. In many cases it is presented in a table format including the following columns: data, lowest level of granularity, description and units.
- **Applied techniques**: The techniques to be used by the service algorithm are identified.
- **Open-source / proprietary software**: The ownership of the software is presented.

The following sections (3.3 to 3.6) present the different services defined for OMEGA-X, divided by Use Case Family. Since 34 services have been identified, and to reduce the deliverable's size, only the Service Template's descriptive sections are included. The technical sections on data and algorithms (Input data requirements, Expected outputs, Applied techniques and ownership of the software) can be found in *Annex B: Service Templates*.

3.3 Renewables UCF Services

3.3.1 Introduction

This use case family focuses on renewable distributed energy resources (DERs) and in particular on solar photovoltaic (PV) energy.

During the recent decades the levelized cost of energy (LCOE) of PV energy has significantly decreased driven by a reduction of both capital expenditure (CAPEX) and operational expenditure (OPEX). In fact, thanks to new digital operational and maintenance services, the maintenance strategy for solar power plants has shifted from a preventive to a predictive maintenance strategy. However, these services are currently developed with limited data from a single company that usually covers specific operating conditions. As a consequence, the generalization capacity of these data analytics services is limited and usually does not behave well when the operating conditions change. Thus, this use case family aims to break the existing silos and leverage data sharing from multiple data providers, enabled by data space technologies, in order to develop more robust services that can improve the performance of existing services and further reduce the OPEX for PV plants.

In addition, due to the reduced LCOE, the penetration of PVs has significantly increased specially at the low voltage part of the grid. Hence, it is getting more difficult for DSOs to maintain the security and quality of electric power supply. To optimize the operation and maintenance of PV plants and to ensure power quality in distribution grids, this use case family aims to develop data analytic services that enable a seamless integration of distributed PV power plants into the smart grids by leveraging data space technologies, which allow to exchange data and services from different actors (prosumers, DSOs, aggregators, service providers, etc.).

In particular, Table 2 presents the services that will be developed in this use case family.

Table 2. List of Data Analytic Services - Renewable Use Case family.

Service	Category	Service Developers (SDs)	Data Providers (DPs)
Predictive Maintenance for large PV plants	Ren-PV O&M optimisation	*TECNALIA*	ENGIE and EDF
Benchmarking	Ren-PV O&M optimisation	*TECNALIA*	ENGIE and EDF
Compare actual production versus expected	Ren-PV O&M optimisation	*SENER ING*	ENGIE and EDF
PV Cleaning Advisor	Ren-PV O&M optimisation	*SENER ING*	ENGIE and EDF
Shading Analysis	Ren-PV O&M optimisation	*SENER ING*	ENGIE and EDF
Tracking algorithm check	Ren-PV O&M optimisation	*SENER ING*	ENGIE and EDF
Detect measurement errors	Ren-PV Smart Grid Integration	*UPC*	EyPESA
Detect non-technical losses	Ren-PV Smart Grid Integration	*UPC*	EyPESA
Congestion detection	Ren-PV Smart Grid Integration	*UPC*	EyPESA
Detection of the volatility of voltage in grids with high renewable penetration	Ren-PV Smart Grid Integration	*UPC*	EyPESA
Plan grid reinforcements for future renewable scenarios	Ren-PV Smart Grid Integration	UPC	EyPESA
Energy Generation Forecast	Ren-PV Smart Grid Integration	METEO	EyPESA

Service	Category	Service Developers (SDs)	Data Providers (DPs)
Digital Twin BIPV Self Consumption Systems in Buildings	Ren-PV Smart Grid Integration	TECNALIA	EyPESA

The following subsections explain in more detail each of the services including the input data requirements, expected output and the way they will leverage the data space that will be developed and implemented in OMEGA-X.

3.3.2 Predictive Maintenance for large PV plants

Pilots: ENGIE and EDF (interpilot).

Category: Renewable Energy Generation – Operation and Maintenance

Description: More efficient and low-cost operation and maintenance requires better characterization of PV system performance. With this objective, the Predictive Maintenance service has been developed applying data analysis techniques and Machine Learning methods to the data generated by the PV SCADA (Supervisory Control and Data Acquisition) system to obtain a set of different normality mechanistic models. This approach is based on knowledge of the physics of photovoltaic devices, with a second approach, based on the use of data analysis techniques, including machine learning and statistical models. This hybrid model is designed with the goal of taking advantage of both approaches. On the one hand, physical models make characterization and possible deviations without long dataset analysis, thanks to the use of PV plant design information (i.e., PV module characteristics provided by the manufacturer); on the other hand, statistical/machine learning models provide greater sensitivity, boosting the quantity and quality of information extracted from the PV installation. The proposed approach uses a disaggregated analysis of the dependence of maximum current power point and voltage on operating conditions.

The Predictive Maintenance service proposed in OMEGA-X is based on the use of a Failure Detection and Diagnosis (FDD) tool, which consists of advanced data analytics able to early detect some failure modes of PV arrays, modules and inverters, and determine their potential causes.

In addition, it is necessary to have an adequate characterization of the PV generator based on the use of representation models that are sufficiently reliable, precise and complete, capable of representing the behaviour of the system in ideal operating conditions, but also in degraded conditions. With this objective, it is necessary to resort to the generation of these array state parameters (ASPs) that combine the domain expert knowledge contained in the physical models with advanced data analytics techniques capable of extracting additional knowledge from the monitoring information of the real system.

The FDD comprises the following functional blocks:

1. Data preconditioning is always necessary: for formatting, reliability cross-check, selection, adjustment and filtering of input data, in order to minimize errors and reduce dispersion of normality models to be developed. For that, the FDD tool contains different algorithms to verify the quality of the data and clean or correct this type of deficiencies. In the first place, there is a general statistical conditioning related to the PV domain and then, an individual

conditioning to feed the different hybrid models proposed for the maintenance of photovoltaic systems.

2. Normality mechanistic models of PV module, PV array and inverter parameters (Voc, Vmpp, Impp, Rseries, Rshunt, degradation, soling, conversion efficiency, inverter internal temperature, etc.) from input variables as independent as possible to avoid interdependencies and error propagation. They are structured in 5 groups:
   - The first model estimates the open circuit voltage of the photovoltaic array in standard conditions. It determines those periods of time in which the photovoltaic inverter is not delivering energy, and the measured voltage corresponds to the open circuit voltage (Voc) of the photovoltaic field. Combined with the identification of the operating conditions (irradiation level and module temperature), a multiparametric model is adjusted that allows the evaluation of the open circuit voltage in any other operating condition. From this multiparameter model, a set of ASPs such as the open circuit voltage under standard test conditions (STC) or the temperature coefficient of Voc are extracted. These parameters combined with other ASPs are then used for the identification of possible failure modes, such as bypass diode short circuit or potential induced degradation (PID), among others.
   - A second model estimates the coefficient temperature of the maximum voltage at the maximum power point. It analyses the dependence of the voltage of the maximum power point (Vmpp) with the thermal conditions of the photovoltaic field, which are defined by the temperature of the module and the wind speed. The temperature coefficient of Vmpp is evaluated under different thermal conditions to detect possible hidden overheating of the PV array or a discrepancy between the measured and real PV array temperatures. The overheating of the photovoltaic array leads to a decrease in performance and causes the appearance of different failure modes. The power losses related to hidden overheating of the PV array can be evaluated using this procedure.
   - A third model estimates the behaviour of the series resistance and the "Voltage Fill Factor", in charge of evaluating the relationship between the maximum power point voltage and the open circuit voltage, the so-called "Voltage Fill Factor" (FFv). The dependence of FFv on the current maximum power point (Impp) of the PV array and the module temperature is analysed and described using a multiparameter model. From these multiparametric models, a set of array parameters such as the FFv under STC conditions, or the rate of change of FFv with Impp are obtained. Based on the rate of change FFv *vs.* Impp, the series resistance of the photovoltaic array (Rseries) is estimated for any operating condition. The evolution of the Rseries is used for the identification and analysis of different failure modes such as corrosion in the internal electrical circuit of the photovoltaic modules, interconnection problems between photovoltaic modules, etc.
   - A fourth model evaluates the shunt resistance of the photovoltaic array, providing information on failure modes, such as PID. This algorithm is dedicated to the analysis of the FFv of the photovoltaic array under low irradiation operating conditions. The FFv is evaluated for any operating condition from SCADA data and previous ASP and multi-parameter models. The dependence of FFv on the irradiance level fits a physical model that describes its behaviour as a function of the shunt resistance of the photovoltaic field (Rshunt). The evolution of the Rshunt function is used for the identification of possible failure modes, such as potential induced degradation or short circuit of PV modules.

  - A fifth model analyses the relationship between the current of the maximum power point and the irradiance of plane of the array, providing information on failure modes, such as dirt or yellowing of the encapsulant.
3. Early failure detection and diagnosis detecting potential instantaneous and long-term deviations in mechanistic models along the time with different sliding time windows and initially in comparison to design specifications and neighbours and correlating these with specific failure modes and their potential causes. In turn, these deviations are correlated to specific failure modes and their possible causes. The tool provides a list of triggered alarms in case these PV ASPs show a significant deviation from normal operation. The alarms include information about the affected components in the photovoltaic plant and the possible failure causes. This output must be complemented with the information collected in the maintenance logbook to interpret it, in case of deviations derived from maintenance activities.

In OMEGA-X, the existing tool will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In particular, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common Interoperability framework.
- Validate and improve the robustness of the tool by leveraging data from different data providers (ENGIE, EDF) at different operating conditions for different PV plants.
- Improve the overall output by combining the outputs of the services provided by the tool with the output from other service providers (SENER ING).

Description of stakeholders involved:

- Service user: PV plant owners/Original Equipment Manufacturers (OEMs)
- Data provider: PV plant owners/OEMs.
- Algorithm developer: TECNALIA.

3.3.3 Benchmarking Service

Pilots: ENGIE and EDF (interpilot).

Category: Renewable Energy Generation – Operation and Maintenance.

Description: Business intelligence service on operating magnitudes of PV installations, applying machine learning techniques (advanced time series analysis). This machine learning techniques will allow to process the real operation data with hybrid models, generating the KPIs [Voc, Vmpp, Impp, Rseries, Rshunt, degradation, soling, conversion efficiency, inverter internal temperature, etc. that indicate the state of health of the photovoltaic arrays / modules and inverters of the different technologies (bifacial, monofacial, BIPV), type of tracker or different applications (floating, etc.)] to allow the operation efficiency comparison between different PV plants from different owners. For that, the Failure Detection and Diagnosis (FDD) tool will be used in the different PV plants to compare different indicators related to normality mechanistic models (Voc, Vmpp, Impp, Rseries, Rshunt, degradation, soling, conversion efficiency, inverter internal temperature, etc.) between different PV plants using different technologies and applications in similar conditions to select the best investment option. The description of the FDD tool is in the description of predictive Maintenance Services.

Based on the monitoring information, performance models of the installation are usually used, such as the performance factor or Performance Ratio (PR), to try to assess whether if a PV plant is operating as expected. The PR is a very simplified model that relates the output power of the installation with that which would be produced by an ideal PV generator without losses for given irradiance conditions. If the model also considers how the existing operating temperature conditions impact performance, it is referred to as PR corrected by temperature. This is the main key performance indicator (KPI) in large installations and is usually used as the reference to ensure the quality of the installation in the contractual conditions established between owners and maintainers. Through the normality mechanistic previous indicators, it will be possible to calculate the PR for each plant and determine the performance of each of them, favouring decision-making to select the best investment option.

In OMEGA-X, the existing tool will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data from different providers. In particular, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common interoperability framework.
- Validate and improve the robustness of the tool by leveraging data from different data providers (ENGIE, EDF) at different operating conditions for different PV plants.
- Allow PV plant owners (ENGIE, EDF) to share some aggregated data (KPIs) that can be used to assess the real performance of their plant compared to a similar plant as an additional source to the expected values provided by the manufacturer. This will allow a win-win situation for plant owners, without sharing business critical information.

Description of stakeholders involved:

- Service user: PV plant owners/OEMs
- Data provider: PV plant owners/OEMs.
- Algorithm developer: TECNALIA.

3.3.4 Compare actual production versus expected

Pilots: ENGIE and EDF (interpilot).

Category: Renewable Energy Generation – O&M.

Description: Solar power plant financial models rely on accurate simulations of the energy yield based on typical meteorological year (TMY) and several simplifications and assumptions. But during operation, the behaviour of the power plants can differ from the foreseen models.

This service compares real production of PV installations versus the simulated production of the facility under the expected specification, providing notification in case the deviation exceeds stablished values.

Simulation models replicate the expected operation condition of the facility and compare the results both in daily performance and in sub-hourly operation.

In OMEGA-X, the existing tool will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In particular, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common interoperability framework.
- Validate and improve the robustness of the tool by leveraging data from different data providers (ENGIE, EDF) at different operating conditions for different PV plants.
- Improve the overall output by combining the outputs of the services provided by the tool with the output from other service providers (TECNALIA).

Description of stakeholders involved:

- Service user: Utility scale solar power plants.
- Data provider: Utility scale solar power plants.
- Algorithm developer: SENER ING.

3.3.5 PV Cleaning Advisor

Pilots: ENGIE and EDF (interpilot).

Category: Renewable Energy Generation – O&M.

Description: Dirtiness accumulation on the photovoltaic modules surface decreases the solar radiation penetration to the PV cells and, eventually, the power production from the PV system. To prevent this kind of power losses, PV systems require maintenance cleaning tasks, the frequency of which depends on the geographical location, site conditions and module surface treatments.

This service compares production of each string of the plant versus the production of the clean string reference, allowing the optimization of the cleaning tasks.

In OMEGA-X, the existing tool will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In particular, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common interoperability framework.
- Validate and improve the robustness of the tool by leveraging data from different data providers (ENGIE, EDF) at different operating conditions for different PV plants.
- Improve the overall output by combining the outputs of the services provided by the tool with the output from other service providers (TECNALIA).

Description of stakeholders involved:

- Service user: Utility scale solar power plants.
- Data provider: Utility scale solar power plants.
- Algorithm developer: SENER ING.

3.3.6 Shading Analysis

Pilots: ENGIE and EDF (interpilot).

Category: Renewable Energy Generation – O&M.

Description: The performance of a photovoltaic plant is affected by temperature, solar irradiation, shading and electrical configuration. Often, PV arrays get shadowed (completely or partially), by the passing clouds, near obstacles or other PV modules. The situation is of particular interest in case of large PV installations such as those used in distributed power generation schemes. Under partially shaded conditions, the PV characteristics get more complex with multiple peaks. Yet, it is very important to understand and predict them, in order to extract the maximum possible power. This service assesses the performance loss of each string due to shading between rows.

In OMEGA-X, the existing tool will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In particular, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common interoperability framework.
- Validate and improve the robustness of the tool by leveraging data from different data providers (ENGIE, EDF) at different operating conditions for different PV plants.
- Improve the overall output by combining the outputs of the services provided by the tool with the output from other service providers (TECNALIA).

Description of stakeholders involved:

- Service user: Utility scale solar power plants.
- Data provider: Utility scale solar power plants.
- Algorithm developer: SENER ING.

3.3.7 Tracking algorithm check

Pilots: ENGIE and EDF (interpilot).

Category: Renewable Energy Generation – O&M.

Description: This service will evaluate the tracking angle with respect to the standard flat terrain tracking algorithm both with and without back-tracking. Based on the theorical model of the trackers, the annual production loss will be quantified.

In OMEGA-X, the existing tool will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In particular, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common interoperability framework.
- Validate and improve the robustness of the tool by leveraging data from different data providers (ENGIE, EDF) at different operating conditions for different PV plants.
- Improve the overall output by combining the outputs of the services provided by the tool with the output from other service providers (TECNALIA).

Description of stakeholders involved:

- Service user: Utility scale solar power plants.
- Data provider: Utility scale solar power plants.
- Algorithm developer: SENER ING.

3.3.8 Detect measurement errors

Pilots: EyPESA, ENGIE and EDF (interpilot).

Category: Smart grid data-driven services.

Description: This service seeks to identify, detect and eventually solve anomalies, such as errors or missing values coming from data sources, in order to ensure the optimum data quality and usability for the services. Depending on the type of anomaly detected, a correction will be automatically performed, using tools ranging from statistics to artificial intelligence.

The Detect Measurement Errors (DME) service follows the workflow presented in Figure 2: the first diagram (top) shows the available options for non-big data files cleaning process. The algorithm makes sure that no data is missing in the case of time series. However, if the dataset needed is used for training tasks (data for training models' option), the missing data rows are eliminated instead of imputing a non-true value since the missing data records account for a small amount of the total. In data for operation, the next step notifies the service user, the row/columns with missing data, if requested in the DME setup.

The following step is data analysis to detect outliers which do not correspond with the rest of the data. Depending on the service requesting the DME, the anomaly is eliminated and/or reported. For example, the fraud detection or predictive maintenance service will choose to be notified of anomalous data, which may imply that a fraud is being committed in the network or the malfunctioning of a piece of equipment.

Then, the DME performs the task of imputing or deleting rows with missing values, depending on whether the dataset is used for training tasks. By imputing values, the dataset size remains the same. Only the empty records which were either missing or deleted due to anomalies are now filled with predicted values.

The second diagram (bottom) describes the DME process for big data files. Spark [26] is an open-source parallel processing framework for running large-scale data analytics applications across clustered computers. It can handle both batch and real-time analytics and data processing workloads. In this context, DME uses Spark to process and clean the data, returning clean datasets with the required quality.

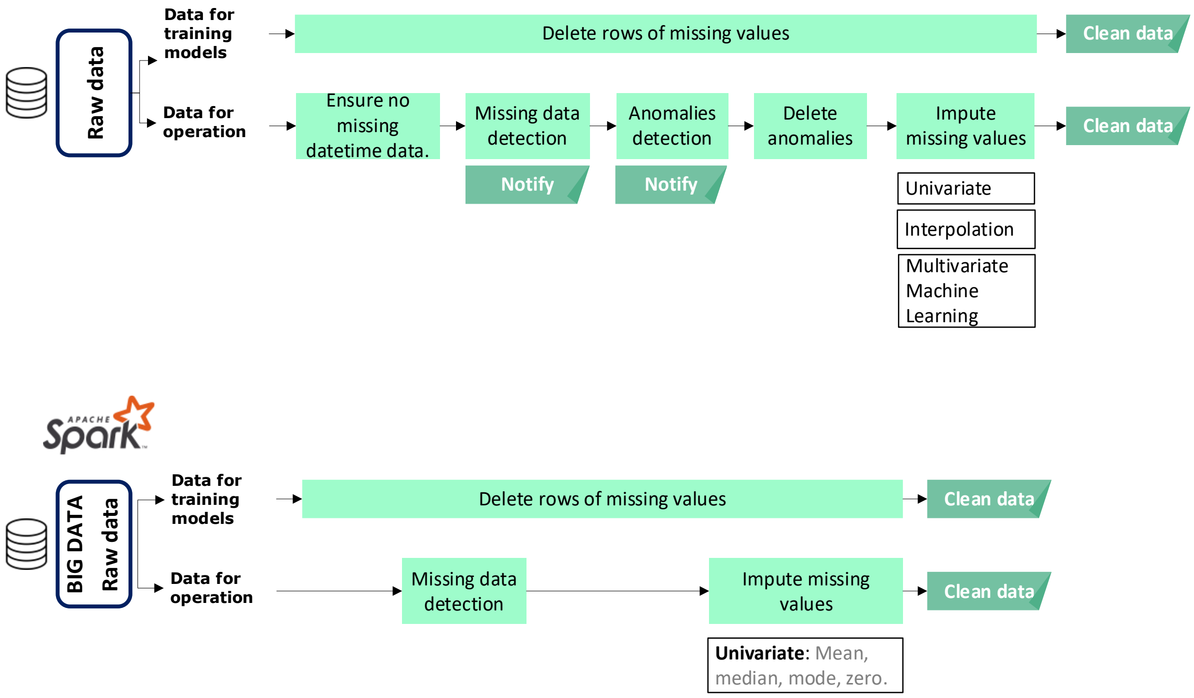

Figure 2. Detect Measurement Errors flowchart. Source: UPC and BD4OPEM

Some functionalities have been eliminated when dealing with big data files, since not all the libraries used for smaller files (first diagram) allow to work computationally in a distributed manner, as Spark requires.

In OMEGA-X, the existing tool will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In particular, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common interoperability framework.
- Validate and improve the robustness of the tool by leveraging different data from different providers (ENGIE, EDF, EyPESA).
- Improve the overall output by combining the outputs of the services provided by the tool with the output from other service providers (e.g. MFE, TECNALIA).

Description of stakeholders involved:

- Service user: It can be used by services and pilots (TBD).
- Data provider: Depends on the service that requests DME; it could be any data sources available (smart meters, sub-metering, SCADA, phasor measurement unit (PMUs), etc.).
- Algorithm developer: UPC.

3.3.9 Detect non-technical losses

Pilots: For the moment it will be applied only in EyPESA, although this service will also be applied in other pilots of the LECs UC.

Category: PV smart grid integration.

Description: Non-Technical Losses (NTL) are a direct loss of revenue for DSOs and usually are due to fraud. Analysing the power that goes through the LV side of the transformers and the data obtained from the smart meters allows calculating the total losses profile and creating consumption and fraud patterns.

Knowing the grid topology (or the percentage of losses due to Joule's effect) helps estimate the technical losses and thus facilitates the obtention of the NTL curve. The literature classifies these fraudulent losses into two categories as to their origin: the most typical is direct connection, which occurs in front of the meter; the other category is meter tampering, increasing importance due to cyber-attacks, such as false data injection (FDI).

With the help of classification algorithms, this service can let the grid operators know which type of fraud (marihuana plantation, crypto mining, squatting or unclear) is predominant in each case by analysing the patterns of the NTL curve.

Subsequently, the algorithm will review the consumption of the different SMs, and by using the short-pattern clustering technique, it will flag suspicious customers in case of meter tampering fraud.

Thus, in summary, the objectives of this service are:

- Detect if there is fraud (NTL > 5%) in a low voltage (LV) or medium voltage (MV) distribution grid.
- Classify the fraud into marijuana plantation, cryptocurrency mining, squatting or other.
- Provide information of the fraud.
- In case of possible losses due to meter tampering, flag the possible fraudsters.

In OMEGA-X, the existing service will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In this aspect, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common interoperability framework.
- Incorporate a wider range of frauds by applying the service in different pilots, since the type of fraud depends on the social and geographical reality of each country. For the moment, due to the type of pilot, it will be applied only to EyPESA, although the application in some other pilot of the LECs UC will be considered. Also, the algorithm will be adapted to be able to nurture pilots from all over Europe in the future.

Description of stakeholders involved:

- Service user: DSOs.
- Data provider: DSOs.
- Algorithm developer: UPC.

3.3.10 Congestion detection

Pilots: EyPESA from the Renewable UC and ASTEA from the LEC UC (interpilot).

Category: PV Smart Grid Integration.

Description: At MV level, distribution networks can be subjected to line or transformer congestions due to load variations and the presence of intermittent power generation at different locations. This service aims to identify the possible congestion scenarios in a MV distribution grid and LV grid to be used by DSOs as a tool for congestions forecast based on day-ahead demand prediction, possible congestion scenarios produced by fluctuating power demand, and new distributed generation facilities, etc. The objective is to identify when power lines and transformers in the distribution grid are prone to overload given the forecasted demand. This will allow the DSOs to predict when and where to perform activation of flexibility strategies to distress the grid.

The service will identify the possible congestions scenarios and grid disturbances related to voltage quality, allowing the DSO to monitor and forecast these scenarios, and provide a quick response when needed to comply with secure service.

The service comprises two algorithms, which are used in two phases: training and operation:

- **Training algorithm:** The grid is analysed and studied with the probabilistic power flows, evaluating the congestions condition in such grid. Further, machine learning models are trained to develop a congestion model able to detect congestion given a demand forecast.
- **Operation algorithm:** This section is where the forecast of demand interacts with the congestion model, giving the output of the probability of congestions to the service user in a daily-ahead schedule.

In OMEGA-X, the existing service will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In this aspect, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common interoperability framework.
- Validate and improve the robustness of the tool by leveraging data from different data providers LV grids belonging in different Use Case Families (EyPESA from the Renewable UC and ASTEA from the LEC UC).
- Improve the overall output by combining the outputs of the services provided by the tool with the output from other service providers (ODIT-E in the Flexibility UC).

Description of stakeholders involved:

- Service user: DSOs.
- Data provider: DSOs.
- Algorithm developer: UPC.

3.3.11 Detection of the volatility of voltage in grids with high renewable penetration

Pilots: EyPESA (intrapilot).

Category: Smart grid data-driven services.

Description: DSOs oversee the operation and maintenance of power grids, but most of all, they are supervising a good quality of service. The quality of service embraces the continuity of supply and a good wave quality of the voltage. If we look closer to this second feature, the voltage can be altered easily with a fluctuation of the consumption and/or the generation. This situation becomes worse with an increasing penetration of intermittent RES, such as solar or wind. Thus, the problem of voltage volatility becomes something that DSOs have to oversee for the proper operation of the grid.

The main objective of this service is to visualize voltage fluctuations in real time based on inputs such as consumption data at different points of the grid (active/reactive powers), electrical quantities measured at the different transformer substations, and generation data from prosumers/PV plants; and, only as the last possible resource, the topology of the grid. Together with the DSO, a range of values will be defined to indicate when there is volatility and of what magnitude it is. Two different approaches are considered at this stage of development:

- Regression algorithm: the output of the algorithm will be a learning-based approximation of the power flow calculation, to determine the overvoltage/undervoltage that might happen in a specific area.
- Classification algorithm: a traffic light approach might indicate the probability of formation of voltage violations or the approximated voltage quality state of the zonal power system. It could be used to give a practical understanding to the operator of the current situation.

Both approaches allow to localize and quantify approximately the amount of voltage variability.

The prediction of the voltage volatility can be done in different time windows (15 minutes, 1 hour, 1 day), feeding the algorithm with renewable generation and demand forecasts. As with any prediction, the closer the time window, the higher the accuracy.

Finally, strategies for the management of this voltage volatility in the short and long term will be proposed, through the interaction with other services such as "Estimate the probability of congestions" or "Plan Grid Reinforcements for Future Renewable Scenarios".

In OMEGA-X, the service will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In this aspect, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common interoperability framework.
- Improve the overall output by combining the outputs of the services provided by the tool with the output from other service providers such as generation forecasting (METEO and SENER ING) and demand (ODIT-E and UPC).

Description of stakeholders involved:

- Service user: DSOs.
- Data provider: DSOs.
- Algorithm developer: UPC.

3.3.12 Plan grid reinforcements for future renewable scenarios

Pilots: EyPESA from the Renewables UC and ASTEA from the LEC UC (interpilot).

Category: PV smart grid integration.

Description: In order to ensure that future grid expansions will be able to host new renewable generation, utilities must apply appropriate investment strategies to plan grid reinforcements. Making the right investment decisions is a challenge for DSOs, that need to consider technical and economic criteria.

The objective of this service is to reduce reinforcements costs of DSOs with an automated planning tool that minimizes costs of capacity upgrades and operating costs of feeder lines, transformers and/or Li-ion batteries, considering the expected renewable generation in a long-term horizon.

Two optimal planning approaches will be explored. Firstly, a traditional approach will be used, which consists of replacing or reinforcing the network considering traditional assets, including feeder lines and transformers as decision variables. Secondly, a combined approach will be used, considering traditional assets and flexible assets in active distribution networks, such as electric vehicles, energy storage systems and/or energy demand response, which will be included as decision variables to provide a solution for energy congestions or over/under voltage problems. In both cases, the objective function will be to minimize the costs of investment and operation.

The outputs of the service will be:

- Lower Capital Expenditure Costs (CAPEX).
- Lower Operation Expenditure Costs (OPEX).
- List of feeder lines to be reinforced, including geographic information system (GIS) location and capacity upgrades.
- List of transformers to be reinforced or replaced, including GIS location and capacity upgrade.
- List of new batteries installation, including GIS location and capacity.

In OMEGA-X, the existing service will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In this aspect, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common interoperability framework.
- Validate and improve the robustness of the tool by leveraging data from different data providers LV grids belonging in different Use Case Families (EyPESA from the Renewables UC and ASTEA from the LEC UC).

Description of stakeholders involved:

- Service user: DSOs.
- Data provider: DSOs.
- Algorithm developer: UPC.

3.3.13 Energy Generation Forecast

Pilots: EyPESA from the Renewables UCF and EDP from the LEC UCF (intrapilot).

Category: PV Smart Grid Integration.

Description: Energy generation forecasts for the renewable energy power plants are an important aspect to provide grid stability and do not require ancillary services to cover their uncertainty due to the meteorological uncertainty.

With the help of predictive algorithms adjusted for each specific site, the utility scale power plants can automate their energy output declaration to the system operator of the grid as the uncertainty increases with the time horizon.

So, it is mandatory for the energy transition to the renewable energy power plants to have accurate energy generation forecasts and improve them with updates on real-time markets such as the European Cross-bid (XBID), introduced to boost the digitalization and integration of renewable energies to the grid, and not only provide a day-ahead daily declaration.

Thus, in summary, the objectives of this service are:

- Accurate energy generation forecasts for solar power plants to reduce deviation costs.
- Day-ahead and intraday forecast updates.
- XBID hourly update for grid stability.

In OMEGA-X, the service will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In particular, the objectives are the following:

- Integrate the service into the Energy Data Space according to the defined common interoperability framework.
- Integrate the service for the different data providers from both Renewables use case family (EyPESA) and Local Energy Communities use case family (EDP).
- Deploy the data space connector on top of the service to allow other data or service providers take advantage of the energy generation forecast service.

Description of stakeholders involved:

- Service user: Utility scale solar power plants.
- Data provider: Utility scale solar power plants (total inverter energy or total meter energy).
- Algorithm developer: METEO.

3.3.14 Digital Twin BIPV Self Consumption Systems in Buildings

Pilots: For the moment, it will be implemented in EyPESA pilot but it could be considered to be implemented in other pilots (e.g. from LEC use case family).

Category: PV smart grid integration.

Description: Currently, most of BIPV systems are not supervised or their supervision is limited to the remote monitoring provided by PV inverters. This means that most of failure modes go unnoticed until they severely impact on energy yield and, even then, identifying the causes and the corresponding maintenance activity to solve it may not be a straightforward question. Thus, BIPV systems normally show lower performance ratios in comparison to large PV plants, where energy yield is maximized, not only at design stage, but also during their whole lifetime with a well-defined O&M plan, including supervision tasks based on advanced remote monitoring and on-field inspection.

The hybrid Digital Twin BIPV Self Consumption Systems in Buildings comprises advanced monitoring, modelling and data analytics to optimise the energy yield production with two main functions:

1. Energy generation forecast,
2. Predictive maintenance.

On the one hand, the current service allows to predict the energy generation in self-consumption systems based on weather conditions and system health status. The hybrid digital twin can also be used for generating synthetic data for simulating different scenarios for similar building favouring decision-making to select the best investment option.

On the other hand, it enables automatic supervision of the BIPV system, triggering an alarm in case a significant deviation occurs, identifying their possible causes and allowing the user to carry out the suitable corrective action to enhance the energy yield whenever this is considered feasible and worthwhile.

The hybrid digital twin combines a digital twin of the photovoltaic asset (BIPV DT) and numerical modules based on machine learning techniques, including optimization techniques.

The core of the tool comprises the following functional blocks:

- Data pre-conditioning: this component is aimed to filter and transform weather and SCADA data as required by the rest of tool components.
- Initialization of digital twin (DT): this component, can simulate the BIPV system production, and it also calculates a set of parameters and the irradiance reached to each PV module. It needs as input the Building Information Modelling (BIM) project. This project contains a simplified version of the three-dimensional Computer Aided Design (3D-CAD) file with data of the location and orientation of each BIPV module, as well as of other physical objects of the environment that could influence the PV production (for example through shadowing). Additionally, the optoelectronic properties of the BIPV modules, electrical characteristics of the PV inverter, or electrical installation design (series/parallel connection, wiring length, etc.) are included to later asses the energy production. Precisely, the energy production is modulated at a given time instant (t) by the operation conditions of the BIPV installation, and hence, by the weather conditions. If the production estimated by the DT matches with the one measured by SCADA, no failures are searched. Otherwise, the other three components of the tool must be executed to find a diagnosis.
- DMPs estimation: a hybrid model (data-driven and model-based) will calculate a set of the PV array performance indicators, called ASPs, which are also calculated at module level, Diode Model Parameters (DMPs). DMPs are used as input to the optimisation loop.
- DMPs tuning: this component is aimed to find the most likely DMP that achieve a simulated production closely as possible to real production.
- Fault detection and diagnosis: this component estimates the most likely failures that have caused the divergences between the digital twin and the SCADA data.

This service has been previously developed for the analysis of the state of health of PV arrays integrated in Buildings, in which TECNALIA has been working within the framework of the Work Package 7 of the European founded project H2020- BIPVBOOST [27], bringing down costs of BIPV multifunctional solutions and processes along the value chain, enabling widespread nearly zero-emission building (nZEB) implementation.

In OMEGA-X, the existing service will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In particular, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common interoperability framework.

- Validate and improve the robustness of the tool by leveraging data from different data providers from both Renewables UCF (EyPESA) and LECs UCF (EDP).
- Improve the overall output by combining the outputs of the services provided by the tool with the output from other service providers (METEO and UPC).

Description of stakeholders involved:

- Service user: Prosumers.
- Data provider: Prosumers.
- Algorithm developer: Tecnalia.

3.4 Local Energy Communities UCF Services

3.4.1 Introduction

This use case family is led by RINA-C and will be made up of four different pilot sites, located in Spain, France, Italy and Serbia. It involves four LEC energy managers (IMPULSA - EyPESA, RINA-C, ASTEA and PUPIN, as LEC operators and data providers), one SME (REVOLT as data user and service provider), one energy consultancy (RINA-C as data user and service provider), two Universities/RTOs (Tecnalia and UPC as data user and service provider), and a cross-demo actor (EDF) in charge of developing a methodology for analysing compliance of proposed services with regulatory frameworks existing and envisaged in various countries in Europe. Energy Communities (ECs) arise from the increasingly concrete need to manage their own energy supply chain. In particular, this kind of local approach to a clean energy society helps the engagement of local citizens to become active contributors and to influence other consumers to adopt similar behaviour. On that purpose, Local energy Communities (LECs) have a primary objective of providing environmental, economic, and social community benefits. Community energy can foster citizens' participation and control over decision-making in renewable energy and support the energy transition, but also reinforce positive social norms. Community energy initiatives are offering new opportunities for citizens to increase the share of renewables in local areas with limited impact on the public grid and an enhancement of energy efficiency as bottom-up actions. It is fundamental to become more proactive in the energy production–consumption process and in the achievement of sustainability targets.

An example of this involvement is the participation of citizens and communities as partners in energy projects. Its social innovation potential also resides in the ability to integrate consumers, regardless of their income and access to capital, ensuring that the benefits of decentralization are also shared with those that cannot participate.

Interoperability among data driven platforms is fundamental to improve coordination between these actors. In addition to achieve the ambitious objectives of carbon neutrality and collective self-consumption, it is crucial that data and associated representations are shared/used in a secure way.

Energy communities can bring a lot of benefits to the energy systems. For example, they can support system operations by providing services locally and alleviating the need for traditional network upgrades. On that purpose, different services will be developed, implemented and validated, leveraging data from different community actors. Furthermore, some of the services will also be shared amongst communities and even with other actors such as RES and flexibility families

Table 3 shows the services that will be developed within this UCF benefitting from the OMEGA-X energy data space:

Table 3. List of Data Analytic Services – Local Energy Communities UCF.

Service	Category	Service Developers (SDs)	Data Providers (DPs)
Local Energy Communities Designer	Planning	*REVOLT*	ASTEA and other to be confirmed
Gamification for electrical energy savings	Gamification	*REVOLT*	ASTEA and other to be confirmed
Thermal Losses Detection and Benchmarking at LEC level	LEC optimization	*REVOLT*	ASTEA
Water Losses Detection and Benchmarking at LEC level	LEC optimization	*REVOLT*	ASTEA
Estimate the probability of congestions	LEC optimization	*UPC*	EyPESA and other to be confirmed
Electrical losses detection and benchmarking at LEC level	LEC optimization	*UPC*	EyPESA and other to be confirmed
Reinforcement Plan of LEC for future renewable scenarios	Planning	*UPC*	EyPESA and other to be confirmed
Optimizing self-consumption of renewable energy at LEC level	LEC optimization	*Tecnalia*	To be confirmed
Optimizing sharing coefficients in collective self-consumption	LEC optimization	*Tecnalia*	To be confirmed
Planning services	Planning	*Tecnalia*	To be confirmed

The following subsections explain in more detail each of the services implemented in OMEGA-X within the LEC UCF.

3.4.2 Local Energy Communities Designer

Pilots: Pilots where the energy consumption and production will be provided partially/totally by local energy community. ASTEA, EDP and other pilots to be confirmed (intrapilot and interpilot).

Category: Planning services.

Description: In recent years local energy communities are continuously rising as a potential solution to deal with challenges related to global climate targets. These communities involve consumers which join forces for power production, distribution and use, and thus provide the local community with environmental, economic and social benefits. The structure and the number of actors of a local energy community may vary from small local groups and neighbourhoods producing their own renewable energy locally, to larger projects where external actors like power companies or contractors establish a local energy community to save money on the services that are offered.

The design of a local energy community in terms of the number of consumers to engage in relation to feasible power plant sizes is a challenge to tackle in order to efficiently project and realise a LEC.

This service aims to define, considering the possible power plant size and consumers' consumption, which is the optimal consumers configuration to maximise auto consumption while increasing economic benefits, avoiding energy losses and limiting unnecessary costs. Thus, an optimisation algorithm based on consumption data is developed to this aim.

In OMEGA-X, the existing tool will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In particular, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common interoperability framework.
- Validate and improve the tool by leveraging data from different data providers (ASTEA, EDP, EyPESA, PUPIN), testing the LEC Designer tool for different LECs.
- Improve the overall output by combining the outputs of the services provided by the tool with the output of other service providers.

Description of stakeholders involved:

- Service user: DSO, Energy Community Operator.
- Data provider: ASTEA, EDP.
- Algorithm developer: REVOLT.

3.4.3 Gamification for electrical energy savings

Pilots: Pilots where the energy consumption and production will be provided partially/totally by local energy community. ASTEA, EDP and more pilots to be confirmed (intrapilot and interpilot).

Category: Gamification services.

Description: The ongoing worldwide increase of electrical energy consumption and the climate crisis require single individuals, companies and public institutions to make lifestyle changes modifying some daily habits and highlight the necessity of a considerable effort for energy saving in all residential, commercial, transportation and industrial contexts. Reducing energy consumption at all levels becomes crucial, leading to benefits for both domestic and national economies and for the environment, mitigating air and water pollution and conserving natural resources, which in turn guarantees a healthier living environment for the entire world's population and promotes the development in progress and creating jobs.

This service aims to assist final LEC users in taking actions, in order to reduce consumption at home. To this scope, a gamification platform is developed to propose daily challenges to energy users based on their consumption profiles, home appliances, habits and the level of consumer engagement.

Therefore, the solution is a platform that is composed of:

- several dashboards, which enable users to monitor their consumption and to get involved in some challenges proposed by an algorithm that quantifies the probable level of consumer engagement;
- an algorithm which classifies the possible causes of consumption, in order to suggest the best actions to optimise electrical energy usage in order influence end users' consumption patterns.

In OMEGA-X, the existing tool will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In particular, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common interoperability framework.
- Validate and improve the tool by leveraging data from different data providers (ASTEA, EDP, EyPESA, PUPIN), testing the gamification tool for different LECs.
- Improve the user engagement through gamification among different LECs.
- Improve the overall output by combining the outputs of the services provided by the tool with the outputs of other service providers.

Description of stakeholders involved:

- Service user: DSOs, LEC promoter, energy community operator.
- Data provider: ASTEA, EDP.
- Platform developer: REVOLT.
- Intelligent system developer: REVOLT.

3.4.4 Thermal Losses Detection and Benchmarking at LEC level

Pilots: Pilots where the thermal consumption will be partially/totally provided by a local energy community.

Category: Optimization services.

Description: Nowadays, operating heating networks mainly work at temperatures greater than 80 °C heat distributing pressurised water by pipe, reaching heat losses up to 30%. Thus, the reduction of thermal losses is a crucial aspect to be addressed in District Heating Networks

(DHNs). The introduction and the integration of digital technologies represent one of the promising solutions in reducing heat losses.

In the pilot, thermal losses can be estimated by monitoring the flow rates through smart meters with the standard 868 MHz remote reading infrastructure.

The service exploits data-driven approaches to estimate the thermal losses in the local thermal LEC network and the service automatically evaluates KPIs and data analytics including benchmarking services across communities to compare operation in similar operating conditions and identify potential inefficiencies.

Thus, in summary, the objectives of this service are:

- Monitor the thermal consumption of the local communities to provide benchmarking analysis.
- Develop a data-driven tool for thermal losses estimation.
- Implement specific operator interfaces for viewing KPIs and data analytics, in order to minimize the hours spent by technical staff in data analysis and to assist maintenance on the grid at LEC level.

In OMEGA-X, the existing tool will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In particular, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common interoperability framework.
- Validate and improve the tool by leveraging data from different data providers (ASTEA, EDP, EyPESA, PUPIN).
- Improve the overall output by combining the outputs of the services provided by the tool with the output of other service providers.

Description of stakeholders involved:

- Service user: DSO, energy community operator.
- Data provider: ASTEA.
- Algorithm developer: REVOLT.

3.4.5 Water Losses detection and benchmarking at LEC level

Pilots: Pilots where the water consumption will be partially/totally provided by a local energy community.

Category: Optimization services.

Description: The reduction of water losses is a crucial aspect to be addressed in Water Distribution Networks (WDNs) since they account for almost 50% of the global water supply [28].

In the pilot, water losses can be estimated by monitoring the flow rates through smart meters with the standard 868 MHz remote reading infrastructure.

The service exploits data-driven approaches to estimate the water losses in the local water network and the service automatically evaluates KPIs and data analytics including benchmarking services across communities to compare operation in similar operating conditions and identify potential inefficiencies.

Thus, in summary, the objectives and solutions of this service are:

- Monitor the water consumption of the local communities to provide benchmarking analysis.
- Develop a data-driven tool for water losses estimation.
- Implement specific operator interfaces for viewing KPIs and data analytics, in order to minimize the hours spent by technical staff in data analysis and to assist maintenance on the grid at LEC level.

In OMEGA-X, the existing tool will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In particular, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common interoperability framework.
- Validate and improve the tool by leveraging data from different data providers (ASTEA, EDP, EyPESA, PUPIN).
- Improve the overall output by combining the outputs of the services provided by the tool with the output of other service providers.

Description of stakeholders involved:

- Service user: DSO, energy community operator.
- Data provider: ASTEA.
- Algorithm developer: REVOLT.

3.4.6 Estimate the probability of congestions

Pilots: EyPESA and other LECs to be confirmed.

Category: Optimization services.

Description: LECs are a form of stakeholder in the electric power system. These projects aim to promote energy sustainability and social involvement in energy activities. They can be connected to distribution or transmission systems, and they can offer some services to both of the system operators (for example, provide flexibility when there are congestions). With the increasing renewable generation, power systems can be subjected to line or transformer congestions due to load variations and the presence of intermittency of these renewable generation sources at different locations. This service aims to identify the possible congestion scenarios in distribution systems for LECs to be able to provide flexibility service. This tool will be able to forecast congestions based on day-ahead demand prediction, warning for possible congestion scenarios produced by fluctuating power demand, and new distributed generation facilities, etc. The objective is to identify when power lines and transformers in the distribution grid are prone to overload given the forecasted demand. This will allow the DSOs to predict when and where to perform activation of flexibility strategies to distress the grid and further ask this flexibility to LECs.

The service will identify the possible congestions scenarios and grid disturbances related to voltage quality, allowing the system operator to monitor and forecast these scenarios and provide a quick response, where LECs can be able to provide this needed flexibility, to comply with secure service

The service consists of two phases: training algorithm and operation algorithm.

- **Training algorithm:** the grid is analysed and studied with the probabilistic power flows, evaluating the congestions condition in such grid. Further, machine learning models are trained to develop a congestion model able to detect congestion given a demand forecast.
- **Operation algorithm:** this section is where the forecast of demand interacts with the congestion model, giving the output of the probability of congestions to the service user in a daily-ahead schedule.

In OMEGA-X, the existing service will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. The objectives are the same as in the service “Congestion detection” from the Renewables UCF.

Description of stakeholders involved:

- Service user: DSOs and LECs.
- Data provider: DSOs and LECs.
- Algorithm developer: UPC.

3.4.7 Electrical losses detection and benchmarking at LEC level

Pilots: Pilots connected to the electric grid with or without generation assets. EyPESA and other LECs to be confirmed.

Category: Optimization services.

Description: Electricity losses are a direct loss of revenue for LECs and can be divided into technical and NTL. With an energy balance, subtracting the power going out of the system from the power coming into the system, the total losses can be calculated. These can be within an acceptable range (there will always be a minimum of technical losses) or exceed it. In case these losses exceed the range set by the LEC operator, the algorithm will analyse them, with the main objective of reporting the magnitude of the losses and possible causes.

In case the loss profile indicates that NTLs are present, these will also be analysed. The literature classifies these fraudulent losses into two categories as to their origin. The most typical is direct connection, which occurs in front of the meter. The other category is meter tampering, increasing importance due to cyber-attacks, such as FDI.

With the help of classification algorithms, this service can let the LEC operators know which type of fraud (marihuana plantation, crypto mining, squatting or unclear) is predominant in each case by analysing the NTL curve.

Subsequently, the algorithm will review the consumption of the different SMs, and using the short-pattern clustering technique will flag suspicious customers in case of meter tampering fraud.

This service shares algorithms with the Detect non-technical losses service of the Use case family renewables (see section 3.3.9) and introduces some LEC-specific functionalities such as detection of unexpected technical losses and benchmarking.

Thus, in summary, the objectives of this service are:

- Detect if there are abnormal losses (Losses > threshold set by the LEC operator) in a LEC.

- Classify the type of losses into high technical losses or fraud (marijuana plantation, cryptocurrency mining, squatting or other).
- Provide information of the fraud.
- In case of possible losses due to meter tampering, flag the possible fraudsters.

In OMEGA-X, the existing service will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In this aspect, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common interoperability framework.
- Combine the service with the other two losses detection services (developed by REVOLT) to create a BUC that allows LECs detect and benchmark losses in all the different energy vectors.
- Validate and improve the robustness of the tool by leveraging data from different data providers belonging to different use case families (EyPESA from the Renewable UC and other pilots to be confirmed from the LEC UC).

Description of stakeholders involved:

- Service user: LEC operators and DSOs.
- Data provider: LEC operators and DSOs.
- Algorithm developer: UPC.

3.4.8 Reinforcement Plan of Local Energy Communities for Future Renewable Integration

Pilots: EyPESA and other pilots to be confirmed.

Category: Planning services.

Description: In order to ensure that LECs are capable of hosting new renewable generation, utilities must apply appropriate investment strategies to plan grid reinforcements. Making the right investment decisions is a challenge for DSOs that needs to consider technical and economic criteria.

The objective of this service is to apply planning strategies with a tool that minimizes costs of capacity upgrades and operating costs of feeder lines, transformers and/or Li-ion batteries, considering the expected renewable generation in local energy communities.

Two planning strategies will be explored: 1) firstly, a traditional approach will be used, which consists of replacing or reinforcing the network considering traditional assets, including feeder lines and transformers as decision variables; and 2) secondly, a combined approach will be used, considering traditional assets and flexible assets in LEC, such as energy storage systems, which will be included as decision variables to provide a solution for energy congestions or over/under voltage problems. In both cases the objective function will be to minimize the costs of investment and operation.

The outputs of the service will be:

- CAPEX.
- OPEX.

- List of feeder lines to be reinforced, including GIS location and capacity upgrades.
- List of transformers to be reinforced or replaced, including GIS location and capacity upgrade.
- List of new batteries installation, including GIS location and capacity.

In OMEGA-X, the existing service will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In this aspect, the objectives are the following:

- Integrate the existing tool into the Energy Data Space according to the defined common interoperability framework.
- Validate and improve the robustness of the tool by leveraging data from different data providers LV grids belonging in different use case families (ENGIE and ASTEA from the LEC UC and EyPESA from the Renewable UC).

Description of stakeholders involved:

- Service user: DSOs and LEC operators.
- Data provider: DSOs and LEC operators.
- Algorithm developer: UPC.

3.4.9 Optimizing self-consumption of renewable energy at LEC level

Pilots: Pilots where energy management of a PV storage system is required to optimize its performance, belonging to the demonstration of the energy communities.

Category: Optimization services.

Description: This service provides an optimized schedule, on quarter-hourly basis, of the charging/discharging of a battery system associated to a PV generator in a self-consumption application. The schedule is provided for the day after, assuming the availability of the forecasted PV production and consumption profiles on quarter-hourly basis for the same period.

The optimization is carried out targeting the following objectives in order: 1) to maximize self-consumption rate; 2) to limit the electricity consumption from the grid at a given maximum value (peak-shaving), including the potential battery charging from the grid, if needed; 3) to get the most profitability from battery discharging taking advantage of time-of-use tariff structure; and 4) to maximize battery lifetime by means of charging it at lower power rate, when possible.

In OMEGA-X, the existing service will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In particular, the objectives are the following:

- Integrate the existing service into the Energy Data Space according to the defined common interoperability framework.
- Validate and improve the robustness of the service by leveraging data from different data providers (ASTEA, EDP, EyPESA, PUPIN) and different operating and market conditions for different PV plants.
- Improve the overall output by combining the outputs of the services provided by the tool with the output of other service providers.

Description of stakeholders involved:

- Service user: Energy community operator.
- Data provider: Energy community operator (ASTEA, EDP, EyPESA, PUPIN).
- Algorithm developer: Tecnalia.

3.4.10 Optimizing sharing coefficients in collective self-consumption

Pilots: Pilots where energy management of a shared PV system is required to optimize its self-consumption by a group of consumers, belonging to the demonstration of the Energy Communities use case family.

Category: Optimization services.

Description: This service provides the scheduled sharing coefficients, on hourly basis, of a shared PV generator in a collective self-consumption application. The schedule is provided for a certain period, assuming clear-sky conditions for the PV generation and forecasted consumption profiles from the available historic monitoring datasets, on hourly basis, of the consumers.

The optimization is carried out targeting the maximum saving for the energy community as a whole, ensuring a minimum PV generation rate is assigned to every single consumer along the period.

In OMEGA-X, the existing service will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In particular, the objectives are the following:

- Integrate the existing service into the Energy Data Space according to the defined common interoperability framework.

Description of stakeholders involved:

- Service user: Energy community operator (EDP).
- Data provider: Energy community operator (EDP).
- Algorithm developer: Tecnalia.

3.4.11 Planning services

Pilots: Pilots where the building energy consumption will be partially/totally provided by local energy production and have the aim of decarbonization in a short/medium/long therm.

Category: Planning services.

Description: Planning services will define different yearly decarbonization scenarios (until 2050) for the LEC, based on the combination of alternative strategies.

In order to define these scenarios, the planning service will initially calculate an energy baseline for the LEC, considering existing members' energy use, and associated environmental impacts.

This calculation will consider the current characteristics of buildings and energy systems, energy consumption in each point of use, current renewable energy production, geographically representing the data in a GIS interface.

This baseline will be represented in a GIS map, from which different alternatives to consider in 2050 decarbonization scenarios that could be analysed, such as:

- Potential renewable energy generation within the LEC geographical scope and its surroundings.
- Energy efficient and renewable energy systems that can be applied in the LEC.
- Flexibility management strategies (for producers and consumers), and integration of flexibility assets within the LEC.
- Scenarios for enlarging the LEC, analysing surrounding buildings, energy suppliers or energy users that could become part of the LEC.

The economic and environmental performance evaluation resulting from these analyses will support LEC on making decision on potential investments, and for approaching potential new LEC members. It will also guide the LEC on their strategy towards 2050 decarbonization.

In OMEGA-X, the existing service will be integrated in the Energy Data Space and will be enhanced by leveraging multiple data and services from different providers. In particular, the objectives are the following:

- Integrate the existing service into the Energy Data Space according to the defined common interoperability framework.

Description of stakeholders involved:

- Service user: LEC.
- Data provider: Municipality, LEC operator, Tecnalia.
- Algorithm developer: Tecnalia.

3.5 Electromobility UCF Services

3.5.1 Introduction

The Electromobility use case family includes two main use cases: roaming of booking services (led by EDF) and roaming of self-consumption for EV charging (led by ELIA). The objective of these use cases is to set up data exchanges that would simplify or make possible the emergence of services that are collectively useful for the development of electric mobility in Europe, without national borders or technical or commercial networks being an obstacle to exchanges.

The first use case focuses on the evolutions needed to simplify access to charging. Electric vehicle users, especially away from home, need to be reassured about the possibility of finding available charging points on their routes and at their destination. Currently, electric vehicle users already have the possibility to obtain information about charging points in a given area, either through an electromobility service operator with whom they already have a contract or through applications or websites that provide this information free of charge. However, it is clear that this information is often insufficient to answer the essential questions that consumers may have:

- Will the charge point really be available when I arrive to charge?
- What will be the real rate that will be applied to me if I choose to charge at this point of charge?

When the user remains in the network of an electromobility service provider (EMSP) with whom he is contractually linked, the second question is usually answered (sometimes approximately in the case where the charge point is that of a partner with whom his EMSP has a commercial agreement) and the first depends on the possibility of reserving the charge point or having internal information on its status in real time; with a freshness of update that is not the same for all EMSPs.

The roaming of booking services use case proposes to extend the data models that are already currently used by EMSPs and Charging Point Operator (CPOs) to make possible:

- access to the status of charge points in real time,
- access to reliable pricing information for each charge point,
- a charging point reservation mechanism (with all that this implies: modification, cancellation, no show management...).

The second use case focuses on an innovative proposal: to allow users who have solar panels at home to use this energy to charge their electric vehicles while they are at a charging site located elsewhere, perhaps even in another European country. To achieve this, data exchanges between different actors (CPO, EMSP, energy suppliers and network managers), are necessary. The data exchange platform that will be implemented in OMEGA-X will allow to specify and test these exchanges to see how such a service can be offered by EMSPs to end users.

The key elements of this service are:

- the ability to measure and certify the energy exchanges that occur when there is local production or consumption at a load point,
- the functional implementation of the service that considers the user experience.

3.5.2 Roaming of booking services

As stated in the introduction, charging electric vehicles is an activity that can still be complex or stressful for users when they are far from home and from the charging sites where they are used to charge their vehicles.

In order to allow a harmonious development of electric mobility, it seems important to provide services that allow users to: know the actual availability of charging points; know the applicable tariff information for these charging points; reserve a charging slot (with all that this implies). The roaming of booking services that will be specified and for which an implementation is expected, proposes to make these improvements feasible. It will standardize the information models handled by mobility service operators, charging infrastructure operators and other stakeholders; it will also specify data exchange interfaces and exchange dynamics that will enable the implementation of the three functionalities stated above.

Beyond the benefit for the end users, this service will also have an objective interest for the economic actors involved in the recharging of electric vehicles, but also for the community as a whole.

For charging infrastructure operators, this service will enable them to better manage energy consumption on their sites on a dynamic way, to perform smart charging, and to better understand the availability they will have to offer for services to energy markets or network operators.

For network managers, the development of reservation will have an indirect impact: charging infrastructure operators will have a better knowledge of their infrastructure and will better exploit it (load curve) and will therefore be able to request better sized network connections; similarly, users will be better distributed among the existing charging points and will therefore smooth out the overall load observed in an area, while limiting the number of installations required. We can also expect more availability for network services.

For mobility service providers, the proposed functionalities will improve the customer experience and enhance the value of the various agreements they have with infrastructure operators (or their own).

Description of the service:

EV users who want to charge their vehicle have different options: charging at home, at work, or at public stations. In this latter case, the user may wish to reserve a charge point in advance at a given site; and not simply rely on luck to find a free charge point when the time comes.

The service presented here is thus to propose to users of electric vehicles to reserve a charging point on the EU territory.

The service naturally includes all the sub-features expected when talking about reservation: visibility on a set of available infrastructures and tariff conditions associated with each case; possibility to reserve but also to modify or cancel a reservation (with here also a good information of the user on the costs associated with these choices); link with the recharging process itself and its billing.

In a simple version, such a service is easy to set up within the same network of charging infrastructures directly linked to a mobility service provider.

In OMEGA-X, the service benefits from the energy data space and aims at extending and facilitating the availability of charging infrastructures to the largest number of people, by standardizing exchanges and interoperability of exchanges between charging infrastructure operators and mobility service providers.

Description of stakeholders involved:

- Service user: electric vehicle users (EVU).
- Data provider: charging point operators (CPOs), electromobility service providers (EMSPs).
- Other parties involved: EV interoperability service provider (EVISP).

3.5.3 Roaming of self-consumption

Description of the service:

Roaming of self-consumption for EV charging with the help of granular certificates of origin (GCOs). Prosumers receive their own granular certificates of origin and can offer them on a marketplace to claim them for their own EV charging sessions. This allows prosumers to increase self-consumption and to charge fully green. The whole concept relies on self-sovereign identities and is end-to-end verifiable.

Thus, in summary, the objectives of this service are:

- Create GCOs for prosumers (production certificates).
- Create GCOs for EV charging (consumption certificates).
- Matching of GCOs in a simplified marketplace.
- Claim GCOs only if net transfer capacities between bidding zones allow it.

- Application of portable self-sovereign user identity.

The existing service can be build following the Energy Data Space principles and architecture and will be enhanced by leveraging multiple data and services from different providers. In particular, the objectives are the following:

- Creating access to validated information from different sectors and providers leveraging on the enhanced data exchange through the Energy Data Space principles. In particular, we combine e-mobility charging information (time series, location and responsible service provider) with power sector information (time series of renewable energy production devices) to achieve a better integration of both.
- Building new business cases, services and offering higher transparency to the end user through higher availability and portability of such information.
- Avoid data silos and verification effort. The information available in the Energy Data Space is verifiable, ideally through P2P validation, reducing the validation effort. This increases the possibility to automate processes.

Description of stakeholders involved:

- Service user: electric vehicle users, prosumers.
- Data provider: CPO, EMSP, metered data administrator, data hub operator, TSO.
- Other parties involved: GCO registry, GCO platform, ISP.

3.6 Flexibility UCF Services

3.6.1 Introduction

In energy systems, flexibility is linked to the ability a power system has in adjusting its consumption and production to a varying electricity demand, both anticipated and unanticipated. Currently, electricity produced from renewables cannot be stored efficiently, therefore we can, for instance, shave peak hours and smoothen the energy demand by moving some of the consumption to off-peak hours. The end goal is therefore to provide flexibility while optimizing some cost function, such as by minimizing the cost of production through renewable usage.

Flexibility benefits from data spaces, in the sense that it requires a high-level of information integration between different partners, which will be highly accelerated through a common data space easing the burden of many managerial, bureaucratic, and technological aspects of the data exchange. Examples of information exchanged include consumption and production of all network nodes periodically, information from weather stations, information from advanced analytics and predictive services, models already trained, energy consumption/production profiles, information on how to activate and deactivate consumption and production remotely of specific resources, among others.

This use case family will be led by EDP and will have one pilot site located in Portugal, in the city of Maia, in Porto region. In terms of partners, the roles will be:

- MAIA will be a data provider and data user;
- ISMAI (University/RTO) will be a data provider;
- EDP will be a data provider and data user through E-REDES;

- Tecnalia and Odit-e will be service providers;

The Municipality of MAIA is within the metropolitan area of Porto and has 135.000 inhabitants. MAIA is one of the most industrialized municipalities of Portugal and an important transportation hub. MAIA began seriously paving the way to be a sustainable city in 2012, first by tackling energy issues and in 2014 by creating the Sustainable Energy Action Plan addressing the RES penetration, energy efficiency, CO_2 emissions, mobility, citizens' engagement, among others. In the Municipality of MAIA, there are a few production and consumption endpoints that can be used as a source of flexibility that will be the basis for our pilot. These are briefly listed next.

Production sources:

- PV generators (municipal pool and Fórum da MAIA).

Consumption sources:

- Fórum da MAIA (public building): heating, ventilation and air conditioning (HVAC), pumps, and other consumption appliances;
- municipality's EV fleet;
- other public buildings.

Thanks to the Data Space in OMEGA-X, it will be possible to develop the following services:

- Grid observability and network analysis.
- Grid validation platform, real-time.
- Flexibility platform for DER connection, planning.
- Passive consumption baseline prediction service.
- Active consumption resource prediction service.
- Intermittent DER generation resource baseline prediction service.
- Prosumer EMS internal optimization service.
- Flexibility order disaggregation service.
- Aggregated flexibility offer optimization service.

The use cases that will support the services are explained in Table 4 (BUCs) and Table 5 (SUCs), including how each is connected to the services:

Table 4. Flexibility BUC list.

BUC	Description
BUC Flex 1.0 **Flexibility for internal optimization**	In this BUC flexibility is used internally by the Energy Service Company (ESCO) to optimize the energy use in its portfolio of resources, either of production or consumption, owned/operated by a prosumer (as the Municipality of MAIA). The objective of the optimization is either: • to minimize the overall energy bill. • to maximize local renewable self-consumption.

BUC	Description
BUC Flex 2.0 **Flexibility for congestion management with bilateral contracts**	This BUC extends BUC Flex 1.0 so that Flexibility Service Providers (FSPs) provide flexibility services to the DSO by means of long-term bilateral contracts associated to flexible connections to avoid congestions in the distribution grid.
BUC Flex 3.0 **Flexibility for capacity management with market structures**	This BUC extends BUC Flex 2.0 with the ability of the DSO to also procure flexibility from FSPs in a local flexibility market, to manage the distribution grid capacity.

Table 5. Flexibility SUC list.

SUC	Description
SUC Flex 1.0 **Define the context of flexibility management (all BUCs)**	• The DSO calculates the topology of the grid with information coming from the DSO and the historic measurements of the prosumers (*network analysis service from Odit-e*). • The DSO defines (*observability service from Odit-e*) the congestion zones of its grid, being each congestion zone the collection of electrical connection points of resources which have the same effect on grid state. This information will be available to the FSPs, so that each FSP knows which resources it will aggregate in the information provided to the DSO (the baseline and the flexibility offer). • The DSO agrees with each FSP the flexibility options for the collection of DER (limited to generation resources) resources in each congestion zone and signs the associated bilateral connection agreements (*flexibility platform for DER connection from Odit-e*). This just applies to BUC Flex 2.0 and BUC Flex 3.0. This SUC would also involve all the training stages of the predictive services to be delivered by TECNALIA: • Passive consumption baseline prediction service • Intermittent DER generation resource baseline prediction service • Active consumption resource prediction service

SUC	Description
SUC Flex 2.0 **Optimize the baseline of resources (all BUCs)**	• The FSP calculates the baseline of the passive consumption resources that it manages (*Passive consumption baseline prediction service* from TECNALIA). • The FSP calculates the baseline of the DER generation resources that it manages (*Intermittent DER generation resource baseline prediction service* from TECNALIA). • The FSP calculates the optimal baseline operation schedule of each flexible resource based on its internal objective function (*Prosumer EMS internal optimization service* from TECNALIA, which internally calls to the *Active consumption resource prediction service*). The FSP sends to the DSO the baseline of the resources that it manages, aggregated by each congestion zone (BUC Flex 2.0 and 3.0). The DSO calculates the baseline of the resources that is not provided by any FSP, for each congestion zone (*Passive consumption baseline prediction service* from TECNALIA and *Intermittent DER generation resource baseline prediction service* from TECNALIA). (BUC Flex 2.0 and 3.0).
SUC Flex 3.0 **Manage flexibility needs (BUC Flex 2.0 and 3.0)**	• The DSO calculates the need for flexibility in each congestion zone (*observability service* from Odit-e), considering the baselines calculated in SUC Flex 2.0 (Optimize the baseline of resources). • The DSO decides the optimal combination of flexibility options to be activated (i.e. the limit of aggregated production in each time step), among the FSPs with which it has a bilateral contract (*Flexibility platform for DER connection service* by Odit-e). The DSO sends a flexibility order to each selected FSP, so that the FSP activates the corresponding flexibility options. For the remaining uncovered flexibility need, the DSO sends a flexibility request for each congestion zone to the market participants (*Grid validation platform* from Odit-e). This just applies to BUC Flex 3.0.
SUC Flex 4.0 **Optimize flexibility offers (BUC Flex 3.0)**	Each FSP assesses the flexibility request of the DSO, calculates the optimal combination of flexibility provided by the resources in its portfolio and creates the corresponding flexibility offer (*Aggregated flexibility offer optimization service* from TECNALIA). The FSP sends the flexibility offer to the DSO.
SUC Flex 5.0 **Manage flexibility offers (BUC Flex 3.0)**	The DSO decides the optimal combination of flexibility offers to be activated, validating that the combination of the solutions for each congestion zone does not imply any distribution grid violation (*Grid validation platform* from Odit-e). The DSO sends a flexibility order to each FSP, so that the FSP activates the corresponding flexibility offers.

SUC	Description
SUC Flex 6.0 **Activate flexibility orders (BUC Flex 2.0 and 3.0)**	The FSP recalculates the optimal operation schedule of each resource considering the aggregated constraints included in the flexibility order (*Flexibility order disaggregation service by TECNALIA*).

The proposed list of services will be connected to the use cases according to Table 6.

Table 6. Services connection to service developers and use cases.

Service	Category	Service Developers (SDs)	Data Providers (DPs)	BUC	SUC
Grid observability and network analysis	Predictive services	*Odit-e*	MAIA / DSO	BUC Flex 1.0 BUC Flex 2.0 BUC Flex 3.0	SUC Flex 1.0 SUC Flex 3.0
Grid validation platform, real-time	Flexibility management services	*Odit-e*	MAIA / DSO	BUC Flex 2.0 BUC Flex 3.0	SUC Flex 3.0 SUC Flex 5.0
Flexibility platform for DER connection, planning	Flexibility management services	*Odit-e*	MAIA / DSO	BUC Flex 1.0 BUC Flex 2.0 BUC Flex 3.0	SUC Flex 1.0 SUC Flex 3.0
Passive consumption baseline prediction service	Predictive services	*Tecnalia*	MAIA / DSO	BUC Flex 1.0 BUC Flex 2.0 BUC Flex 3.0	SUC Flex 1.0 SUC Flex 2.0
Active consumption resource prediction service	Predictive services	*Tecnalia*	MAIA	BUC Flex 1.0 BUC Flex 2.0 BUC Flex 3.0	SUC Flex 2.0
Intermittent DER generation resource baseline prediction service	Predictive services	*Tecnalia*	MAIA / DSO	BUC Flex 1.0 BUC Flex 2.0 BUC Flex 3.0	SUC Flex 1.0 SUC Flex 2.0
Prosumer EMS internal optimization service	Flexibility management services	*Tecnalia*	MAIA	BUC Flex 1.0 BUC Flex 2.0 BUC Flex 3.0	SUC Flex 2.0
Flexibility order disaggregation service	Flexibility management services	*Tecnalia*	MAIA	BUC Flex 2.0 BUC Flex 3.0	SUC Flex 6.0
Aggregated flexibility offer optimization service	Flexibility management services	*Tecnalia*	MAIA	BUC Flex 3.0	SUC Flex 4.0

3.6.2 Grid observability and network analysis

Pilots: MAIA (intrapilot).

Category: State estimation and predictive services.

Description: Grid observability and network analysis is an analytic service based on smart meters data and SCADA/DMS data. It is twofold:

- Automatic retrieval of the low voltage network topology and associated grid analysis with statistics on the load and voltage distributions of each secondary substation and secondary substation's feeder.
- Estimation of the real time and forecasted grid state (voltage plan and power level at given assets) with 10 minutes time steps and 2 hours of time horizon considering an estimate of the non-measured consumption.

The output of this service allows to identify the area in the network where the operational conditions are close or exceed nominal values and therefore pre-map the area needing flexibilities. The state estimation provides observability on real time and short-term operating conditions and is therefore used for energy community operation in the service Grid Validation Platform RT.

Thus, in summary, the objectives of this service are:

- Improve the knowledge on low voltage network structure and on operating conditions.
- Provide DSOs with real time and forecasted observability on the LV network to anticipate network issues.

Description of stakeholders involved:

- Service user: DSOs.
- Data provider: DSO, meteorological data provider.
- Algorithm developer: Odit-e.

3.6.3 Grid validation platform, real-time

Pilots: MAIA (intrapilot).

Category: Flexibility management services.

Description: The Grid validation platform in real-time is a service which constantly monitors the grid state thanks to the Grid observability and network analysis service. From this information and the models it embeds, all the impacts on the grid of energy exchanges and/or flexibility offers are simulated to give a validation check before being activated (for the day ahead and the next hour with 10' time steps) and take the necessary mitigation measures.

In case of predicted low voltage grid limit excursion, the required flexibilities will be estimated and located for the next hours with 10' time steps. Such information will be sent to the local trading platform for flexibility offer design.

Thus, in summary, the objectives of this service are:

- Ensure the emergence of flexibilities is not worsening the low voltage network operation.
- Using flexibilities as a lever to improve low voltage network operation practices and provide an alternative to reinforcement in cases of low occurrence rate of grid limit excursion.

Description of stakeholders involved:

- Service user: DSOs / local energy community manager.
- Data provider: DSO.
- Algorithm developer: Odit-e.

3.6.4 Flexibility platform for DER connection, planning

Pilots: MAIA (intrapilot).

Category: Planning tool for energy community creation.

Description: The Flexibility platform for DER connection (planning) is a collaborative platform for the creation of local energy communities. The platform allows public institutions to submit calls for expression of interest in order to fulfil targets such as renewables insertion, citizen involvement in the energy transition and energy poverty reduction.

The platform receives expression of interest and computes the optimal collective territorial scenario to align:

- Public institutions interest in relation with the predefined targets.
- RES and flexibility investors interest regarding the financial aspect.
- DSO interests in terms of impact on the grid and subsequent reinforcement work.

The output of the platform is a collaborative offer which includes technical aspects and allows optimal integration of the energy community in the grid. This platform will enable different functionalities, such as for example the optimization of connection requests and the management of collaborative offers.

Thus, in summary, the objectives of this service are:

- Promoting RES insertion through an optimized insertion strategy.
- Improving the economical profitability of the operation through pooling of resources.
- Improving fairness in the allocation of investment costs on the grid for RES insertion.

Description of stakeholders involved:

- Service user: Public Institutions, RES and flexibility investors (professional or individuals) and DSOs.
- Data provider: MAIA / DSOs / RES and flexibility investors.
- Algorithm developer: Odit-e.

3.6.5 Passive consumption baseline prediction service

Pilots: Pilots where consumption prediction of non-controllable resources is needed, belonging to the demonstration of the Energy Communities and Flexibility use case families.

Category: Predictive Services.

Description: This service provides a forecast of the electrical consumption of one or several metering points (e.g. a building, or a collection of buildings), based on a model developed with historic measurements provided by the smart meter at the metering point. Optionally, it can be considered the influence of the ambient temperature as an exogenous variable.

The resources connected downstream the metering point are considered therefore as a whole, as they all are considered passive resources with no controllability by the FSP. In case that there are certain active resources downstreaming a metering point (i.e. their consumption profiles cannot be varied due to the influence of third parties like the FSP or the DSO) it could be estimated its baseline using the *Active consumption prediction service* (if the measurements from the submeters of those resources are available) then the smart meter measurements used for model training will be modified subtracting those submeter measurements. Then, in case that in a certain building there is an active resource, this prediction will not include the baseline consumption of that resource.

The metering points included in a same prediction can depend on the definition of the bidding zones. For instance, if the service user is the DSO, it would get an aggregated prediction of all metering points in each bidding zone. A bidding zone is defined by the Harmonised Electricity Market Role Model (HEMRM) as the largest geographical area within which market participants are able to exchange energy without capacity allocation. In the flexibility use case family, we will use the denomination of bidding zones compliant with the USEF approach, calling them congestion zones. Alternatively, if the location of a metering point in the grid topology has an influence on the analysis of the flexibility needs, the prediction can be decoupled in specific predictions for sets of metering points which influence on the analysis is identical (i.e. the active energy consumed in a metering point has the same effect that the active energy consumed in all the metering points within that group).

The prediction is done for the day after, and the granularity will be 15 minutes, assuming the ability of the available building smart meters to accumulate the energy in those time intervals. Optionally, it can be considered the influence of the ambient temperature as an exogenous variable.

Thus, in summary, the output of this service is a 96-quarter-hour consumption forecast for the following day.

In OMEGA-X, the existing service will be integrated in the Energy Data Space. In particular, the objectives are the following:

- Integrate the existing service into the Energy Data Space according to the defined common interoperability framework.
- Validate and improve the robustness of the tool by leveraging data from different data providers from the Flexibility UCF (MAIA), and also the Local Energy Communities use case family (EDP, ASTEA), as passive resource prediction would be used in many of the BUCs.
- Provide the necessary functionalities to a FSP or a DSO, by coordinating the utilization of this service in conjunction with the other services to be developed by Tecnalia, regarding the prediction of generation and consumption resources.

Description of stakeholders involved:

- Service user: FSP/DSO.
- Data provider: Prosumer (MAIA).
- Algorithm developer: Tecnalia.

3.6.6 Active consumption prediction service

Pilots: Pilots where consumption prediction of controllable (flexible) resources is needed, belonging to the demonstration of the Energy Communities and Flexibility use case families.

Category: Predictive services.

Description: This service provides a forecast of the electrical consumption of an active resource, given some input operating parameters. An active resource is a resource whose consumption profile can be varied due to the influence of third parties like the FSP or the DSO. The consumption profile depends on the operation parameters of the resource, i.e. the control setpoints commanded by the user when utilizing the resource.

In principle, we will consider three types of active resources, which could be developed based on the availability of input data in the MAIA demonstration.

- HVAC: system that provides heat and/or cold to a building, including domestic hot water (DHW)
- Pumps: irrigation pumps used for watering gardens.
- EVs: the prediction is related to the electricity used at a charging post (or a set of charging posts in the same distribution grid zone). Operation parameters could include the charging point utilization schedule (ON/OFF) and a charging point maximum energy limitation schedule. Each one of these resource types implies the development of a different model, which correlates the process specific inputs and operating parameters with the predicted electrical consumption profile.

The prediction is done for the day after, and the granularity will be 15 minutes, assuming the ability of the needed measurements in those time intervals.

Thus, in summary, the objectives of this service are:

- To provide a 96-quarter-hour consumption forecast for the following day.
- Optionally, to provide the estimation of the profile of the operating conditions result of the application of certain operation parameters. For instance, in the case of an HVAC, the estimation of the indoor temperature of each zone, needed to assess if the comfort range is satisfied.

In OMEGA-X, the existing service, will be customized for the typology of the active resources in the MAIA demonstration coordinated by EDP, and will be integrated in the Energy Data Space. In particular, the objectives are the following:

- Integrate the existing service into the Energy Data Space according to the defined common interoperability framework.
- Provide the necessary functionalities to an FSP, by coordinating the utilization of this service in conjunction with the other services to be developed by Tecnalia for the FSP, so that this actor acquires the technical capabilities to manage the prosumer resources in a flexible manner, either for internal optimization or to provide flexibility to the DSO.

Description of stakeholders involved:

- Service user: FSP.
- Data provider: MAIA, flexibility service provider.
- Algorithm developer: Tecnalia.

3.6.7 Intermittent DER generation resource baseline prediction service

Pilots: Pilots where PV prediction is needed, belonging to the demonstration of the Energy Communities, Renewables and Flexibility use case families. In fact, the service will be developed within the activities related to the Renewables use case family, but its application on the Flexibility use case family will be based on the specific requirements of the MAIA demonstration.

Category: Predictive services.

Description: This service provides a forecast of the production of a PV generator, at the output of the inverter, being the inverter connected to a collection of PV modules. The prediction is done for the day after, and the granularity will be 15 minutes, assuming the ability of the available energy measurement device (the smart meter at the PV generator connection point with the distribution grid, a dedicated sub meter or the inverter itself) to accumulate the energy in those time intervals.

The prediction is done ad-hoc for a specific PV generator, based on weather forecast and historic data describing both the production and the weather conditions.

Thus, in summary, the output of this service is a 96 quarter-hour PV production forecast for the following day.

In OMEGA-X, the existing service will be integrated in the Energy Data Space. In particular, the objectives are the following:

- Integrate the existing service into the Energy Data Space according to the defined common interoperability framework.
- Validate and improve the robustness of the tool by leveraging data from different data providers from the Flexibility use case family (MAIA), and also the Renewables use case family (ESTEBANELL) and Local Energy Communities use case family (EDP, ASTEA), as photovoltaic prediction would be used in many of the BUCs.
- Generalize the interoperability of this service standardizing the data interface among the service providers (TECNALIA, METEO).
- Provide the necessary functionalities to an FSP, by coordinating the utilization of this service in conjunction with the other services to be developed by this for the FSP, so that this actor acquires the technical capabilities to manage the prosumer resources in a flexible manner, either for internal optimization or to provide flexibility to the DSO.

Description of stakeholders involved:

- Service user: Flexibility service provider (FSP).
- Data provider: Prosumer (MAIA, EyPESA, ASTEA, EDP).
- Algorithm developer: TECNALIA.

3.6.8 Prosumer EMS internal optimization service

Pilots: Pilots where initial scheduling of active resources is needed, belonging to the demonstration of the Energy Communities and Flexibility use case families.

Category: Flexibility management services.

Description: Service used by the FSP to plan the optimal operation schedule of each active consumption resource of a prosumer. This service is executed therefore for the whole collection of active resources of a prosumer in BUC Flex 1.0, despite there are located in different congestion zones (so no constraints imposed by the DSO are yet considered, as in the BUC Flex 2.0 and BUC Flex 3.0).

The objective of the optimization is either:

- to minimize the overall energy bill, assuming that the prosumer has a supply contract with the same supplier for some of its metering points, we consider the case in which there are contractual conditions applicable to that collection of metering points. Those conditions could define constraints in the aggregated energy consumption (for instance, imposing minimum or maximum values) or could imply billing energy according to predefined ranges.
- to maximize local renewable self-consumption, based on the expected production of the intermittent renewable (such as PV) generators of a prosumer.

In OMEGA-X, the existing service will be integrated in the Energy Data Space. In particular, the objectives are the following:

- Integrate the existing service into the Energy Data Space according to the defined common interoperability framework.
- Provide the necessary functionalities to an FSP, by coordinating the utilization of this service in conjunction with the other services to be developed by Tecnalia for the FSP, so that this actor acquires the technical capabilities to manage the prosumer resources in a flexible manner, in this case for internal optimization.

Description of stakeholders involved:

- Service user: FSP.
- Data provider: Prosumer (MAIA), FSP.
- Algorithm developer: Tecnalia.

3.6.9 Flexibility order disaggregation service

Pilots: Pilots where re-scheduling of active resources is needed due to the order of the DSO, belonging to the demonstration of the Energy Communities and Flexibility use case families.

Category: Flexibility management services.

Description: Service used by the FSP to re-schedule of each active consumption resource of a prosumer included in a bidding zone, due to the flexibility order of the DSO. The disaggregation service addresses the problem of the operation of the active resources in real-time. This service is executed just in case of the reception of the flexibility order, and the initial schedule of the active resources is considered a prerequisite (prosumer EMS internal optimization service). In fact, this service can be considered as a functional extension of that service that 1) is executed for all the active resources of an FSP in a bidding zone and 2) takes into account the flexibility order active energy values (maximum or minimum) as a constraint for the selection of the optimal resources of each resource (optionally including also the available capacity in those 15' periods where flexibility is not ordered, so that the re-scheduling in those periods does not exceed the available capacity).

In principle, the following active resources (both generation and consumption) are considered, which could be developed based on the availability of input data in the MAIA demonstration:

- HVAC.
- Pumps: Irrigation pumps used for watering gardens.
- EVs.
- PVs.

Thus, in summary, the objective of this service is to ensure that the re-schedule of the active consumption is deployed reliably, at the minimum cost and within the technical and comfort limits of the resources. For this purpose, an optimization algorithm will be developed. This service is used in BUC Flex 2.0 and 3.0.

In OMEGA-X, the existing service will be integrated in the Energy Data Space. In particular, the objectives are the following:

- Integrate the existing service into the Energy Data Space according to the defined common interoperability framework.
- Provide the necessary functionalities to an FSP, by coordinating the utilization of this service in conjunction with the other services to be developed by Tecnalia for the FSP, so that this actor acquires the technical capabilities to provide flexibility to the DSO.

Description of stakeholders involved:

- Service user: FSP.
- Data provider: MAIA, FSP, DSO.
- Algorithm developer: Tecnalia.

3.6.10 Aggregated flexibility offer optimization service

Pilots: Pilots where flexibility is provided by aggregators to the DSO following market structures, belonging to the demonstration of the Energy Communities and Flexibility use case families.

Category: Flexibility management services.

Description: This service is needed by FSPs (and Large prosumers) to present flexibility offers in (local) flexibility markets (BUC Flex 3.0). FSPs have a portfolio of prosumer flexible resources, having a contract associated to each resource. Each contract reflects payments both due to the availability and the utilization of the contracted resource.

Each prosumer resource is initially scheduled as a result of the process of internal optimization (BUC Flex 1.0). This information has to be forwarded to the DSO, aggregated for all the resources represented by the same FSP in each congestion zone defined by the DSO, which will take it into account to calculate its flexibility needs.

Additionally, the DSO can present for each congestion zone a flexibility request at the local flexibility market, in case that the process of assessment of flexibility needs predicts a congestion in its grid for the following day. The flexibility request will be for a determined period within the following day, the so-called congestion period.

As a response to the flexibility request presented by the DSO, each FSP can present a flexibility offer (either to increase or to decrease the baseline consumption presented before) for each 15' minute period of the congestion period. Part of this flexibility offer is the price that the aggregator asks for the flexibility activation.

In principle, the following active resources (both generation and consumption) are considered, which could be developed based on the availability of input data in the MAIA demonstration:

- HVAC.
- Pumps: Irrigation pumps used for watering gardens.
- EVs.
- PVs.

Each one of these resource types implies the development of a different model, which correlates the process specific inputs and operating parameters with the predicted electrical consumption profile. This prediction is done for the day after, and the granularity will be 15 minutes, assuming the ability of the needed measurements in those time intervals.

Thus, in summary, the objective of this service is to decide the best flexibility offer that an FSP can present at the flexibility market, given the needs of the DSO (flexibility request) and the price that it would have to pay to the prosumers for the activation of the flexibility.

This is, to provide flexibility offers for the following day for the flexibility request within the congestion period defined by the DSO. For this purpose, an optimization algorithm will be developed. This algorithm will define the bidding strategy to participate in flexibility markets (BUC Flex3.0). determining the amount of flexibility to bid into the market, as well as its price. The objective is to maximize the incomes received by the aggregator from its participation into such markets.

In OMEGA-X, the existing service will be integrated in the Energy Data Space. In particular, the objectives are the following:

- Integrate the existing service into the Energy Data Space according to the defined common interoperability framework.
- Provide the necessary functionalities to an FSP, by coordinating the utilization of this service in conjunction with the other services to be developed by Tecnalia for the FSP, so that this actor acquires the technical capabilities to provide flexibility to the DSO using flexibility market structures.

Description of stakeholders involved:

- Service user: FSP.
- Data provider: DSO / FSP.
- Algorithm developer: Tecnalia.

4 Data Sovereignty, Security and Privacy

4.1 State of the art analysis of existing data sovereignty solutions

Data sovereignty is the capability of a natural person or corporate entity for exclusive self-determination regarding its economic data goods [8].

There are different frames for providing sovereign data exchange. Nowadays, companies share data considering the access control, which restrict what a user can do with data, or which programmes can execute. From the literature review of [29] and [30], we can identify three different perspectives:

1. To restrict the usage on the data receiver's side following the Digital Rights Management (DRM), so the data is protected by intellectual properties [29].
2. To process data without acceding to the raw data: Secure Multiparty Computation (SMC) [30].
3. The usage control "is a generalization of access control to cover authorizations, obligations, conditions, continuity (ongoing controls), and mutability", coined by Park and Sandhu [29].

The usage control is a promising approach to provide trust to sharing sensitive data because it can be applied to peer-to-peer data sharing and federated infrastructure [31]. Data usage control is a key capability of the International Data Spaces (IDS) initiative. IDS provides the information model and policy language to specify usage restrictions and everything that a connector needs to implement and enforce the data usage restrictions at the consumer side.

Furthermore, IDS provides strategies for the negotiation of restrictions between data provider and data consumer and the transformation of independent IDS contracts to technology.

In addition, IDS addresses data provenance as additional concept to data usage control to cope with transparency and accountability.

Based on IDS there are different available solutions to use data usage control, developed by Fraunhofer: MYDATA Control Technologies, LUCON and Degree [32].

A Policy Definition Language is required to define and agree access and usage policies. The defined and agreed policies can be used directly or translated into an executable language, e.g. Rego [33]. We propose to use ODRL as an interoperable standard for the negotiation and acceptance of access and usage policies including the policy negotiation sequence as defined by IDSA in the IDS-RAM [14]. Nevertheless, it might be required to translate ODRL policies into an executable language like Rego during runtime. The enforcement of those policies can be realized as described in the IDS-RAM. To be noted that under the condition that the various grammars allow it, the semantic interoperability between different executable policy engines could be achieved only if common controlled vocabularies are adopted.

4.2 State of the art analysis of existing security & data privacy solutions

Data security and data privacy lie at the core of the Gaia-X principles for data exchange. The participants should be reassured that their data is handled according to agreed conditions of security and privacy. This survey provides an overview of the various approaches to data security and data privacy, in order to identify those directly applicable to the OMEGA-X project and determine how they can be implemented.

4.2.1 Data security

Protecting data is essential in any domain. Whether the data is sensitive or not, data security comprises a wide range of strategies to protect data from unauthorized or involuntary access, alteration or deletion. This subsection lists the most common data security techniques and popular solutions in the market.

4.2.1.1 Data encryption

Data encryption is a technique that converts information into an unreadable format that requires a key to be deciphered. This key can be a digital element, a password, or even a physical device. This technique prevents data from being read in the case of a data leak. We distinguish two general types of data encryption: at rest and in transit.

Data at rest means data residing on a storage device.

*Full-disk encryption (***FDE***)* is a fundamental encryption method where the whole storage volume is encrypted, except for the booting area. Some reasons to opt for a FDE is to have all data systematically encrypted – useful for sensitive data storage – and to be able to immediately destroy data (render it unreadable) by destroying the keys in case that data has been compromised. Major vendors like Apple [34] and Microsoft [35] natively integrate full-disk encryption in their offers; whereas other commercial multi-purpose solutions, like Trellix Data Protection [36] allow to manage disk encryption even in cloud volumes. Veracrypt [37] is an example of a multi-platform free and open-source tool. For Linux based servers, which are increasingly used in industry, there exist FDE open source (LUKS [38]) and proprietary (SecureDoc [39]) solutions, both based on dm-crypt subsystem of the Linux kernel.

*File-Based Encryption (***FBE***)*, on the other hand, allows for encryption of individual files or directories, which offers the flexibility to use different keys for encrypting different structures of the file system, as well as an individual access control management. There are plenty of FBE solutions in the market, both free (e.g. CryFS [40]) and paid (e.g. AxCrypt [41]). More generally, public cloud service providers offer granular data encryption options in their storage solutions (e.g. Amazon S3 [42], Azure Storage [43] or Google Cloud Storage [44]).

Data in transit means data being transferred between two storage devices.

The main mechanisms for encrypting data in transit rely on secure transfer protocols; the most popular of them being, arguably, **SFTP** [45]. SFTP (SSH File Transfer Protocol) provides encrypted data transmission between a client and a server. A symmetric session key is randomly generated, which is used to encrypt and decrypt the data. Moreover, the server authenticates the client using an SSH key pair. Once the authentication is complete, the data is transferred in packets over the encrypted channel. SFTP is generally replacing the former FTPS (FTP secure) which uses the cryptographic protocol TLS (currently in version 1.3 [46]), used as well to secure communication in HTTPS. **TLS** provides a secure channel between two communicating entities. The secure channel fulfils the properties of authentication (at least on the server side), confidentiality (data is encrypted and only visible to both entities), and integrity (data cannot be modified by an attacker without detection. Both SFTP and TLS are widely used current standards for encrypting data in transit over the internet.

IoT data encryption. With the fast-growing use of Internet of Things (IoT) devices in the energy sector, data encryption, both at rest and in transit has become a key security concern. However, a number of constraints inherent to IoT networks influence the designers' decisions with respect to the implementation of a cryptography strategy. Limited power and limited memory, coupled with a potentially large number of devices connected in a single network, make it difficult to use common algorithms such as AES [47] (used also in TLS), which require computing power beyond of what an IoT device can supply. This challenge has been addressed for several years now, proposing lightweight cryptography algorithms for IoT [48]. A recent survey [49] lists and compares a number of algorithms to enable IoT cryptography.

4.2.1.2 Data Backup

Backup is a common method to increase data security, which consists in creating copies of the data that can be used for operational recovery in case of data loss due to involuntary or malicious deletion, or encryption.

Traditional backup techniques consist in copying the data to other storage volumes. The well-known 3-2-1 backup strategy [50] recommends having at least three copies of your data, the original plus two more copies (3); in at least two **different devices** (2), to prevent media failure; and at least one copy **off-site** (1), to prevent site-specific failures, including natural disasters.

Popular OS vendors provide native solutions to backup data manually or create restoring points on a device, and **synchronize** data in the cloud [51] [52]. Needless to say, the best approach for data backup is to have an automated process which takes care of triggering periodic data backups.

Periodic backups can be managed as full backups (mirror) or in an incremental way. Full backups make copies of the whole file system (or file set), whereas incremental backups will store only the files that are different from the previous backup point. A widely used tool for backups in Unix-compatible OS is **rsnapshot** [53]. This tool is based on rsync (utility for incremental file transfer) and runs over SSH. The frequency of snapshots can be easily configured, and the user-defined number of snapshots prevent disk space overload. An alternative tool is **rdiff-backup** [54], in which the target directory becomes a copy of the source directory, but extra reverse diffs are stored as well, so that data recovery of older files is possible. rdiff-backup aims at combining the best features of a mirror and an incremental backup.

In a similar way to data encryption, large cloud providers offer a wide range of (paid) backup options for users of their data storage solutions, from local to global replication.

4.2.1.3 Data Erasure

Also known as data wiping, this technique consists in **overwriting** existing data in a sector of a storage device with random or pseudorandom binary values, with the purpose to render this data irrecoverable. Data erasure is the data security procedure applied when data must be completely destroyed; it usually involves **multiple passes** of sector overwriting. It is an intermediate action between regular data “deletion”, which may just remove the pointers to specific data while leaving them on the storage device, and physical device destruction, which would render the media itself unusable.

Data erasure is useful when data has to be destroyed upon reaching its intended lifetime, or when a device storing highly sensitive data is thought to be compromised or will be reused for other purposes. Two popular solutions used by large companies and financial institutions and certified by several national security and defence organisations are BitRaser [55] and WipeDrive [56]. For Unix-compatible systems, it is possible to use free and simple utilities such as *shred* or *dd*, in less sensitive environments, since the result is the responsibility of the user.

A special combination of data encryption and data erasure is called crypto-shredding, and consists in destroying or overwriting the encryption keys of an already encrypted storage device. This technique renders the data immediately irrecoverable.

Data erasure often follows regulatory compliances and legal obligations of the data holder, in order to protect the interests of the subjects to whom the data belongs. In these cases, the data erasure service provider might produce a detailed report and a certificate for the company's accountability.

4.2.2 Data privacy

Data privacy, part of a more general data management strategy, deals with the handling of sensitive data, particularly personal data, in compliance with applicable territorial laws and regulations. As evoked in Section 2, data privacy in the European Union is ruled by the General Data Protection Regulation (GDPR) as of 2016 (went into effect in May 2018), yet a number of national laws enforced data privacy decades before.

According to Article 4 of the GDPR, 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

Data privacy strategies aim at preserving an individual's personal data from being accessed, modified without their explicit consent.

Privacy rights of the data subject

Chapter 3 (Articles 12-23) of the GDPR concern the rights of a person (data subject) regarding their personal data. In particular, the data subject should be informed about those rights and the procedures to exercise them. In particular, some user rights that should be addressed to technical procedures, are:

- Right of access by the data subject (Art. 15): People have the right to see what personal data you have about them and how you are using and storing it, and for how long. A technical procedure should be in place to ensure people can receive a copy of this information upon verification of their identity.
- Right to rectification (Art. 16): The data subject has the right to rectify inaccurate personal data concerning them. A procedure must ensure that data rectification is possible.
- Right to erasure (Art. 17): People generally (with limited exceptions) have the right to ask for the deletion of all their personal data. The data controller must be able to honor their request and, if possible, send an erasure certificate.
- Right to restriction of processing (Art. 18): Under certain conditions, a data subject can request to restrict or stop processing their data. A procedure should allow the controller to do so, but they are still entitled to keep the data.
- Right to data portability (Art. 19): The data controller should be able to send to the subject their personal data in a commonly machine-readable format either to them or to a third party they designate.

- Right to object (Art. 20): A subject may request to stop processing their data if it is being used for direct marketing or profiling.

Note that the above is not an exhaustive list; the reader is invited to refer to the Chapter 3 of the GDPR [1] to read the full text and specifications.

Data protection by design and by default

Companies and organisations are encouraged to implement technical and organisational measures, at the earliest stages of the design of the processing operations, in such a way that safeguards privacy and data protection principles right from the start ('data protection by design'). By default, companies/organisations should ensure that personal data is processed with the highest privacy protection so that by default personal data isn't made accessible to an indefinite number of persons ('data protection by default') [57]. Data protection by design techniques include the use of pseudonymization (replacing personally identifiable material with artificial identifiers) and encryption. Data protection by default requires, for example, limiting from the beginning the accessibility of a user's profile so that it isn't accessible by default to an indefinite number of persons.

Compliance verification

Data privacy covers a large spectrum of requirements which are not simple to verify automatically. A number of commercial tools propose compliance audits for websites or applications with respect to the data stored by them, including cookies and personal data such as name and email. However, certain procedures to guarantee the rights of the data subject can only be manually audited only manually by certified entities. Some national authorities and EU projects provide toolkits [58] or checklists [59] to self-assess GDPR compliance, yet the safest way to ensure that data privacy is applied correctly is to request legal advice.

4.3 OMEGA-X approach to data sovereignty and Privacy

The Data Spaces Business Alliance (DSBA), in its recent published technical convergence document, proposes an implementation-driven plan of a Minimum Viable Framework (MVF) for the creation of data spaces [60].

DSBA-CTO recommendations for data sovereignty and trust are:

- An eIDAS and EBSI -compatible Trust Anchor framework.
- A decentralized Identity and Access Management (IAM) framework based on:
  - A set of verifiable credential issuing protocols (Self-Issued OpenID Provider v2 (SIOPv2), via DID Comm channel, etc).
  - A set of verifiable presentation protocols (ex: OpenID Connect for Verifiable Presentations (OIDC4VP), Verifiable Presentation Request [61].
  - An ABAC (Attribute Based Access Control) framework comprising components implementing PEP, PDP, PAP/PMP, and PIP functions.

For the OMEGA-X MVF, the first proposal of the implementation-driven plan consists in working through different versions; these, are expected to have a duration between three and six months.

Regarding data usage control, the workstream 2 will incorporate the IDS Connector functions and support to ODRL for the definition of access/usage control policies, the workstream 3 will incorporate the Shared Catalogue and Federated Marketplace services based on TM Forum standards and aligned with Gaia-X and IDS RAM specifications and workstream 4 will incorporate additional IDS architectural elements for usage control.

4.4 Data quality & certification mechanisms

4.4.1 Data Quality measurements

Every modern business relies on its industry-specific data when making most of its far-reaching decisions. Hence, poor and unreliable information, that is based on a low-quality data, can damage business performance and efficiency. Examples of said impact are listed in research [62], and some of them are:

- wasted resources and additional costs;
- damage to the reliability of analytics;
- negative impact on both business reputation and customer trust;
- negative effect on customer experience;
- slowing down of digital transformation, which in turn hinders key business initiatives;
- delays in fulfilment;
- hindering compliance around regulatory obligations.

Considering all the above, it's crucial for businesses to apply methods of data quality measurements.

Data quality is generally defined as an assessment, that represents how a certain dataset is fit for its intended use inside an organization. Although there are some general requirements for data to be considered of high quality, in-depth definition of such quality is highly dependent on specific requirements of business itself.

Measurements, that serve as an instrument to differentiate between high-quality and low-quality data are called Data Quality Metrics. Quality metrics can be objective (based on quantitative measures), or subjective (based on qualitative evaluations, for example – experience, needs and perception of stakeholders). A unified standard for these metrics is not yet established, and defining new measurement systems is always an option for businesses. However, metrics that are considered core by many internet resources including [63], [64] are portrayed in the table below Table 7.

Table 7. Data Quality Metrics

Metric name	Definition	Example
Completeness	Evaluates if a specific dataset includes all the necessary data.	Percentage of data fields that have values entered in them.
Accuracy	Represents how accurately data reflects real-world objects.	Percentage of values that are correct in comparison to the actual value.

Metric name	Definition	Example
Relevancy	Shows whether data is relevant to the specific business industry.	Percentage of data that is relevant to business industry.
Auditability	Allows to track how well databases are managed or data inside of them is used.	Percentage of data that was not overwritten/misused.
Consistency	Ensures that data remains consistent.	Percentage of values that match across multiple sources.
Validity	Measures how well data aligns with required values.	Percentage of data, values in which are within a domain of acceptable values.
Timeliness	Reflects the accuracy of data at the specific point of time.	Percentage of data that can be obtainable within a certain period (weeks or days, for example).
Integrity	Relates to how all the data quality metrics mentioned are maintained while data moves between different systems.	Percentage of data that remained the same across multiple systems after being moved.
Uniqueness	Ensures that data in the dataset is the most relevant and has no duplicates.	Percentage of unique entries in the dataset.

Along with data quality, another important issue to consider when evaluating data is security. Data Security is defined [65] as a set of standards and technologies, that protect data from intentional and accidental destruction, modification and disclosure. According to the article [66], data security ensuring process is mainly focused on:

1. confidentiality – an intention to protect data from unauthorized access;
2. integrity – an intention to protect data from unauthorized changes;
3. availability – an intention to deal with making data accessible to authorized entities and users.

Security breaches can lead to leakage of confidential data, which in its turn may result in litigation cases, fines and damage to an organization's reputation. As mentioned in the report [67], the average number of cybercriminal attacks and security breaches rose to approximately 15.1% in 2021, while the number of material breaches – those generating a large loss, compromising many records, or having a significant impact on business operations – jumped up to 24.5% more incidents than on a previous year. That is why ensuring data security is critical on every device, network or application level of distributed energy resources (DERs).

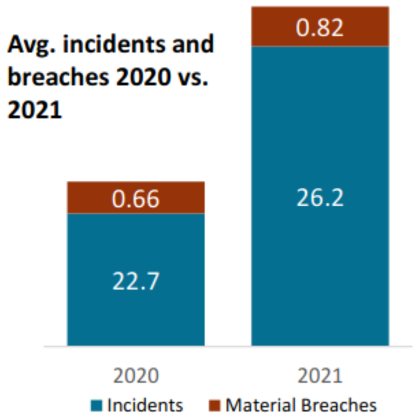

Figure 3. Leakage incidents & breaches

Throughout the years multiple industry standards and guidelines were developed to help businesses handle security issues. For instance, North American Electric Reliability Corporation has developed mandatory and enforceable Reliability Standards, which are subject to Federal Energy Regulatory Commission. Among these standards there are cybersecurity requirements for critical infrastructure protection [68], however these requirements are more applicable to cybersecurity risks of the bulk power system rather than corresponding issues of distributed energy resources.

Another framework to mention is the one developed by National Institute of Standards and Technology (NIST) [69]. NIST cybersecurity framework is meant to help an organization begin (or upgrade) their cybersecurity program. However, this framework also does not address the cybersecurity risks of DERs, as well as DERs' communications with the grid.

In addition, these are cybersecurity standards and tools for power systems that several other organizations have developed:

- The International Electrotechnical Commission 62351 standards for power systems management and associated information exchange [70].
- The Institute of Electrical and Electronics Engineers C37.240 standard cybersecurity requirements for substation automation, protection, and control systems [71].
- Cybersecurity Capability Maturity Model, developed by The U.S. Department of Energy as a tool to help organizations evaluate their cybersecurity capabilities and optimize security investments [72].

In conclusion, well-known living cybersecurity frameworks are mainly focusing on managing issues of the entire distribution grid. In order to successfully protect both data and communications of distributed energy resources, industry definitely needs some sufficient procedures, standards and policies.

4.4.2 Data Certification Mechanisms

In order to explore data certification mechanisms, it's necessary to take a closer look into standards and initiatives that are well-known in the industry, such as IDS Certification [73].

IDS Certification ensures that all IDS components and possible participants – individuals and organizations who express a wish of taking part in data sharing process – meet the highest security standards. Said standards are an industry-proven security criteria, which are derived from ISO/IEC 27001 (international standard for information security management) and IEC 62443 (cybersecurity for operational technology in automation and control systems). In addition, IDS Certification puts into practise security criteria, which are described in the Cloud Computing Compliance Criteria Catalogue (C5) and CSA Cloud Controls Matrix. Finally, IDS certification includes some IDS specific criteria.

The certification criteria are formalised in IDS Certification criteria catalogues. As it is stated in the resource [15], IDS certification has two complementary parts: the first part is defined for a participant to prove that the IDS component they are using is built in a correct way; the second part is intended to make sure that the operational environment of an IDS component aligns with all requirements and rules in terms of data sharing support.

Core component certification

This evaluation offers two types of assessments to allow more flexibility for companies and organizations, which have differences in terms of data sharing needs, budgets and timelines.

- The IDS assurance level 1 - checklist approach - touches basic requirements for participant to take part in IDS. It is prepared by participants themselves and allows them to prove that their IDS Connector is interoperable with other connectors in the data space and goes along with an IDS Reference Architecture Model.
- The IDS assurance level 2 - concept review certification - serves as a proof that participant's IDS component meets all requirements in terms of security and functionality. This level of certification is done by an independent third party, which is IDS evaluation facilities. An external concept review implies the examination of documentation along with actual testing of the connector. This review is performed to make sure that the implementation of both functional and security requirements is correct.

Operational environment certification

This evaluation is made in order to provide an assessment of the overall trustworthiness of the processes, physical environment and organisational rules. Operational environment certification allows small companies to try out data spaces participation. It is prepared by participants themselves by filling the questionnaire provided on IDS Certification Portal [73], and includes the self-assessment, that checks if basic requirements for participating in IDS are met.

Taking into consideration all the above, the next execution step involves evaluating how the type of certification process mentioned can be applied to the components that form the OMEGA-X architecture.

5 Conclusions

This report documents the work performed under WP3 tasks *T3.2 Definition of analytic services to be used as baseline for use case demonstration* and *T3.3 Analysis of requirements for interoperability, security, privacy and data sovereignty and certification*. These tasks aim to identify the energy services that will benefit from the OMEGA-X Data Space and explore the requirements of the project's platform.

Regarding federated infrastructures, data spaces and data sovereignty and privacy, the contribution of this deliverable is determining a set of requirements that the OMEGA-X platform must satisfy. These results are summarised in Table 9, included in Annex A: Data Space Requirements. Table 9 collects all the identified functional and non-functional requirements, as well as some suggested open-source technologies that can cover them. It also includes some comparisons on their main advantages against other competitors.

Regarding data analytic services, a total of 33 services have been defined, which are classified into the four use case families of the project (as shown in

Table 8). This document describes the services, specifying how they can benefit from OMEGA-X Data Space. It also identifies the input and output data, the involved stakeholders, the techniques they use, and the ownership of each of them.

Table 8. OMEGA-X data analytic services and Use Case Families

Use case family	Data analytic service
Renewables	Predictive Maintenance for large PV plants
	Benchmarking
	Compare actual production versus expected
	PV Cleaning Advisor
	Shading Analysis.
	Tracking algorithm check
	Detect t measurement errors
	Detect non-technical losses
	Congestion detection
	Detection of the volatility of voltage in grids with high renewable penetration
	Plan grid reinforcements for future renewable scenarios
	Energy Generation Forecast
	Digital Twin BIPV Self Consumption Systems in Buildings

Use case family	Data analytic service
Local Energy Communities	Local Energy Communities designer
	Gamification for electrical energy savings
	Thermal losses detection and benchmarking at LEC level
	Water losses detection and benchmarking at LEC level
	Estimate the probability of congestions
	Electrical losses detection and benchmarking at LEC level
	Reinforcement plan of LECs for future renewable Integration
	Optimizing self-consumption of renewable energy at LEC level
	Optimizing sharing coefficients in collective self-consumption
	Planning services
Electromobility	Roaming of booking services
	Roaming of self-consumption
Flexibility	Grid observability and network analysis
	Grid validation platform, real-time
	Flexibility platform for DER connection, planning
	Passive consumption baseline prediction
	Active consumption prediction
	Intermittent DER generation resource baseline prediction
	Prosumer EMS internal optimization
	Flexibility order disaggregation
	Aggregated flexibility offer optimization

At this phase of the project, the level of maturity of the 33 services is not uniform. Hence, the degree of detail of each service differs from one to another. However, this deliverable provides sufficient information to enable and foster interaction between service developers and pilots, which is essential to develop the services within tasks *T5.1 Development of analytic services* and *T5.2 Digital Twins* from WP5 – Data Space Marketplaces.

The algorithms described in this document and developed in WP5 will be finally tested in the pilots of the different UCFs, within tasks *T6.2 Renewables Use case family*, *T6.3 Energy communities and sector integration Use case family*, *T6.4 Collaboration among Electromobility actors Use case family* and *T6.5 Flexibility Use case family* from WP6 – Demonstration.

In the ongoing (T3.1 and T3.4) and subsequent (T3.5) tasks of the WP3, the focus will lie on the definition of the System Use Cases (SUCs) and the Business Use Cases (BUCs), starting from the services defined in this document. This activity will be carried on by *T3.1 Use case identification*. Also, requirements related to interoperability, security, privacy and data sovereignty will provide the foundation for *T3.4 Full system architecture and building blocks activities*, and a set of validation metrics (KPIs) will be defined in task *T3.5 KPI elicitation for use case validation*, which will enable the assessment of the use cases (T3.1) and reference architecture (T3.4). These three tasks (T3.1, T3.4 and T3.5) will report their first results in the deliverable OMEGA-X_*D3.1 Use Cases and Architecture living report*. *First release* [M12] (April 2023)..

6 References

[1] "REGULATION (EU) 2016/679," *Official Journal of the European Union,* 2016.

[2] "A European strategy for data," February 2020. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0066. [Accessed October 2022].

[3] "Mastodon Social networking that's not for sale.," [Online]. Available: https://joinmastodon.org/. [Accessed October 2022].

[4] "EU Voice," [Online]. Available: https://social.network.europa.eu/public.

[5] "EU Video," [Online]. Available: https://tube.network.europa.eu/. [Accessed October 2022].

[6] "[Matrix] An open network for secure, decentralized communication," [Online]. Available: https://matrix.org/. [Accessed October 2022].

[7] "NGSI-LD," [Online]. Available: https://ngsi-ld-tutorials.readthedocs.io/en/latest/.

[8] OPENDEI, "Design Principles for Data Spaces – Position Paper," [Online]. Available: https://design-principles-for-data-spaces.org/.

[9] "Open Source Identity and Access Management," [Online]. Available: https://www.keycloak.org/. [Accessed October 2022].

[10] "European Commission, Data protection in the EU," [Online]. Available: https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en.

[11] "European Commission, A European Green Deal," [Online]. Available: https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en.

[12] "IDSA," [Online]. Available: https://internationaldataspaces.org.

[13] "Gaia-X," [Online]. Available: https://gaia-x.eu.

[14] "Reference Architecture Model," 2019. [Online]. Available: https://internationaldataspaces.org//wp-content/uploads/IDS-Reference-Architecture-Model-3.0-2019.pdf. [Accessed October 2022].

[15] "GAIA-X: A Pitch Towards Europe Status. Report on User Ecosystems and Requirements," 2020. [Online]. Available: https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Publications/gaia-x-a-pitch-towards-europe.pdf?__blob=publicationFile&v=6. [Accessed October 2022].

[16] “GAIA-X and IDS,” 2021. [Online]. Available: https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Position-Paper-GAIA-X-and-IDS.pdf. [Accessed October 2022].

[17] R. M. M. M. M. A. Irakli Nadareishvili, Microservice Architecture - Aligning principles, practices, and culture, O'Reilly Media, Inc., 2016.

[18] “Docker Develop faster. Run anywhere.,” [Online]. Available: https://www.docker.com/. [Accessed October 2022].

[19] “FIWARE Foundation,” [Online]. Available: https://www.fiware.org/foundation/.

[20] “Technical Convergence - Discussion Document,” Data Spaces Business Alliance, 2022.

[21] “The OAuth 2.0 Authorization Framework,” October 2012. [Online]. Available: https://www.rfc-editor.org/rfc/rfc6749. [Accessed October 2022].

[22] “ Security Assertion Markup Language (SAML) 2.0 Profile for OAuth 2.0 Client Authentication and Authorization Grants,” May 2015. [Online]. Available: https://www.rfc-editor.org/rfc/rfc7522. [Accessed October 2022].

[23] “OpenID Connect Specifications,” [Online]. Available: https://openid.net/developers/specs/. [Accessed October 2022].

[24] “Self-Issued OpenID Provider v2,” September 2022. [Online]. Available: https://openid.net/specs/openid-connect-self-issued-v2-1_0.html. [Accessed October 2022].

[25] I. E. Commission, “IEC TR 62559-1:2019. Use Case methodology - Part 1: Concept and processes in standardization,” 29 January 2019. [Online]. Available: https://webstore.iec.ch/publication/32220.

[26] “Spark Apache,” [Online]. Available: https://spark.apache.org/.

[27] “BIPVBOOST H2020,” [Online]. Available: https://bipvboost.eu/.

[28] E. Ociepa, M. Mrowiec and I. Deska, “Analysis of Water Losses and Assessment of Initiatives Aimed at Their Reduction in Selected Water Supply Systems,” *MDPI - Water,* Vols. 11, 1037, no. Water Resources Management, Policy and Governance, 2019.

[29] J. P. a. R. Sandhu, “The ucon abc usage control model,” *ACM Transactions on Information and System Security,* vol. 7, no. 1, pp. 128-174, 2004.

[30] H. G. E. V. D. a. L. B. S. M. J. Atallah, “Secure supply-chain protocols,” *EEE International Conference on E-Commerce,* pp. 293-302, 2003.

[31] S. B. H. S. E. D. S. K. M. P. M. Dalmolen, “Maintaining control over sensitive data in the Physical Internet: Towards an open, service oriented, network-model for infrastructural data sovereignty,” in *6th International Physical Internet Conference (IPIC)*, London, 2019.

[32] I. Fraunhofer, “Usage Control in the International Data Spaces,” March 2001. [Online]. Available: https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Position-Paper-Usage-Control-in-the-IDS-V3..pdf.

[33] O. P. Agent, “Policy Language - What is Rego?,” [Online]. Available: https://www.openpolicyagent.org/docs/latest/policy-language/.

[34] “How does FileVault encryption work on a Mac?,” [Online]. Available: https://support.apple.com/en-gb/guide/mac-help/flvlt001/12.0/mac/12.0.

[35] “BitLocker Windows Security,” [Online]. Available: https://learn.microsoft.com/en-us/windows/security/information-protection/bitlocker/bitlocker-overview.

[36] Trellix, “Complete Data Protection,” [Online]. Available: https://www.trellix.com/en-us/products/complete-data-protection.html .

[37] “VeraCrypt,” [Online]. Available: https://www.veracrypt.fr/en/Home.html .

[38] “Linux Unified Key Setup,” [Online]. Available: https://gitlab.com/cryptsetup/cryptsetup/ .

[39] “Linux endpoint protection with SecureDoc,” [Online]. Available: https://winmagic.com/en/products/linux-endpoint-protection/ .

[40] “CryFS: A cryptographic filesystem for the cloud,” [Online]. Available: https://www.cryfs.org/ .

[41] “AxCrypt – File Security Made Easy,” [Online]. Available: https://axcrypt.net/ .

[42] “Enabling Amazon S3 Default Bucket Encryption,” [Online]. Available: https://docs.aws.amazon.com/AmazonS3/latest/userguide/default-bucket-encryption.html .

[43] “Azure Storage encryption for data at rest,” [Online]. Available: https://learn.microsoft.com/en-us/azure/storage/common/storage-service-encryption .

[44] “Default encryption at rest,” [Online]. Available: https://cloud.google.com/docs/security/encryption/default-encryption .

[45] “SFTP protocol,” [Online]. Available: https://www.ssh.com/academy/ssh/sftp .

[46] “The Transport Layer Security (TLS) Protocol Version 1.3,” [Online]. Available: https://www.rfc-editor.org/rfc/rfc8446 .

[47] NIST, “Advanced Encryption Standard (AES),” [Online]. Available: https://www.nist.gov/publications/advanced-encryption-standard-aes .

[48] “Lightweight Cryptography Applicable to Various IoT Devices,” [Online]. Available: https://www.nec.com/en/global/techrep/journal/g17/n01/170114.html .

[49] M. Rana, Q. Mamun and R. Islam, “Lightweight cryptography in IoT networks: A survey,” *Future Generation Computer Systems,* vol. 129, pp. 77-89, 2022.

[50] “What is a 3-2-1 Backup strategy?,” [Online]. Available: https://www.seagate.com/gb/en/blog/what-is-a-3-2-1-backup-strategy/ .

[51] “Backup and restore in Windows,” [Online]. Available: https://support.microsoft.com/en-us/windows/backup-and-restore-in-windows-352091d2-bb9d-3ea3-ed18-52ef2b88cbef .

[52] “How to backup your Mac,” [Online]. Available: https://support.apple.com/mac-backup .

[53] “rsnapshot,” [Online]. Available: https://rsnapshot.org/ .

[54] “rdiff-backup,” [Online]. Available: https://rdiff-backup.net/ .

[55] “Certified Data Erasure Software for Data Destruction,” [Online]. Available: https://www.bitraser.com/.

[56] “The Most Trusted for Data Erasure,” [Online]. Available: https://www.wipedrive.com/ .

[57] EC, “What does data protection ‘by design’ and ‘by default’ mean?,” [Online]. Available: https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/obligations/what-does-data-protection-design-and-default-mean_en .

[58] CNIL, “GDPR toolkit,” [Online]. Available: https://www.cnil.fr/en/gdpr-toolkit .

[59] “GDPR compliance checklist,” [Online]. Available: https://gdpr.eu/checklist/ .

[60] F. G.-X. a. I. BDV, “Data Spaces Business Alliance: Unleasing the Data Economy,” 26 09 2022. [Online]. Available: https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/Data-Spaces-Business-Alliance-Technical-Convergence.pdf.

[61] “Verifiable Presentation Request v0.2,” September 2022. [Online]. Available: https://w3c-ccg.github.io/vp-request-spec/. [Accessed October 2022].

[62] “2021 Global data management research,” 2021. [Online]. Available: https://edqcdncmsprod.azureedge.net/491de5/globalassets/white-papers/global-data-management-research-2021.pdf. [Accessed October 2022].

[63] “6 Data Quality Metrics You Can't Afford To Ignore,” 2020. [Online]. Available: https://www.cloverdx.com/blog/6-data-quality-metrics-you-cant-ignore. [Accessed October 2022].

[64] “8 dimensions of data quality,” 2022. [Online]. Available: https://www.cloverdx.com/blog/8-dimensions-data-quality. [Accessed October 2022].

[65] “What is Data Security?,” [Online]. Available: https://www.forcepoint.com/cyber-edu/data-security. [Accessed October 2022].

[66] M. A. E. K. A. a. E. N. Alha, “Big Data: Trade-off between Data Quality and Data Security,” *Procedia Computer Science,* vol. 151, pp. 916-922, 2019.

[67] "Cybersecurity Solutions for a Riskier World," Thoughtlabgroup.com, 2022. [Online]. Available: https://thoughtlabgroup.com/cyber-solutions-riskier-world/. [Accessed 19 September 2022].

[68] "US Reliability Standards," [Online]. Available: https://www.nerc.com/pa/Stand/Pages/USRelStand.aspx. [Accessed 19 September 2022].

[69] "NIST Cybersecurity Framework," NIST, [Online]. Available: https://www.nist.gov/cyberframework. [Accessed 19 September 2022].

[70] "IEC 62351:2022 SER - Power systems management and associated information exchange - Data and communications security," 17 August 2022. [Online]. Available: https://webstore.iec.ch/publication/6912. [Accessed 19 September 2022].

[71] "IEEE Standard Cybersecurity Requirements for Substation Automation, Protection, and Control Systems," 2015. [Online]. Available: https://ieeexplore.ieee.org/document/7024885. [Accessed 19 September 2022].

[72] "Cybersecurity Capability Maturity Model (C2M2)," Office of Cybersecurity, Energy Security, and Emergency Response, 2022. [Online]. Available: https://www.energy.gov/ceser/cybersecurity-capability-maturity-model-c2m2. [Accessed 19 September 2022].

[73] "Certification - International Data Spaces," International Data Spaces, [Online]. Available: https://internationaldataspaces.org/use/certification/. [Accessed 19 September 2022].

[74] FIWARE, "NGSI-LD Step-By-Step," [Online]. Available: https://ngsi-ld-tutorials.readthedocs.io/en/latest/.

Annex A: Data Space Requirements

Table 9. OMEGA-X Data Space Requirements

Requirement	Type	Description	Proposed Technology	Advantages
GDPR Compliance	Non-Functional	OMEGA-X Data Space must be compliant with the current European regulation in terms of data protection and privacy.		
Decentralisation	Non-Functional	OMEGA-X Data Space must be able to be deployed in multiple sites, clouds or locations.		
Interoperability	Non-Functional	OMEGA-X Data Space must be able to share data and services with other Data Spaces.		
Standard Communication Channels	Non-Functional	OMEGA-X Data Space must offer standard communication channels for input/output.	RESTful APIs, Apache Kafka, MQTT, AMQP…	In this case multiple technologies shall coexist together each one providing their connection capabilities. No need to select one over the others as the greater number of technologies the more interconnection capabilities the platform shall have. (This initial set of technologies might be extended in the future according to the new requirements)
Data Accessibility	Non-Functional	OMEGA-X Data Space must facilitate access to data it contains.	NGSI-LD	NGSI-LD promoted by the FIWARE foundation [74] looks like a promising solution as it supports linked data, relationships and semantics. Moreover, it was standardised by ETSI.

Requirement	Type	Description	Proposed Technology	Advantages
Content Catalogue	Functional	OMEGA-X Data Space must offer data and service discovery capabilities in the form of content catalogues	To be decided	Due to the early stage of development the existing solution might be prone to change and is better to delay this decision.
Open-Source Solution	Non-Functional	OMEGA-X Data Space source code must be available to everyone under open-source license.	To be decided (GitHub, GitLab, Bitbucket, etc)	Source code must be stored in a public code repository. There are multiple equivalent alternatives (all based on git) that can be used in the project.
Data Sharing	Functional	OMEGA-X Data Space must offer data sharing capabilities.	To be decided	Due to the early stage of development the existing data-space connectors might be prone to change and is better to delay this decision.
Service Sharing	Functional	OMEGA-X Data Space must offer service sharing capabilities.	To be decided	Due to the early stage of development the existing data-space connectors might be prone to change and is better to delay this decision.
Data Sovereignty & Governance	Functional	OMEGA-X Data Space must ensure data sovereignty and governance to data owners.		
Privacy	Functional	OMEGA-X Data Space must ensure data privacy.		
Security	Functional	OMEGA-X Data Space must ensure data security.		
Trustworthiness	Functional	OMEGA-X Data Space must offer protocols in order to be trusted by data owners and users.		

Requirement	Type	Description	Proposed Technology	Advantages
Value Creation	Non-Functional	OMEGA-X Data Space must use stored data to create value as new services.		
Microservice Architecture	Non-Functional	OMEGA-X Data Space must follow microservice architecture rather than monolithic approach.		
Containerisation	Non-Functional	OMEGA-X Data Space must take advantage of containerisation technologies to build their services.	Docker	Although there are some alternatives to Docker it remains as the industry standard and the more spread and sensible solution.
Orchestration	Non-Functional	OMEGA-X Data Space must take advantage of orchestration technologies to automate shipping and maintaining the different services.	Kubernetes	Kubernetes is the industry standard and it supports multicloud deployments both for public and private clouds.
User Identification	Functional	OMEGA-X Data Space must be able to identify users following strong protocols.	Keycloak	Keycloak is an open-source identity manager that simplifies managing all the protocols needed for user identification and role-based access control: OpenID Connect, OAuth 2.0, SAML…
Role-Based Access Control	Functional	OMEGA-X Data Space must be able to limit data and service access to users based on a set of predefined roles	Keycloak	Keycloak is an open-source identity manager that simplifies managing all the protocols needed for user identification and role-based access control: OpenID Connect, OAuth 2.0, SAML…

Requirement	Type	Description	Proposed Technology	Advantages
Data Quality Metrics	Non-Functional	OMEGA-X Data Space must be able to grade data quality based on some predefined metrics		
Data Confidentiality	Functional	OMEGA-X Data Space must ensure data confidentiality using strong cryptographic algorithms and protocols	To be decided	Cipher suites and protocols that ensure state-of-the-art data protection must be selected. There shall be different symmetric and asymmetric protocols to maximise security and interoperability.
Data Integrity	Functional	OMEGA-X Data Space must ensure data integrity using strong cryptographic algorithms and protocols		
Data Availability	Functional	OMEGA-X Data Space must ensure data availability minimising the downtimes and optimising disaster recovery.	Microservice architecture, Docker, Kubernetes, etc.	To ensure data availability the data space must be stable and scalable. Thus, the selected architecture must follow microservice paradigms and take advantage of state-of-the-art containerization and orchestration technologies.

Annex B: Service Templates

B.1 Renewables UCF Services

B.1.1 Predictive Maintenance for large PV plants

Service to be offered in real-time/offline:

Preferably online but can also be run offline.

Execution frequency: Weekly, monthly or on-demand.

Input data requirements:

Table 10. Execution input data for Predictive Maintenance for large PV plants service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
DC and AC voltages	< 15 min	Monitoring Information from inverters and module level electronics – Raw Data.	V	< 1 day	ENGIE, EDF, EyPESA	R
DC and AC currents	< 15 min		A	< 1 day	ENGIE, EDF, EyPESA	R
Internal temperature	< 15 min		ºC	< 1 day	ENGIE, EDF, EyPESA	R
Operating modes	< 15 min		N/A	< 1 day	ENGIE, EDF, EyPESA	O
Alarms	< 15 min		N/A	< 1 day	ENGIE, EDF, EyPESA	O
Global Irradiance on PoA (Plane of Array)	< 15 min	Meteorological Information– Raw Data.		< 1 day	ENGIE, EDF, EyPESA	R
Direct Irradiance (dni), diffuse irradiance above the horizontal (dhi), global irradiance above the horizontal (ghi)	<15 min			< 1 day	ENGIE, EDF, EyPESA	O

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Ambient and Module temperature	< 15 min			< 1 day	ENGIE, EDF, EyPESA	R
Rainfall events	< 15 min			< 1 day	ENGIE, EDF, EyPESA	O
Wind speed and direction	< 15 min			< 1 day	ENGIE, EDF, EyPESA	O

Table 11. Training data for Predictive Maintenance for large PV plants service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
DC and AC voltages	< 15 min	Monitoring Information from inverters and module level electronics – Raw Data.	V	≥ 1month	ENGIE; EDF and EyPESA	R
DC and AC currents	< 15 min		A	≥ 1month	ENGIE; EDF and EyPESA	R
Internal temperature	< 15 min		ºC	≥ 1month	ENGIE; EDF and EyPESA	R
Operating modes	< 15 min		N/A	≥ 1month	ENGIE; EDF and EyPESA	O
Alarms	< 15 min		N/A	≥ 1month	ENGIE; EDF and EyPESA	O

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Global Irradiance on PoA (Plane of Array)	< 15 min	Meteorological Information– Raw Data.		≥ 1month	ENGIE; EDF and EyPESA	R
Direct Irradiance (dni), diffuse irradiance above the horizontal (dhi), global irradiance above the horizontal (ghi)	< 15 min			≥ 1month	ENGIE; EDF and EyPESA	O
Ambient and Module temperature	< 15 min			≥ 1month	ENGIE; EDF and EyPESA	R
Rainfall events	< 15 min			≥ 1month	ENGIE; EDF and EyPESA	O
Wind speed and direction	<15 min			≥ 1month	ENGIE; EDF and EyPESA	O
Maintenance Logbook	< 15 min	Maintenance Logbook with conducted maintenance actions	N/A	≥ 1 month	ENGIE; EDF and EyPESA	O
Latitude and longitude	Static	Coordinates of the PV installation		Static	ENGIE; EDF and EyPESA	R

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Number of modules per string and strings per array	Static	Number of modules per string and strings per array connected to each inverter input (in series and in parallel).		Static	ENGIE; EDF and EyPESA	R
Characteristic parameters of the PV modules	Static	Characteristic parameters available in the module manufacturer's technical sheet: Voc, Isc, Vmp Imp, temperature coefficients, the number of cells in series for the calculation of the ideality factor and the type of module and assembly for the thermal model of the temperature of operation		Static	ENGIE; EDF and EyPESA	R
Characteristic parameters of the PV inverters	Static	The maximum power point voltage operating range for each input section of the MPPT. Maximum current and power values for each input section of the MPPT.		Static	ENGIE; EDF and EyPESA	R
Position of the irradiance sensor	Static	The elevation and orientation of the irradiance sensor		Static	ENGIE; EDF and EyPESA	R

Expected outputs:

Table 12. Output data for Predictive Maintenance for large PV plants service.

Data	Lowest level of granularity	Description	Units
PV Characterization	< 15 min	Characterization of PV modules, PV arrays and inverters performance parameters.	Not Applicable (NA)
Alarms	< 15 min	List of triggered alarms in the case these performance parameters present a significant deviation. For each case, the affected component or subsystem will be identified, the specific failure modes related to this underperformance will be reported and the potential causes will be exposed	NA

Applied Techniques:

The AI techniques applied in this service are:

- Data Mining and Feature Engineering.
- Clustering algorithms to identify different operating conditions.
- Regression algorithms.

Open-source/Proprietary software: Proprietary offered as a service.

B.1.2 Benchmarking Service

Service to be offered in real-time/offline:

Preferably online but can also be run offline.

Execution frequency: Weekly, monthly or on-demand.

Input data requirements:

Table 13. Execution input data for Benchmarking Service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
DC and AC voltages	< 15 min	Monitoring Information from inverters and module level electronics – Raw Data.	V	< 1 day	ENGIE, EDF, EyPESA	R
DC and AC currents	< 15 min		A	< 1 day	ENGIE, EDF, EyPESA	R
Internal temperature	< 15 min		ºC	< 1 day	ENGIE, EDF, EyPESA	R
Operating modes	< 15 min		N/A	< 1 day	ENGIE, EDF, EyPESA	O
Alarms	< 15 min		N/A	< 1 day	ENGIE, EDF, EyPESA	O
Global Irradiance on PoA (Plane of Array)	< 15 min	Meteorological Information– Raw Data.		< 1 day	ENGIE, EDF, EyPESA	R
Direct Irradiance (dni), diffuse irradiance above the horizontal (dhi), global irradiance above the horizontal (ghi)	< 15 min			< 1 day	ENGIE, EDF, EyPESA	O

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Ambient and module temperature	< 15 min			< 1 day	ENGIE, EDF, EyPESA	R
Rainfall events	< 15 min			< 1 day	ENGIE, EDF, EyPESA	O
Wind speed and direction	< 15 min			< 1 day	ENGIE, EDF, EyPESA	O

Table 14: Training data for Benchmarking Service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
DC and AC voltages	< 15 min	Monitoring Information from inverters and module level electronics – Raw Data.	V	≥ 1year	ENGIE; EDF and EyPESA	R
DC and AC currents	< 15 min		A	≥ 1year	ENGIE; EDF and EyPESA	R
Internal temperature	< 15 min		ºC	≥ 1year	ENGIE; EDF and EyPESA	R
Operating modes	< 15 min		N/A	≥ 1year	ENGIE; EDF and EyPESA	O
Alarms	< 15 min		N/A	≥ 1year	ENGIE; EDF and EyPESA	O

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Global Irradiance on PoA (Plane of Array)	< 15 min	Meteorological Information– Raw Data.		≥ 1year	ENGIE; EDF and EyPESA	R
Direct Irradiance (dni), diffuse irradiance above the horizontal (dhi), global irradiance above the horizontal (ghi)	< 15 min			≥ 1year	ENGIE; EDF and EyPESA	O
Ambient and Module temperature	< 15 min			≥ 1year	ENGIE; EDF and EyPESA	R
Rainfall events	< 15 min			≥ 1year	ENGIE; EDF and EyPESA	O
Wind speed and direction	< 15 min			≥ 1year	ENGIE; EDF and EyPESA	O
Maintenance Logbook	< 15 min	Maintenance Logbook with conducted maintenance actions	N/A	≥ 1year	ENGIE; EDF and EyPESA	O
Latitude and longitude	Static	Coordinates of the PV installation		Static	ENGIE; EDF and EyPESA	R

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Number of modules per string and strings per array	Static	Number of modules per string and strings per array connected to each inverter input (in series and in parallel).		Static	ENGIE; EDF and EyPESA	R
Characteristic parameters of the PV modules	Static	Characteristic parameters available in the module manufacturer's technical sheet: Voc, Isc, Vmp Imp, temperature coefficients, the number of cells in series for the calculation of the ideality factor and the type of module and assembly for the thermal model of the temperature of operation		Static	ENGIE; EDF and EyPESA	R
Characteristic parameters of the PV inverters	Static	The maximum power point voltage operating range for each input section of the MPPT. Maximum current and power values for each input section of the MPPT.		Static	ENGIE; EDF and EyPESA	R
Position of the irradiance sensor	Static	The elevation and orientation of the irradiance sensor		Static	ENGIE; EDF and EyPESA	R

Expected outputs:

Table 15: Output data for Benchmarking Service.

Data	Lowest level of granularity	Description	Units
PV Characterization	< 15 min	Characterization of PV modules, PV arrays and inverters performance parameters between different PV plants.	NA

Applied Techniques:

The AI techniques applied in this service are:

- Data Mining and Feature Engineering.
- Clustering algorithms to identify different operating conditions.
- Regression Algorithms.
- Hybrid Digital Twins for KPI generation.
- Visual Analytics.

Open-source/Proprietary software: Proprietary offered as a service.

B.1.3 Compare actual production versus expected

Service to be offered in real-time/offline:

Online.

Execution frequency: daily.

Input data requirements:

Table 16: Execution input data for Compare actual production versus expected service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Energy	Hourly or less Recom: 10 min	Sum of energy on the granularity period	kWh	> 1 year	EDF/ENGIE	R
Global Horizontal Irradiance (GHI)	Hourly or less Recom: 10 min	Global Horizontal Irradiance measured by a meteorological station on site	W/m²	> 1 year	EDF/ENGIE	R

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Diffuse Horizontal Irradiance (DHI)	Hourly or less Recom: 10 min	Diffuse Horizontal Irradiance measured by a meteorological station on site	W/m²	> 1 year	EDF/ENGIE	R
Ambient temperature	Hourly or less Recom: 10 min	Ambient temperature measured by a meteorological station on site	ºC	> 1 year	EDF/ENGIE	R
Wind speed	Hourly or less Recom: 10 min	Wind speed measured by a meteorological station on site	m/s	> 1 year	EDF/ENGIE	R
Soiling	Daily	Soiling measure	%	> 1 year	EDF/ENGIE	R
Availability	Daily	Plant availability considering Strings, power stations and transformers	%	> 1 year	EDF/ENGIE	R
Grid restrictions	Hourly or less Recom: 10 min	Grid restrictions to be considered	MW	> 1 year	EDF/ENGIE	R

Table 17. Training data for Compare actual production versus expected service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Plant layout UTM coordinates	As built	Coordinates of the PV fixed structure/tracker	M		EDF/ENGIE	R
PVsyst model	As built	Simulation model with main characteristics			EDF/ENGIE	R

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Energy	Hourly or less Recom: 10 min	Sum of energy on the granularity period	kWh	> 1 year	EDF/ENGIE	R
Power per string	Hourly or less Recom: 10 sec	Average power per string on the granularity period	kW	> 1 year	EDF/ENGIE	R
Soiling	Daily	Soiling measure	%	> 1 year	EDF/ENGIE	R
Availability	Daily	Plant availability considering Strings, power stations and transformers	%	> 1 year	EDF/ENGIE	R
Grid restrictions	Hourly or less Recom: 10 min	Grid restrictions to be considered	MW	> 1 year	EDF/ENGIE	R
Degradation	Daily	Plant cumulated degradation	%	> 1 year	EDF/ENGIE	O
Intensity and voltage per string	Hourly or less Recom: 10 sec	Intensity and voltage per string on the granularity period	A and V	> 2 months	EDF/ENGIE	O
Tracking mode	Hourly or less Recom: 10 min	Tracking or back-tracking mode specified for tracker if different	-	> 1 year	EDF/ENGIE	O
Global Horizontal Irradiance (GHI)	Hourly or less Recom: 10 min	Global Horizontal Irradiance measured by a meteorological station on site	W/m²	> 1 year	EDF/ENGIE	R

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Diffuse Horizontal Irradiance (GHI)	Hourly or less Recom: 10 min	Diffuse Horizontal Irradiance measured by a meteorological station on site	W/m²	> 1 year	EDF/ENGIE	R
Ambient temperature	Hourly or less Recom: 10 min	Ambient temperature measured by a meteorological station on site	ºC	> 1 year	EDF/ENGIE	R
Wind speed	Hourly or less Recom: 10 min	Wind speed measured by a meteorological station on site	m/s	> 1 year	EDF/ENGIE	R

Expected outputs:

Table 18. Output data for Compare actual production versus expected service.

Data	Granularity	Description	Units
Expected performance of the plant	Hourly or less	Sum of energy on the granularity period	kWh
Daily cumulated production	Daily	Cumulated energy of the simulation day	kWh
Deviation	Daily	% of deviation between the simulation and the real energy generation	%

Applied Techniques:

The techniques applied in this service are:

- Performance Modelling of PV plants.
    - Precise determination of annual Plant Performance, considering:
        - Actual shadowing, via ray-tracing methods.
        - I-V curves behaviour of cells+panels+strings+inverters.
- Fine tuning of PV plant model to represent real life behaviour.
- Ray tracing techniques.

- Data analytics.
- Regression algorithms.

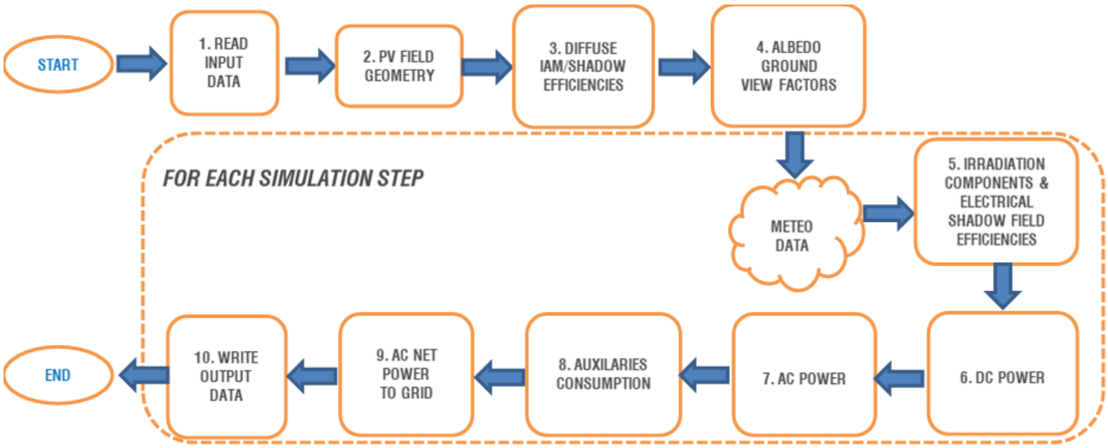

Figure 4. Service Pipeline. Source: SENER ING.

Open-source/Proprietary software: Proprietary software delivered through an API service.

Additional Information: Only two plants are simulated.

B.1.4 PV Cleaning Advisor

Service to be offered in real-time/offline:

Offline.

Execution frequency: daily.

Input data requirements:

Table 19. Execution input data for PV Cleaning Advisor service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Energy per string	10 min	Sum of energy on the granularity period	kWh	> 1 year	EDF/ENGIE	R
Clean string energy	10 min	Sum of energy on the granularity period	W/m²	> 1 year	EDF/ENGIE	R

Table 20. Training data for PV Cleaning Advisor service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Plant layout UTM coordinates	As built	Coordinates of the PV fix structure/tracker	M		EDF/ENGIE	R
Cleaning loss	As built	Cleaning loss % below the cleaning is recommended	%		EDF/ENGIE	R

Expected outputs:

Table 21. Output data for PV Cleaning Advisor service.

Data	Granularity	Description	Units
Soiling loss per string	Hourly or less	Energy loss on the granularity period	kWh
Soiling loss	Daily	Cumulated energy loss due to soiling	kWh
Cleaning advise	Daily	Cleaning signal for string	-

Applied Techniques:

The techniques applied in this service are:

- Advanced data analytics.

Open-source/Proprietary software: Proprietary software delivered through an API service.

B.1.5 Shading Analysis

Service to be offered in real-time/offline:

Offline.

Execution frequency: daily.

Input data requirements:

Table 22. Execution input data for Shading Analysis service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Energy per string	10 min	Sum of energy on the granularity period	kWh	> 1 year	EDF/ENGIE	R
Unshaded string energy	10 min	Sum of energy on the granularity period	W/m²	> 1 year	EDF/ENGIE	R

Table 23. Training data for Shading Analysis service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Plant layout UTM coordinates	As built	Coordinates of the PV fix structure/tracker	M		EDF/ENGIE	R

Expected outputs:

The outputs of the service:

Table 24. Output data for Shading Analysis service.

Data	Granularity	Description	Units
Unshaded string performance	Hourly or less	Sum of energy on the granularity period	kWh
Shaded string performance	Hourly or less	Sum of energy on the granularity period	kWh
Shading loss per string	Daily	% of performance loss estimate due to shading	%

Applied Techniques:

The techniques applied in this service are:

- Advanced data analytics

Open-source/Proprietary software: Proprietary software delivered through an API service.

B.1.6 Tracking algorithm check

Service to be offered in real-time/offline:

Offline.

Execution frequency: daily.

Input data requirements:

Table 25. Execution input data for Tracking algorithm check service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Tracking angle of the trackers	1min or less	Tilt angle of the tracker. Reference: (+) east, (-) west	º	> 1 year	EDF/ENGIE	R
Tracking mode	1min or less	Tracking or back-tracking mode	-	> 1 year	EDF/ENGIE	R

Table 26. Training data for Tracking algorithm check service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Plant location	As built	Latitude and longitude of the plant	º		EDF/ENGIE	R
Plant layout UTM coordinates	As built	Coordinates of the PV fix structure/tracker	M		EDF/ENGIE	R

Expected outputs:

The outputs of the service:

Table 27. Output data for Tracking algorithm check service.

Data	Granularity	Description	Units
Tracking accuracy	Hourly or less	Plant tracking compared to the theorical tracking considering the tracking mode on the granularity period	%

Data	Granularity	Description	Units
Performance loss	Hourly or less	Production loss estimate with respect to tracking error	kWh
Daily tracking deviation	Daily	Daily percentage of deviation of the trackers	%

Applied Techniques:

The techniques applied in this service are:

- Astronomic tracking algorithm.
- Flat terrain backtracking for the avoidance of shadows.
- Advanced analytics.

Open-source/Proprietary software: Proprietary software delivered through an API service.

B.1.7 Detect measurement errors

Service to be offered in real-time/offline:

Offline and online. Execution frequency: on-demand. When a service requires to clean its input data, this service will be requested.

Input data requirements:

The inputs depend on the service that contracts this cleaning service. They must be time series with at least one datetime column.

Expected outputs:

- Clean dataset with no errors nor missing values. (Excel, csv file).
- Outliers report, txt file (if requested by the Service User).
- Missing data report, txt file (if requested by the Service User).

Applied Techniques:

Missing imputation service:

The missing values should be replaced with rational records in order to offer the data-driven services a complete dataset. The approach of handling missing values is called Imputation. Several Imputation techniques are applied and offered in this service, ranging from Statistics to Artificial Intelligence.

- Statistic methods:
  - Univariate imputation: The following methods are used as the imputation value for the Statistical estimation of the missing data.
    - Mean.
    - Median.
    - Mode.

    - Zero values.
    - Most frequent value (categorical data).
  - Interpolation: Interpolation is a mathematical method that adjusts a function to the dataset and uses this function to extrapolate the missing data. The simplest type is the linear interpolation, but polynomial is also available in DME, indicating the degree.
    - Linear.
    - Polynomial (order quadratic, cubic...).
- Machine Learning methods
  - Multivariate Imputation: For this imputation technique, a distributed set of observed data is used to estimate a set of imputation values for the missing data. In this method, a Multiple Imputation by Chained Equations (MICE method) is applied.
  - Random Forest.
  - KNNeighbours.
  - Bayesian Ridge.

Outliers detection:

This service offers two methods to detect outliers.

- Statistic methods
  - IQR method: The Interquartile Range (IQR) is often used to detect outliers in data, following these steps:
    - First, calculate the interquartile range for the data.
    - Multiply the interquartile range (IQR) by a parameter set up by the service used. Usually, the value 1.5 is set to discern outliers.
    - Add 1.5 x (IQR) to the third quartile (Q3). Any number more significant than this is considered an outlier.
    - Subtract 1.5 x (IQR) from the first quartile (Q1). Any number less than this is considered an outlier.
  - Mean-std difference method: This method deletes the outliers that are above and below the mean value minus the standard deviation, multiplied by a parameter:

$$\text{Outlier} = (\text{mean} \pm \text{std}) * \text{Constant parameter}$$

Open-source/Proprietary software: Open-source.

B.1.8 Detect non-technical losses

Service to be offered in real-time/offline:

Offline. Execution frequency: Weekly, monthly or on-demand.

Input data requirements:

Table 28. Execution input data for Detect non-technical losses service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
SM Active Power	Hourly	Hourly active power recorded by the SMs under a specific CT	kW	> 1 month	EyPESA	R
SM Reactive Power	Hourly	Hourly reactive power recorded by the SMs under a specific CT	kW	> 1 month	EyPESA	O
SM event logs	Daily	Daily recording of SM event logs. These indicate errors and alarm arising from SMs	ID integer numbers	> 1 month	EyPESA	O
Transformer Active Power	Hourly	Hourly active power recorded the CT meter	kW	> 1 month	EyPESA	R
Transformer Reactive Power	Hourly	Hourly reactive power recorded the CT meter	kvar	> 1 month	EyPESA	O
Topology	-	Grid topology including line impedances to calculate Technical Losses	-	-	EyPESA	O

Table 29. Training data for Detect non-technical losses service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Historical Detected Frauds	-	Information regarding the historical detected fraud: type of fraud, day of inspection, magnitude of fraud, etc.	-	N/A	EyPESA	O

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Historical SM Active Power	Hourly	Hourly active power recorded by the SMs under a specific CT that experienced fraud	kW	N/A	EyPESA	O
Historical SM Reactive Power	Hourly	Hourly reactive power recorded by the SMs under a specific CT that experienced fraud	kvar	N/A	EyPESA	O
Historical SM event logs	Daily	Daily recording of SM event logs. These indicate errors and alarm arising from SMs.	ID integer numbers	N/A	EyPESA	O
Historical Transformer Active Power	Hourly	Hourly active power recorded the CT meter	kW	N/A	EyPESA	O
Historical Transformer Reactive Power	Hourly	Hourly reactive power recorded the CT meter	kvar		EyPESA	O

Expected outputs:

Table 30. Output data for Detect non-technical losses service.

Data	Lowest level of granularity	Description	Units
NTL detected	When executed	Sent a notification if NTL > X%	YES / NO
Fraud data: Type	When executed	Classification between Marihuana Plantation, Crypto Mining, Squatting and Unclear	Classification
Fraud data: Probability	When executed	Probability of correct prediction of type of fraud and fraudster.	%

Data	Lowest level of granularity	Description	Units
Fraud data: Magnitude	When executed	Magnitude of fraud (95% percentile)	kW
Fraud data: Duration	When executed	Duration of fraud in days.	days
SM ID	When executed	SM ID of the consumer who is potentially committing fraud	SM ID

Applied Techniques:

The AI techniques applied in this service are

- Classification Algorithms (e.g., Random Forest Classifier, Gradient Booster Classifier and Support Vector Classifier) to find the type of fraud based on the NTL profile.
- Fuzzy c-Means as a clustering algorithm to detect SMs with suspicious shifts in their consumption patterns.
- Data Mining and Feature Engineering to extract features and information from active power curves.

Open-source/Proprietary software: Open-source.

B.1.9 Congestion detection

Service to be offered in real-time/offline:

Offline.

Execution frequency: Day-ahead.

Input data requirements:

Table 31. Execution input data for Congestion detection service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
SM Active Power	Hourly	Hourly active power recorded by the SMs under a specific CT	kW	> 1 month	EyPESA	R
SM Reactive Power	Hourly	Hourly reactive power recorded by the SMs under a specific CT	kvar	> 1 month	EyPESA	O
Transformer Active Power	Hourly	Hourly active power recorded the CT meter	kW	> 1 month	EyPESA	R

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Transformer Reactive Power	Hourly	Hourly reactive power recorded the CT meter	kvar	> 1 month	EyPESA	O
Topology	-	Grid topology including line impedances	-	-	EyPESA	R

Table 32. Training data for Congestion detection service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
SM Active Power	Hourly	Hourly active power recorded by the SMs under a specific CT	kW	> 1 month	EyPESA	R
SM Reactive Power	Hourly	Hourly reactive power recorded by the SMs under a specific CT	kvar	> 1 month	EyPESA	O
Transformer Active Power	Hourly	Hourly active power recorded the CT meter	kW	> 1 month	EyPESA	R
Transformer Reactive Power	Hourly	Hourly reactive power recorded the CT meter	kvar	> 1 month	EyPESA	O
Topology	-	Grid topology including line impedances	-	-	EyPESA	R

Expected outputs:

Table 33. Output data for Congestion detection service.

Data	Lowest level of granularity	Description	Units
Probability of congestions in lines and/or bus voltages.	When executed	Probability of a line or bus voltage	kA, V p.u. or % of overload
Forecast of demand	When executed	Demand of loads for the day-ahead	kW

Applied Techniques:

The techniques applied in this service are:

- Probabilistic Power Flows, as a method to analyse the grid and scenario creation.
- Machine Learning regression methods such as: random forest, support vector machines, and neural networks.

Open-source/Proprietary software: Open-source.

B.1.10 Detection of the volatility of voltage in grids with high renewable penetration

Service to be offered in real-time/offline:

Real-time or day-ahead.

Execution frequency: Daily.

Input data requirements:

Table 34. Execution input data for Detection of the volatility of voltage in grids with high renewable penetration service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
SM Active Power	15 min (minimum)	Active power recorded by the SMs under a specific CT	kW	> 1 day	DSO	R
CT Active Power		Active power recorded by the CT				
SM Reactive Power	15 min (minimum)	Reactive power recorded by the SMs under a specific CT	kvar	> 1 day	DSO	R
CT Reactive Power		Reactive power recorded by the CT				
SM Voltage	15 min (minimum)	Voltage recorded by the SMs under a specific CT	V	> 1 day	DSO	R
CT Voltage		Voltage recorded by the CT				
Grid Topology	-	Grid topology including line impedances to calculate voltage drop	-	-	DSO	O

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Day-ahead Forecast of SM Active Power	15 min (minimum)	Active power recorded by the SMs under a specific CT	kW	> 1 month	DSO	R
Day-ahead Forecast of SM Reactive Power	15 min (minimum)	Reactive power recorded by the SMs under a specific CT	kvar	> 1 month	DSO	R
Day-Ahead Forecast of Renewable energy generation	15 min (minimum)	Renewable energy generation	kW	> 1 month	DSO	R

Table 35. Training data for Detection of the volatility of voltage in grids with high renewable penetration service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
SM Active Power	15 min (minimum)	Active power recorded by the SMs under a specific CT	kW	> 1 year	DSO	R
CT Active Power		Active power recorded by the CT				
SM Reactive Power	15 min (minimum)	Reactive power recorded by the SMs under a specific CT	kvar	> 1 year	DSO	R
CT Reactive Power		Reactive power recorded by the CT				
SM Voltage	15 min (minimum)	Voltage recorded by the SMs under a specific CT	V	> 1 year	DSO	R
CT Voltage		Voltage recorded by the CT				

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Grid Topology	-	Grid topology including line impedances to calculate voltage drop	-	-	DSO	O
Renewable energy generation	15 min (minimum)	Renewable energy generation	kW	> 1 month	DSO	R

Expected outputs:

The service output is a multi-class classification on the voltage volatility of the distribution grid downstream the transformer substation. In particular, three different states have been chosen:

1. Red State: High voltage volatility.
2. Orange State: Medium voltage volatility.
3. Green State: Low voltage volatility.

If, instead of following the classification approach, the regression one is used, the output will be the prediction of overvoltage or undervoltage in a certain area.

Applied Techniques:

The AI techniques applied in this service are:

- Classification Algorithms (e.g., Random Forest Classifier, Gradient Booster Classifier and Support Vector Classifier) to determine the voltage volatility class.
- Regression Algorithms (depending on the final approach).

Open-source/Proprietary software: Open-source.

B.1.11 Plan grid reinforcements for future renewable scenarios

Service to be offered in real-time/offline:

Offline.

Execution frequency: Weekly, monthly or yearly.

Input data requirements:

Table 36. Execution input data for Plan grid reinforcements for future renewable scenarios service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Transformer Active Power	Hourly	Hourly active power recorded the CT meter	kW	> 3 years	EyPESA	R
Transformer Reactive Power	Hourly	Hourly reactive power recorded the CT meter	kvar	> 3 years	EyPESA	O

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Topology	-	Grid topology including line impedances, transformer parameters and bus voltages	-	-	EyPESA	R
Asset Costs	-	Investment and operational costs of line types, transformer types and Li-ion battery systems (for different capacity levels).	Euros (€)	-	EyPESA	R
GIS Data	-	Geographic coordinates of grid assets	Latitude and Longitude	-	EyPESA	O
Asset Age	-	Current life-time of line and transformers	Years	-	EyPESA	O
Expansion plans	-	Location of future load and generation expansions according to user requests.	kW	-	EyPESA	O

Expected outputs:

The service reports optimal planning actions for long term horizon scenarios, aiming to minimize asset investment costs for DSOs using traditional planning strategies (replacement/reinforcement of feeder lines and transformers) and flexible planning strategies (installation of Li-ion batteries). Once the optimal planning strategy is found, the operator can obtain, new loading percentages (%) per asset and total investment costs.

The following table summarizes the different service outputs.

Table 37. Output data for Plan grid reinforcements for future renewable scenarios service.

Data	Lowest level of granularity	Description	Units
CAPEX	Depends on execution frequency	Capital Investment Costs of new lines, transformers or batteries.	Euro (€)

Data	Lowest level of granularity	Description	Units
OPEX	Depends on execution frequency	Operational Investment Costs of new lines, transformers or batteries	Euro (€)
Asset Improvements	Depends on execution frequency	Improvement of Lines Loadings, Transformer Loadings and Bus Voltages.	%

Applied Techniques:

The techniques applied in this service are:

- Optimization model to determine best size and location of new lines, transformers or Li-ion batteries to reinforce the grid minimizing CAPEX and OPEX.
- Data driven techniques to reduce order of scenarios.
- Machine learning forecasting techniques for future renewable generation.

Open-source/Proprietary software: Open-source.

B.1.12 Energy Generation Forecast

Service to be offered in real-time/offline:

Real-time.

Execution frequency: Hourly, daily or on-demand.

Input data requirements:

Table 38. Execution input data for Energy Generation Forecast service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Power	Hourly or less Recom: 5 min	Average power on the granularity period	kW	> 1 year	EyPESA	R
Energy	Hourly or less	Sum of energy on the granularity period	kWh	> 1 year	EyPESA	O
Global Horizontal Irradiance (GHI)	Hourly or less	Global Horizontal Irradiance measured by a meteorological station on site	W/m²	> 1 year	EyPESA	O

Table 39. Training data for Energy Generation Forecast service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Power	Hourly or less Recom: 5 min	Average power on the granularity period	kW	> 1 year	EyPESA	R
Energy	Hourly or less	Sum of energy on the granularity period	kWh	> 1 year	EyPESA	O
Global Horizontal Irradiance (GHI)	Hourly or less	Global Horizontal Irradiance measured by a meteorological station on site	W/m²	> 1 year	EyPESA	O

Expected outputs:

The outputs of the service:

- **Power:** Power generation by the site for the following 144 hours (6 days).
- **Uncertainty:** Percentiles 10, 25, 75 and 90 to define the uncertainty respect to the most probable power generation.

The following table summarizes the different service outputs.

Table 40. Output data for Energy Generation Forecast service.

Data	Lowest level of granularity	Description	Units
Power	5, 10, 15, 30 or 60 minutes	Most probable accurate forecast for each data point of the 144 hours forecast	kWh [float]
Uncertainty	5, 10, 15, 30 or 60 minutes	Uncertainty percentiles 10%, 25%, 75% and 90% to define the trust of the forecast for each data point.	kWh [float]

Applied Techniques:

The AI techniques applied in this service are:

- Feature Engineering to create added value for each site to improve the forecast accuracy (e.g. Aperture Normal Irradiance feature for horizontal 1 axis tracking solar fields).
- Genetic algorithms to train the predictive models and define the most accurate Machine Learning predictive model for each site.

- Machine learning to generate the most probable forecast and an ensemble to define the uncertainty of the forecast for each data point predicted.

Open-source/Proprietary software: Proprietary software delivered as a service.

B.1.13 Digital Twin BIPV Self Consumption Systems in Buildings

Service to be offered in real-time/offline:

Preferably online but can also be run offline.

Execution frequency: Weekly, monthly or on-demand.

Input data requirements:

Table 41. Execution input data for Digital Twin BIPV Self Consumption Systems in Buildings service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
DC and AC voltages	< 15 min	Monitoring Information from inverters and module level electronics – Raw Data.	V	< 1 day	EyPESA	Required
DC and AC currents	< 15 min		A	< 1 day	EyPESA	Required
Internal temperature	< 15 min		ºC	< 1 day	EyPESA	Required
Operating modes	< 15 min		N/A	< 1 day	EyPESA	Optional
Alarms	< 15 min		N/A	< 1 day	EyPESA	Optional
Global Irradiance on PoA (Plane of Array)	< 15 min	Meteorogical Information– Raw Data.		< 1 day	EyPESA	Required
Direct Irradiance (dni), diffuse irradiance above the horizontal (dhi), global irradiance above the horizontal (ghi)	< 15 min			< 1 day	EyPESA	Optional

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Ambient and Module temperature	< 15 min			< 1 day	EyPESA	Required
Rainfall events	< 15 min			< 1 day	EyPESA	Optional
Wind speed and direction	< 15 min			< 1 day	EyPESA	Optional

Table 42. Training data for Digital Twin BIPV Self Consumption Systems in Buildings service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
DC and AC voltages	< 15 min	Monitoring Information from inverters and module level electronics – Raw Data.	V	≥ 1 month	EyPESA	Required
DC and AC currents	< 15 min		A	≥ 1 month	EyPESA	Required
Internal temperature	< 15 min		ºC	≥ 1 month	EyPESA	Required
Operating modes	< 15 min		N/A	≥ 1 month	EyPESA	Optional
Alarms	< 15 min		N/A	≥ 1 month	EyPESA	Optional
Global Irradiance on PoA (Plane of Array)	< 15 min	Meteorological Information– Raw Data.		≥ 1 month	EyPESA	Required

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Direct Irradiance (dni), diffuse irradiance above the horizontal (dhi), global irradiance above the horizontal (ghi)	< 15 min			≥ 1 month	EyPESA	Optional
Ambient and Module temperature	< 15 min			≥ 1 month	EyPESA	Required
Rainfall events	< 15 min			≥ 1 month	EyPESA	Optional
Wind speed and direction	< 15 min			≥ 1 month	EyPESA	Optional
Maintenance Loogbok	< 15 min	Maintenance Logbook with conducted maintenance actions	N/A	≥ 1 month	EyPESA	Optional
PV Plant design	Static	BIM model or information required for building and PV Plant design		Static	EyPESA	Required
Latitude and longitude	Static	Coordinates of the PV installation		Static	EyPESA	Required

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Number of modules per string and strings per array	Static	Number of modules per string and strings per array connected to each inverter input (in series and in parallel).		Static	EyPESA	Required
Characteristic parameters of the PV modules	Static	Characteristic parameters available in the module manufacturer's technical sheet: Voc, Isc, Vmp Imp, temperature coefficients, the number of cells in series for the calculation of the ideality factor and the type of module and assembly for the thermal model of the temperature of operation		Static	EyPESA	Required
Characteristic parameters of the PV inverters:	Static	The maximum power point voltage operating range for each input section of the MPPT. Maximum current and power values for each input section of the MPPT.		Static	EyPESA	Required
Position of the irradiance sensor	Static	The elevation and orientation of the irradiance sensor		Static	EyPESA	Required

Expected outputs:

Table 43. Output data for Digital Twin BIPV Self Consumption Systems in Buildings service.

Data	Lowest level of granularity	Description	Units
Improved BIPV model with adjusted parameters providing more reliable and accurate output AC power	N/A	Characterization of PV modules, PV arrays and inverters performance parameters between different PV plants.	N/A
Energy Prediction	< 15 min	Forecast of produced energy by the BIPV panels.	kW
Alarms	< 15 min	list of triggered alarms in the case these adjusted parameters present a significant deviation. For each case, the affected component or subsystem will be identified, the specific failure modes related to this underperformance will be reported and the potential causes will be exposed.	N/A

Applied Techniques:

The AI techniques applied in this service are:

- Data Mining and Feature Engineering.
- Clustering algorithms to identify different operating conditions.
- Hybrid Digital Twins.
- Regression.
- Bayesian optimization using Gaussian Processes.
- Sequential optimisation using decision trees.

Open-source/Proprietary software: Proprietary offered as a service.

B.2 Local Energy Communities UC Services

B.2.1 Local Energy Communities Designer

Service to be offered in real-time/offline:

Offline.

Execution frequency: On-demand.

Input data requirements:

Table 44. Algorithm input data for Local Energy Communities Designer service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
energy_kWh	Hourly	Energy consumed in kilo-Watt-hours (kWh)	kWh	> 1 month		R
date	Daily	Date of the recording in YYYY-MM-DD	date	> 1 month		R
day	Daily	Day of the week	string	> 1 month		O
participantID	-	ID code identifying which is the specific building/consumer	ID integer numbers	> 1 month		R
hour	Hourly	Hour of the recording from 01 to 24.	integer (01-24)	> 1 month		R

Table 45. Consumer description data for Local Energy Communities Designer service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
participantID	-	ID code identifying which is the specific building/consumer/participant	ID integer numbers			R
Coverage	-	The percent of non-missing readings. A value of 1.000 is 100%	integer numbers (0-1000)			O

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
BuildingType	.	Type of the building related to the consumer	string			R

Expected outputs:

The outputs of the service indicate which is the best possible configuration among the consumers provided as inputs. Moreover, it contains information about another possible configuration even if not optimum. The following table summarizes the service outputs.

Table 46. Output data for Local Energy Communities Designer service.

Data	Lowest level of granularity	Description	Units
optimal_configuration	When executed	Lists all the participants needed to optimise auto consumption in relation to power plant size	list of participantID
suboptimal_configuration	When executed	Several suboptimal configurations of LEC participants	array of list of participantID

Applied Techniques:

The AI techniques applied in this service are:

- Optimization technique (heuristic model).

Open-source/Proprietary software: Proprietary software.

B.2.2 Gamification for electrical energy savings

Service to be offered in real-time/offline:

Real-time.

Execution frequency: on-demand.

Input data requirements:

Input data are summarised in the following tables. The second table represents optional questionnaire data, which consumers can compile to assist classification algorithms on recognising how mainly the energy is used.

Table 47. Input data for Gamification for electrical energy savings service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
energy_kWh	Hourly	Energy consumed in kilo-Watt-hours (kWh)	kWh	> 1 month		R

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
date	Daily	Date of the recording in YYYY-MM-DD	date	> 1 month		R
day	Daily	Day of the week	string	> 1 month		O
consumerID	-	ID code identifying which is the specific consumer	ID integer numbers	> 1 month		R
hour	Hourly	Hour of the recording from 01 to 24.	integer (01-24)	> 1 month		R

Table 48. Home appliances data for Gamification for electrical energy savings service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
consumerID	-	ID code identifying which is the specific consumer	ID integer numbers			R
refrigerator	-	Appliances present	integer numbers			O
freezer	-	Appliances present	integer numbers			O
dishwasher	-	Appliances present	integer numbers			O
oven	.	Appliances present	integer numbers			O
microwave oven	-	Appliances present	integer numbers			O
washing machine	-	Appliances present	integer numbers			O
clothes dryer	.	Appliances present	integer numbers			O
air conditioner		Appliances present	integer numbers			O

Expected outputs:

The possible outputs of the service can be classified as follows:

- **consumer notification:** The service notifies whether or not a specific challenge is overcome and when the consumer receives rewards from the system due to his actions;

- **consumption analysis**: energy usage is analysed and visualised in graphical dashboards with optional quantitative information (KPI) relative to consumption and transferred to the user;
- **saved energy:** the amount of saved energy and reduced costs are returned to consumer

The following table summarizes the different service outputs.

Table 49. Output data for Gamification for electrical energy savings service.

Data	Lowest level of granularity	Description	Units
KPI	When executed	Key performance indicator at different time ranges	kWh
saved_energy	When executed	Energy saved from single consumer	kWh

Applied Techniques:

The AI techniques applied in this service are

- Classification Algorithms (e.g., Random Forest Classifier, Gradient Booster Classifier and Support Vector Classifier) to recognise appliances consumption profile.
- Data Mining and Feature Engineering to extract features and information from active consumption curves.

Open-source/Proprietary software: Proprietary software.

B.2.3 Thermal Losses Detection and Benchmarking at LEC level

Service to be offered in real-time/offline:

Real-Time.

Execution frequency: Daily/Weekly/Monthly.

Input data requirements:

Table 50. Execution input data for Thermal Losses Detection and Benchmarking at LEC level service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Ec,h	Hourly/Daily	Hourly/Daily thermal energy consumed by end-users	kWh	> 1 year	ASTEA	R
Es,h	Hourly/Daily	Hourly/Daily thermal energy supplied to all end-users	kWh	> 1 year	ASTEA	R
Topology	-	Grid topology of District Heating Network for thermal losses isolation				O

Table 51. Training data for Thermal Losses Detection and Benchmarking at LEC level service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Ec,h	Hourly/Daily	Hourly/Daily thermal energy consumed by end-users	kWh	> 1 year	ASTEA	R
Es,h	Hourly/Daily	Hourly/Daily thermal energy supplied to all end-users	kWh	> 1 year	ASTEA	R

Expected outputs:

The outputs of the service are

- **Elosses**: Daily estimated thermal losses.
- Statistics related to thermal losses (e.g., number of detected thermal losses in different periods, % of smart meters offline, etc.).
- Benchmarking.

The following table summarizes the different service outputs.

Table 52. Output data for Thermal Losses Detection and Benchmarking at LEC level service.

Data	Lowest level of granularity	Description	Units
Elosses	Daily	Daily estimated thermal losses	m^3/h

Applied Techniques:

The AI techniques applied in this service are:

- Clustering algorithms to discriminate different consumption patterns.
- Data Mining and Feature Engineering to extract information from thermal consumption curves.

Open-source/Proprietary software: Proprietary software.

B.2.4 Water Losses detection and benchmarking at LEC level

Service to be offered in real-time/offline:

Real-Time.

Execution frequency: Daily.

Input data requirements:

Table 53. Execution input data for Water Losses Detection and Benchmarking at LEC level service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Qc,h	Hourly	Hourly flow rate consumed by end-users	m^3	> 1 year	ASTEA	R
Qs,h	Hourly	Hourly flow rate supplied to all end-users	m^3	> 1 year	ASTEA	R
Topology	-	Grid topology of Water Distribution Network for water losses isolation				O

Table 54. Training data for Water Losses Detection and Benchmarking at LEC level service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Qc,h	Hourly	Hourly flow rate consumed by end-users	m^3	> 1 year	ASTEA	R
Qs,h	Hourly	Hourly flow rate supplied to all end-users	m^3	> 1 year	ASTEA	R

Expected outputs:

The outputs of the service are

- **Qlosses**: Daily estimated water losses.
- Statistics related to water losses (e.g., number of detected water losses in different periods, % of smart meters offline, etc.).
- Benchmarking.

The following table summarizes the different service outputs.

Table 55. Output data for Water Losses Detection and Benchmarking at LEC level service.

Data	Lowest level of granularity	Description	Units
Qlosses	Daily	Daily estimated water losses	m^3/h

Applied Techniques:

The AI techniques applied in this service are:

- Clustering algorithms to cluster different consumption pattern.
- Data Mining and Feature Engineering to extract information from water consumption curves.

Open-source/Proprietary software: Proprietary software.

B.2.5 Estimate the probability of congestions

Service to be offered in real-time/offline:

Offline.

Execution frequency: Day-ahead.

Input data requirements:

Table 56. Execution input data for Estimate the probability of congestions service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
SM Active Power	Hourly	Hourly active power recorded by the SMs under a specific CT	kW	> 1 month	LEC operator or DSO	R
SM Reactive Power	Hourly	Hourly reactive power recorded by the SMs under a specific CT	kvar	> 1 month	LEC operator or DSO	O
Transformer Active Power	Hourly	Hourly active power recorded the CT meter	kW	> 1 month	LEC operator or DSO	R
Transformer Reactive Power	Hourly	Hourly reactive power recorded the CT meter	kvar	> 1 month	LEC operator or DSO	O
Topology	-	Grid topology including line impedances	-	-	LEC operator or DSO	R

Table 57. Training data for Estimate the probability of congestions service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
SM Active Power	Hourly	Hourly active power recorded by the SMs under a specific CT	kW	> 1 month	LEC operator or DSO	R
SM Reactive Power	Hourly	Hourly reactive power recorded by the SMs under a specific CT	kvar	> 1 month	LEC operator or DSO	O
Transformer Active Power	Hourly	Hourly active power recorded the CT meter	kW	> 1 month	LEC operator or DSO	R

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Transformer Reactive Power	Hourly	Hourly reactive power recorded the CT meter	kvar	> 1 month	LEC operator or DSO	O
Topology	-	Grid topology including line impedances	-	-	LEC operator or DSO	R

Expected outputs:

The possible outputs of the service are as follows:

- Probability of congestions in lines and/or transformers and/or bus voltages.
- Forecasted demand of the loads.

The following table summarizes the different service outputs.

Table 58. Output data for Estimate the probability of congestions service.

Data	Lowest level of granularity	Description	Units
Probability of congestions in lines and/or bus voltages.	When executed	Probability of a line or bus voltage	kA, v p.u. or % of overload
Forecast of demand	When executed	Demand of loads for the day-ahead	kW

Applied Techniques:

The techniques applied in this service are:

- Probabilistic Power Flows, as a method to analyse the grid and scenario creation.
- Machine Learning regression methods such as: random forest, support vector machines, and neural networks.

Open-source/Proprietary software: Open-source.

B.2.6 Electrical losses detection and benchmarking at LEC level

Service to be offered in real-time/offline:

Offline.

Execution frequency: Weekly, monthly or on-demand.

Input data requirements:

Table 59. Execution input data for Electrical losses detection and benchmarking at LEC level service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
SM Active Power	Hourly	Hourly active power recorded by the SMs under a specific CT	kW	> 1 month	DSO or LEC operator	R
SM Reactive Power	Hourly	Hourly reactive power recorded by the SMs under a specific CT	kvar	> 1 month	DSO or LEC operator	O
SM event logs	Daily	Daily recording of SM event logs. These indicate errors and alarm arising from SMs	ID integer numbers	> 1 month	DSO or LEC operator	O
Transformer Active Power	Hourly	Hourly active power recorded the CT meter	kW	> 1 month	DSO or LEC operator	R
Transformer Reactive Power	Hourly	Hourly reactive power recorded the CT meter	kvar	> 1 month	DSO or LEC operator	O
Topology	-	Grid topology including line impedances to calculate Technical Losses	-	-	DSO or LEC operator	O

Table 60. Training data for Electrical losses detection and benchmarking at LEC level service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Historical Detected Frauds	-	Information regarding the historical detected fraud: type of fraud, day of inspection, magnitude of fraud, etc.	-	N/A	DSO or LEC operator	O

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Historical SM Active Power	Hourly	Hourly active power recorded by the SMs under a specific CT that experienced fraud	kW	N/A	DSO or LEC operator	O
Historical SM Reactive Power	Hourly	Hourly reactive power recorded by the SMs under a specific CT that experienced fraud	kvar	N/A	DSO or LEC operator	O
Historical SM event logs	Daily	Daily recording of SM event logs. These indicate errors and alarm arising from SMs.	ID integer numbers	N/A	DSO or LEC operator	O
Historical Transformer Active Power	Hourly	Hourly active power recorded the CT meter	kW	N/A	DSO or LEC operator	O
Historical Transformer Reactive Power	Hourly	Hourly reactive power recorded the CT meter	kvar	N/A	DSO or LEC operator	O

Expected outputs:

The possible outputs of the service, depending on whether or not there is fraud and the type of fraud, can be classified as follows:

- **Abnormal losses:** The service notifies whether or not there are abnormal losses in the analysed grid.
- **Losses data:** If there are abnormal losses in the grid, these are analysed, and some descriptive information is obtained and transferred to the user (type of losses, magnitude, duration and probability).
- **Culprit:** If there are SMs with anomalous behaviour, they are detected and assigned a probability of committing fraud.

The following table summarizes the different service outputs.

Table 61. Output data for Electrical losses detection and benchmarking at LEC level service.

Data	Lowest level of granularity	Description	Units
Abnormal losses detected	When executed	Sent a notification if NTL > X%	YES / NO

Data	Lowest level of granularity	Description	Units
Losses data: Type	When executed	Classification between high technical losses or fraud (Marihuana Plantation, Crypto Mining, Squatting and Unclear)	Classification
Losses data: Probability	When executed	Probability of correct prediction of type of losses and fraudster.	%
Losses data: Magnitude	When executed	Magnitude of fraud (95% percentile)	kW
Losses data: Duration	When executed	Duration of abnormal losses in days.	days
SM ID	When executed	SM ID of the consumer who is potentially committing fraud	SM ID

Applied Techniques:

The AI techniques applied in this service are:

- Classification Algorithms (e.g., Random Forest Classifier, Gradient Booster Classifier and Support Vector Classifier) to find the type of losses based on their load profile.
- Fuzzy c-Means as a clustering algorithm to detect SMs with suspicious shifts in their consumption patterns.
- Data Mining and Feature Engineering to extract features and information from active power curves.

Open-source/Proprietary software: Open-source.

B.2.7 Reinforcement Plan of Local Energy Communities for Future Renewable Integration

Service to be offered in real-time/offline:

Offline.

Execution frequency: Weekly, monthly or yearly.

Input data requirements:

Table 62. Execution input data for Reinforcement Plan of LEC for Future Renewable Integration service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Transformer Active Power	Hourly	Hourly active power recorded the CT meter	kW	> 3 years	DSO or LEC operator	R

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Transformer Reactive Power	Hourly	Hourly reactive power recorded the CT meter	kW	> 3 years	DSO or LEC operator	O
Topology	-	Grid topology of LEC including line impedances, transformer parameters and bus voltages	-	-	DSO or LEC operator	R
Asset Costs	-	Investment and operational costs of line types, transformer types and Li-ion battery systems (for different capacity levels).	Euros (€)	-	DSO or LEC operator	R
GIS Data	-	Geographic coordinates of grid assets	Latitude and Longitude	-	DSO or LEC operator	O
Asset Age	-	Current life-time of line and transformers	Years	-	DSO or LEC operator	O
Expansion plans	-	Location of future load and generation expansions according to user requests.	kW	-	DSO or LEC operator	O

Expected outputs:

The service reports planning actions for future renewable scenarios in LEC, aiming to minimize asset investment costs for DSOs using traditional planning strategies (replacement or reinforcement of feeder lines and transformers) and flexible planning strategies (installation of Li-ion batteries). Once the optimal planning strategy is found, the operator can obtain, new loading percentages (%) per asset and total investment costs.

The following table summarizes the different service outputs.

Table 63. Output data for Reinforcement Plan of LEC for Future Renewable Integration service.

Data	Lowest level of granularity	Description	Units
CAPEX	When executed	Capital Investment Costs of new lines, transformers or batteries.	Euro (€)
OPEX	When executed	Operational Investment Costs of new lines, transformers or batteries	Euro (€)
Asset Improvements	When executed	Improvement of Lines Loadings, Transformer Loadings and Bus Voltages.	%

Applied Techniques:

The techniques applied in this service are

- Optimization model to determine best size and location of new lines, transformers or Li-ion batteries to reinforce the grid minimizing CAPEX and OPEX.
- Data driven techniques to reduce order of scenarios.
- Machine learning forecasting techniques for future renewable generation.

Open-source/Proprietary software: Open-source.

B.2.8 Optimizing self-consumption of renewable energy at LEC level

Service to be offered in real-time/offline:

The frequency of execution of the service will be daily, and the time of execution will depend on the renewal of PV production and consumption forecasts.

Input data requirements:

Table 64. Execution input data for Optimizing self-consumption of renewable energy at LEC level service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Predicted active energy production	15’	Predicted active energy production at the PV generator output	Wh	1 day	LEC operator	R
Predicted active energy consumption	15’	Predicted active energy consumption at the PCC of the prosumer	Wh	1 day	LEC operator	R

Expected outputs:

The output of the service is the 96 quarter-hour PV battery schedule for the following day.

The following table summarizes the different service outputs.

Table 65. Output data for Optimizing self-consumption of renewable energy at LEC level service.

Data	Lowest level of granularity	Description	Units
Battery charging/discharging active energy schedule	15'	Scheduled battery charging/discharging active energy at the battery inverter output	Wh

Applied Techniques:

The AI techniques applied in this service are:

- Optimization Greedy algorithm.

Open-source/Proprietary software: Proprietary software.

B.2.9 Optimizing sharing coefficients in collective self-consumption

Service to be offered in real-time/offline:

The frequency and time of execution of the service will be under demand depending on operating and market conditions. For instance, the current Spanish regulation requires determination of these sharing coefficients for all the year, allowing changes twice a year.

Input data requirements:

Table 66. Execution input data for Optimizing sharing coefficients in collective self-consumption service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Historic active consumed energy	1 h	Historic active energy consumed by every consumer in the Energy Community	Wh	> 1 year	Energy Community Operator (EDP)	R

Expected outputs:

The output of the service is the hourly sharing coefficients schedule of a shared PV generator in a self-consumption application for a certain period.

The following table summarizes the different service outputs.

Table 67. Output data for Optimizing sharing coefficients in collective self-consumption service.

Data	Lowest level of granularity	Description	Units
Sharing coefficients of all the consumers in the collective self-consumption system for the determined period	1 h	Scheduled sharing coefficients of all the consumers in the collective self-consumption system	%

Applied Techniques:

The AI techniques applied in this service are:

- Optimization Greedy algorithm.

Open-source/Proprietary software: Proprietary software.

B.2.10 Planning services

Service to be offered in real-time/offline:

The frequency of execution of the service would be yearly, and the time of execution will depend on the fuel energy markets and measures implementation revision in the demonstrator. The time of execution would also depend on the demonstration decarbonisation KPIs revision.

Input data requirements:

Table 68. Execution input data for Planning services.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Building geometry		Building geometry and geolocated (preferably in a GIS format)			Municipality (Public or Private information)	R
Property registry information		Building information included in a shape file or related to Building ID (Construction year, Use, Height)			Municipality (Public or Private information)	O
Additional building information		Energy Performance Certificates, Retrofitted buildings, Building Energy System, Ownership			Municipality (Public or Private information), LEC Operator	O
Historic energy consumption	60'	Provided by the LEC operator (by energy us, by fuel type). If not available, estimated by the planning service	kWh	1 year	LEC Operator	R/O
Socioeconomic data		Average income per household, per person, age of residents, educational level, etc..		1 year	Municipality (Public or Private information)	O

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Other additional information related to relevant building energy generation systems		Hourly profiles, equipment performance, installed power		1 year	LEC Operator	R/O
Historic local renewable production	60’	Hourly profiles	kWh	1 year	LEC Operator	R/O
Renewable equipment information		Equipment performance, installed power, CAPEX, OPEX, embedded impact			LEC Operator	R/O
Historic energy prices	60’	Prices agreed with the supplier. Includes the energy and price constraints agreed with the supplier.	€/kWh	1 year	LEC Operator	R
Decarbonisation objectives	yearly	Decarbonisation objectives will be defined (kgCO_2/year). Economical restrictions should be also defined (€invested/year)			LEC Operator, Municipality	R

The expected outputs provided by the Planning Service will vary considering the information available or provided by the LEC operator.

Expected outputs:

Table 69. Output data for Planning services.

Data	Lowest level of granularity	Description	Units
Energy consumption information	60'	Demand per end-use, consumption by end-use, cost of energy consumed, emissions per end-use, degree of self-consumption, etc.	kWh
Local renewable generation (potential generation)	60'	Energy Produced, consumed and surplus. Cost of energy produced, consumed and surplus.	kWh
Local renewable interventions (alternative scenarios)	Yearly (2020-2050)	Type of intervention applied, configuration, cost of interventions, impact embedded on interventions	

Results can be provided in different units considering the data input provided: per kWh consumed, per m^2, per person, per average income, per euro invested, for the global LEC, etc.

Alternative scenarios combining renewable interventions will be proposed by the Planning Service considering KPIs about decarbonisation objectives, energy consumption information and local renewable generation considering socioeconomic impacts (ex. avoiding energy poverty).

Applied Techniques:

The AI techniques applied in this service are:

- GIS techniques (TBD).
- Hourly energy simulation based on hourly-degree day calculation.

Open-source/Proprietary software: Proprietary software. Some developments will be deployed in QGIS (open-source software).

B.3 Electromobility UC Services

B.3.1 Roaming of booking services

Service to be offered in real-time/offline:

Real-time.

The service will be available on-demand.

It will rely on a set of data exchanges that can take place at different time scales, depending on the needs:

- Updating of static information (on event or regularly).
- dynamic information update on the real availability of the infrastructures (as close as possible to real time).
- transfer of billing and monitoring information (frequency to be defined according to the billing policy between actors).

Input data requirements:

Table 70. Static data for Roaming of booking services

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
charging infrastructures description	-	localization of the stations of a CPO for each station, description of the infrastructure on the field (charging points and parking lots) for each charging point, technical characteristics of the device (outlets, power, communication protocols available…)	longitude and latitude address (text) kWh outlet type (text)		CPO	Required
agreed tariffs description	-	for each relation between a CPO and an EMSP, description of the tariffs (this data will have to be set at a charging point level, i.e., the costs associated to the reservation and the use of a given CP)	€ per unit of time, per action, per kWh		CPO EMSP the EVISP will also have this data	Required
booking service tariffs description	-	Set of reservations (time slot, charging point, + additional information) with their status (request in progress, valid, use in progress, cancelled...)	multiple units, which will be described in the data model on which the messaging will be based		EMSP CPO	Required

Expected outputs:

- Interface with real time availability of the charging points in a dedicated area.
- Pricing data applicable to a recharge that considers local conditions, agreements between parties and the way the end user is solicited.
- Reservation data set (with different statuses) for a given mobility service provider; and its counterparts at infrastructure operators.

Applied Techniques: Data modelling and standardization.

Open-source/Proprietary software: Open source.

B.3.2 Roaming of self-consumption

Service to be offered in real-time/offline:

Real-time.

The service is available on-demand. It relies on a set of data exchanges that might have different time scales:

- Meter readings for production certificates are available in a range from close to real time to 24 h *ex-post*.
- Meter readings for consumption certificates are available shortly after the charging session.

Input data requirements:

Table 71. Execution input data for Roaming of self-consumption service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
DID or user and involved parties	1/party	Unique identifier			Self-sovereign	
Time series household PV injection (kWh)	15 min	Measurement of excess PV generation at head meter level	kWh per 15 min		User	R
Production (household) information	1/Meter	Installed capacity PV panel (kw), physical address Meter type (utility, submeter), meter serial number MeterID (PV installation), meter ID (head meter), Responsible Metered Data Administrator, DSO Market Location ID	kW, address Tbd ID DID ID		Prosumer/Metering Point Operator	R

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Charging infrastructure description	1/EVSE	localization of the stations of a CPO incl. head meter ID, meter ID of charging points, Market Location ID Responsible DSO / Metered Data Administrator for each station, description of the infrastructure on the field (charging points and parking lots) for each charging point, technical characteristics of the device (outlets, power, communication protocols available…)	Physical address	Number of EVSEs	CPO of the EVSE	R
Consumption (EVSE) information	1-5 min	Time series (session updates with kWh and timestamps)	kWh time	Number of charging sessions	CPO	R
Market clearing price	15 min	Differences in electricity market prices in production and consumption bidding zone	€/MWh	96 (92, 100) per day	Publicly available	R

Expected outputs:

The output of the service is the matching of production and consumption certificates:

- Consumption and production time series are signed with the DID of the data owner and trustable for the data recipient.
- The GCO registry creates granular certificates of origin for the Data Hub Operator production and eMSP consumption data. Granular energy certificates that are issued on consumed energy contain a reference to the user (DID identifier). It will have a unique id, metadata about its origin – and for production certificates a fuel code, tech code etc., which gives the certificate its qualities.
- Each certificate represents a fixed start timestamp and a duration of 15 minutes. The start timestamp and consumption volumes are the bases for temporal matching.
- The marketplace request to redeem the production GCO that match the production GCO.
- The GC system itself requires temporal matching, meaning certificates must be based the same time-period and owned by the same user.
- The GCO registry redeems the certificates and adapts the ledger.
- The GCO registry allow cross-border exchange. If a cross-border matching is requested, an exchange request is sent to the cross-border module operated by TSO. Only if sufficient trading capacities are available cross-border trades are allowed. As a result, the user is not always able to claim his/her certificates if the grid situation does not allow it.
- The Granular Certificate of Origin Registry confirms to the Granular Certificate of Origin Platform the redemption of certificates on the user account.

Applied Techniques:

Application of self-sovereign identity standards:

- Creation of decentralized identifier for all involved parties.
- Storage, issuance and verification of verifiable credentials.

Open-source/Proprietary software: Open-source.

B.4 Flexibility UC Services

B.4.1 Grid observability and network analysis

Service to be offered in real-time/offline:

Analysis offline, Observability real time.

Execution frequency: On demand.

Input data requirements:

Table 72. Execution input data for Grid observability and network analysis service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Smart Meter (SM) – ID	NA	Identification number of the SMs	NA	NA	DSO	R
Secondary Substation (SS) – ID	NA	Identification number of the SSs	NA	NA	DSO	R
SM – Type	NA	{“Single phase”; “Three-phase"}	NA	NA	DSO	R
Single phase SM Active Power load curve	10 min timestep, daily batch	10 minutes time step active power load curve recorded by the SMs under a specific SS	kW	> 1 month	DSO	R
Three phase SM Active Power load curve – Phase 1						
Three phase SM Active Power load curve – Phase 2						
Three phase SM Active Power load curve – Phase 3						
Single phase SM Reactive Power load curve	10 min timestep, daily batch	10 minutes time step reactive power load	kvar	> 1 month	DSO	O

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Three phase SM Reactive Power load curve – Phase 1		curve recorded by the SMs under a specific SS				
Three phase SM Reactive Power load curve – Phase 2						
Three phase SM Reactive Power load curve – Phase 3						
Single phase SM Voltage Profile	10 min timestep, daily batch	10 minutes time step voltage profile recorded by the SMs under a specific SS	V	> 1 month	DSO	R
Three phase SM Voltage Profile – Phase 1						
Three phase SM Voltage Profile – Phase 2						
Three phase SM Voltage Profile – Phase 3						
SM – Location	NA	GPS coordinate of each SM	NA	NA	DSO	O
SS – Location	NA	GPS coordinate of each SS	NA	NA	DSO	O

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Primary Substation (PS) - Location	NA	GPS coordinate of each SM	NA	NA	DSO	R
Nominal voltage limits	NA	Regulatory voltage limit for electricity supply	V	NA	DSO	R
SM – SS connection	NA	A priori SM – SS connection as recorded in legacy systems	NA	NA	DSO	R
SM – LV feeder connection	NA	A priori SM – LV feeder connection as recorded in legacy systems	NA	NA	DSO	R
SS – LV fuse rating	NA	Rating of each fuse of a specific SS	A	NA	DSO	R
SS – PS connection	NA	SS – PS connection as recorded in legacy systems	NA	NA	DSO	R
SS – PS feeder connection	NA	SS – PS feeder connection as recorded in legacy systems	NA	NA	DSO	R
PS – Active power	10 min timestep provided in real time	10 minutes averaged active power recorded at PS level	kW	Every 10 minutes	DSO	R

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
PS – Voltage profile	10 min timestep provided in real time	10 minutes averaged voltage recorded at PS level	V	Every 10 minutes	DSO	R
Meteorological data	10 min timestep provided in real time	Real-time meteorological data (temperature, sunshine)	NA	Every 10 minutes	Meteorological data provider	R

The required training data is similar to the Execution input data but with two months of historical value, which will grow longer with the execution.

Expected outputs:

The output of the service is a deep analysis of the network's operating conditions and a model which provides a state estimation:

- Reliable low voltage network topology from a priori connections.
- Set of KPIs depicting low voltage assets health level.
- User interface to navigate in the analysis.
- Real-time and forecasted state estimation.

The following table summarizes the different service outputs.

Table 73. Output data for Grid observability and network analysis service.

Data	Lowest level of granularity	Description	Units
LV topology	When executed	SM – SS's feeder – SS connection	NA
LV KPIs	When executed	Set of KPIs (load level, unbalancing level, …)	NA
SM – RT voltage	Every 10 minutes	Voltage estimation at SM level	V
SM – Forecasted Voltage	Every 2 hours for the time horizon	Voltage profile estimation at SM level during the time horizon	V
Asset – RT load	Every 10 minutes	Load estimation at Asset level	V
Asset – Forecasted load	Every 2 hours for the time horizon	Voltage profile estimation at SM level during the time horizon	V

Applied Techniques:

- Optimization algorithms.
- Classification and clustering algorithms.
- Regression algorithms.
- Neural networks models.

Open-source/Proprietary software: Proprietary software.

B.4.2 Grid validation platform, real-time

Service to be offered in real-time/offline:

Real-time with execution frequency of 10 minutes.

Input data requirements:

Table 74. Execution input data for Grid validation platform, real-time service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Smart Meter (SM) – ID	NA	Identification number of the SMs	NA	NA	DSO	R
Secondary Substation (SS) – ID	NA	Identification number of the SSs	NA	NA	DSO	R
SM – Type	NA	{“Single phase”; “Three-phase"}	NA	NA	DSO	R
Single phase SM Active Power load curve	10 min timestep, daily batch	10 minutes time step active power load curve recorded by the SMs under a specific SS	kW	> 1 month	DSO	R
Three phase SM Active Power load curve – Phase 1						
Three phase SM Active Power load curve – Phase 2						
Three phase SM Active Power load curve – Phase 3						
Single phase SM Reactive Power load curve	10 min timestep, daily batch	10 minutes time step reactive power load curve recorded by the SMs under a specific SS	kvar	> 1 month	DSO	O
Three phase SM Reactive Power load curve – Phase 1						
Three phase SM Reactive Power load curve – Phase 2						

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Three phase SM Reactive Power load curve – Phase 3						
Single phase SM Voltage Profile	10 min timestep, daily batch	10 minutes time step voltage profile recorded by the SMs under a specific SS	V	> 1 month	DSO	R
Three phase SM Voltage Profile – Phase 1						
Three phase SM Voltage Profile – Phase 2						
Three phase SM Voltage Profile – Phase 3						
SM – Location	NA	GPS coordinate of each SM	NA	NA	DSO	O
SS – Location	NA	GPS coordinate of each SS	NA	NA	DSO	O
Primary Substation (PS) - Location	NA	GPS coordinate of each SM	NA	NA	DSO	R
Nominal voltage limits	NA	Regulatory voltage limit for electricity supply	V	NA	DSO	R
SM – SS connection	NA	A priori SM – SS connection as recorded in legacy systems	NA	NA	DSO	R
SM – LV feeder connection	NA	A priori SM – LV feeder connection as recorded in legacy systems	NA	NA	DSO	R

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
SS – LV fuse rating	NA	Rating of each fuse of a specific SS	A	NA	DSO	R
SS – PS connection	NA	SS – PS connection as recorded in legacy systems	NA	NA	DSO	R
SS – PS feeder connection	NA	SS – PS feeder connection as recorded in legacy systems	NA	NA	DSO	R
PS – Active power	10 min timestep provided in real time	10 minutes averaged active power recorded at PS level	kW	Every 10 minutes	DSO	R
PS – Voltage profile	10 min timestep provided in real time	10 minutes averaged voltage recorded at PS level	V	Every 10 minutes	DSO	R
Meteorological data	10 min timestep provided in real time	Real-time meteorological data (temperature, sunshine)	NA	Every 10 minutes	Meteorological data provider	R

Open-source/Proprietary software: Proprietary software.

B.4.3 Flexibility platform for DER connection, planning

Service to be offered in real-time/offline:

Offline and On-demand.

Input data requirements:

Table 75. Execution input data for Flexibility platform for DER connection, planning service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Smart Meter (SM) – ID	NA	Identification number of the SMs	NA	NA	DSO	R
Secondary Substation (SS) – ID	NA	Identification number of the SSs	NA	NA	DSO	R
SM – Type	NA	{“Single phase”; “Three-phase"}	NA	NA	DSO	R
Single phase SM Active Power load curve Three phase SM Active Power load curve – Phase 1 Three phase SM Active Power load curve – Phase 2 Three phase SM Active Power load curve – Phase 3	10 min timestep, daily batch	10 minutes time step active power load curve recorded by the SMs under a specific SS	kW	> 1 month	DSO	R
Single phase SM Reactive Power load curve Three phase SM Reactive Power load curve – Phase 1 Three phase SM Reactive Power load curve – Phase 2	10 min timestep, daily batch	10 minutes time step reactive power load curve recorded by the SMs under a specific SS	kvar	> 1 month	DSO	O

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Three phase SM Reactive Power load curve – Phase 3						
Single phase SM Voltage Profile	10 min timestep, daily batch	10 minutes time step voltage profile recorded by the SMs under a specific SS	V	> 1 month	DSO	R
Three phase SM Voltage Profile – Phase 1						
Three phase SM Voltage Profile – Phase 2						
Three phase SM Voltage Profile – Phase 3						
SM – Location	NA	GPS coordinate of each SM	NA	NA	DSO	O
SS – Location	NA	GPS coordinate of each SS	NA	NA	DSO	O
Nominal voltage limits	NA	Regulatory voltage limit for electricity supply	V	NA	DSO	R
SM – SS connection	NA	A priori SM – SS connection as recorded in legacy systems	NA	NA	DSO	R
SM – LV feeder connection	NA	A priori SM – LV feeder connection as recorded in legacy systems	NA	NA	DSO	R
SS – LV fuse rating	NA	Rating of each fuse of a specific SS	A	NA	DSO	R

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Targets	NA	Specification of expected outcome from the call	NA	NA	Public Institutions	R
Area	NA	Specification of eligible area for the call	NA	NA	Public Institutions	R
Asset – ID	NA	Identification number of the asset	NA	NA	RES & flexibility investors	R
Asset – type	NA	Per asset, the type such as “Battery”, “PV panels”, “controllable load” …	NA	NA	RES & flexibility investors	R
Asset – locations	NA	List of SM – ID under which the asset is expected could be connected	NA	NA	RES & flexibility investors	R
Asset – size	NA	Sizing of the asset, (MIN / EXPECTED / MAX), unit according to its type (kWp for PV panels, (kW, kWh) for batteries…)	NA	NA	RES & flexibility investors	R

Expected outputs:

The output of the service is an optimized implementation plan of the asset proposed by the RES & flexibility investors. The optimization weights:

- The targets from the public institution.
- The expectations from the RES & flexibility investors.
- The constraints of the DSO.

The following table summarizes the different service outputs.

Table 76. Output data for Flexibility platform for DER connection, planning service.

Data	Lowest level of granularity	Description	Units
Asset – Optimal size	When executed		Same as “Asset – Size”
Asset – Optimal location	When executed	Selected SM – ID to connect the asset	NA
Asset – Operating Constraints	When executed	List of rules the asset must follow	

Applied Techniques:

- Optimization algorithms.
- Classification and clustering algorithms.
- Regression algorithms.

Open-source/Proprietary software: Proprietary software.

B.4.4 Passive consumption baseline prediction service

Service to be offered in real-time/offline:

The frequency of execution of the service will be daily, and the time of execution will depend on the market’s organization in the country where it will be demonstrated.

Input data requirements:

Table 77. Execution input data for Passive consumption baseline prediction service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Predicted ambient temperature	15’	Predicted ambient temperature at the location of the resources	ºC	1 day	Prosumer	O

Table 78. Training data for Passive consumption baseline prediction service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Historic active energy	15’	Historic active energy consumed by the passive resources	Wh	> 1 year	Prosumer	R
Historic ambient temperature	15’	Historic ambient temperature at the location of the resources	ºC	> 1 year	Prosumer	O

Expected outputs:

The output of the service is the 96-quarter-hour consumption forecast for the following day.

Table 79. Output data for Passive consumption baseline prediction service.

Data	Lowest level of granularity	Description	Units
Predicted active energy consumption	15’	Predicted active energy consumption	Wh

Applied Techniques:

The AI techniques applied in this service are:

- Preprocessing techniques to remove outliers, scale the dataset and clear faulty data.
- Regressive Algorithms such as Random Forest.

Open-source/Proprietary software: Proprietary software.

B.4.5 Active consumption prediction service

Service to be offered in real-time/offline:

The frequency of execution of the service will be daily, and the time of execution will depend on the market’s organization in the country where it will be demonstrated.

Input data requirements:

Table 80. Execution input data for HVAC Resources. Active consumption prediction service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Predicted ambient temperature	15’	Predicted ambient temperature at the location of the HVAC building	ºC	1 day	MAIA	R
Indoor temperature setpoint schedule	15’	Indoor temperature setpoint schedule at each control zone of the building (the zones for which the HVAC temperature setpoint is independent)	ºC	1 day	Flexibility service provider	R
HVAC utilization schedule	15’	HVAC utilization schedule (i.e. time periods in which the HVAC is either ON or OFF)	ON/OFF	1 day	Flexibility service provider	R

Table 81. Training data for HVAC Resources. Active consumption prediction service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Historic active energy	15’	Historic active energy consumed by the HVAC	Wh	> 1 year	MAIA	R

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Historic ambient temperature	15’	Historic ambient temperature at the location of the HVAC building	ºC	> 1 year	MAIA	R
Historic indoor temperature	15’	Historic indoor temperature at each thermal zone of the building (the rooms for which we would have internal temperature measurement)	ºC	> 1 year	MAIA	O
Historic indoor temperature setpoint	15’	Historic indoor temperature at each control zone of the building (the zones for which the HVAC temperature setpoint is independent)	ºC	> 1 year	MAIA	R
Historic HVAC utilization	15’	Historic data of when the HVAC has been ON or OFF	ON/OFF	> 1 year	MAIA	R

Table 82. Execution input data for pumps. Active consumption prediction service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Predicted Irradiance	15’	Predicted global Irradiance	W/m^2	1 day	MAIA	O
Predicted ambient temperature	15’	Predicted ambient temperature	°C	1 day	MAIA	O
Predicted precipitation	15’	Predicted precipitation	mm	1 day	MAIA	O
Start time	N/A	Start time(s) when the pumping system begins to work	Time	Discrete values	Flexibility service provider	R

Table 83. Training data for pumps. Active consumption prediction service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Historic active energy	15’	Historic active energy consumed by the pumps	Wh	> 1 year	MAIA	R

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Historic soil moisture	15'	Historic soil moisture (if the pump control system depends on soil moisture values)	%	> 1 year	MAIA	O
Historic Irradiance	15'	Historic global Irradiance	W/m^2	> 1 year	MAIA	O
Historic ambient temperature	15'	Historic ambient temperature	°C	> 1 year	MAIA	O
Historic precipitation	15'	Historic precipitation	mm	> 1 year	MAIA	O

Table 84. Execution input data for EVs. Active consumption prediction service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Utilization schedule	15’	Time periods during which the charging post is disabled (OFF). Being ON just implies that whenever an EV is connected, the fast charge process is allowed to begin.	ON/OFF	1 day	Flexibility service provider	R

Table 85. Training data for EVs. Active consumption prediction service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Historic active energy	15’	Historic active energy consumed at each charging post	Wh	> 1 year	MAIA	R
Historic EV identification	15’	Historic data about which EV is connected to the post (or there is no EV connected)	EV identifier	> 1 year	MAIA	R

Expected outputs:

The output of the service is the 96-quarter-hour consumption forecast for the following day.

Table 86. Output data for Active consumption prediction service.

Data	Lowest level of granularity	Description	Units
Resource predicted active energy consumption	15’	Predicted active energy consumption of a resource	Wh

Applied Techniques:

- The AI techniques applied in this service are:
- Pre-processing techniques to remove outliers, scale the dataset and clear faulty data.
- Regressive Algorithms such as Random Forest or Neural Networks.
- Classification Algorithms such as the K-nearest neighbour.

Open-source/Proprietary software: Proprietary software.

B.4.6 Intermittent DER generation resource baseline prediction service

Service to be offered in real-time/offline:

The frequency of execution of the service will be daily, and the time of execution will depend on the market’s organization in the country where it will be demonstrated.

Input data requirements:

Table 87. Execution input data for Intermittent DER generation resource baseline prediction service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Predicted Irradiance	15’	Predicted global Irradiance on PoA (Plane of Array) at the location of the PV generator	Wh/m^2	1 day	Prosumer	R
Predicted ambient temperature	15’	Predicted ambient temperature at the location of the PV generator	°C	1 day	Prosumer	R

Table 88. Training data for Intermittent DER generation resource baseline prediction service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Historic Irradiance	15’	Historic global Irradiance on PoA (Plane of Array) at the location of the PV generator	Wh/m^2	> 1 month	Prosumer	R
Historic ambient temperature	15’	Historic ambient temperature at the location of the PV generator	°C	> 1 month	Prosumer	R
Historic active energy	15’	Historic active energy produced by the PV generator	Wh	> 1 month	Prosumer	R

Expected outputs:

The output of the service is the 96 quarter-hour PV production forecast for the following day. The following table summarizes the different service outputs.

Table 89. Output data for Intermittent DER generation resource baseline prediction service.

Data	Lowest level of granularity	Description	Units
Predicted active energy production	15’	Predicted active energy production at the PV generator output	Wh

Applied Techniques:

The AI techniques applied in this service are:

- Preprocessing techniques to remove outliers, scale the dataset and clear faulty data.
- Regressive Algorithms such as Random Forest.

Open-source/Proprietary software: Proprietary software.

B4.7 Prosumer EMS internal optimization service

Service to be offered in real-time/offline:

The frequency of execution of the service will be daily, and the time of execution will depend on the market’s organization in the country where it will be demonstrated.

Input data requirements:

Table 90. Execution input data for Prosumer EMS internal optimization service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Predicted local renewable production	15'	Calculated by the service "Intermittent DER generation resource baseline prediction service"	kWh	1 day	FSP	R
Predicted consumption baseline of passive consumption	15'	Calculated by the service "Passive consumption baseline prediction service"	kWh	1 day	FSP	R
Estimated consumption of active consumption resources	15'	Calculated by the service "Active consumption resource prediction service"	kWh	1 day	FSP	R
Energy prices	15'	Prices agreed with the Supplier. Includes the energy and price constraints agreed with the supplier.	€/kWh	1 day	Prosumer	R

Expected outputs:

The output of the service is the 96 quarter-hour resource operation parameters schedule for the following day (one schedule for each flexible resource) and the predicted active energy consumption profile associated to that set of operation parameters.

Table 91. Output data for Prosumer EMS internal optimization service.

Data	Lowest level of granularity	Description	Units
Resource operation parameters schedule	15’	Depending on the kind of resource the resource operation schedule will be composed of different variables.	Pumps and EV charging posts --> ON/OFF, HVAC --> ON/OFF and room setpoint temperatures.
Resource active energy consumption profile	15’	Active energy profile of a resource, derived from the application of a set of operating parameters schedule.	kWh.

Applied Techniques:

The AI techniques applied in this service are:

- Optimization techniques (TBD).

Open-source/Proprietary software: Proprietary software.

B4.8 Flexibility order disaggregation service

Service to be offered in real-time/offline:

The frequency of execution of the service will be when the flexibility request from the DSO is received.

Input data requirements:

Table 92. Execution input data for Flexibility order disaggregation service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Predicted local renewable production	15’	Calculated by the service “Intermittent DER generation resource baseline prediction service”	kWh	1 day	FSP	R
Predicted consumption baseline of passive consumption	15’	Calculated by the service "Passive consumption baseline prediction service”	kWh	1 day	FSP	R

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Estimated consumption of active consumption resources	15'	Calculated by the service "Active consumption resource prediction service"	kWh	1 day	FSP	R
Energy prices	15'	Prices agreed with the Supplier. Includes the energy and price constraints agreed with the supplier.	€/kWh	1 day	MAIA	R
Energy constraints	15'	Energy constraints in the flexibility order sent by the DSO.	kWh	1 day	DSO	R
Available capacity	15'	Available capacity in the flexibility order sent by the DSO.	kWh	1 day	DSO	R

Expected outputs:

The output of the service is the 96 quarter-hour resource operation parameters schedule for the following day (one schedule for each flexible resource) and the predicted active energy consumption profile associated to that set of operation parameters.

Table 93. Output data for Flexibility order disaggregation service.

Data	Lowest level of granularity	Description	Units
Operation parameters schedule	15'	Depending on the kind of resource the control schedule will be composed of different variables.	Pumps and EV charging posts --> ON/OFF, HVAC --> ON/OFF and room setpoint temperatures.
Active energy consumption profile	15'	Active energy profile derived from the application of a set of operating parameters schedule.	kWh.

Applied Techniques:

The AI techniques applied in this service are:

- Optimization techniques (TBD).

Open-source/Proprietary software: Proprietary software.

B.4.9 Aggregated flexibility offer optimization service

Service to be offered in real-time/offline:

The frequency of execution of the service will be daily, and the time of execution will depend on the market's organization in the country where it will be demonstrated.

Input data requirements:

Table 94. Execution input data for Aggregated flexibility offer optimization service.

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Baseline consumption of the flexible resources in the FSP portfolio	15'	Provided by the FSP as the result of the "Prosumer EMS internal optimization service"	kWh	1 day	Flexibility Service Provider	R
Flexibility Request issued by the DSO	15'	Amount of active energy to be reduced or increased in a congestion zone (compared to the previously sent baseline calculated by the Active consumption prediction service)	kWh	1 day	DSO	R
Contract conditions agreed between FSP and Prosumer	N/A	Applicable both for the availability and for the utilization of each flexible resource	N/A	N/A	Flexibility Service Provider	R
Estimated consumption of active consumption resources	15'	Calculated by the service "Active consumption resource prediction service"	kWh	1 day	Flexibility Service Provider	R

Data	Granularity	Description	Units	Volume	Owner	Optional/ Required
Estimated passive consumption baseline prediction	15'	Calculated by the service "Passive consumption baseline prediction service "	kWh	1 day	FSP	O
Energy prices	15'	Prices agreed with the Supplier. Includes the energy and price constraints agreed with the supplier.	€/kWh	1 day	MAIA	R

Expected outputs:

The output of the service is the flexibility offer that will be sent to the DSO in response to the Flexibility Request.

Table 95. Output data for Aggregated flexibility offer optimization service.

Data	Lowest level of granularity	Description	Units
Flexibility offer volume	15'	Offered amount of active energy to be reduced or increased in the congestion zone defined by the Flexibility Request (compared to the previously sent baseline), from each FSP	kWh
Flexibility offer price	15'	Price of the offered amount of active energy to be reduced or increased in the congestion zone defined by the Flexibility Request (compared to the previously sent baseline), from each FSP.	€/kWh

Applied Techniques:

The AI techniques applied in this service are:

- Optimization techniques (TBD).

Open-source/Proprietary software: Proprietary software.