Grant Agreement N° 872592

PLATOON

Digital platform and analytic tools for energy

Deliverable D3.1

Data governance, security and privacy requirements for the energy sector

Contractual delivery date:
M12

Actual delivery date:
21st December 2020

Responsible partner:
P02: TECNALIA, Spain

Project Title	PLATOON – Digital platform and analytic tools for energy
Deliverable number	D3.1
Deliverable title	Data governance, security and privacy requirements for the energy sector
Author(s):	Valentín Sánchez

Responsible Partner:	P02 - TECNALIA
Date:	21.12.2020
Nature	R
Distribution level (CO, PU):	PU
Work package number	WP3 – Data Governance, Security and Privacy

Work package leader	IAIS, Germany
Abstract:	This report provides an accurate description of the captured data governance, security and privacy requirements for the PLATOON platform from the pilots. Additionally, the document maps these requirements into the IDS Reference architecture, roles and stake holders.
Keyword List:	Requirements, Platform, Reference Architecture, Interoperability, Data Governance, security and privacy requirements.

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

Editor(s):	Valentín Sánchez (TECN)
Contributor(s):	IAIS, ENG, IND.
Reviewer(s):	Hantong Liu (IAIS) Philippe Calvez (ENGIE) Erik Maqueda (TECN)
Approved by:	Philippe Calvez (ENGIE)
Recommended/mandatory readers:	Mandatory readers WP2-WP7 WP and Task Leaders.

Document Description

Version	Date	Modifications Introduced	
		Modification Reason	Modified by
v0.1	11/12/2020	First version	Valentín Sánchez (TECN)
V0.2	17/12/2020	Revision from IAIS	Hantong Liu (IAIS
V0.3	18/12/2020	Internal revision from TECNALIA	Erik Maqueda
V0.4	21/12/2020	Internal revision from ENGIE	Philippe Calvez
V0.5	23/12/2020	Updates document after Internal revision from ENGIE and IAIS.	Erik Maqueda

Table of Contents

List of Figures

List of Tables

Terms and abbreviations

API	Application Programming Interface
CA	Consortium Agreement
CD	Continuous Development
CI	Continuous integration
CQ	Continuous Quality
CO	Confidential
COSMAG	Comprehensive Architecture for Smart Grids
DAPS	Dynamic Attribute Provisioning Service
DoA	Declaration of Action
DM	Dissemination Manager
DoA	Description of Action
EC	European Commission
EM	Exploitation Manager
EtB	Ethical Board
GA	Grant Agreement
IA	Innovation Action
IDS	International Data Spaces
IoT	Internet of Things
IM	Innovation Manager
KR	Key Result
NFV	Network Function Virtualization
MEC	Multi-access Edge Computing
PM	Project Manager
PU	Public
QA	Quality Assurance
RDF	Resource Description Framework
RE	Restricted
SAREF	Smart Appliances REFerence Ontology
SC	Steering Committee
SGAM	Smart Grid Architecture Model
TA	Technical Annex
TM	Technical Manager
WP	Work package
WPL	Work package Leader

Executive Summary

Tomorrow's energy grids consist of heterogeneous interconnected systems, of an increasing number of small-scale and of dispersed energy generation and consumption devices, generating huge amounts of data. The electricity sector, in particular, needs big data tools and architectures for optimized energy system management under these demanding conditions.

The objective of PLATOON project is to develop a big data platform relying on interoperable ecosystem for the energy sector to leverage data and provide new analytics tools that enable the development of new services and business models that boost the decarbonization of the energy sector in line with the European Green Deal.

In the current context, it is necessary to build a cyber-secure digital platform that allows for large-scale multi-party data exchange, processing and monetisation, all governed by a clear data governance policy.

PLATOON platform is presented as a breakthrough COSMAG compliant platform with flexible capabilities covering three main pillars:

- **Interoperability** to ensure multiparty data exchange and deal with a wide spectrum and heterogenous data sources, formats, interfaces.
- **Data governance** to meet data/app providers requirements regarding security, privacy and sovereignty.
- **Data Analytics Toolbox** to extract value from the data and that can be easily used by energy domain experts without deep coding skills and mathematical knowledge.

This report provides an accurate description of the **Data governance, security and privacy requirements** for the PLATOON platform. It gathers the data governance, security and privacy requirements from the use cases, considering both the technical and business perspectives, and the state of the art.

The requirements gathered are mapped to the Security and Governance perspectives of the IDS reference architecture model and the set of applicable open standards, models and tools have been identified and analysed.

1 Introduction

The objective of PLATOON (Digital PLAtform and analytical TOOls for eNergy) project is to develop a big data platform relying on interoperable ecosystem for the energy sector to leverage data and provide new analytics tools that enable the development of new services and business models that boost the decarbonization of the energy sector in line with the European Green Deal.

PLATOON platform is presented as a breakthrough COSMAG compliant platform with flexible capabilities covering three main pillars:

- **Interoperability** to ensure multiparty data exchange and deal with a wide spectrum and heterogenous data sources, formats, interfaces.
- **Data governance** to meet data/app providers requirements regarding security, privacy and sovereignty.
- **Data Analytics Toolbox** to extract value from the data and that can be easily used by energy domain experts without deep coding skills and mathematical knowledge.

It is generally accepted that it is almost impossible to have a single big data platform that fits perfectly the necessities for all the stakeholders of the energy sector due to various reasons:

On the one hand, most of the PLATOON partners and the companies in general already have their own (usually more than one) legacy systems, business digital platforms, expert tools, etc. that have been specifically, over time and organization, tailored to meet their business and process requirements. Deploying new tools, or large platforms becomes difficult due to multiple specific constraints such as economic (cost), time to market and internal and external expectations, direct and indirect risks, technical existing legacy systems'contraints, etc. that will require time and money that companies are not willing to spend.

On the other hand, having a single platform would require the introduction of a third-party company that is responsible for managing the PLATOON platform and that would incur in further costs. In this sense, companies continuously manifest that they do not want to rely on a single platform provider due to the so-called vendor lock-in, but they want to have control over their platform and services instantiations and be able to change the platform provider if necessary.

Therefore, PLATOON aims to create what is known as a federated platform. A federated platform is a decentralized platform formed of different platforms from different companies that are able to exchange data and services with each other.

A governance framework will lay the grounds for secure data exchange to take place between stakeholders (consumers and providers) equipped with a compliant endpoint, or data connector. This exchange environment will be evolved into an open and trusted data marketplace.

Regarding security and privacy, PLATOON will mainly focus on developing the solutions to ensure security and privacy in the data exchange between platforms. As part of the project no specific security and privacy component will be developed for protecting the individual platforms from the different partners. Instead a "Platform Security and Privacy" guideline will be defined based on already available solutions.

This report provides an accurate description of the **Data governance, security and privacy requirements** for the PLATOON platform. It gathers the data governance, security and privacy

requirements from the use cases, considering both the technical and business perspectives, and the state of the art.

The requirements gathered are mapped to the Security and Governance perspectives of the IDS reference architecture model and the set of applicable open standards, models and tools have been identified and analysed.

These requirements are in line with the use case, data exchange and business exchange requirements defined in tasks T1.1, T1.2 and T1.4 defined in deliverables D1.1 and D1.2.

The report is structured in 4 main sections and a conclusion section.

- **Section 1** sets the context and purpose of this deliverable.
- **Section 2** includes a summary of the generic data governance, security and privacy requirements gathered in WP1, which will serve as key inputs for the requirements gathering process.
- **section 3** presents the IDS Data governance, security and privacy approach.
- **Section 4** includes the use cases data governance, security and privacy requirements.
- **Section 5** presents some conclusions.

2 Generic data governance, security and privacy requirements

This section presents a summary of the main findings in WP1 regarding data governance, security and privacy requirements., including the business perspectives. These findings set the base to analyse and gather the pilot specific requirements, taking into account each particular business model.

2.1 Data economy

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

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

Three main distinct types of big data business roles are identified: data users, data suppliers, and data facilitators.

- **Data users** are organisations that use data either for informing business decisions, or as an input into other products and services such as credit reports or targeted advertising campaigns.
- **Data suppliers** are organisations that either generate data that is of intrinsic value and therefore marketable, or else serve a kind of brokerage role by providing access to an aggregation of first- and third-party data.

- **Data facilitators** encompasses the range of activities that support third parties that are lacking in infrastructure or expertise including advice on how to capitalise on big data, the provision of physical infrastructure, and the provision of outsourced analytics services[ii].

2.2 Data economy key factors

The EC has published a proposal for the so called "Data Governance Act", which is a regulation on the subject of data sharing and the creation of the European data space[1]. The problem that this initiative addresses is that data sharing in the EU remains limited, in spite of the potential benefits of such sharing for the economy and for society. Three main reasons for this have been identified:

- low trust in data sharing
- difficulties in reusing certain public-sector data and collecting data on altruistic grounds
- technical obstacles to reusing data.

The instrument aims to foster the availability of data for use by increasing trust in data intermediaries and by strengthening data-sharing mechanisms across the EU.

The instrument would address the following situations among others:

- Sharing of data among businesses, against remuneration in any form.
- Allowing personal data to be used with the help of a 'personal data-sharing intermediary', designed to help individuals exercise their rights under the General Data Protection Regulation (GDPR).

According to the recent communication from the European Commission "A European strategy for data[2]", several issues are holding the EU back from realising its potential in the data economy:

- **Availability of data**: The value of data lies in its use and re-use. Currently there is not enough data available for innovative re-use, including for the development of artificial intelligence. Specifically, in the B2B data-sharing context (sharing and use of privately-held data by other companies), in spite of the economic potential, data sharing between companies has not taken off at sufficient scale. This is due to a ***lack of economic incentives*** (including the fear of losing a competitive edge), ***lack of trust between economic operators*** that the data will be used in line with contractual agreements, imbalances in negotiating power, the ***fear of misappropriation of the data by third parties***, and a lack of legal clarity on who can do what with the data (for example for co-created data, in particular IoT data).
- **Data Governance:** It is necessary to reinforce the governance of data use in society and the economy. For these data spaces to become operational, organisational approaches and structures (both public and private) are needed that enable data-driven innovation on the basis of the existing legal framework.
- **Empowering individuals to exercise their rights*:*** Individuals value the high level of protection granted by the **GDPR** and ePrivacy legislation. However, they suffer from the absence of technical tools and standards that make the exercise of their rights simple and not overly burdensome.

[1] Proposal for a Regulation on European data governance (Data Governance Act) | Shaping Europe's digital future (europa.eu)
[2] https://ec.europa.eu/digital-single-market/en/european-strategy-data

- **Cybersecurity:** The new data paradigm where less data will be stored in data centres, and more data will be spread in a pervasive way closer to the user 'at the edge', brings new challenges for cybersecurity. It will be **essential to preserve data security when data are being exchanged**. Ensuring the continuity of access controls (i.e. how security attributes of data are managed and respected) across data value chains will be a key, but demanding, pre-requisite to foster data sharing and ensure trust among the different actors of European data ecosystems.

The Platoon strategy regarding data governance, security and privacy tackles these issues by providing an integrated data governance framework that ensures data security and respects data privacy and sovereignty, as a basis for secure data sharing (including the exchange of data and data services) to be carried out in PLATOON. The governance model and concepts promoted by IDS will be adapted to fit the project's energy-specific data and stakeholders.

It will design and implement a data security model and framework capable of managing the security and privacy requirements. The security and privacy model will include generic data security issues (secure communication, identity management), data access control (authorization management) and data usage control aspects, inspired by the most advanced state-of-the-art concepts (particularly the IDS reference architecture).

Security Policies will enable data providers to attach metadata to the data exchanged (directly or using a tool or service) to ensure data sovereignty is respected at the required level of granularity (e.g., no re-sharing of data with third parties, limited use of data to specific recipient devices, or within specific geographical locations, etc.).

Security models will also ensure that personal data can either be shared by the person that they represent directly and securely in line with legal constraints and the GDPR, or that data that can directly or indirectly identify a person can be fully depersonalised (anonymization, obfuscation, etc.) ahead of being shared by third parties.

2.2.1 Barriers to data exchanges

Participation in data sharing and exchange remains difficult as data producers feel their ownership rights cannot be guaranteed. Different strategies and governance models have been proposed to overcome this challenge, including different ownership models and data rights management.

There are technical and legal compliance challenges that we have compiled and are detailed in the table below:

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

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

Table 1: Data exchange barriers

2.3 Data security requirements

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

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

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

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

Following sections describes the three main aspects of Data Security: Secure communication, Data access control and Data usage control.

2.3.1 Secure communication

Secure communication among parties it is crucial to avoid data leakage. Unsecure communication may result in the capture of plain-text credential by an attacker.

Several techniques can be adopted to ensure secure communication among parties. In order to guarantee, for instance, the identity of the communication parties and the authenticity of the transmitted contents it is recommended to rely on a Public Key Infrastructure (PKI).

Applications of PKI are, among the others, digital signature that are used to ensure authenticity and integrity for sharing documents and HTTPS that is used to exchange encrypted data and assure confidentiality, authenticity and integrity of the data.

Other approaches that can guarantee secure communication are VPNs by which the data transported is encrypted and the endpoints that belongs to the VPN are authenticate and authorised among each other.

2.3.2 Data access control

Authentication and authorization capabilities are critical aspect to support services and applications. An access control policy is defined as sets of conditions that define whether users have access granted

to a protected resource. The authorization function can support different mechanisms, such as Access Control List (ACL), Role Based Access Control (RBAC), Attribute Based Access Control (ABAC), etc. Among the authentication and authorization standard solutions we can mention Security Assertion Markup Language (SAML), eXtensible Access Control Markup Language (XACML), Open Digital Rights Language (ODRL), JSON Web Token (JWT), OAuth2 and OpenID.

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

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

2.3.3 Data usage control

Usage control is an extension to traditional access control. It is about the specification and enforcement of restrictions regulating what must (not) happen to data. Thus, usage control is concerned with requirements that pertain to data processing (obligations), rather than data access (provisions). Usage control is relevant in the context of intellectual property protection, compliance with regulations, and digital rights management.

Figure 1 Data usage control (*Source: IDS Reference Architecture 3.0***)**

Next figure shows an example of data usage enforcement flow:

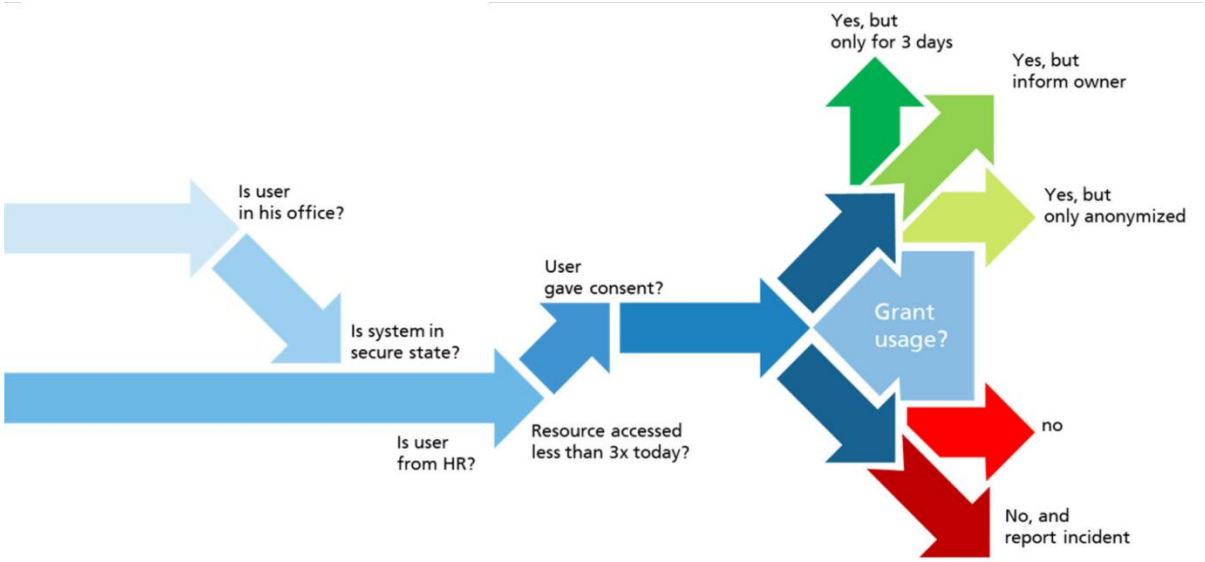


Figure 2 Data Usage flow

2.4 Data governance requirements

A data governance model defines a framework of decision-making rights and processes with regard to the definition, creation, processing, and use of data.

Data is a valuable resource in any digital, data-driven business and is necessary to enable participants to leverage the potential of their data within a secure and trusted business ecosystem.

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

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

In the context of data sharing, Data Governance is concerned with data lifecycle management decisions that ensure the safe, fair and secure (determined by pre-defined rules, legislative requirements, etc.) handling of data within and across a network of nodes over which data is passed on to fulfil pre-identified data value chains. In a data ecosystem that involves data exchange between distinct entities to fulfil such data value chains

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

IDS define the following data governance perspectives:

1. Data as an "Economic Good"
2. Quality
3. Provenance
4. Ownership
5. Sovereignty

2.5 Data privacy requirements

When it comes to data privacy the main requirement is the General Data Protection Regulation (GDPR) compliance. The GDPR is a European Union law that was implemented May 25, 2018 and requires organizations to safeguard personal data and uphold the privacy rights of anyone in EU territory.

The regulation includes seven principles of data protection that must be implemented and eight privacy rights that must be facilitated:

1. **Lawfulness, fairness and transparency** — Processing must be lawful, fair, and transparent to the data subject.
2. **Purpose limitation** — You must process data for the legitimate purposes specified explicitly to the data subject when you collected it.
3. **Data minimization** — You should collect and process only as much data as absolutely necessary for the purposes specified.
4. **Accuracy** — You must keep personal data accurate and up to date.
5. **Storage limitation** — You may only store personally identifying data for as long as necessary for the specified purpose.
6. **Integrity and confidentiality** — Processing must be done in such a way as to ensure appropriate security, integrity, and confidentiality (e.g. by using encryption).
7. **Accountability** — The data controller is responsible for being able to demonstrate GDPR compliance with all of these principles.

The GDPR recognizes a set of privacy rights for data subjects, which aim to give individuals more control over the data they loan to organizations:

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

In the context of PLATOON two main models of dealing with data privacy issues have been identified:

1. Avoiding personal data transfer by data anonymization techniques.
2. Assuring GDPR compliance via data access and usage restrictions.
    a. Ensuring that the data provider has the necessary user consent to use the personal data.
    b. Compliance with GDPR user rights is ensured via data access and data usage enforcement by the IDS connectors.

3 IDS Data governance, security and privacy approach

PLATOON has identified the IDS as the of-choice 'de facto' standard for data sharing ecosystems, the project will adhere to the IDS governance model, which governs and determines usage rights of data exchanged within IDS-compliant ecosystems.

The international data spaces (IDS) is a virtual data space leveraging existing standards and technologies, as well as governance models well-accepted in the data economy, to facilitate secure and standardized data exchange and data linkage in a trusted business ecosystem. It thereby provides a basis for creating smart-service scenarios and facilitating innovative cross-company business processes, while at the same time guaranteeing data sovereignty for data owners.

The overall goal is to make more data available to more organizations and ecosystems being aware, that data availability and exchange become a critical success factor for national and international economies. The key idea is to attach data sovereignty to data to enable even the exchange of sensible data, which often is some of the most valuable. For this reason, the International Data Spaces Association (IDSA) has defined a data sharing scheme including a reference architecture and a set of agreements to be used for creating and operating virtual data spaces.

The requirements apply to the data exchange process. The internal data governance, security and privacy requirements are not covered by IDS. The following topics are covered by the IDS sharing schema.

1. Data security
   - Confidentiality
   - Integrity
   - Availability
   - Authorized users

2. Data governance perspective
   - Data as an economic good
   - Data provenance
   - Data Quality
   - Data ownership and sovereignty.
     a) Data usage constraints

3. Data privacy. Although, it is not specifically mentioned, it can be treated as follows:
   - GDPR compliancy: Modelling protection of personal data (e.g., according to GDPR) using the W3C Data Privacy Vocabulary
   - Anonymization (IDS Data app integrated in the connector)
   - Data usage constraints: include the consent from the user
   - Managing user consent: CAPE, MyData.

Next sections provide an overview of the IDS approach regarding data security, governance and privacy.

3.1 IDS infrastructure

To ensure Governance and compliance, IDS defines roles, functions, and processes that need to be met by the business ecosystem to achieve secure and reliable interoperability. Next diagram shows the basic interaction taking place in the IDS ecosystem.

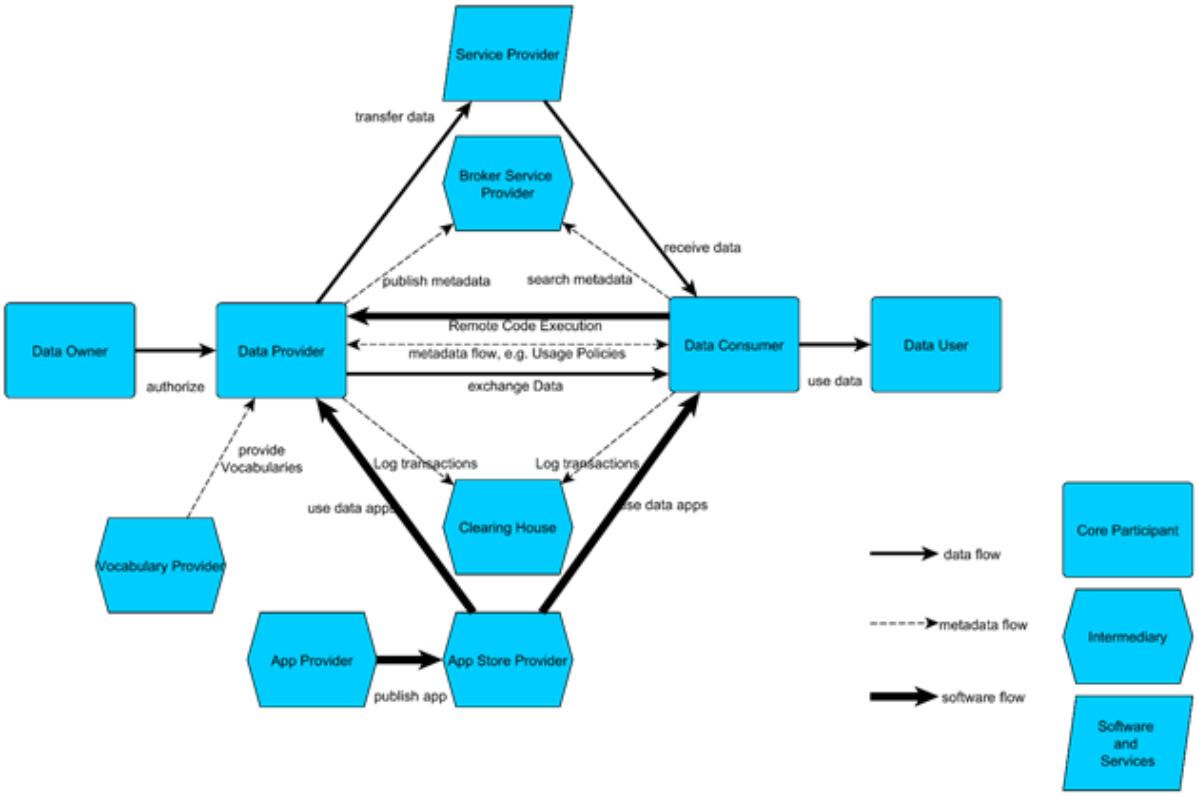


Figure 3 General structure of IDS Reference Architecture Model

In IDS, the fundamental mode of communication between a Data provider and Data Consumer happens via an **IDS Connector**, a dedicated software component, which acts as a communication interface for both data providers/consumers and app providers. They use a common language, **Information Model**, to communicate with each other. The information model facilitates compatibility and interoperability, thereby all connectors can exchange data. Furthermore, the vocabulary provider can be used to integrate **domain-specific data vocabularies**.

Next figure shows the internal structure of the connector, including the data access/usage module. In the figure IDSCP protocol is used for communication, which is the most secure protocol option in IDS. Currently, IDS is analysing the inclusion of other communication protocols like for example a REST based protocol. The addition of this new protocol is interesting for PLATOON since the Analytic toolboxes are going to expose their functionalities via OpenAPI REST interfaces and it will facilitate the adoption of IDS for data transfer needed for data analytic services.

The Data services implement the link between the IDS connector and the data sources inside the company (backend) and also the data processing applications provided by the App providers.

The core container provides the core IDS connector functionalities, including routing capabilities, communication protocol deployment, security and identity management (see Figure 5 IDS Identity management processes) and data access and usage capabilities.

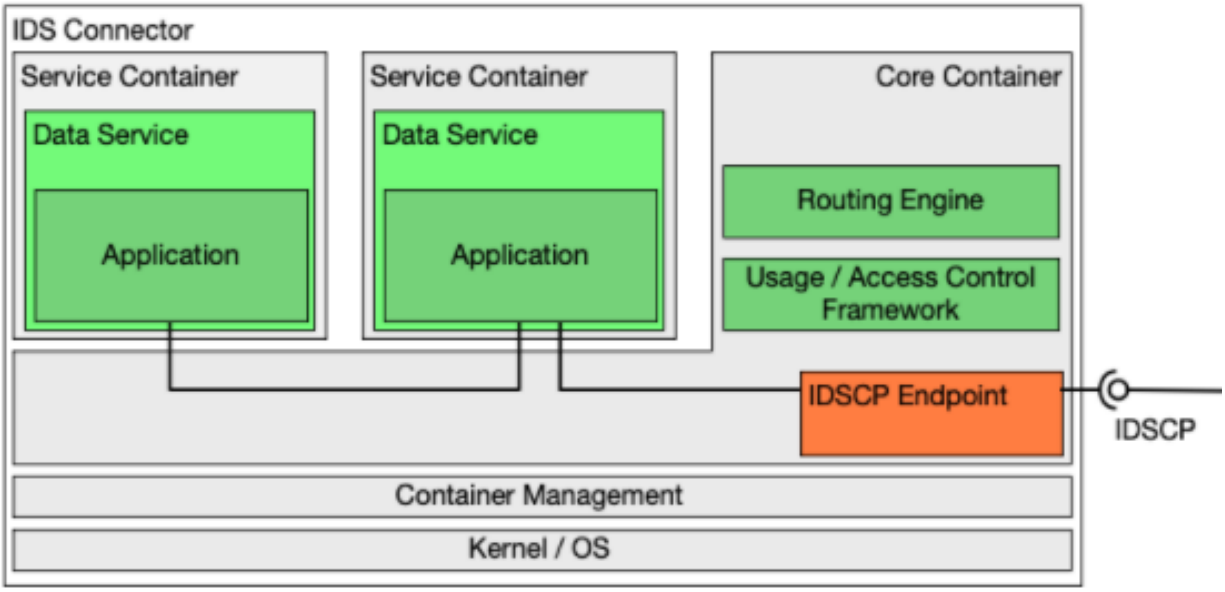


Figure 4 IDS Connector internal architecture.

Next, the main roles and components and services related to data governance, security and privacy are described:

DATA OWNER/PROVIDER:

The Data Provider makes data available for being exchanged between a Data Owner and a Data Consumer. The Data Provider is in most cases identical with the Data Owner, but not necessarily. To submit metadata to a Broker, or exchange data with a Data Consumer, the Data Provider uses software components that are compliant with the Reference Architecture Model of the International Data Spaces.

Exchanging data with a Data Consumer needs not necessarily be the only activity of the Data Provider. At the end of a data exchange transaction completely or partially executed, for example, the Data Provider may log the details of the successful (or unsuccessful) completion of the transaction at a Clearing House (see below) to facilitate billing or resolve a conflict.

In order to realise an IDS-compliant governance model, the following list of requirements needs to be met by the data provider:

- needs to define the usage constraints for data resources
- transfer data with usage constraints linked to data
- send/receive information about data transaction from Clearing House
- bill data, if required
- monitor policy enforcement
- manage data quality

- describe data source and provide the necessary data models for the data set.
- authorize data provider, if data provider is not the data owner

DATA USER/CONSUMER:

The Data Consumer receives data from a Data Provider. From a business process modelling perspective, the Data Consumer is the mirror entity of the Data Provider; the activities performed by the Data Consumer are therefore similar to the activities performed by the Data Provider.

Like a Data Provider, the Data Consumer may log the details of a successful (or unsuccessful) data exchange transaction at a Clearing House, use Data Apps to enrich, transform, etc. the data received, or use a Service Provider to connect to the International Data Spaces (if it does not deploy the technical infrastructure for participation itself). In order to realise an IDS-compliant governance model, the following list of requirements needs to be met by the data consumer:

- use data in compliance with usage constraints
- send/receive information about data transaction from Clearing House
- monitor policy enforcement

CLEARING HOUSE

The Clearing House is an intermediary that provides clearing and settlement services for all financial and data exchange transactions.

The Clearing House logs all activities performed in the course of a data exchange. After a data exchange, or parts of it, has been completed, both the Data Provider and the Data Consumer confirm the data transfer by logging the details of the transaction at the Clearing House. Based on this logging information, the transaction can then be billed. The logging information can also be used to resolve conflicts (e.g., to clarify whether a data package has been received by the Data Consumer or not). The Clearing House also provides reports on the performed (logged) transactions for billing, conflict resolution, etc.

The clearing house plays a central role to provide data provenance, data quality management and data monetization capabilities.

IDENTITY PROVIDER

Even though they are not included explicitly in the architecture, there are a set of Essential Services needed to deploy the IDS ecosystem:

- Certification Body (CB)
- Certification Authority (CA) (provisioning of X.509 certificates)
- Dynamic Attribute Provisioning Service (DAPS) (OAuth compatible)
- Participant Information System (ParIS)

- Dynamic Trust Management (DTM)

All these services have been included in the role called **IDENTITY PROVIDER.**

The Identity Provider should offer a service to create, maintain, manage, monitor, and validate identity information of and for participants in the International Data Spaces. This is imperative for secure operation of the International Data Spaces and to avoid unauthorized access to data.

The Identity Provider consist of a **Certification Authority** (managing digital certificates for the participants of the International Data Spaces), and a **Dynamic Attribute Provisioning Service** (DAPS, managing the dynamic attributes of the participants), a service named **Dynamic Trust Monitoring** (DTM) for continuous monitoring of the security and behavior of the network and the **Participant Information Service** manage the participants (companies) related information and certification.

Finally, four more roles complete the whole IDS picture:

- The Broker Service provider
- App Store Provider and the App provider.
- The Vocabulary provider

Even though these roles are not directly related to the data governance, security and privacy requirements when transferring data, they provide some intermediation services needed to deploy the IDS whole vision.

BROKER SERVICE PROVIDER: The Broker Service Provider is an intermediary that stores and manages information about the data sources available in the International Data Spaces. The activities of the Broker Service Provider mainly focus on receiving and providing metadata. The Broker Service Provider must provide an interface for Data Providers to send their metadata. The metadata should be stored in an internal repository for being queried by Data Consumers in a structured manner. While the core of the metadata model must be specified by the International Data Spaces, a Broker Service Provider may extend the metadata model to manage additional metadata elements.

Metadata includes the so-called **security profile** which includes the set of security guarantees claimed by a Connector. Connectors may evaluate their mutual technical reliability and trustworthiness by evaluating each other's security profile.

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

APP STORE PROVIDER AND THE APP PROVIDER: App Providers develop Data Apps to be used in the International Data Spaces. These are applications that can be deployed inside the Connector, the core technical component required for a participant to join the International Data Spaces. Data Apps facilitate data processing workflows. Data Apps can be certified by a Certification Body in order to increase trust in these applications (especially with regard to Data Apps processing sensitive information).

Each Data App must be published in the App Store for being accessed and used by Data Consumers and Data Providers. App Providers should describe each Data App using metadata with regard to its semantics, functionality, interfaces, etc.). The App Store is responsible for managing information about

Data Apps offered by App Providers. The App Store should provide interfaces for publishing and retrieving Data Apps plus corresponding metadata.

Although Data Apps could implement some data analytics or machine learning functionalities, they are linked to the data transfer process and they must be deployed within the IDS connector. Therefore, due to those limitations, in PLATOON the analytic toolboxes are not going to be implemented as IDS data apps.

However, some data processing facilities could be implemented as data apps, including for example anonymization for personal data, data filtering and cleaning or data aggregation functionalities.

VOCABULARY PROVIDER: The Vocabulary Provider manages and offers vocabularies (i.e., ontologies, reference data models, or metadata elements) that can be used to annotate and describe datasets. In particular, the Vocabulary Provider provides the Information Model of the International Data Spaces, which is the basis for the description of data sources. In addition, other domain specific vocabularies can be provided. In Platoon, the deployment of a vocabulary provider to provide the Energy data models being defined is still under analysis.

3.2 IDS Security architecture

According to the IDS Reference Architecture the IDS Security approach addresses seven key security concepts:

1) secure communication,
2) identity management,
3) trust management,
4) trusted platform,
5) data access control,
6) data usage control and
7) data provenance tracking.

3.2.1 Secure communication

To ensure confidentiality and authenticity of data transfers, communication between Connectors must be protected. When using the IDS Connector, two layers of security are in place:

Point to point encryption (between Connectors), using an encrypted tunnel: State of the art TLS encryption of all messages is required by the IDS. Non-TLS requests, for instance plain HTTP, must always be rejected. Furthermore, certificates and keys used for the communication encryption must be valid and signed by a trustworthy CA. A client must not expect a reasonable response for non-encrypted requests, even though the server should respond with a redirect to the HTTPS URL.

End-to-end authorization (authenticity and authorization): The Security aspect covers, among other things, the authorization features of the client (e.g., JSON Web token) and references to the contract underlying the interaction. The IDS transfers and validates identity claims using the Dynamic Attribute Token (DAT). The DAT contains, among other attributes, the signed identity of an IDS Connector. Each IDS Connector involved in an interaction, both the origin server and the user agent, must verify the opposite's DAT by interacting with the Identity Provider (DAPS).

These two layers covers the two main security requirements for data transfer:

Confidentiality: This means that it allows individuals to see only the data which they are supposed to see. Confidentiality has several different aspects:

- Privacy of Communications
- Secure Storage of Sensitive Data
- Authenticated Users
- Granular Access Control

Integrity: A secure system ensures that the data it contains is valid. Data integrity means that data is protected from deletion and corruption, both while it resides within the database, and while it is being transmitted over the network.

3.2.1.1 IDSCP protocol security

The IDSCP serves as a secure tunnel that supports three phases:

- Metadata exchange. including Dynamic Attribute Token (DAT) and Information Model serialization (Connector Self-Description).
- Remote Attestation (Software integrity verification bidirectionally between server and client.
- Payload transmission

Where the payload can be arbitrary streams of data. This way, the integrity verification and authentication are performed once at connection setup and transmitted data is bound to the channel. This avoids the possibility of man-in-the-middle attacks.

3.2.2 Identity management

The IDS identity concept provides means to handle dynamically changing attributes without the need for certificate revocation and reissuance. The identity concept for devices is a pluggable and modular concept, building on top of a traditional PKI foundation. Dynamic attributes (such as geographic location) are provided by the Dynamic Attribute Provisioning Service (DAPS). Identity certificates can be created for persons, devices and services.

IDS identity management is designed to create trust chains that support the creation, evaluation and acceptance of software artefacts by creating personal signatures after each step (publication and evaluation stages). The concept is designed with flexibility in mind: Device Certification Authorities (CAs), which are responsible for issuing, validating and revoking digital certificates and identities, are supported by either using a central infrastructure CA or by using company and vendor specific CAs. eIDAS support for users is also part of the concepts, incorporating the possibility of introducing Trust Service Providers into the scheme.

Next figure shows the Identity management processes.

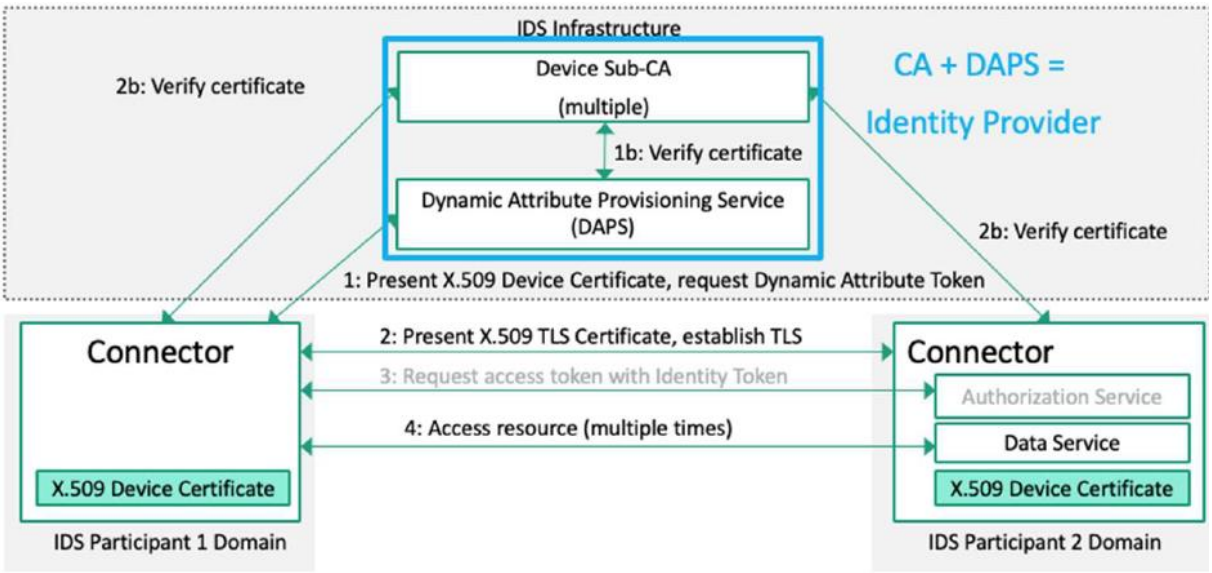


Figure 5 IDS Identity management processes

Each International Data Spaces connector has a private key with a corresponding X509v3 certificate (device certificate).

In contrast to conventional PKI-based enterprise IDM systems, these static certificates are however used for authentication only and not for the exchange of identity attributes.

Instead, these are exchanged using dynamic tokens that the connectors obtain from an attribute server. It administers self-descriptions and attested (certified) attributes of the connectors and issues tokens as needed for the required attributes of a connector.

Issuing the static X509v3 certificates is therefore decoupled from identity attributes which may change over time (for example due to certification).

3.2.2.1 Certifications

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

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

At this time, it is expected that no certification will be obtained during the project.

For validation purposes a “fictitious” Certification Authority can be created that will act as the certification body. Formal certification will be sought after project completion by individual partners that are interested in exploiting IDS components in the PLATOON ecosystem.

3.2.2.2 Dynamic Attribute Provisioning Service (DAPS)

DAPS is an attribute server that issues OAuth2 access tokens to International Data Spaces connectors. The connectors need these to access the services and data of other connectors.

The Fraunhofer DAPS can be accessed at https://daps.aisec.fraunhofer.de and implements RFC7523 JWT bearer client authentication for OAuth2.

This protocol allows the connectors to authenticate themselves on DAPS with their X509v3 certificate and to receive an access token in exchange which they can use to access other connectors.

Here, the decision about permitted access is not made by DAPS but always by the requested connector itself. DAPS merely administer the International Data Spaces attributes of the registered connectors.

3.2.3 Connector security profiles

The Connector shall provide a sufficiently high degree of trust and security regarding the integrity, confidentiality and availability of information exchanged in the IDS.

On the other hand, to ensure a low entry barrier specifically suitable for SME's and a scalable certification to meet high information security requirements, three different security levels, with an increasing extent of the security requirements that need to be fulfilled, have been defined:

- The Base Security Profile offers basic security features to protect against attackers from outside, to ensure integrity and availability. It is therefore designed for use in scenarios with only low security requirements. A Connector meeting this profile is suitable for exchanging data with limited trust and security needs, for exchange of data in a contained environment or for demonstration purposes.

- The Trust Security Profile includes strict container isolation, integrity-protected logging, encryption of all persisted data, protection against accidental misuse by administrators. This profile is used for scenarios in which the protection of the processed and transmitted data is essential.

- In comparison to the Trust profile, the Trust+ Security Profile also offers additional protection against misuse of privileged access, i.e. manipulation by administrators. This includes the protection against insider attacks as well as against external attackers who could gain privileged access. This is achieved by actively monitoring users and data on behalf of the data owner.

3.2.4 IDS Security infrastructure

Today, the IDSA Support Organization is conducted as an activity in the IDSA Head Office to support the following processes:

- Act as Certification Body
- Enabling the essential services:
    - CA
    - DAPS
    - ParIS

Each partner deploying an IDS connector must

- Request a valid certificate from IDSA.
- Register the connector in DAPS
- Configure the connector to use both the certificate and DAPS.

3.3 Data governance

Governance in the International Data Spaces comprises five aspects: data as an economic good, data ownership, data sovereignty, data quality, and data provenance.

Data Quality: Quality is commonly interpreted as "fitness for use", emphasizing the contextual nature of quality. In IDS, data Consumers can assess the fitness of a data offering for their needs based on quality statements supplied alongside with the Resource. These are, among other things, quality assessments according to a multidimensional model (e.g. ISO/IEC 25012 data quality model34), a certificate of quality, or any form of community feedback.

Data as an economic good: Today, data is traded in the market like a commodity; it has a price, and many companies monitor the costs incurred for data management. IDS tracking capabilities provide the means to audit data exchanges and assign economic value to those transfers. The International Data Spaces covers the information system perspective and provides the components that enable participants to define individual business cases.

Data ownership and sovereignty: According to the IDS Reference Architecture, data sovereignty is the capability of an entity (natural person or corporate) of being entirely self-determined with regard to its data, i.e. the ability to retain control over how data is used.

From the technical point of view data sovereignty must be assured all along the data life cycle:

- **In-Transit**: TLS encryption
- **At-Rest**: Storage encryption, client-side encryption of data, Access control
- **During computation**: Confidential Computing
- **When sharing data**: IDS connector / Usage Control

Data provenance: Data provenance tracking is closely related, but also complementary to distributed data usage control. It has its origins in the domain of scientific computing, where it was introduced to trace the lineage of data. Data provenance tracking thereby allows finding out when, how and by whom data was modified, and which other data influenced the process of creating new data items. The focus of data provenance tracking is on transparency and accountability.

3.3.1 Data usage control

In IDS the usage control is considered as an extension of the access control[vii]. Moreover, the usage control goal is to enforce the execute policies to data after the access has been granted controlling how data is processed, aggregated or forwarded.

Data usage control in the IDS basically works by attaching data usage policy information to data being exchanged and continuously controlling the way data is processed, aggregated, or forwarded to other endpoints. This data-centric perspective allows Data Providers to continuously control data flows,

rather than accesses to services. At configuration time, data usage policies support developers and administrators in setting up correct data flows. At runtime, data usage control enforcement prevents IDS Connectors from handling data in an undesired way (for example, by forwarding personal data to public endpoints).

Thus, data usage control is both a tool for system integrators to ensure they are not building an architecture that violates security requirements, and an audit mechanism providing evidence of compliant data usage.

A Usage Contract formalizes the expectations regarding the behaviour of Participants involved in a data exchange transaction in a declarative, technology-agnostic way. A data provider could share the data with a data consumer, provided they agree upon a **Data Usage Policy**. Data Usage policy specifies a Contractual agreement, which defines rules and conditions on how to use the data from Data Owner, which needs to be established between a Data Provider and a Data Consumer to exchange data in the PLATOON ecosystem.

3.3.1.1 Data usage IDS Infrastructure

In order to use the Data usage functionality of IDS each pilot must:

- Deploy a connector with data usage functionality.
- Define the data usage policy model using the IDS contracts and policy model templates available: https://github.com/International-Data-Spaces-Association/InformationModel/tree/develop/examples/contracts-and-usage-policy/templates/

3.3.2 Data provenance/quality

Data provenance (also referred to as “data lineage”) is metadata that is paired with records that details the origin, changes to, and details supporting the confidence or validity of data. Data provenance is important for tracking down errors within data and attributing them to sources. Additionally, data provenance can be useful in reporting and auditing for business and research processes.

Put most simply, the provenance of data helps to answer questions like “why was data produced,” “how was data produced,” “where was data produced,” “when was data produced,” and “by whom was data produced.” An example model of common provenance types can be seen below.

Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness.

The primary metadata points related to data provenance include:

- Data origin
- How data is transformed or augmented
- Where data moves over time

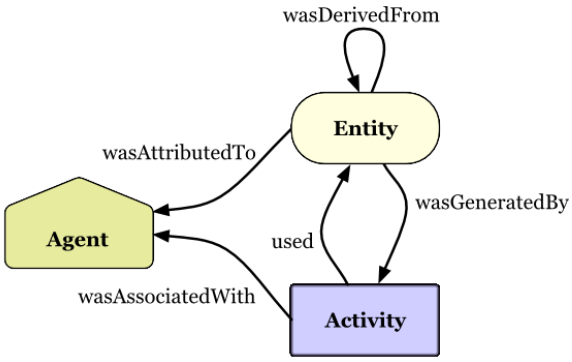


Figure 6: Data provenance/quality

By creating transparency and offering clearing functionality, the International Data Spaces provides a way to track the provenance and lineage of data. This is strongly linked to the topics of data ownership and data sovereignty. Data provenance tracking can be implemented with local tracking components integrated into IDS Connectors and a centralized provenance storage component attached to the Clearing House, which receives all logs concerning activities performed in the course of a data exchange transaction, and requests confirmations of successful data exchange from the Data Provider and the Data Consumer. In doing so, data provenance is always recursively traceable. In, addition provenance information can be integrated into the IDS Vocabulary, so as to enable the participants to maintain data provenance as part of the metadata during the process of data exchange.

Additionally, in IDS, data Consumers can assess the fitness of a data offering for their needs based on **quality statements** supplied alongside with the Resource. This quality statements can be tracked by the **clearing house**.

The International Data Spaces thereby provides the possibility to implement and use appropriate concepts and standards. However, it does not force participants to use these concepts and standards.

3.3.3 IDS data governance related models

The IDS information model includes the *Commodity* concern which helps to assess the value and utility of a Resource as an obtainable asset regarding a client's needs. It addresses questions like:

- Does the Resource origin from a reliable source?
- What level of quality does the Resource have?
- What are the restrictions regarding the use of the Resource?
- How much does it cost to use the Resource?

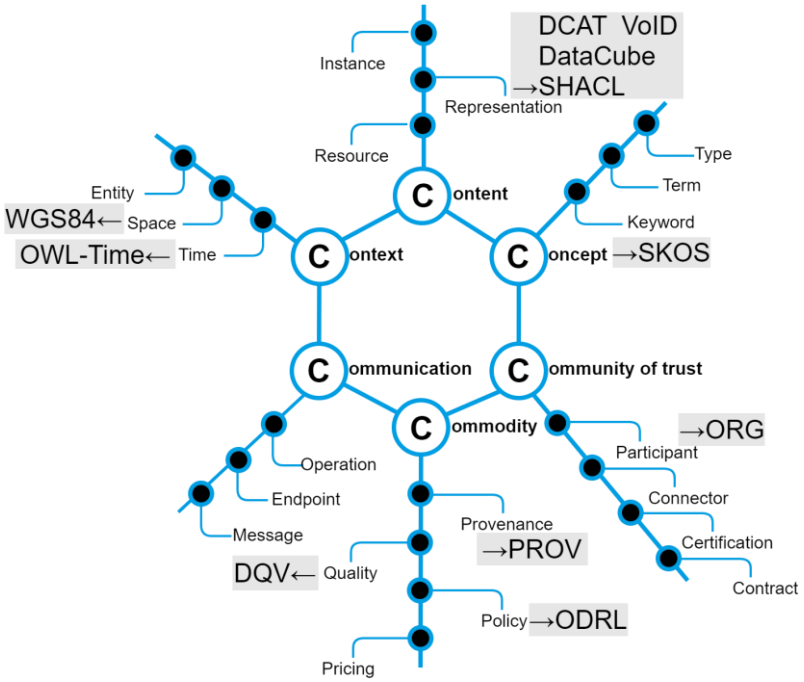


Figure 7 IDS Information model

The IDS information model[3] includes a specific model for each aspect of the commodity concern:

Data provenance: Explicates the context of the Resource's creation and its history of modification. Data provenance IDS model reuses the W3C PROV-DM specification. According to PROV-DM, provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness. PROV-DM is the conceptual data model that forms a basis for the W3C provenance (PROV) family of specifications.

https://github.com/International-Data-Spaces-Association/InformationModel/blob/develop/docs/provenance/provenance-en.ttl

Data quality: It may be assessed by means of tests, quality of service (QoS) parameters, and ratings from previous users in the community. Data provenance IDS model uses the W3C Data on the Web Best Practices: Data Quality Vocabulary (DQV).

https://github.com/International-Data-Spaces-Association/InformationModel/tree/develop/model/quality

Data usage and pricing: The policy determines the conditions for using the Resource, including Pricing, in a formal way supporting contract negotiation and (automated) contract enforcement. IDS

[3] https://github.com/International-Data-Spaces-Association/InformationModel

information model includes a set of policy templates must be used to formalize the different policy classes:

https://github.com/International-Data-Spaces-Association/InformationModel/tree/develop/examples/contracts-and-usage-policy

Regarding **pricing**, data marketplaces enable completely new business models where data can be considered as a commodity to be bought and sold in the market[viii]. There are different data pricing strategies utilized by organizations/data owners for maximizing profit. Muschalle et al.[ix] present four main categories of data pricing models: 1) free data usage model, 2) usage-based pricing model, 3) package pricing model, and 4) flat fee tariff model. In free data usage model data is available in public storage and marketplace for free, in hope that it attracts customers/consumers into buying or paying for the complete data or other premium services on it. Usage based pricing model considers the measurement of each single data commodity and time counts for pricing, e.g., fee varies on the number of API calls at peak time to during normal hours. Package pricing model refers to a pricing model that offers a customer a certain amount of data or API calls for a fixed fee. The flat fee tariff is one of the simplest pricing models with minimal transaction costs based on time as the only parameter. Hence, it provides safety in planning future activities. On the other hand, it lacks flexibility for data consumers.

Other pricing models can be used by combining the four main pricing models. For instance, a two-part tariff pricing model is a combination of package pricing with a flat fee tariff model. In this scenario consumers pay a fixed basic fee and on top of that an additional 'fee per unit consumed'. *Another* pricing model is known as Freemium which provides access to basic or limited services for free but charges for premium content and services. The payment model for additional content and services can be according to one of the pricing models.

The Pricing strategies of data marketplaces apply likewise to IDS resources. The Free strategy does not charge the usage of resources. The Freemium strategy exposes a limited part (or capabilities) of the resource at no cost, while additional parts are charged Pay-per-Use, or based on a Flat Rate. The Pay-per-Use strategy relies on a particular metrics (volume, access count, download) to define a charged instance of usage, while the Flat Rate strategy charges usage per quantitative slot (time, volume, credit), optionally associated with a tiered cost model according to the configuration of the retrieved resource.

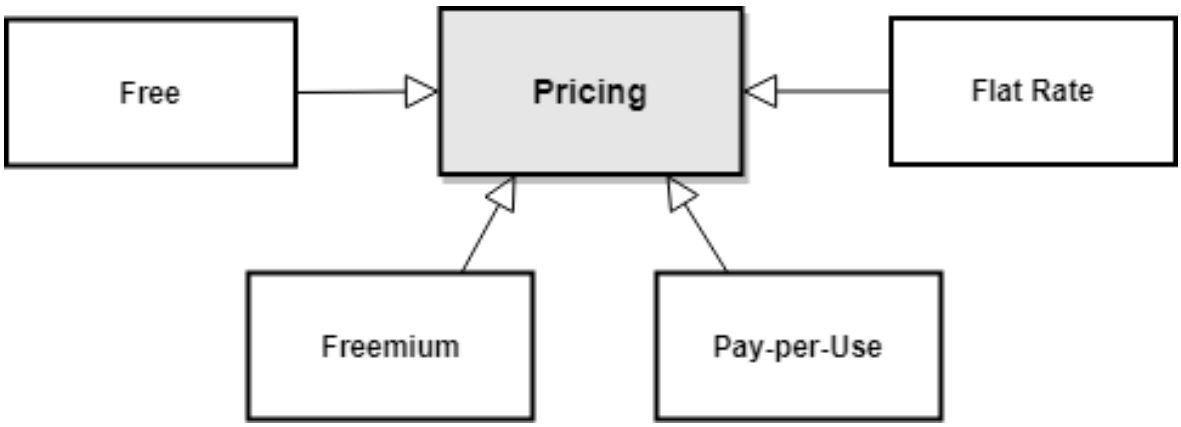


Figure 8 Taxonomy of Product Pricing concepts[4]

3.4 Data privacy

IDS does not deal explicitly with data privacy issues. However, data privacy issues are mentioned in the data usage context. According to the IDS reference architecture, data usage facilitates enforcement of legal or technical requirements, such as service level agreements (SLAs) or **data privacy regulations**.

The following examples included in the reference architecture illustrate data privacy requirements that cannot be achieved by data access control, but require data-centric usage control:

- **Anonymization by data aggregation**: Personal data may be used only in an aggregated form by untrusted parties. To do so, a sufficient number of distinct data records must be aggregated in order to prevent deanonymization of individual records.

- **Anonymization by Data Substitution**: Data allowing personal identification (e.g., faces in video files) must be replaced by an adequate substitute (e.g., pixelized) in order to guarantee that individuals cannot be deanonymized.

Currently, IDS is working in the definition of the policy data model and some of the policy classes can be used for assuring the GDPR privacy rights for data subjects, which aim to give individuals more control over the data they loan to organizations.

For example:

- **Purpose-restricted Data Usage Policies.**

This category represents the class of policies that restrict the usage of data assets limited to specific purposes. “If the purpose is risk management, then allow the usage of data and else if the purpose is marketing, then inhibit the usage of data” is an example policy that is instantiated from this policy

[4] https://international-data-spaces-association.github.io/InformationModel/docs/index.html

class. The W3C Data Privacy Vocabulary includes a vocabulary of purposes and processing methods that can be used.

- **Use data and delete it after. (Right to be forgotten)**

Even though the policy requires the deletion when the usage period is over, this has to be understood as the obligation to prevent the reappearance of the data asset, using reasonable actions.

- **Modify data (in transit) (Anonymize the data)**

There might be cases where data must be modified or partially anonymized before it is allocated to the user. The data modification must be done before the permission to use the data is granted. This class of policy represents the Data Usage Control policies demanding to modify the data in transit; a Data Usage Control application intercepts the data that is transmitted and applies the modifications on them.

Another possibility is to use a **consent-based and user-centric component** for management and enforcement of Personal Data Usage Consents for those services having the role of Data Consumer and/or Data Provider. This component is in charge, in case of data requests involving end user personal data, of ensuring compliancy with GDPR regulation, by enforcing data usage/processing Consents given by the Data Owner.

In particular, the component acts as an intermediary and as a tool of communication between data subjects and controllers/processors, supporting the entire end-to-end process of generation and management of dynamic consents. The end user privacy is ensured by enabling him to grant and withdraw consents to third parties for accessing and processing his own data. Consent authorizes a Data Source to provision data to Data Consumer and authorizes Data Requester to process that data. Consent must be linked with a Data Usage Policy to be formalized.

Optionally, the Consent Manager can even handle the actual data requests between components, by issuing authorization tokens to their own services regulated by a given Consent.

Summarizing, using IDS, data privacy can be treated as follows:

- GDPR compliancy: Modelling protection of personal data (e.g., according to GDPR) using the W3C Data Privacy Vocabulary
- Anonymization (IDS Data app integrated in the connector)
- Data usage constraints
- Managing user consent

3.5 IDS specifications and tools development status

IDS final specifications are still under development and some decisions proposals are still ongoing. These are the decisions currently open which could affect the IDS components development during PLATOON.

- Decision about what protocols should be included as IDS protocols: Multipart & IDS-LDP & IDSCP protocols.

- Decision about the query Language for the IDS Metadata Broker
- Decision about the connector security profiles for certification regarding the usage control functionality

Currently, IDS is analysing the inclusion of other communication protocols like for example a REST based protocol. The addition of this new protocol is interesting for PLATOON since the Analytic toolboxes are going to expose their functionalities via OpenAPI REST interfaces and it will facilitate the adoption of IDS for data transfer needed for data analytic services.

4 Data governance, security and privacy requirements for the energy sector

This section gathers the specific data governance, security and privacy requirements for each of the pilots. The pilots represent a wide range of problems and needs of the energy sector, including:

- To increase the efficiency and reliability of the operation of the electricity network (e.g. by predictive maintenance)
- To optimize the management of assets connected to the grid (small-scale/renewable electricity generation and those used for demand response)
- To increase the efficiency and comfort of buildings, and to de-risk investments in energy efficiency (e.g. by reliably predicting and monitoring energy savings).

Considering the objective of pilots, they have been classified in four categories:

1. Predictive maintenance in renewables (Wind Farm).
2. Distribution grids efficient operation and assets life extension.
3. Efficient End Use of Energy, peak avoidance and demand side response
4. There is a fourth service so-called Optimum Energy Management in a Microgrid, which in fact considers a lot of the previous services.

In order to present homogeneous information about the pilots, a common structure has been defined:

1) First, a brief description of the pilot and its scenarios is presented.
2) The general technical architecture of the pilots, showing the data transfer among the involved stakeholders (systems or companies). In some of the scenarios where several companies participate, IDS connectors are used to transfer data.
3) Business model associated to data transfer via the IDS connectors, including for example:
   a) **Transfer of data necessary for analytical task**: One company sends data to the analytical tool needed to provide the service to a third party.
   b) **Data transfer associated with the use of an "analytical toolbox":** A client of a analytical toolbox" sends the data needed and get the result.
   c) **Data transfer for the training of a model**. A company transfer data to the analytical toolbox provider in order to train the analytical models
   d) **Comparison of models between pilots**. Pilots covering the same functionality share data to compare their results.
4) Answers to a very simple questionnaire filled by the pilots regarding the data governance, security and privacy requirements. The information gathered is related to the data governance perspectives in IDS: data sovereignty and ownership, data quality, data provenance and data monetization. Additionality, a question about data privacy has been included. The answers to these questions have been gathered during a couple of workshops with each of the pilots.

   a) How many different companies are included in the pilot?
   b) Are data usage constraints applicable? What kind of data usage constrains are envisaged?
   c) Is data quality a requirement?
   d) Is data provenance a requirement?
   e) Is monetization of data a requirement?
   f) Is personal data going to be used by the pilots?

4.1 Pilot 1a: VUB Predictive Maintenance of Wind Turbines

Pilot #1 focusses on predictive maintenance in wind energy. Today, the O&M cost of wind energy is substantial. Having better failure prediction models can help in substantially reducing these costs by allowing to perform maintenance during clustered and well-planned repair campaigns. The industry is therefore looking for accurate monitoring and diagnostics tools that are able to continuously process data to a large number of wind turbines and pinpoint to root-causes of failures.

The main goal of this pilot use case is to achieve early detection of faults in the electrical components of the powertrain: generator and converter. Different modelling approaches are used to predict behaviour of the electrical components.

On the one hand a physics based digital twin model is used; on the other hand, data-driven normal behaviour models are considered. These models will serve anomaly detection schemes. Through reasoning based on semantics the fault will be diagnosed based on these anomalies and potential root cause pinpointed out. In addition to batch processing at cloud level, edge computing will be used to perform processing of the high frequency current data of the generator for signature extraction. In this pilot focus is on offshore and onshore wind turbines equipped with a doubly fed induction generator.

Next figure shows a simplified view of the technical architecture of the pilot, including the companies involved and the IDS connection needed:

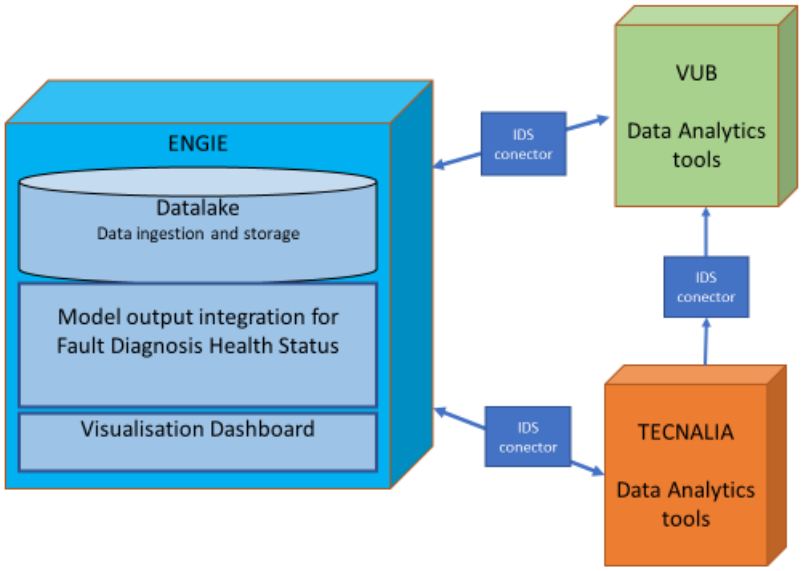


Figure 9: Pilot 1a architecture

According to the architecture, in the pilot three different companies will exchange data, and three IDS based connections will be deployed:

- **TECNALIA - ENGIE**: Bidirectional
  - From ENGIE to TECNALIA: Data needed to train the model
  - From TECNALIA to ENGIE: Results of the predictive maintenance model.

- **TECNALIA → VUB**: Unidirectional. TECNALIA sends data to VUB needed to improve their service.
- **VUB - ENGIE**: Bidirectional
    - From ENGIE to VUB: Data needed to train the model
    - From VUB to ENGIE: Results of the predictive maintenance model.

Model Training:

- Historical High frequency data: detailed current, voltages and powers for normality validation with a frequency of 500Hz. Dataset will contain a small set of parameters for few months for a small number of turbines.
- Historical SCADA: Data from SCADA system or normality and abnormality validation with a frequency of 1 measure every 10 minutes. Dataset will contain a wide set of parameters for few years for a small number of turbines.
- Historical Status codes: Dataset will contain the corresponding fault messages generated for the same turbines and timerange of the SCADA dataset.

Model validation:

- Historical SCADA data containing the same parameters as for model training SCADA dataset but formed of a wider set of turbines and larger time range.

Model execution:

- "Real Time" SCADA data containing the same parameters as for model training and validation SCADA dataset but containing data for the whole fleet with a frequency of a value every 10 min.

Answers to the data governance and privacy questionnaire:

1) **Are data usage constraints applicable?** Yes
    - **Role based**: to ensure that data can only be accessed by specific departments/people in the company.
    - **Purpose**: control different datasets for different functions of the model (e.g. train, operation)
2) **Is personal data going to be transferred?** NO
3) **Is data quality assurance a requirement?** NO
4) **Is data provenance information a requirement?** YES. To track data exchanges between companies for audit reasons and potential monetisation as part of WP8.

Is monetization of data a requirement? Only to define the pricing model.

- **Pricing model**: define the pricing model and business model but not implement it.

4.2 Pilot 2a: IMP Electricity Balance and Predictive Maintenance

PLATOON Pilot #2a will take place in Serbia. The **overall goal** of the pilot is to integrate and deploy different PLATOON analytical services with the Institute Mihajlo Pupin (IMP) proprietary VIEW4 Supervisory control and data acquisition (system). The VIEW4 SCADA is deployed at many parts in the energy value chain in Serbia, starting from control on the production side (in the large hydro and thermal power systems, as well as RES), via transmission management to distribution and electricity dispatching.

Pilot 2a scope is:

- Integrate PLATOON Analytical Tools for Smart Grids with PUPIN SCADA for Improved management of Serbian Smart Grid and power plants;
- Better prediction of production and demand forecast at regional/national level

PLATOON methodology was applied to model in detail the following scenarios:

- **UC P-2a-03 Demand forecast on transmission level**
- **UC P-2a-04 RES (Wind generation) forecasters**
- **UC P-2a-05 Effects of Renewable Energy Sources on the Power System (distribution level)**
- **UC P-2a- 07 Predictive maintenance in RES power plants**

Next figure shows a simplified view of the architecture of the pilot, including the companies involved.

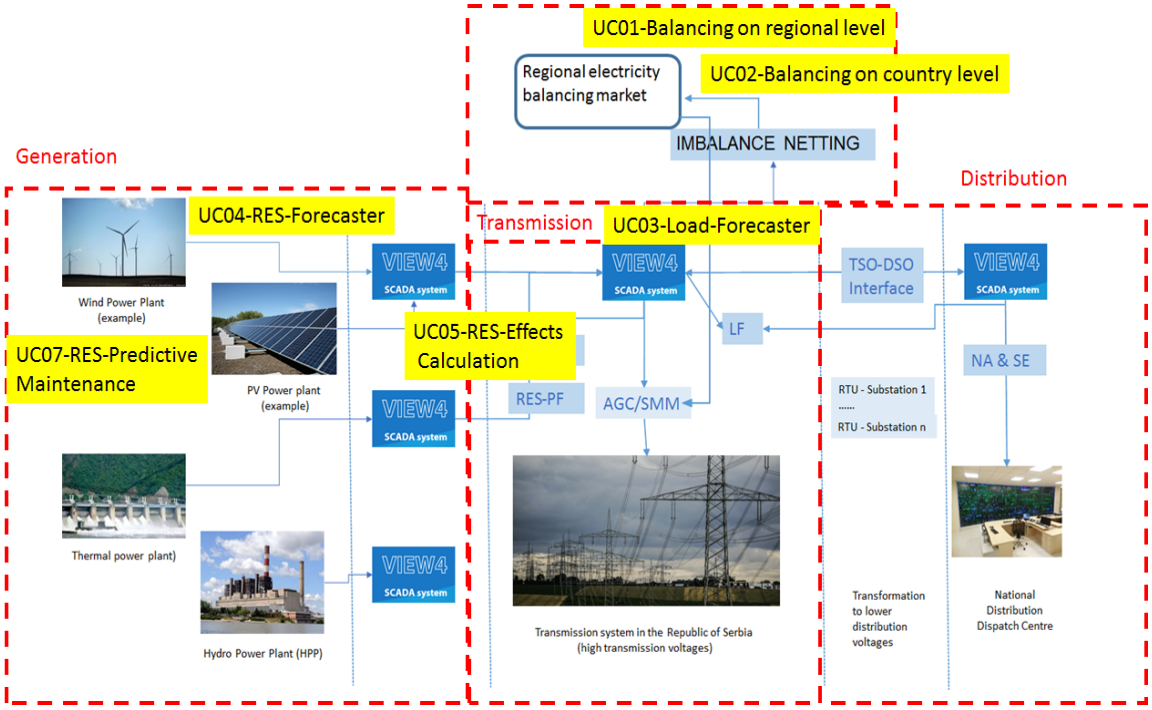


Figure 10: Pilot 2a ICT architecture

According to the architecture, three companies are involved: IMP, TSO and Balance Service Provider (BSP). Data exchange foreseen:

- [UC04 RES Forecaster] Info about Plant generation from a Plant to Power Industry of Serbia (Unidirectional)
- [UC05 RES Effects Calculation] from a Plant to Power Industry of Serbia (Unidirectional)

Next figure shows the services diagram and data exchange among services:

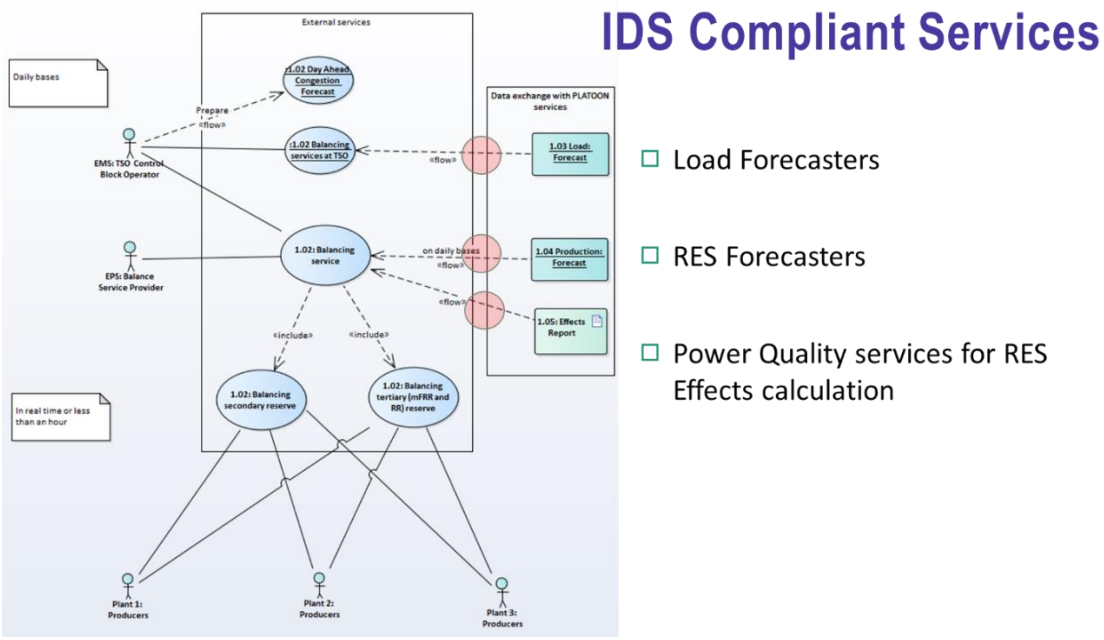


Figure 11 Pilot 2a IDS compliant services

Furthermore, the RES Forecaster and the LOAD Forecaster could be in the future offered as a service through the marketplace. In this case, an IDS connection will be used to transfer the necessary data.

The business model is **Data transfer associated with the use of an "analytical toolbox":** A client of an analytical toolbox" sends the data needed and get the result.

In **UCP-2a-05** (Effects of Renewable Energy Sources on the Power System) and **UC P-2a- 07** Predictive maintenance in RES power plants) an IDS connection is not needed.

Next figure shows the architecture for this scenario:

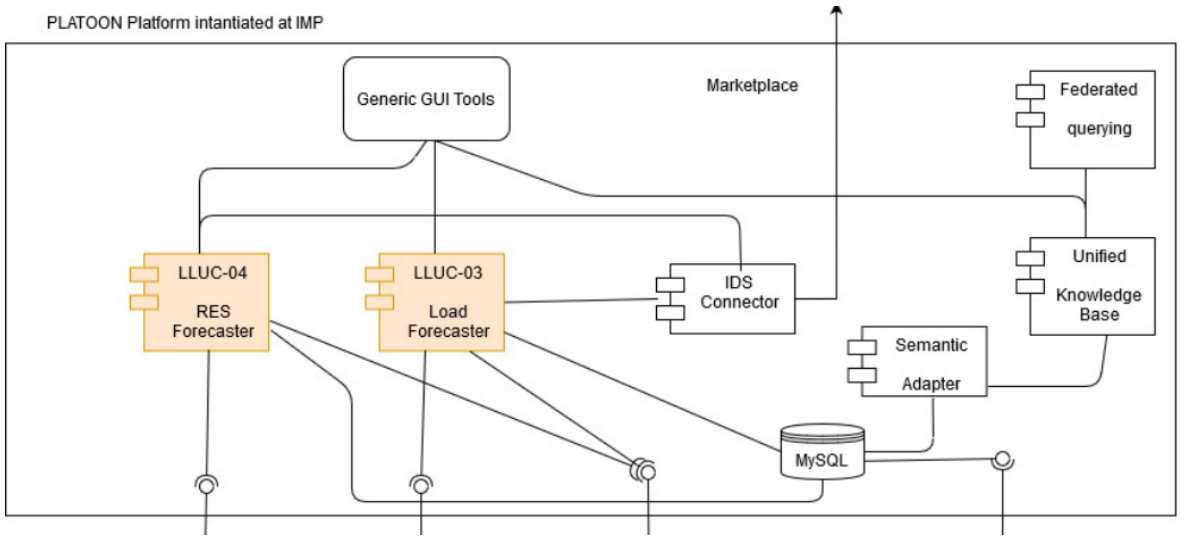


Figure 12 Pilot 2a RES Forecaster and Load Forecaster IDS deployment

Answers to the data governance and privacy questionnaire:

1) **Are data usage constraints applicable?** Yes
   - **Role based**: to ensure that data can only be accessed by specific departments/people in the company.
2) **Is personal data going to be transferred?** NO.
3) **Is data quality assurance a requirement?** NO
4) **Is data provenance information a requirement?** It could be, when the data is sent from a Plant to Balancing services.
5) **Is monetization of data a requirement?** Only to define the pricing model.
   - **Pricing model**: define the pricing model and business model but not implement it.

4.3 Pilot 2b: Electricity Grid Stability, Connectivity, And Life Cycle

Parc Bit is a technological park located in Mallorca where its electrical grid will be under study in pilot 2b. Two different use cases have been defined described in the following list:

- **UC P-2b- 01 - Predictive Maintenance for MV/LV Transformers**

This UC will focus on LV/MV transformer predictive maintenance, estimating transformer components health and its maintenance costs, planning maintenance actions, monitoring transformer alarms and studying different grid scenarios in case of replacing old transformers or adding complementary transformers. This tool will use available data in LV/MV transformers, which usually have a small budget for monitoring and maintenance. Maintenance actions are based in the Remaining Useful Life (RUL), in this project, different failures modes of the transformer critical components will be estimated and reflected to the health index of the transformer. Once it is obtained the health index and considering maintenance and failure costs, the transformer maintenance plan will be defined. Finally, a prescriptive analytics tool will be developed. This tool will allow to evaluate the effect of different operational actions in the grid O&M cost sheet.

- **UC P-2b- 02- Non-technical loss detection in Smart Grids**

The main objective of this use case is to develop a tool for the quantification of losses in the distribution grid of a DSO and the detection of non-technical losses (NTL), using the available smart meter data from Sampol´s smart grid in ParcBit, Majorca (Spain). The main output of this UC is to develop a solution for NTL detection using data analytics, which just requires measurement data available from the Automatic Metering Infrastructure-AMI, and optionally information on the grid topology.

Next figure shows a simplified view of the technical architecture of the two use cases, including the companies involved and the IDS connection needed:

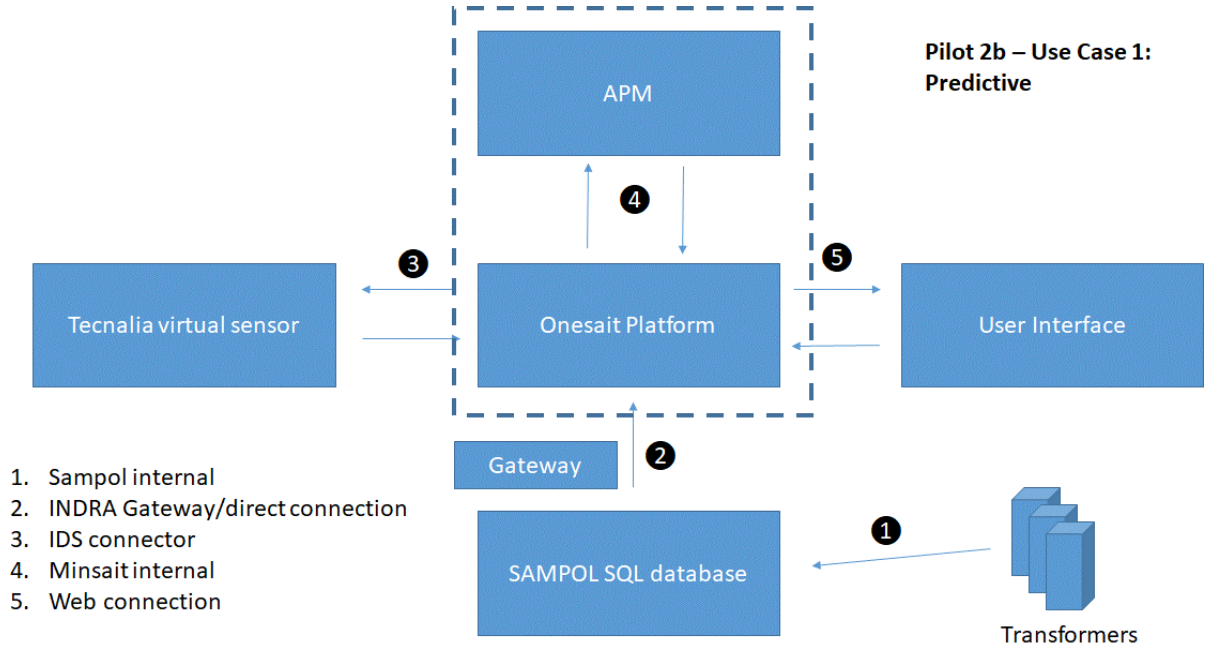


Figure 13 Pilot 2b – Use case 1 architecture

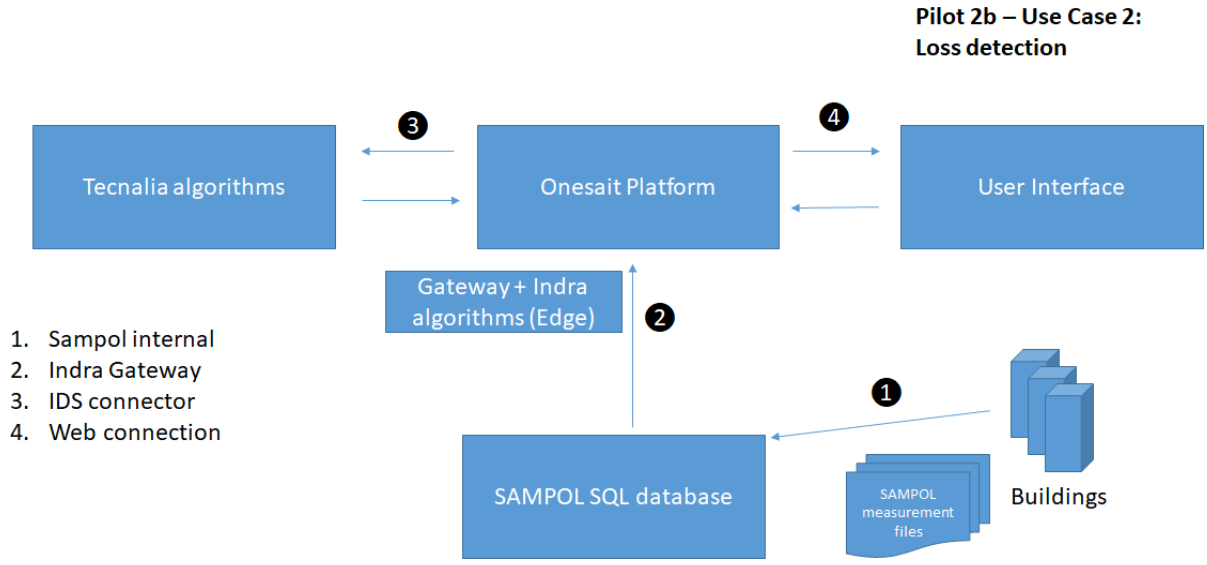


Figure 14 Pilot 2b – Use case 2 architecture

According to the architectures, three companies that will exchange data: Tecnalia, Indra and Sampol. However only one IDS based connection will be developed since no connection is expected between Tecnalia and Sampol and Indra - Sampol connection will be done on a gateway basis ,since Indra already has contracts and NDAs with Sampol to regulate all data governance, security and privacy aspects.

- **TECNALIA - INDRA**: Bidirectional
    - From INDRA to TECNALIA: Data needed to train the model for the virtual sensor.
    - From TECNALIA to INDRA: Virtual sensor measurement.

Answers to the data governance and privacy questionnaire:

1) **Are data usage constraints applicable?** Yes
    - **Role based**: to ensure that data can only be accessed by specific departments/people in the company.
    - **Purpose**: control different datasets for different functions of the model (e.g. train, operation)
    - **Expiration date**. Data is only available only in a specific period.
2) **Is personal data going to be transferred?** NO. No strong restrictions except for some cases that have to be clarified. The information should be scrambled to avoid problems. MANDAT International will be consulted as part of T3.5.
3) **Is data quality assurance a requirement?** NO
4) **Is data provenance information a requirement?** YES. Interesting for pay-per-use, data tracking and auditing purpose.
5) **Is monetization of data a requirement?** Only to define the pricing model.
    - **Pricing model**: define the pricing model and business model but not implement it.

4.4 Pilot 3a: Office Building: Operation Performance Thanks To Physical Models And IA Algorithms

Pilot #3a concerns an office building with a focus on developing Use Cases to optimize the HVAC system performance or provide new kind of services (to help in particular with the grid management). Two main use cases have been identified to be implemented on the pilot:

- **LLUC P-3a-01- Optimizing HVAC control regarding occupancy**
  The use case aims at providing a smart module for an office building that optimize HVAC operation in function of real occupancy. Occupancy data are available via dedicated sensors, and the comfort and HVAC controls are available via the Building Management System (BMS) of the building. Using historical data, some learning algorithm are implemented to predict occupancy and anticipate heating and cooling period in the building and its different zones. A first optimization loop can be implemented to control the overall building occupancy planning and HVAC operation. A second optimization loop is used to adapt HVAC controls in the different zones of the building. The building manager can supervise and update some parameters in the system and access some regular assessment of the system controls. It also collects data from occupancy sensor to map the occupancy in the different zones of the building.
- **LLUC P-3a- 02 - Providing Demand Response Service through HVAC control**
  The use case aims at providing a smart module to supervise the implementation of Demand Response services in an office Building using HVAC control and building inertia. Through the supervision of the building parameters and weather forecast, the module developed is providing predictions of the HVAC load and the potential flexibility available in the building, considering that a certain thermal comfort level must be maintained. These predictions are regularly transmitted to an aggregator that is then able to engage reliable flexibility services with the grid operator. The aggregator can then send or plan orders to stop the HVAC system of the building for a given time. If the orders are validated (within conditions of the contract and minimum comfort parameter are respected), they are implemented in the BMS. Feedback and KPI are shared with the aggregator concerning the load shifting operations.

Next figures show a simplified view of the technical architecture of the two use cases:

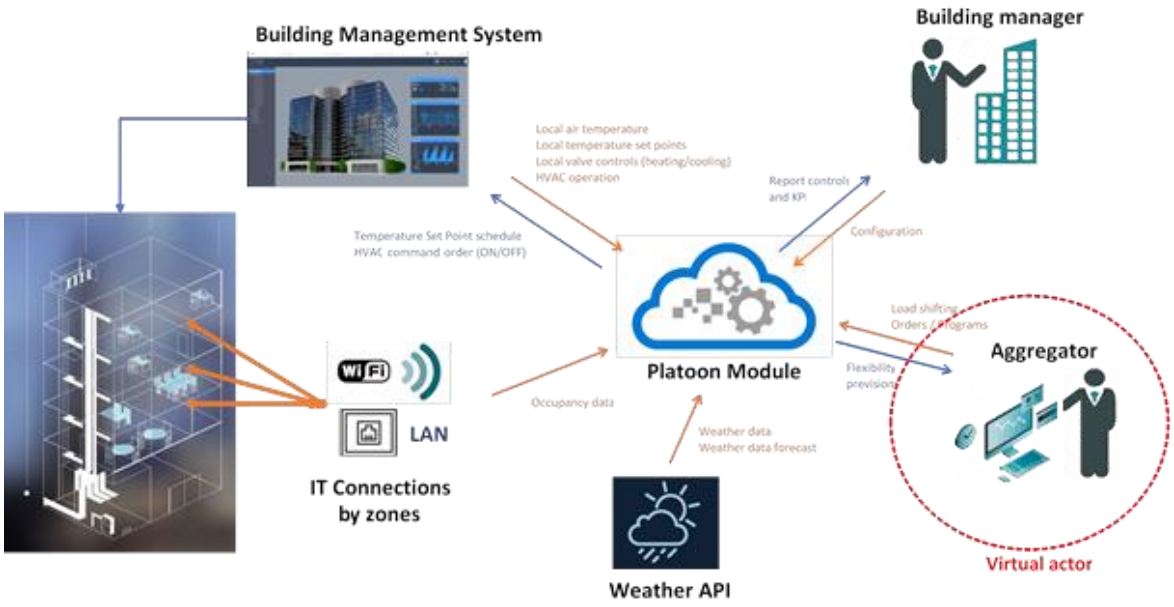


Figure 15 Pilot 3a – Use case 1 architecture

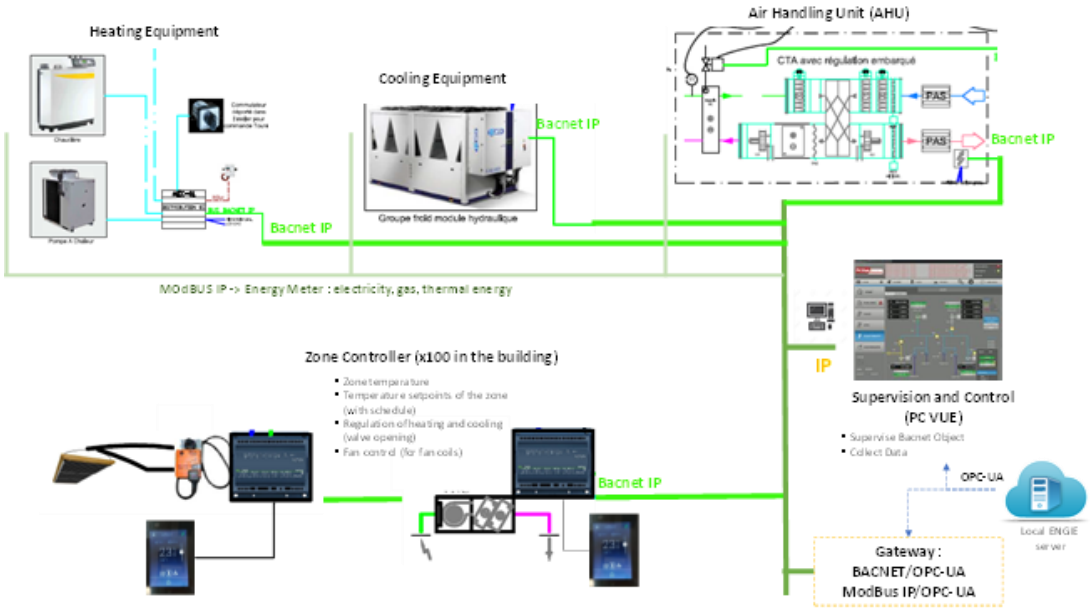


Figure 16 Pilot 3a – Use case 2 architecture

According to the architecture of the two uses cases, all the infrastructure belongs to ENGIE, so only one company is involved and there is no need for an IDS connection.

However, there is the possibility of including data from other buildings, so it could be used to provide services to other companies or to improve the training of the analytic models. In this case:

- **ENGIE – External company or different business units/subsidiaries**: Bidirectional

- From External company to ENGIE: Data needed to train the model
- From ENGIE to External company: Results of the predictive maintenance model.

Additionally, a plan has already been put in place to further explore potential synergies amongst these pilots. In fact, so far there have been identified two potential synergies:

1) co-develop a tool amongst ENGIE, PI and ROM for pilots 3a and 3b.
1. Tecnalia could offer their tool as a service to validate HVAC load forecasting and optimisation tool in pilots 3a and 3b apart from 3c. These potential synergies are still under discussion and will highly depend on the agreements reached by individual partners and the available time. In this case a new IDS connection should be deployed:

- **ENGIE – TECN**: Bidirectional
  - ENGIE sends raw data to TECN and TECN sends results of the HVAC load forecasting and optimisation tool to ENGIE.

Answers to the data governance and privacy questionnaire:

1) **Are data usage constraints applicable?** Yes
   - **Role based**: to ensure that data can only be accessed by specific departments/people in the company.
   - **Purpose**: Only access to specific buildings
2) **Is personal data going to be transferred?** NO
3) **Is data quality assurance a requirement?** YES
4) **Is data provenance information a requirement?** YES. Interesting for pay-per-use, data tracking and auditing purpose.
5) **Is monetization of data a requirement?** Only to define the pricing model.
   - **Pricing model**: define the pricing model and business model but not implement it.

4.5 Pilot 3b: ROM/PI

PLATOON Pilot #3b will take place in Rome (Italy) and the **overall goal** of the pilot is to acquire, aggregate and process energy consumptions and related data of different buildings (often in various form not coherent between them) to make energy domain specific data analysis as consumption forecasting, predictive maintenance, benchmarking and so on. Pilot 3b is actually formed of two different sub pilots from different companies:

- **LLUC P-3b-01-03: Poste Italiane**
- **LLUC P-3b-04: Roma Capitale**

PLATOON methodology was applied to model in detail the scenarios that will be tested in WP6 framework and are related to analytical services that will be developed in WP4 framework, as follows:

- **LLUC P-3b-01 (PI) Building Heating & Cooling consumption Analysis and Forecast:** The correlation with external weather conditions, building characteristics and past performances together with

benchmark with similar building, represent an area of optimization for both cooling and heating systems. Sensors, meters and other hardware produce information that, through processing with forecasting algorithms and machine learning techniques, could be used to predict plants consumption and for the energy efficiency benchmarking.

- **LLUC P-3b-02 (PI) Predictive maintenance of cooling & heating systems:** Using systems energy consumption data and historical information about fault and maintenance it will be possible to identify anomalies and predict failures in the systems.
- **LLUC P-3b-03 (PI) Lighting Consumption Estimation & Benchmarking**: The objective is to estimate the specific building lighting consumption, in order to benchmark, plan optimization actions and detect anomalies and outliers. Estimating the lighting consumption will also be possible to better compare the new performance with the previous lighting technology where new installations are made.
- **LLUC P-3b-04 Monitoring and Analysis of energy meters data of ROM large asset:** Pilot 3b-ROM-04 is related to Roma Capitale large asset of municipal buildings assuming the energy consumptions data coming from the power and gas meters and the structural data of the buildings to be processed with the following general scopes:
    - Improving energy efficiency: increasing building energy performances (EP) reducing Energy Consumptions (EC)
    - Increasing the responsiveness of the municipal offices
    - improving their awareness in terms of energy consumptions and efficiency dynamics
    - correctly planning the increase/optimization of the RES PV plants on the buildings
    - reducing losses of money and time

Five different Services are required to implement in the toolboxes for the pilot as follows:

Energy Consumption Forecaster: Service for prediction of energy consumption of a building or cluster of building due to Heating, Cooling and lighting systems;

Predictive maintenance: Service for prediction of fault by taking into account the scheduled maintenance, the failures occurred and the consumption of heating and cooling systems. An alarm is required if time to failure (TTF) is less than the scheduled maintenance. Time to Failure (TTF) will be the most important parameter determined by the App.

Energy consumption Benchmarking: Service which compare buildings consumptions in a cluster or the performance of the same building in different time windows in order to improve energy efficiency of the buildings.

Lighting Consumption Evaluator: Service which estimate lighting consumption of a building or set of buildings in absence of accurate data (starting from general consumption data or data consumption of different types of systems and comparison with real lighting consumption data extracted from other buildings).

Consumptions and Occupancy Profile Correlator: Service for understanding the impact of occupancy on energy consumption.

Next figure shows a simplified view of the technical architecture of the first three use cases corresponding to Poste Italiane (PI):

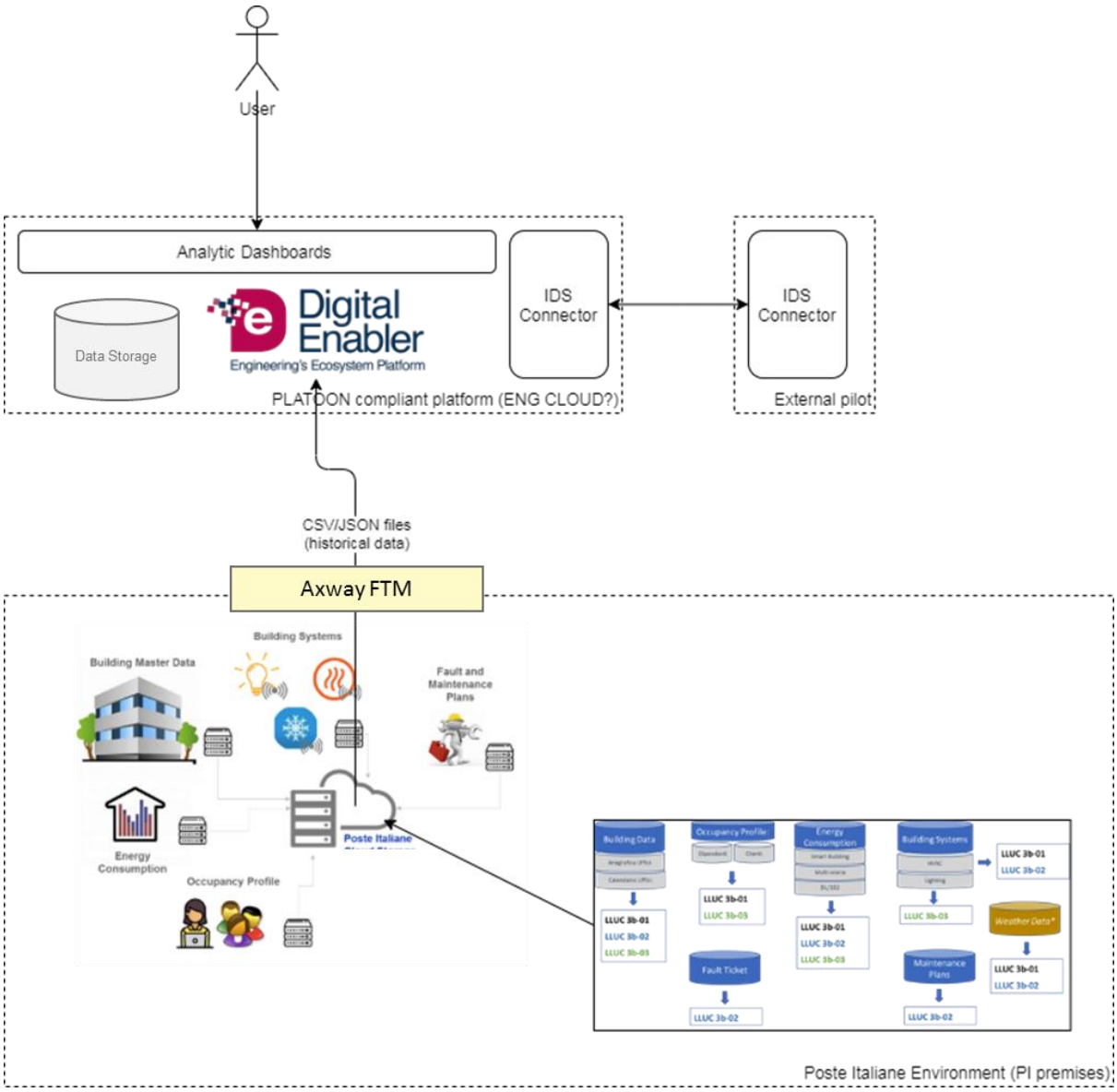


Figure 17 LLUC P-3b-01, 02 and 03 architecture

According to the architecture, the connection among Poste Italiane environment and the Engineering platform is already defined and implemented (Axway FTM) so there is no need for an IDS connection.

Axway FTM deals with the configuration of services and procedures for the interconnection and transfer of data between the systems of the Post Office and internal / external customers. The main components are the Synchrony Transfer CFT, the Synchrony Gateway and the Synchrony Gateway Secure Relay. On these environments, configurations, products and the creation of bath procedures are developed to complete the data transport activities.

However, a plan has already been put in place to further explore potential synergies amongst these pilots. In fact, so far there have been identified a potential synergy: TECNALIA could offer their tool as a service to validate HVAC load forecasting and optimisation tool in pilots 3a and **3b** apart from 3c. This

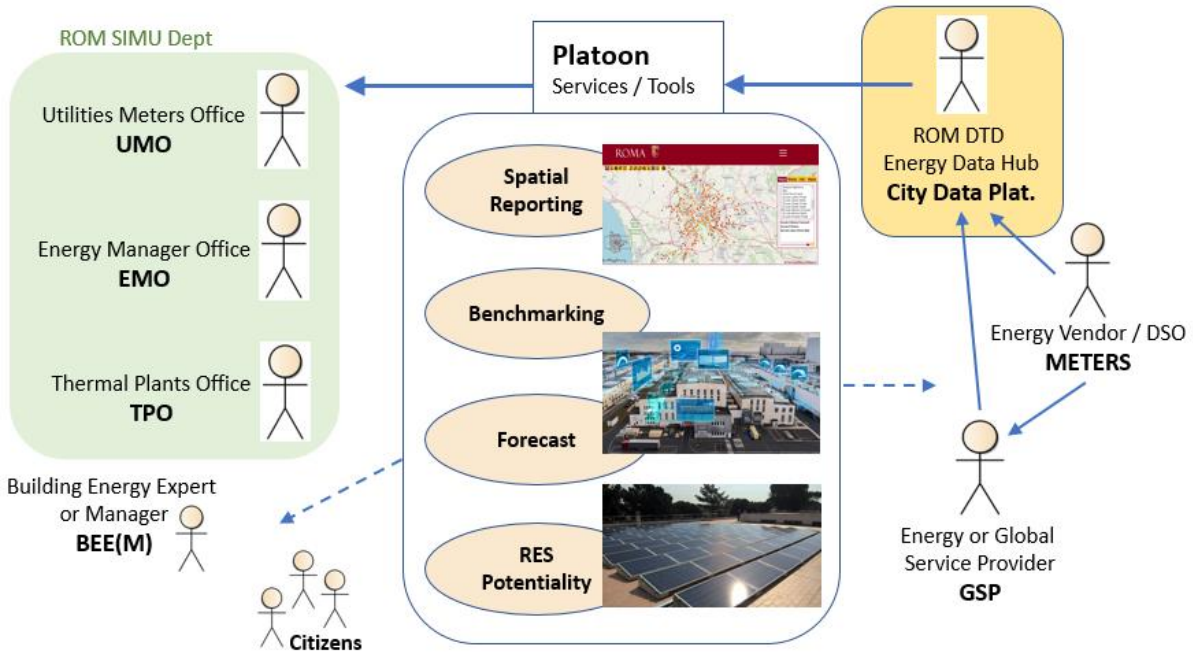
potential synergy is still under discussion and will highly depend on the agreements reached by individual partners and the available time. In this case a new IDS connection should be deployed:

- **TECNALIA – ENGINEERING**: Bidectional
  - ENGINEERING sends raw data to TECN and TECN sends results of the HVAC load forecasting and optimisation tool to ENGINEERING.

Answers to the data governance and privacy questionnaire:

1) **Are data usage constraints applicable?** Yes
   - **Role based**: to ensure that data can only be accessed by specific departments/people in the company.
2) **Is personal data going to be transferred?** NO
3) **Is data quality assurance a requirement? NO,** Data quality is controlled by the source system
4) **Is data provenance information a requirement?** YES. Track data exchanges between companies
5) **Is monetization of data a requirement?** NO.

Next figure shows a simplified view of the technical architecture of the fourth use case LLUC P-3b-04 Monitoring and Analysis of energy meters data of ROM large asset:



According to the architecture several companies are involved in the project (DSOs, BEE and GSP) along with the ROMA City Data Platform. However, these external companies have already deployed the communication channels with the ROMA Capitale platform, so there is no need of an IDS based connection. Regarding the platform to deploy de analytic services, ROM is also analysing the use of

the Digital Enabler for data processing. However, there is the possibility of including data from another DSO, so IDS could be used for the new connection.

Similarly to the PI pilot, a plan has already been put in place to further explore potential synergies amongst these pilots. In fact, so far there have been identified a potential synergy: TECNALIA could offer their tool as a service to validate HVAC load forecasting and optimisation tool in pilots 3a and **3b** apart from 3c. This potential synergy is still under discussion and will highly depend on the agreements reached by individual partners and the available time. In this case a new IDS connection should be deployed:

- **TECNALIA – ENGINEERING**: Bidectional
    - ENGINEERING sends raw data to TECN and TECN sends results of the HVAC load forecasting and optimisation tool to ENGINEERING.

4.6 Pilot 3c: GIR/SIS

Pilot #3c takes place in Nanogune, a tertiary sector smart building dedicated to nanotechnology research, based in San Sebastian (Spain). This building is divided into different areas such as offices and laboratories and has both thermal and electrical meters to differentiate the areas.

LLUC P-3c 01 Advanced EMS for Tertiary Buildings: The Advanced EMS will optimize the local renewable energy resources (RES) and HVAC operation as function of building and RES characteristics, building comfort constraints, ambient conditions and energy market price following a multi-objective pattern which targets to reduce the overall energy bill and maximize the usage of RES. For this the study area will be the second floor of offices in the Nano area and we will use both comfort sensors, such as electrical and thermal constants in this area, as well as the consumption of the AHU that heat this area and finally the PV panels that are installed in Nanogune.

LLUC P-3c 02 Predictive Maintenance in Smart Tertiary Building Assets: Development and implementation of predictive maintenance tools for the thermal control assets of smart tertiary buildings (specifically chillers, pumps and distribution rings). The objective is to improve the maintenance policy increasing the availability and useful life of these assets and reducing the general maintenance costs. For this use case we will use the signals that are currently integrated into the SCADA of this equipment, as well as the history of the maintenance tasks carried out on all the chillers that GIROA manages and the maintenance ranges that it applies.

Next figure shows a simplified view of the technical architecture of the two use cases:

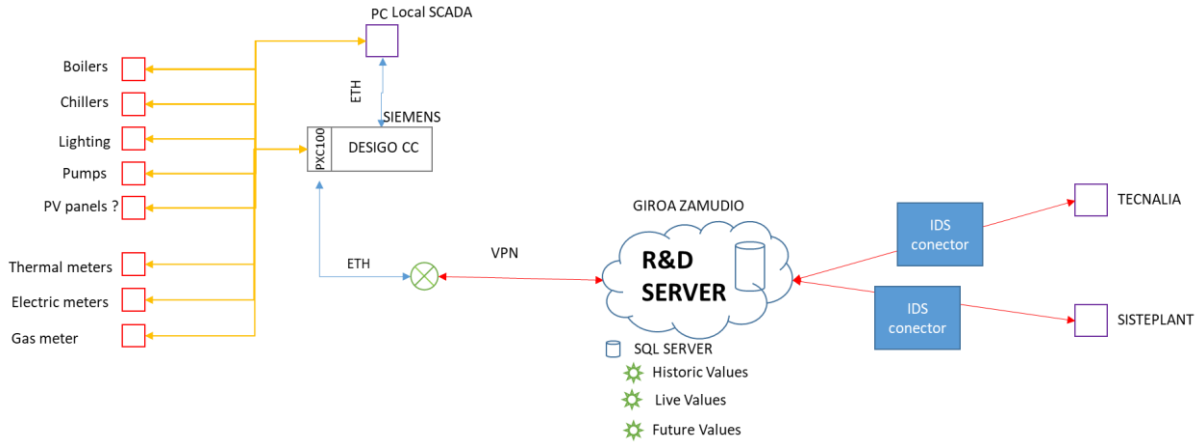


Figure 18 Pilot 3c architecture

According to the architectures, three companies that will exchange data: TECNALIA, GIROA and SISTEPLANT. However, it doesn´t make sense to put a connector between GIROA and SISTEPLAN and they are already using the platform and they have a legal contracts, NDAs and licenses already in place with GIROA. It makes sense though to put a connector between GIROA and TECN. And potentially also a connector if in the end there is going to be data transfer between TECN and SIS.

TECNALIA - GIROA: Bidirectional

- From GIROA to TECNALIA: Data needed to train the model.
- From TECNALIA to GIROA: Results of analytic tool.

There might also be a necessity to stablish a connector between TECN and SIS to visualize the results through the PROMIND platform but this still under debate and it will be decided as part of T6.1 implementation and validation plan.

Answers to the data governance and privacy questionnaire:

1) **Are data usage constraints applicable?** Yes. Between GIR and TECN.
   - **Role based**: to ensure that data can only be accessed by specific departments/people in the company.
   - **Purpose**: Only access to specific buildings.
   - **Access based**: output from TECN tools never write automatically in the final equipment always require a check from the technician.
2) **Is personal data going to be transferred?** NO. No strong restrictions except for some cases that have to be clarified. The information should be scrambled to avoid problems. MANDAT International must be consulted as part of task T3.5.
3) **Is data quality assurance a requirement?** NO. In GIR database they already do a quality check and they put a flag to see if the data is not valid (e.g. power must be always positive, energy must increase…)
4) **Is data provenance information a requirement?** YES. to track data exchanges between companies for audit reasons and potential monetisation as part of WP8. They have two-time register (original read/received data and write/modify data).

5) **Is monetization of data a requirement?** Only to define the pricing model.

- **Pricing model**: define the pricing model and business model but not implement it.

4.7 Pilot 4a: PDM Energy Management of Microgrids

PLATOON Pilot #4a will take place in Milano, Italy. The Politecnico di Milano's Multi-Good Micro-Grid Laboratory (MG2lab) is an experimental facility for real-life scale research, simulation and test purposes, thus, allowing to study new data-driven paradigms for energy management able to deal with increased complexity of the energy systems and to assess the advantages of innovative strategies.

LLUC P-4a- 01 - Energy Management of Microgrids

The goal of the functionality described in the current use case is to study data-driven energy management able to deal with increased complexity of the energy systems and to assess the advantages of innovative strategies, by means of EMS with real-time processing and optimization for small-scale/renewable electricity generation, based on power generation and load forecasts.

Indeed, EMS play a crucial role in the management, real-time processing and optimization of assets connected to the grid (in particular small-scale/renewable electricity generation and those used for demand response): the development of new microgrids control and management strategies involves the integration of data analytics toolboxes and optimization platforms on the EMS that manages the microgrid operation.

The EMS development includes on one side the incorporation of predictive algorithms able to forecast renewables production and load profile and on the other the exploitation of an Optimal Power Flow ('OPF') algorithm, able to consider the fluctuation of Renewable Energy Resources (RES) and to optimize the economic unit dispatch, the reliability of the operation of the electricity network (e.g. by predictive maintenance) and energy efficiency (e.g. by reliably predicting and monitoring energy savings).

Optimization Algorithms and Control Predictive functions, for both off-grid and on grid applications, require accurate load and energy production profiles to exploit the potentiality: AI-based models are employed to improve the accuracy of the forecast. Models including on-site measurement and sky-images are currently developed to further improve the PV and load forecast and will be tested in the micro-grid.

Next figures show a simplified view of the technical architecture of the pilot:

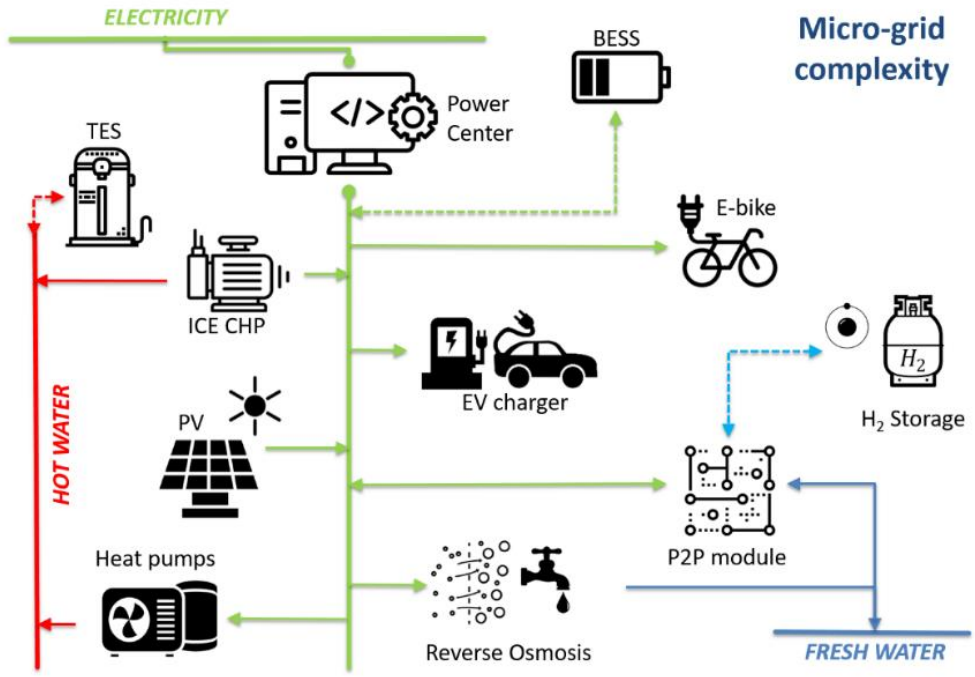


Figure 19 Micro grid architecture

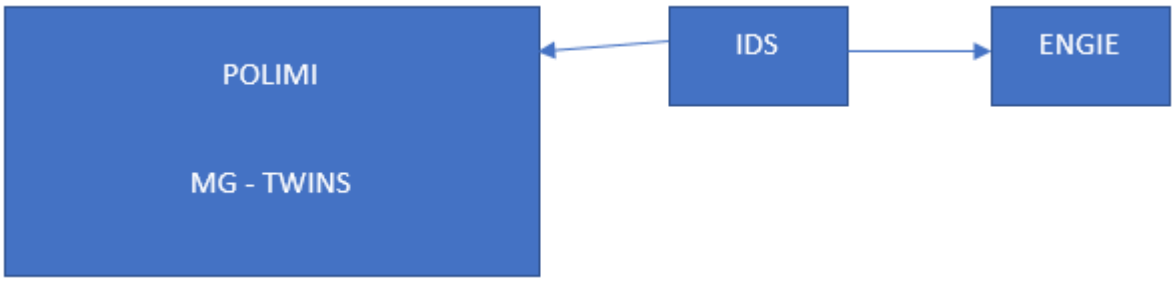


Figure 20 Pilot 4a architecture

According to the architecture, two companies that will exchange data: POLIMI and ENGIE, so one IDS connection will be developed.

POLIMI – ENGIE: Bidirectional

- From POLIMI to ENGIE: Data needed to train the model.
- From ENGIE to POLIMI: Results of analytic tools.
    - Solar energy production forecaster
    - HVAC load prediction
    - RES operation optimization

Answers to the data governance and privacy questionnaire:

1) **Are data usage constraints applicable?** Yes

- **Role based**: to ensure that data can only be accessed by specific departments/people in the company.
- **Purpose**: control different datasets for different functions of the model (e.g. train, operation)
    - Model Training (WP4): Historical High freq data: detailed current, voltages and powers for normality validation. Just waiting for NDA. Need to define how long. VUB to confirm which range of data. Probably there is one month data is more interesting (have more dynamics). Data freq: 500Hz
    - Model validation (WP6): Historical SCADA data for General fleet wide data will be used for validation of the developed models in WP4. Data every 10 min.
    - Model execution (WP6): "Real Time" SCADA data for General fleet wide data every 10 min

2) **Is personal data going to be transferred?** NO.
3) **Is data quality assurance a requirement?** NO.
4) **Is data provenance information a requirement?** YES. to track data exchanges between companies for audit reasons and potential monetisation as part of WP8.
5) **Is monetization of data a requirement?** Only to define the pricing model.
    - **Pricing model**: define the pricing model and business model but not implement it.

5 Conclusions and open issues

This section presents some conclusions drown during the requirement gathering process. The conclusions have been organized in two main topics:

- IDS infrastructure in PLATOON and IDS strategy and usage by the pilots.
- Business case for data transfer

IDS infrastructure in PLATOON

Next table provides information about the IDS infrastructure to be used by the pilots.

Component /service	Provider	Functionality/ Purpose	Improvements needed
Certification Authority	IDSA	Provide connector certificates	
Connector	ENG	Data transfer protocol Secure communication channel	• Data usage module • Personal data consent management • New IDS REST protocol • Data Apps deployment
DAPS	IAIS:	Identity provider. Provides security tokens.	

Broker/App store (metadata registry)	IAIS	Look for data and Apps metadata.	
Clearing House	IAIS	Track data exchange. • Data provenance • Data quality • Data monetization	
Vocabulary provider	TECN	Integrate Energy Date models with the IDS Information model	In development but need to define with IAIS more about its functionality.

PLATOON has identified a set of implementations of the IDS components needed by the pilots. However, the IDS final specifications are still under development and some decisions proposals are still ongoing. Some of those decisions could affect the IDS components development during PLATOON.

For example, currently, IDS is analysing the inclusion of other communication protocols like for example a REST based protocol. The addition of this new protocol is interesting for PLATOON since the Analytic toolboxes are going to expose their functionalities via OpenAPI REST interfaces and it will facilitate the adoption of IDS for data transfer needed for data analytic services.

Therefore, WP3 will focus its efforts on improving current component implementations, adding new functionalities (such as data usage management) or adapting them to new IDS specifications as they evolve.

Regarding the use of the IDS infrastructure by the pilots, most of them are interested in deploying IDS connectors with data usage functionalities for data transfer, at least for one of its scenarios. Data provenance is also a requirement. However, data quality and monetization are not included except for the definition of the data model.

Business cases for data transfer

In principle, several business models associated to data transfer via the IDS connectors have been identified by the pilots:

a) **Data transfer associated with the use of an "analytical toolbox":** A client of an analytical toolbox" sends the data needed and get the result.
b) **Transfer of data necessary for analytical task**: One company sends data to the analytical tool needed to provide the service to a third party.
c) **Data transfer for the training of a model**. A company transfer data to the analytical toolbox provider in order to train the analytical models
d) **Comparison of models between pilots**. Pilots covering the same functionality share data to compare their results.

However, in most cases the pilot selected business model is only related to the first option: Data transfer associated with the use of an "analytical toolbox". Furthermore, long term sustainability of the services, once the project has finished, and monetization are still open issues to be analysed during the project.

The realisation of the so-called European Energy Data Space, which is one of the main objectives of PLATOON will depend on the development of a trusted community of companies willing to share data.

6 References

[i] Brownlow, J., Zaki, M., Neely, A., & Urmetzer, F. (2015). Data and analytics-data-driven business models: A blueprint for innovation. Cambridge Service Alliance.

[ii] Schroeder, Ralph. "Big data business models: Challenges and opportunities." Cogent Social Sciences 2.1 (2016): 1166924.

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

[iv] https://www.cryptopp.com/wiki/GCM_Mode

[v] https://www.cryptopp.com/wiki/EAX_Mode

[vi] https://docs.wso2.com/display/IS530/XACML+Architecture

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

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

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