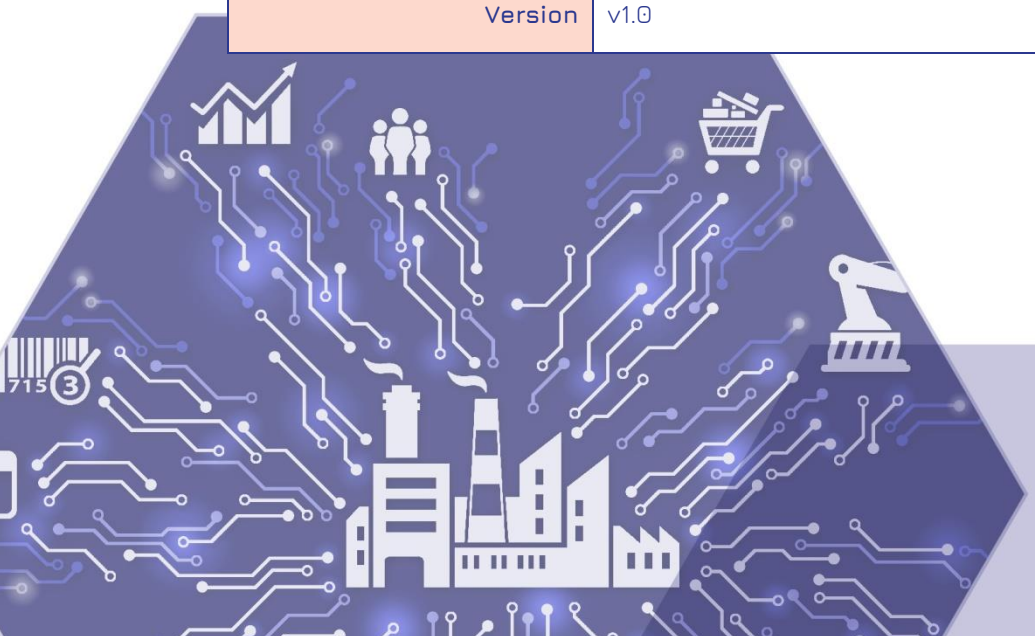
Boost 4.0
BIG DATA FOR FACTORIES

Big Data Value Spaces for Competitiveness of European Connected Smart Factories 4.0

Horizon 2020 EU Grant Agreement 780732

Title	D3.1 – Industrial Data Space v1
Document Owners	Christoph Mertens – IDSA
Contributors	Fabiana Fourier – IBM, Juan Jose Hierro – FIWARE, Anna Kasprzik – LUH, Christoph Lange – UBO, Christoph Mertens – IDSA, Alexandros Nizamis – CERTH, Konstantinos Sispas INTRASOFT, et.al
Dissemination	Public
Date	15/10/2018
Version	v1.0



Version history

01/06/2018	BOOST4.0 D3.1 v0.1 - ToC
22/06/2018	BOOST4.0 D3.1 v0.2 – Vision
13/07/2018	BOOST4.0 D3.1 v0.3 – Enabling Technologies and Infrastructures
10/08/2018	BOOST4.0 D3.1 v0.4 – Enabling Technologies Interaction
28/09/2018	BOOST4.0 D3.1 v0.5 – Document updates, alignment with D3.3. and conclusions
28/09/2018	BOOST4.0 D3.1 v0.6 – Ready for review
15/10/2018	Final version

Document Fiche

Authors	Christoph Mertens – IDSA, Fabiana Fourier – IBM, Juan Jose Hierro – FIWARE, Anna Kasprzik – LUH, Christoph Lange – UBO, Christoph Mertens – IDSA, Alexandros Nizamis – CERTH, Konstantinos Sispas INTRASOFT, et.al
Internal Reviewers	INNO
Workpackage	WP1
Task	T3.1, T3.2, T3.3, T3.4
Nature	Other
Dissemination	PU

Project Partners

Participant organisation name	Acronym
Asociación de Empresas Tecnológicas Innovalia	INNO
Volkswagen Autoeuropa, Lda *	VWAE
Visual Components	VIS
Automatismos y Sistemas de Transporte Interno S.A.U.	ASTI
Telefónica Investigación y Desarrollo SA	TID
Volkswagen AG. *	VW
UNINOVA	UNINO
FILL GmbH. *	FILL
TTTECH Computertechnik AG	TTT
RISC Software GmbH	RISC
PHILIPS Consumer Lifestyle B.V. *	PCL
PHILIPS Electronics Nederland	PEN
Interuniversitair Micro-Electronicacentrum VZW	IMEC
Centro Ricerche Fiat S.C.p.A. *	CRF
SIEMENS S.p.A.	SIEMENS
Prima Industries S.p.A	PRIMA
Politecnico di Milano	POLIMI
AUTOTECH ENGINEERING, AIE *	GESTAMP
Fundació Privada I2CAT, Internet I Innovació Digital A Catalunyai2cat	I2CAT
TRIMEK S.A.	TRIMEK
CAPVIDIA N.V,	CAPVIDIA
Volvo Lastvagnar AB *	VOLVO
Chalmers Tekniska Hoegskola AB	CHAL
Whirlpool EMEA SpA *	WHIR
SAS Institute Srl	SAS
Benteler Automotive GmbH *	BAT
It.s OWL Clustermanagement	OWL
Fraunhofer Gesellschaft Zur Foerderung Der Angewandten Forschung E.V.	FhG
Atlantis Engineering	AE
Agie Charmilles New Technologies SA *	+GF+
Ecole Polytechnique Federale De Lausanne	EPFL
Institut Für Angewandte Systemtechnik Bremen GmbH	ATB
Rheinische Friedrich-Wilhelms-Universitat Bonn	UBO
Ethniko Kentro Erevnas Kai Technologikis Anaptyxis (CERTH)	CERTH

The University of Edinburgh	UED
Institute Mines Telecom	IMT
International Data Spaces e.V.	IDSA
FIWARE Foundation e.V.	FF
GEIE ERCIM EEIG	ERCIM
IBM ISRAEL – Science and Technology LTD	IBM
ESI Group	ESI
Eneo Tecnología, S.L	ENEO
Software Quality Systems S.A.	SQS
Consultores de Automatización y Robótica S.A.	CARSA
INTRASOFT International	INTRA
United Technologies Research Centre Ireland, Ltd *	UTRC-I
Fratelli Piacenza S.p.A. *	PIA
RiaStone - Vista Alegre Atlantis SA *	RIA
Unparallel Innovation, Lda	UNP
Gottfried Wilhelm Leibniz Universität Hannover	LUH

*LHF 4.0 – Lighthouse Factory 4.0 * RF – Replication Factory 4.0

Executive Summary

This deliverable represents the first result of WP3 (tasks T3.1, T3.2, T3.3, T3.4) of the BOOST4.0 Project. This document is the first release of the Industrial Data Space specification. The deliverable presents the 4 reference data-value chain scenarios to be supported by the European Industrial Data Space (EIDS) technologies at various levels; namely (1) all products of one manufacturer, (2) all machines of one production line, (3) all suppliers and consumers of one product type and (4) all data sources for one business activity. The deliverable introduces also key elements on the digital infrastructures for the development of the EIDS. The deliverable introduces the FIWARE, IDS, Hyperledger fabric and Big Data Europe (BDE) framework and explains the plans and current development status of those initiatives to leverage an aligned solution that allows and easier, more cost-effective and trusted development of big data analytic services for Industry 4.0 services. The document introduces the EIDS vision, how it contributes to the development of the reference layered model for data-driven smart service development in Industry 4.0 and how Boost 4.0 plans to adopt and contribute to the Digital Shopfloor Alliance (DSA) solution development framework for cyber physical production solutions planned by the Boost 4.0 pilots.

The deliverable presents how smart big data features and vocabulary support features from BDE initiative can be exploited to support a usable implementation of the IDS business architecture. Along this line, the document presents how FIWARE technology can be used to implement such business architecture with Open Source Software (OSS) components and which is the preferred implementation option from the Hyperledger Fabric blockchain initiative to integrate distributed ledger functionalities in the IDS.

The document presents the laboratories and infrastructures (HPC and networking) that will be the baseline to define the extensions required to operate EIDS services and connectors over such high performance digital infrastructures.

The document concludes that there is a good alignment among the various open initiatives and that they can jointly contribute to the implementation of an open source reference framework for implementation of the EIDS business architecture. The document concludes that the first Proof of Concepts (PoC) developed are of good value to industry and that additional features in terms of better integration is necessary to ensure better usability of technologies and a full coverage of the 20% data sharing scenarios that are requested 80% of the time by industries willing to engage in inter-company or cross-company data sharing and exploitation.

Keywords: digital infrastructure, data sharing, industrial data space, blockchain, smart big data, semantics, vocabularies, context broker, FIWARE, Hyperledger.

Disclaimer

Acknowledgement

This document is a deliverable of Boost 4.0 project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 780732.

Table of contents

List of Figures

1 Introduction

1.1 Scope

Data sovereignty in trusted data sharing remains as one of the main challenges faced by industry in general and manufacturing in particular when it comes to implementation of advanced data value chains and realisation of data-driven business models and operations.

Boost 4.0 represents a concerted effort among equipment providers, digital platform providers, data analytics software providers, manufacturing industry, standardisation bodies, certification service stakeholders and digital innovation multipliers to set up the rules, standards and reference frameworks that should allow the free flow of data assets (interoperability), the growth of both a large big data market and manufacturing capabilities data-driven transformation in the context of Industry 4.0 globally and the emergence of a vibrant European big data-centric ecosystem to support Industry 4.0 servitisation.

Boost 4.0 ambition is to contribute to the vision that any device, any machine, any robot, any digital manufacturing platform and any company, anywhere in Europe will be able to connect, share, process and exploit in a trusted manner manufacturing data across a web of autonomous systems and to develop innovative and valuable big data-driven smart products and services to support human-centric, massively customised, energy efficient zero-defect and zero-breakdown advanced manufacturing processes and digital business models.

The main problems faced by industry today are related with the time, effort, usability, knowledge required and associated costs for exploiting data across the complete lifecycle leveraging a new generation of smart connected services. Industry should address critical barriers in order to ensure access to aligned and robust:

1. **world class high performance communication and computing big data digital infrastructures** with (far) edge data processing and analytics capabilities.
2. **multi-homed open big data spaces** with shared semantic models, vocabularies, data registries, certified data connectors and standardised context-information management components which can be widely applied to build open big data pipelines and middleware.
3. **Distributed and secure technologies** for implementation of industrial big data smart analytic services across complex and highly interconnected data ecosystems.

D3.1 contributes to this ambition providing the first version of the business reference architecture of such shared model; i.e. the European Industrial Data Space (EIDS), which will empower the European industry with the digital ability of dealing with industrial data and building advanced analytic services with full sovereignty and control. Boost 4.0 relies on the Industrial Data Space (IDS) vision and key European open initiatives to leverage the necessary tools and services to make a practical implementation of the EIDS principles across industrial pilots in particular and manufacturing industry in general.

Boost 4.0 project is organised as shown in the Figure below. WP3 is setting up and aligning digital infrastructures, big data service marketplaces and big data platform capabilities to the Boost 4.0 smart data and European Industrial Data Space principles for data sovereignty.

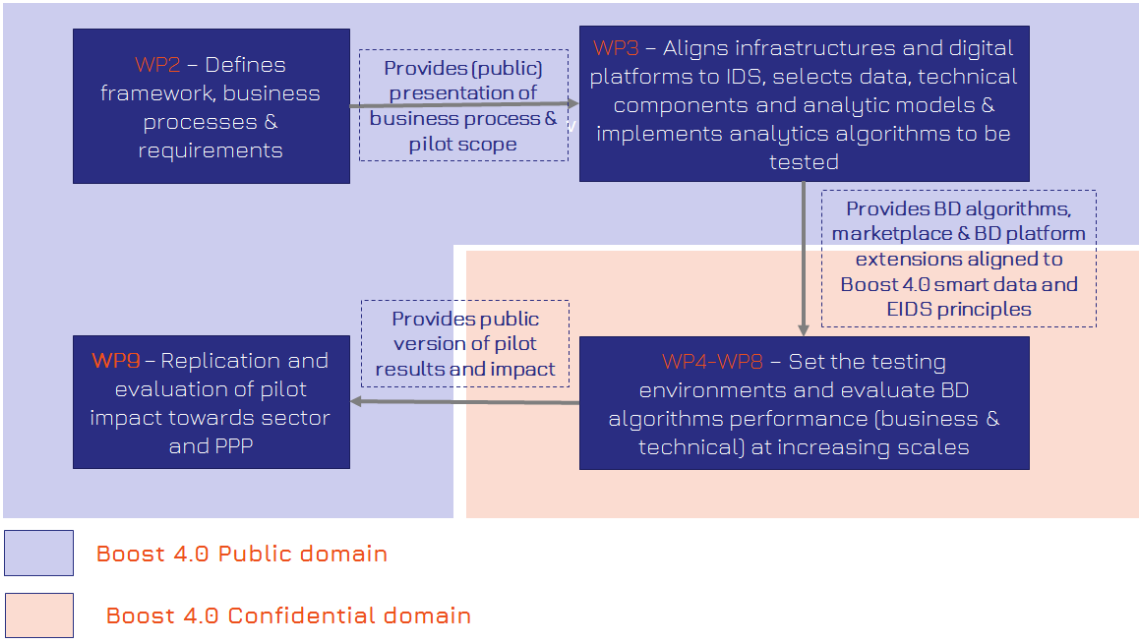


Figure 1 BOOST 4.0 Project Workpackage organisation

The aim of this document is to provide a first introduction to the various technologies and initiatives considered by Boost 4.0 in terms of enabling a European multi-homed open big data space for Industry 4.0 smart data applications. This document aims also at providing a first vision on the synergies and integration roadmaps for the key open initiatives considered under the Boost 4.0 umbrella; i.e. Industrial Data Space (IDS), FIWARE, Hyperledger Fabric and Big Data Europe (BDE).

1.2 Document Structure

The document is organised as follows. Section 2 provides a first introduction to the vision and elements for the implementation of the European Industrial Data Space (EIDS). Next,

Section 3 provides a first introduction to the EIDS building blocks (IDS, Blockchain technologies, BDE smart big data and vocabulary development technologies, FIWARE context broker technology; as well as briefly establish a first context for the interaction with the second pillar of the EIDS the big data analytic algorithms and big data digital manufacturing platforms. Afterwards, Section 4 presents the high performance computing and networking digital infrastructures that will support the implementation and operation of the EIDS vision and technologies. Finally, Section 5 presents the scenarios for interaction and current developments of EIDS enabling technologies. With Section 6, the document drives the main conclusions.

2 Boost 4.0 Vision for the European Industrial Data Space (EIDS)

One of the appointed goals of BOOST 4.0 is to develop a platform that works as enabler for Big Data technologies, the so called European Industrial Data Space (EIDS). The following section aims to present the vision of the EIDS and points out advantages of a data economy, which is enabled by the IDS (cf. chapter 3.1).

2.1 High level vision of the EIDS.

The EIDS creates an ecosystem, in which data is no longer just a simple sequence of numbers and letters, but data arises to an economic good that carries a specific value.

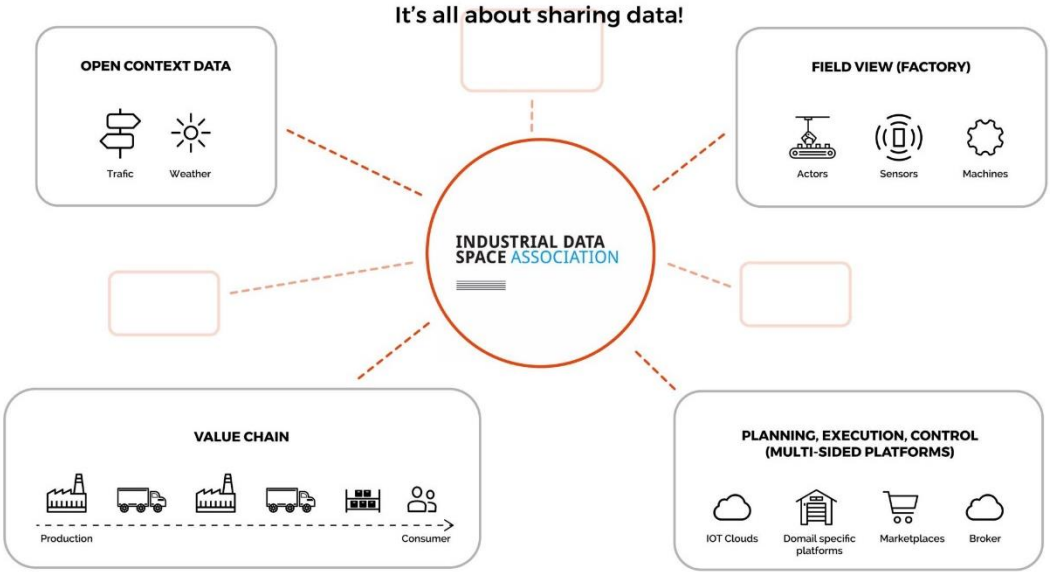


Figure 2: Context of EIDS ecosystem development

As illustrated above, to build an ecosystem like the EIDS basically two things are required: 1) a huge number of sources for raw data and 2) Big Data (BD) applications, which add value to the raw data. Nevertheless the mere existence of both components will not be sufficient to enable a data economy. There is the additional need to make the sources of data easily accessible on the one hand and on the other hand to give the promise to the owner of data that the data are only used in their sense. There must be trust between the different participants of the EIDS, which is why the fundamental idea of the EIDS is based on the IDS (cf. chapter 3.1). The difference is that the EIDS serves the special interest of the Big Data community, whereas the IDS is a domain-agnostic standard.

A BD application in general has the need for huge amounts of raw data from various sources. Reaching from a single sensor on a machine (data in motion) up to data bases of standard ERP systems (data in rest). In order to connect to these resources, nowadays individual interfaces have to be implemented. That leads to a high number of interfaces that all have to be developed individually. In order to make use of the network effect, the EIDS will make use of the concept of **Connectors** proposed by the IDS. Figure 3 depicts the EIDS ecosystem with participants of a greatly simplified supply chain and service providers from the field of BD technologies as well as service providers from the area of infrastructure (set up of production lines etc.). All these participants are connected to the EIDS via Connectors and therefore make the information that are available by their digital twin accessible to other participants of the EIDS. Even though the BD service provider is not part of the supply chain, he can make use of the data that are available from the partners of the supply chain anyway, because he is part of the same EIDS ecosystem. At the same time the owner of data has full control over his data and can decide which other participants are enabled to access which data – data sovereignty prevails.

Besides the Connector as a gateway towards the EIDS, there are several other architectural components which deliver various functionalities. The **Broker** enables BD service providers to search for data on the one hand and on the other hand serves as marketplace for providers of data. The **App Store** is the marketplace for BD applications. It can be used by participants of the EIDS to search the market for specific functionalities. The **Clearing House** is used for transaction logging and therefore can be used to implement different payment models (pay per use etc.). The **Identity Provider** ensures that each participant truly is who he pretends to be and is one key stone for trust in the EIDS. The **Vocabulary Provider** is the instance of the EIDS where ontologies are held in order to generate semantical descriptions of data. At the end there will be **Security** components that ensure safety and security for all participants of the EIDS. One central aspect in means of security are blockchains that grant the logging of various kinds of data, e. g. data that are transferred between supply chain partners like customer orders. The EIDS will most likely have a blockchain app that provides this security profile as a functionality to EIDS participants.

Besides the example where almost all EIDS participants belong to one supply chain, participants can come from all different kinds of domains and levels of abstraction. An EIDS participant can be a single sensor which generates a constant stream of ever changing data, as well as an IT system, like an ERP system, which contains master data that change very rarely. Participants of the EIDS can either be providers of public data, like e. g. weather data, or providers of data, which can only be used by specific participants. The variety of possibly available data makes the EIDS a very important standard for the BD community, as it is the enabler for big data.

It is foreseen, that the EIDS can be joined by new participants easily. This will lead to a higher interest for new companies to be part of the ecosystem and will set the boundaries for entering the EIDS low. The definition of processes regarding the usability of the EIDS will be part of the project.

Figure 3: Vision of EIDS with Supply Chain as example

2.1.1 Boost 4.0 approach to realise the EIDS vision

The previous Section has presented the high-level view and main building blocks for the development of a European Industrial Data space for data-sharing and big data application development. The implementation of the EIDS is well aligned with most recent data-driven frameworks for Industry 4.0; e. g. Smart Service Welt – see figure below.

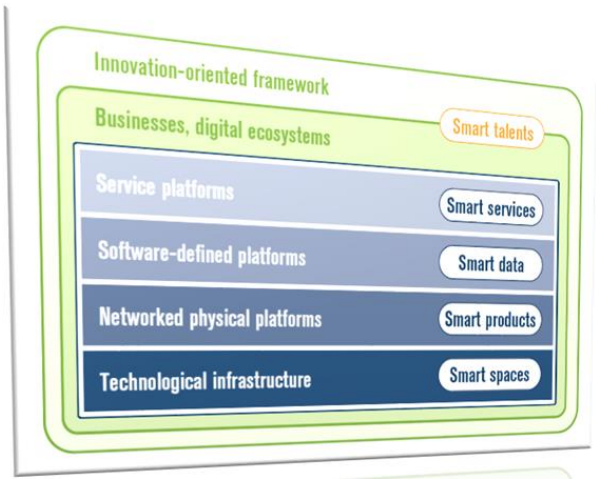


Figure 4: Smart Service Welt data-driven layered reference framework

The data-driven framework relies on the development and smartisation of 4 main domains:

- **Smart Services** – Data-driven added value apps.
- **Smart Data** – Software defined data-driven platforms

- **Smart Products** – Networked Physical products, machines…
- **Smart Spaces** – Infrastructure

The realization of the EIDS relies on the implementation of such 4 layers under a shared model for data-driven operation of Factories 4.0. The implementation of the **Smart Space layer** will imply that both networking and computing infrastructures will need to be adapted and extended to support both the business logic and the technical solutions behind the EIDS operations. This demands not only the successful technical operation of the technologies at hand, but also the definition of clear procedures and supporting tools for the modelling, design, configuration and deployment of EIDS enabled data services, digital manufacturing platforms and/or data connectors and vocabularies. The realization of the EIDS vision at the Smart Space layer calls for taking advantage of micro services and advanced virtualization technologies; as well as open technologies such as OpenFog that will enable an easier support for the operation of smart analytic services over a wide variety of digital infrastructure (edge nodes, HPC centers, cloud platforms).

The implementation of the **Smart Product layer** implies an extension of the product and machine capabilities to operate in the context of the EIDS. In particular, the approach will focus on providing standardized mechanisms for the secure and trusted connection and data exchange protocols of manufacturing equipment (coordinate measurement machines, injection machines, milling machining centers) and the shared industrial data spaces. Moreover, product information is critical for a more agile operation of factories 4.0. In Boost 4.0 this will translate in the support of standardized and advanced Industrial IoT capacities and the adoption of standard vocabularies, implementation of data adapters, connectors and the support of standard connection protocols supported in the context of Industry 4.0; e. g. OPC-UA.

The implementation of the **Smart Data layer** implies the development of specific data-driven services for data registry, semantic data lifting, security, traceability, context-management, data app store ensuring that the various data assets can be easily brought together. In Boost 4.0 this is intimately related to the integration and extension of IDS, FIWARE, Hyperledger Fabric and BDE technologies (data preparation, processing) to set the foundations for the implementation of open big data pipelines across the full industry 4.0 lifecycle.

The implementation of the **Smart Service layer** implies the development of specific data-driven algorithms and models and a marketplace to access such cognitive capabilities as part of advanced smart connected services. In boost 4.0 this relates to the development of mission driven algorithms and capabilities for analysis of different types of data (time series, point-clouds, data lakes, data streams etc.) and leveraging different digital abilities

in the shop floor (visualization, analysis, prediction, autonomy). The implementation of this layer will also entail the demonstration of how very advanced digital manufacturing capabilities leveraged by digital manufacturing platforms could be operated within the trusted environment of the EIDS connector and overall EIDS data space.

2.1.2 EIDS big data driven smart application development

The EIDS vision is intended to serve the vision described in the previous section and to deal with the ecosystem and a business value chain that considers various roles:

- "**Smart Service User:** Provider of specific context information and service co-creator responsible for final value generation from physical and data-driven digital services/goods.
- **"Smart Service" Provider:** In charge of added value services and the dynamic configuration of services for users in their specific situations. Superior user experience (Augmented Reality) and generation of attractive business models.
- **Smart Platform Operator**, in charge of organizing connectivity, processing and analysis of data, defines rules for collaboration between provider and user and ensures security and privacy. Data fusion and -processing (Smart Data) and trustworthy distribution of data.
- **Makers of "Smart Products"**, which cater for the connection and the interaction between intelligent products to platforms. Efficient generation of Big Data

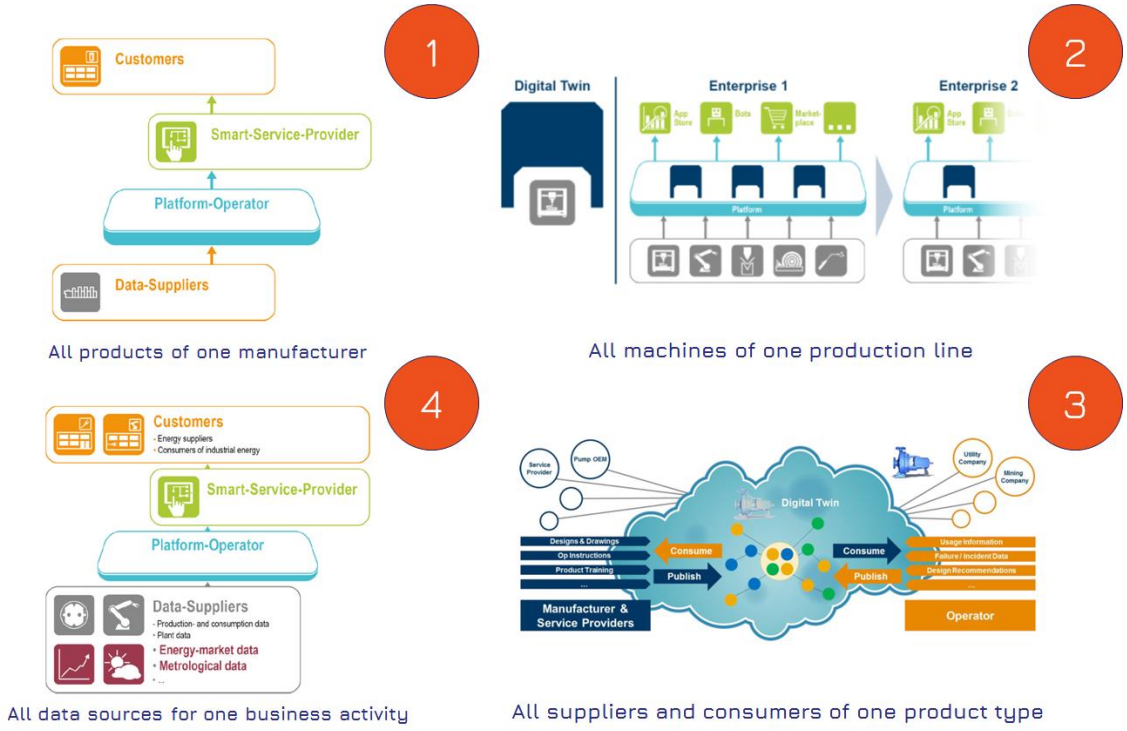


Figure 5: EIDS data value network scenarios.

The EIDS vision is intended to cover increasingly complex and diverse scenarios for data-collaboration – see Figure 5.

1. All products of one manufacturer
2. All machines of one production line
3. All suppliers and consumers of one product type
4. All data sources for one business activity

As already discussed in the previous section, central to the implementation of the EIDS support of the various data value chains is the IDS Connector. In this context, it is worth noting that for the development of data-driven smart applications, the IDS Connector will fulfill two very distinct functions; namely the data sharing across internal data silos (the so called Internal IDS Connector) and across data clouds (the so called External IDS Connector) – see Figure 6. While both connectors will share a same reference model, the business logic and business value will be very distinct; as well as the use cases to be supported. As illustrated by the figure below, the EIDS model allows for coexistence with non-IDS communications as well as exploitation of open data beyond the industrial one.

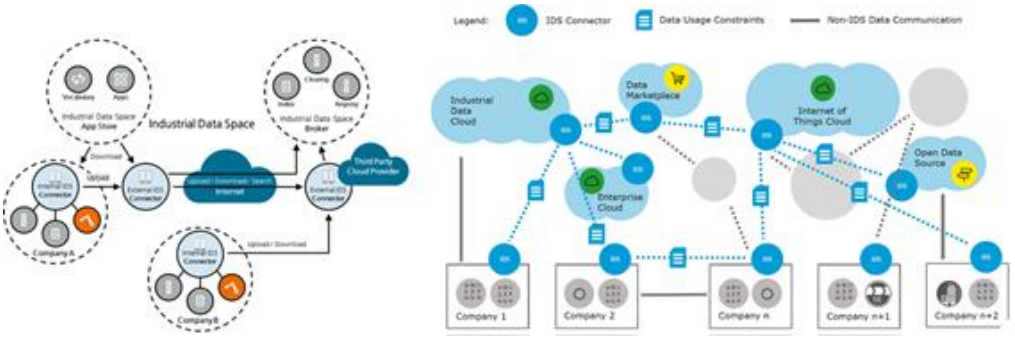


Figure 6: EIDS data-driven smart application development context.

In order to support a holistic approach to smart application development based on big data, Boost 4.0 is adopting the European Factories of the Future Research Association[1] (EFFRA) migration paths developed in the context of the ConnectedFactories initiative.

[1] https://www.effra.eu/

Figure 7: Factory 4.0 Digital Transformation Path (Adapted from Acatech 2018)..

The pathway for the smart autonomous factory is shown in Figure 8. The pathway defines an evolution of the digital architecture as well as a digital manufacturing platform, manufacturing equipment and Shop floor automation capabilities. The pathway is fully aligned with the Acatech migration pathway for Industry 4.0 – see Figure 7.

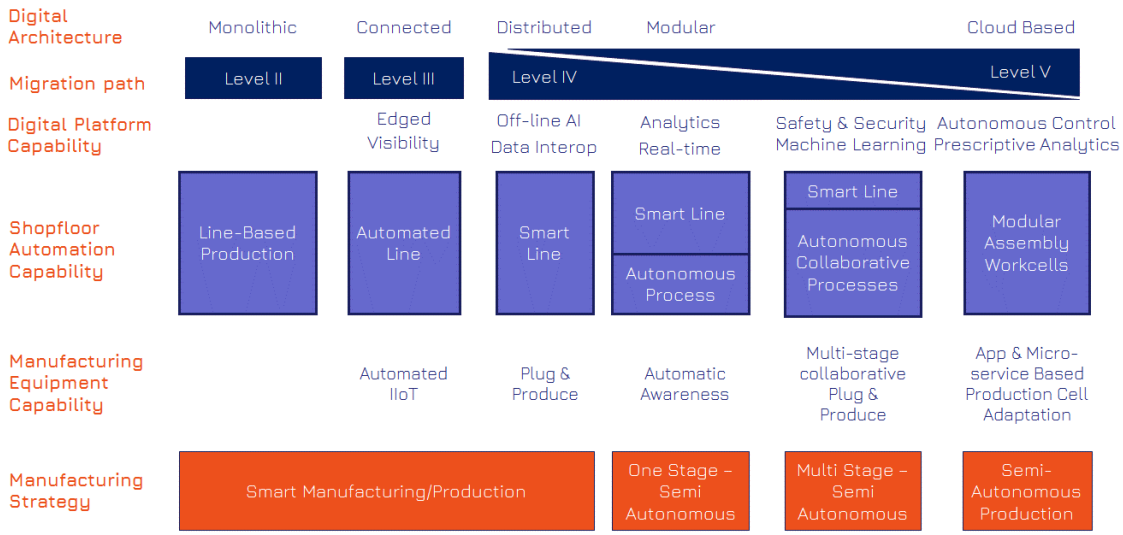


Figure 8: ConnectedFactories smart autonomous factory migration pathway.

Boost 4.0 EIDS contribution will mainly contribute to up to Level IV pathway implementation with specific contributions to the Edge, visibility, data interoperability and analytics/real-time operation capabilities of the digital platforms. In the manufacturing equipment capability dimension Boost 4.0 will contribute to increased IIoT support and plug &produce features in relation to data sharing and vocabulary development. Boost 4.0 will therefore make a significant contribution to the foundation and realisation of this Factories 4.0 pathway for data-driven operations.

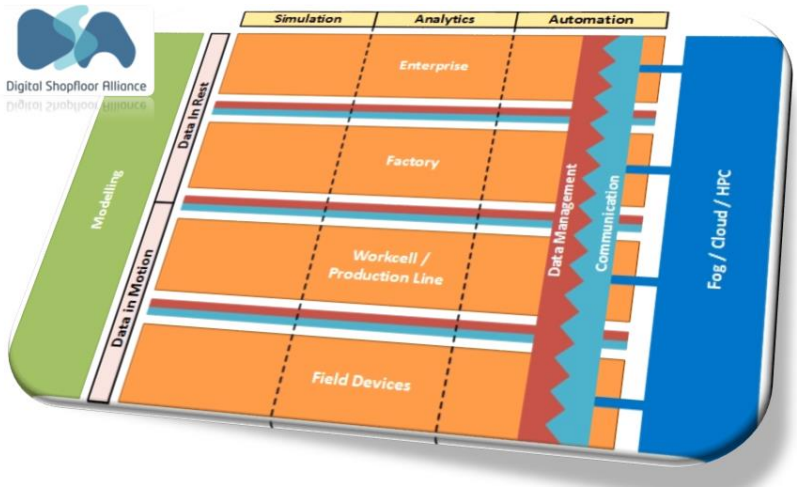


Figure 9: Digital Shopfloor Alliance Data-driven Smart Solution Development Framework.

For the development of the EIDS data-driven smart applications, Boost 4.0 approach is to follow the Digital Shopfloor Alliance (DSA) smart solution development framework. Boost 4.0 smart connected big data driven solutions implemented in the pilots with the support of the EIDS approach will map to the DSA smart solution framework.

The goal of the DSA solution development framework is to have a broad industrial applicability, map applicable technologies to different areas, and to guide technology and standard development. From a structural perspective, the DSA solution development framework covers two different areas denoted as domains:

- **Design domain**: it describes the design and development methods, tools and services for designing Cyber Physical Production Systems & Solutions (CPPS). The components of the design domain enable users to intuitively design the applications (the so called automatic awareness digital ability usability services).
- **Runtime domain:** it includes all the systems that support the execution and operation of the smart autonomous CPPS.

The DSA solution development framework has four layers/levels – see Figure 9:

- **Enterprise:** The enterprise layer is the top layer of the DSA solution development framework and encompasses all enterprise's systems; as well as interaction with third parties and other factories.
- **Factory:** At the factory layer, a single factory is depicted. This includes all the various workcells or production lines available for the complete production.
- **Workcell / Production Line:** The workcell layer represents the individual production line of cell within a company. Nowadays, a factory typically contains multiple production lines (or production cells), where individual machines, robots etc. are located in or become a part of.

- **Field Devices:** The field devices layer is the lowest level of the reference architecture, where the actual machines, robots, conveyer belt, etc., but also controllers, sensors and actuators are positioned.

To uphold the concept of Industry 4.0 and to move from the old-fashioned automation pyramid (where only communication was mainly possible within a specific layer and to establish communication between the different layers, complicated interfaces were required), the communication and data management concept is a "Pillar" to cover all the mentioned layers. The communication pillar enables direct communication between the different layers. The Pillar is named **Fog/Cloud/HPC** and foresees the use of wired (e. g. IEEE 802.1 TSN) and wireless communication to create direct interaction between the different layers by using Fog/Cloud/HPC concepts (blue column in Figure 9 above. In good alignment with this paradigm this pillar is also responsible for data persistence and potentially distributed transaction management services across the various components of the smart autonomous digital manufacturing system.

Finally, the last part of the DSA solution development framework focuses on the actual **engineering, modelling, programming and configuration** of the different technical components inside the different layers (green column in Figure 9 above). On each layer, different tools or services are applied and for all of them different modelling approaches are available. The goal of these modelling approaches is to ease the end user / system developer / system integration developing the tools or technologies for the different levels. Additionally, it could be possible to have modelling approaches that take the different layers into account and make it easier for the users to model the interaction between the different layers.

The DSA solution development framework, also represents the two **data domains** that the architecture anticipates, namely the data in motion and data in rest domains. These layers are also matched in the architecture with the **type of services** automation, analysis, learning/simulation that are also pillars of the RA. The model also represents the layers of the RA where such services could be executed with the support of the fog/cloud computing and persistence services (blue pillar in Figure 9).

The DSA solution development framework for data-driven cyber physical production systems (CPPS) can be extended in a more detailed view of the layered architecture as shown in Figure 10.

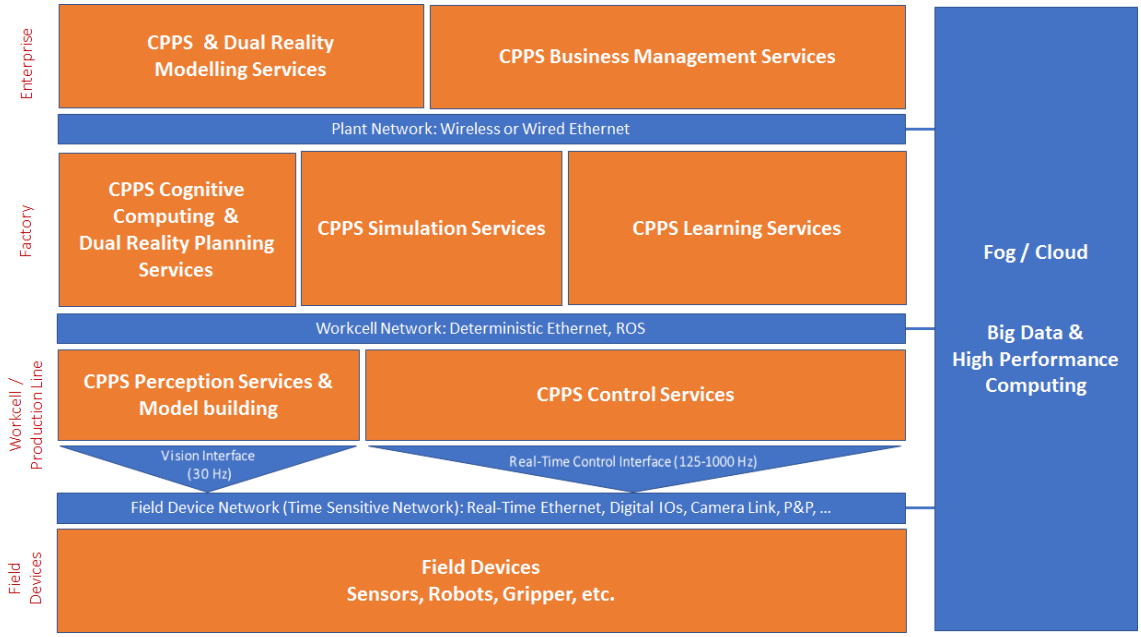


Figure 10: Digital Shopfloor Alliance Data-driven Smart Solution Development Framework (Detailed view).

The DSA solution development framework provides an integrated framework where different data services, manufacturing platforms, data analytic algorithms and communication and computing infrastructures can be put together with field devices & manufacturing equipment to realise big data driven CPPS.

The DSA solution development framework takes into account various digital capabilities in the **Field layer** mainly related to the sensoring and monitoring capabilities of connected products and machines. In particular, this layer considers IIoT technologies as baseline elements for sensing the manufacturing shop floor as well as the smart product. Similarly, this layer hosts the services related to embedded intelligence, connectivity and analytics provided by connected manufacturing equipment.

The DSA solution development framework proposes two types of digital capabilities in the **workcell/production line layer.** In particular (1) the CPPS perception services & model building related with the tools and technologies used to generate automatically awareness about the context of manufacturing operation and the continuous adaptation or a fast application of models for decision making or actuation. (2) the CPPS control services related to tools and technologies used to adapt the operation of equipment in the shop floor in accordance with well-orchestrated manufacturing plans.

The DSA solution development framework proposes three types of digital capabilities in the **Factory layer**. These three digital capabilities are directly linked to tools and services used for generating intelligence and smartness in the digital shop floor. These three capabilities are intended to support a more cost-effective and precise insight generation

allowing both the individual and the integrated operation of machine learning algorithms operating on big data, formalised knowledge application by means of advanced simulation frameworks and digital twin graphs. Each of those capabilities are linked to (1) CPPS learning services operating on massive data sets (2) CPPS simulation services extracting insights from formalised knowledge and (3) CPPS dual reality planning services and cognitive computing algorithms taking advantage from digital twin graphs and data models. Boost 4.0 techniques and algorithms for model development will enhance the capabilities of this layer.

The DSA solution development framework proposes, finally, two types of digital capabilities in the **Enterprise layer**. On one hand, the services and tools related to visual analytics & manufacturing process and CPPS model generation and refinement in the CPPS & Dual reality modelling services. Boost 4.0 will be instrumental in the development of advanced algorithms and features in big data manufacturing platforms for engineering, commissioning, planning, operations and after sale services as part of this digital capability. On the other hand, this layer also hosts CPPS Business Management Services related mainly to data interoperability and collaboration across IT and inter-cloud operations. Boost 4.0 will contribute to the development of this digital capability extending and adopting IDS and FIWARE context broker technologies as part of the Plant and workcell data sharing spaces and communication infrastructures.

3 EIDS Enabling Technologies

The following subsections will describe the basic technologies on which the EIDS will be built on. The fundamental architecture is based on the architecture of the reference architecture model of the **International Data Spaces**. The other subsections are descriptions of important corner stones of the EIDS, like the **FIWARE Context Broker** technology that connects different kinds of participants to the EIDS and brings specific functionality them, **Block Chain** that plays a role in context of Safety and Security in the EIDS, **Smart Big Data and Vocabularies** that are the basis to enable semantical and therefore machine readable descriptions of data to the EIDS and **Big Data Services and Platforms** that bring big data functionality to the EIDS ecosystem.

3.1 International Data Spaces (IDS)

The International Data Spaces Association (IDSA)[2] is an association with almost 100 members from all different kind of industries and sizes. The aim of the IDSA is to establish a worldwide standard for the exchange of data. The constantly growing amount of data and the raising of the data economy call for standards, which ensure the sovereignty of data. That is what the so called IDS standard will deliver, by defining an architecture for a whole new ecosystem, which enables the exchange of data in a secured and sovereign way in a therefore trusted environment.

Digitisation is both, an enabler and a driving force behind innovative business models. A key ability for innovative business models is to be able to combine data in one ecosystem.

- Services are decoupled from physical platforms and products
- The architecture levels are decoupled
- Products become platforms and vice versa
- Ecosystems develop around platforms
- Innovation takes place cooperatively

Data as strategic resource, is an enabler for smart services, smart products and our desired lifestyle of the future, which stands for transparency and simplification of everyday tasks and processes amongst others. Data will become an economic asset, and to some extend are already. Therefore the key focus for a data-driven economy and new evolving business models is linking data, which will lead to a whole new level of information availability.

[2] https://www.internationaldataspaces.org/

When we take a look at the history of data sharing, the International Data Spaces (IDS) seems to be the logical evolution of concepts and technologies that have emerged in recent decades. It all started off with electronic data interchange (EDI) in the 1960's. EDI is a standard that continues to be widely used. On top of EDI came eBusiness in the 1990's, a collective term that comprises all kinds of information and communication technologies. These technologies are enabler for digital business processes that go beyond company boundaries. This applies in particular to sales and procurement processes. With the broad acceptance of IoT in combination with RFID around the 2010s, it became possible to collect and exchange supply chain event data. For a long time now, data is no longer only collected and used within the company's own boundaries, but is also widely used in the entire supply chain. All this logically leads to the need to exchange data in a way that allows you, as the originator of the data, to retain full sovereignty over the data flow. That is, what the IDS ecosystem stands for.

Value of data and provenance, dynamic linking of data endpoints, high amount and rate of data processed. (What is the statement behind slide 7?)

Data sovereignty is the basis for trust between partners of the ecosystem. It is the ability of natural or legal persons to exclusively and sovereignly decide concerning the usage of data as an economic asset. Data sovereignty on the one hand is about ownership, security and value of data and on the other hand about interoperability, data exchange, "sharing economy" and data centric services.

There are certain obstacles concerning the extensive sharing of data, today. 57 % of companies worry about revealing valuable data and business secrets to others. 59 % fear the loss of control over their data. 55 % feel inconsistent processes and systems as a (very) big obstacle. 32 % fear that platforms do not reach the critical mass, so that data exchange will be interesting[3].

IDS brings improvements with regards to these obstacles, e. g. data security is ensured, sovereignty is improved and processes and cost structures are optimised. Furthermore IDS has the opportunity to enable access to yet untouched data treasures in companies, while staying in control over its flow and usage. This can be achieved in five steps:

1. Make data available. This can be done dynamically or on demand. In order to do so, data have to be described before they can be exposed by ontologies.
2. Link with partners of the ecosystem (cf. Figure 11). First you build a connection, then match data demand and data offer and afterwards start interpreting the data.

[3] Source: PwC study "Data exchange as a first step towards data economy", 2018, p.24, available at https://www.pwc.de/en/digitale-transformation/data-exchange-as-a-first-step-towards-data-economy.pdf

3. Control the access to your data by making use of the usage control.
4. Start to create value. This can be done by implementing apps, granting remote software execution or aggregation of information.

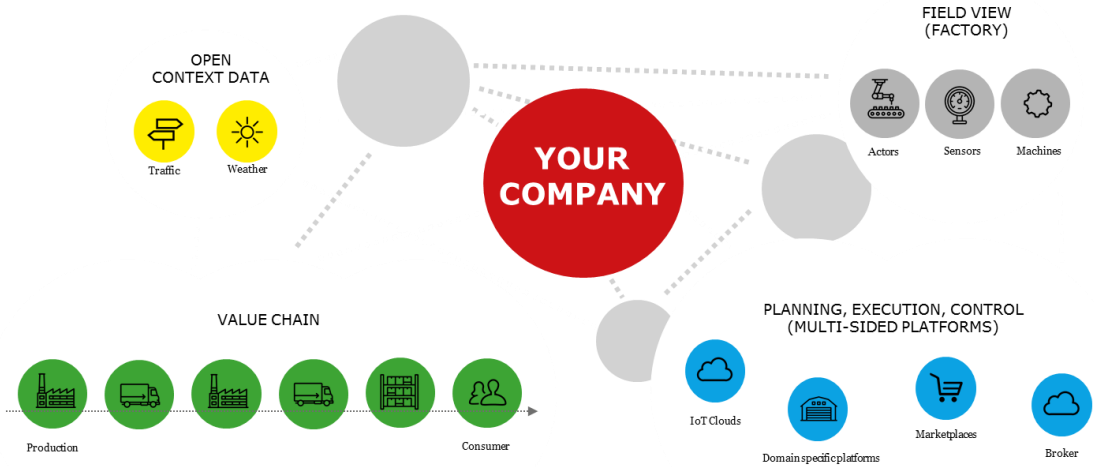


Figure 11: Overview of the IDS ecosystem

This procedure works equally well for all industries and domains. It grants a self-determined control of data flows and leads to the following characteristics:

1. Endless connectivity. IDS is a standard for data flows between all kinds of data endpoints.
2. Trust between different security domains. IDS comes with comprehensive security functions, providing a maximum level of trust.
3. Governance for the data economy. IDS implements usage control and enforcement for data flows.

IDS basically is a peer-to-peer network of industrial data. All actors oblige themselves to play by the rules of IDS. To grant, that all actors and components follow the rules, they must be certified by a trusted authority. The IDS provides usage control for data and different tailor-made levels of trust. Its architecture leads to certain characteristics:

- Economies of scale due to networking effects,
- Open approach because IDS is neutral and user-driven,
- Trust due to the fact, that all participants must be certified,
- Decentral approach due to the distributed architecture,
- Sovereignty over data and services,
- Data governance dictate the “rules of the game”,
- Network of platforms and services.

An overview of the general structure of the IDS Reference Architecture Model (RAM) is given in Figure 12. It defines five layers, each coming with three perspectives. The different layers will be described briefly in the following sections in order to give an overview[4].

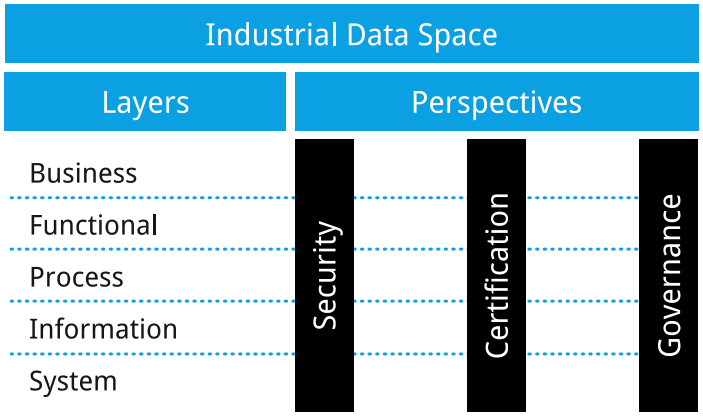


Figure 12: General Structure of IDS Reference Architecture Model

3.1.1 Business Layer

The Business Layer specifies and describes the different roles that participants of the IDS ecosystem can assume. Furthermore the Business Layer specifies the main activities and interactions connected with each of those roles. Due to the fact that the business layer provides an abstract description of the roles in the IDS, it can be considered a blueprint for the other more technical layers. The roles that can be assumed in the IDS ecosystem can be seen in Figure 13.

The **Data Owner** holds all legal rights of its data and has complete control over it. The Data Owner not necessarily is the one who makes the data available. Therefore, the role of a **Data Provider** is foreseen. In most of the cases the Data Owner and Provider are one single IDS participant.

The **Data Consumer** receives the data of a Data Provider. And again the Data Consumer does not necessarily need to be the one who processes the data or makes use of it. Hence there is the role of the **Data User**, who has the legal right to use the data as specified by the user policy.

If a company does not match the technical requirements to be participant in the IDS ecosystem, they can transfer the data to a **Service Provider**. The Service Provider serves for two purposes, 1) to make the data available towards other IDS participants and 2) to

[4] https://www.internationaldataspaces.org/publications/ids-ram2-0/

eventually process the raw data before transferring it with its own data services towards a Data Consumer and Data User.

In case a Data Consumer does not know its counterpart, the Data Provider, there is the possibility to search for data via a **Broker Service Provider**. The Broker Service Provider stores and manages information about the data sources available in the IDS. The activities of the Broker Service Provider are focused on receiving and providing metadata about the actual data, that later will be interchanged.

The **Clearing House** is an instance that logs all activities performed in the course of data exchange in the IDS and it therefore provides clearing and settlement services for all financial and data exchange transactions. Based on the logging information the transaction of data can be billed and possible conflicts can be resolved.

The **App Store** is the instance that provides data apps that can for instance be used for data cleansing, integration or analysis. The functionality of app can range from the simple transfer of data over scheme transformation, data aggregation, and analysis to entire systems that implement complex business processes. The apps themselves are built by an **App Provider**, who publishes its apps via the App Store.

The **Vocabulary Provider** is the instance that manages and offers vocabularies (i. e. ontologies, reference data models and metadata elements) that are used to annotate and describe datasets, in order to publish data descriptions in form of metadata at a Broker.

Further roles in the IDS are the **Identity Provider** that ensures the identity of each IDS participant, the **Software Provider** that provides software for the App Store and the **Certification Body and Evaluation Facility**, which are in charge of the certification of the participants and the technical core components in the IDS.

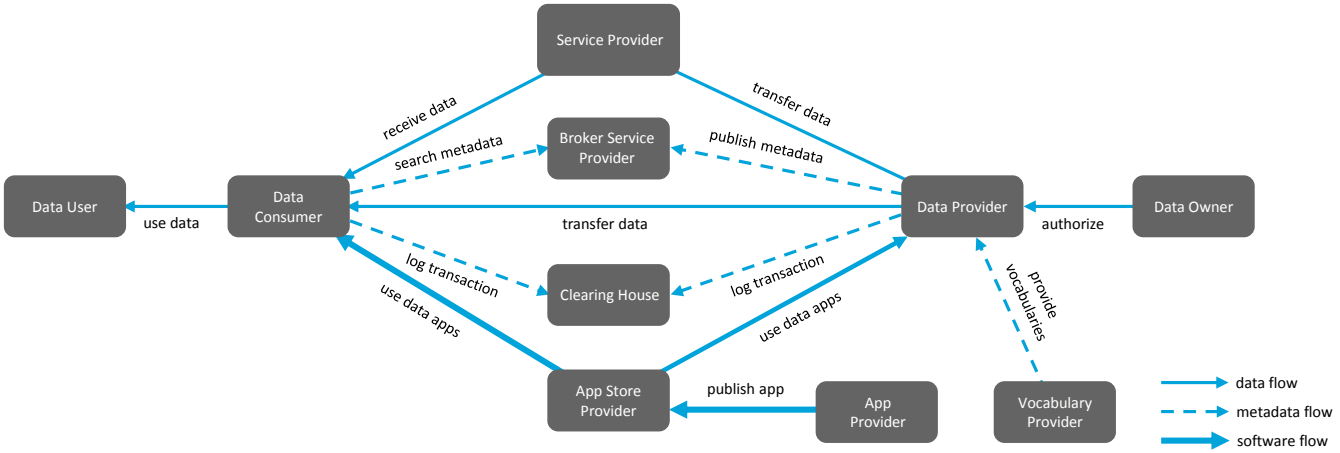


Figure 13: Roles and Interactions in the IDS Business Layer

3.1.2 Functional Layer

The Functional Layer defines the functional requirements of the IDS and the features to be implemented resulting from these requirements. Figure 14 shows the functional architecture of the IDS, which is divided into six functional entities that have to be delivered by the IDS. Each of this six entities comes with specific requirements. The requirements presented here are only an extract of all relevant requirements.

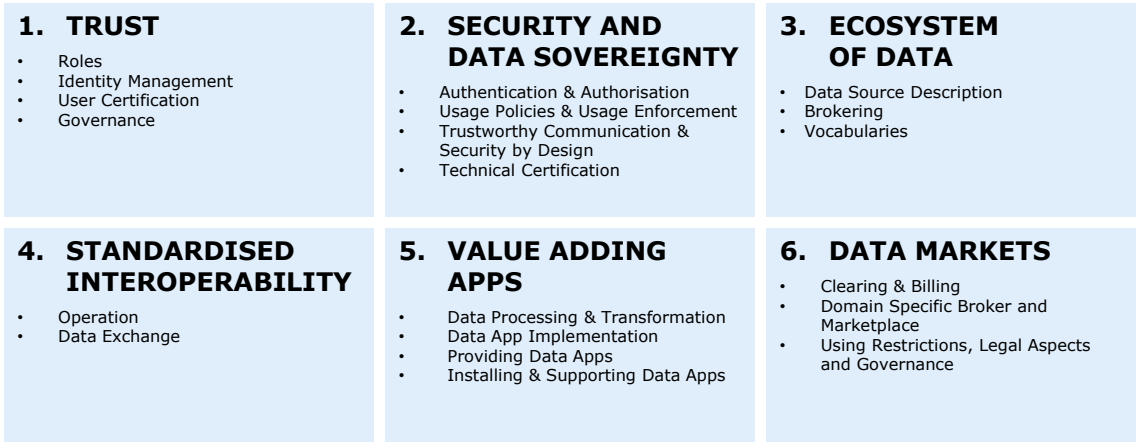


Figure 14: Functional Architecture of the IDS

The first functional entity **Trust** can be split into four main aspects, which are:

- ***Roles***: The different roles in the IDS (cf. Section 3.1.1) have differing needs regarding the level of trust they have to deliver. For example, the role of a broker can only be assumed by a trustworthy participant, whereas a data provider depending on the use case not necessarily has to be trustworthy (e. g. in case of a weather provider).
- ***Identity Management***: In order to create trust among the IDS participants, every connector has to have a unique identifier and a valid certificate. Each connector must be able to verify the identity of other connectors.
- ***User Certification***: The users connected to the IDS require certification in order to create trust.
- ***Governance***: To make sure, that participants follow the rules of the IDS, governance is required.

The second functional entity deals with **Security and Data Sovereignty** and can be split into the following aspects:

- ***Authentication and Authorisation***: Each connector has to have a valid X.509 certificate, which enables IDS participants to identify each other's identity. On top certain conditions, like security profiles can be applied here.
- ***Usage Policies and Usage Enforcement***: Each data owner is able to define usage control policies for their data that are attached to the outbound data.

Therefore, Data Owners can be sure, that their data are treated according to their policies by the Data User.

- *Trustworthy Communication and Security by Design*: Connectors, the App Store, and Brokers can check if the Connector of the connecting party is running a trusted and therefore certified software stack. Any communication between Connectors can be encrypted and integrity protected.
- *Technical Certification*: The core components of the IDS and especially the Connector require certification from the certification body in order to establish trust among all participants.

The third functional entity is about the **Ecosystem of Data** and is split in three aspects:

- *Data Source Description*: Participants of the IDS have to be able to describe, publish, maintain and manage different versions of metadata. Therefore metadata must describe the syntax and serialisation as well as the semantics of the data source. Additionally, it must describe the application domain of the data source. Further aspects that can be described by the metadata are the price of the offered data, the price model and the usage policy.
- *Brokering*: The operator of a Connector must be able to provide an interface for data and metadata access. Each Connector must be able to transmit metadata of its data sources to one or more brokers. Every participant must be able to browse and search metadata in the metadata repository of a Broker, considering the participant has the right to access the metadata. Furthermore each participant of the IDS must be able to browse the list of participants registered at a Broker.
- *Vocabularies*: In order to create metadata Vocabularies are needed. It is possible to either use standard Vocabularies, create own vocabularies or to work together with other participants on the creation of further vocabularies.

The fourth functional entity **Standardised Interoperability** is split into two aspects:

- *Operation*: Participants are enabled to run a Connector in their own IT environment. Alternatively they can make use of a mobile or embedded connector. The operator of a Connector must be able to define the data workflow inside the Connector. Users of the Connector must be identifiable and manageable.
- *Data Exchange*: The sending Connector must receive data from a backend system, either by a push or a pull mechanism. The data then can be provided to other participants of the IDS via an interface or pushed directly towards the receiving participant. The interchange of data only works, if both Connectors involved are uniquely identifiable. After receiving the data on the side of the Data Consumer, it can be written into the backend system of the consumer (if the usage policy allows to do so).

The fifth functional entity **Value Adding Apps** is split into two aspects:

- *Data Processing and Transformation*: A data processing app must provide a clearly defined functionality, which is applied on the input data, and then produce a defined output. A data transformation app must be able to transfer an input with a specific format into an output with a different output format, in order to match the requirements of the Data Consumer/User. The transformation app does not change the information itself.
- *Data App Implementation*: Developers of a data app must describe the app's functionality and interfaces, pricing model and licensing amongst others by metadata.
- *Providing Data Apps*: Any authorised developer of Data Apps is allowed to publish Data Apps on an App Store. The Apps to be published can be undertaken a certification process that is performed by the Certification Body. The App Store supports participants in searching for Data Apps that fit the need of the participant.
- *Installing and Supporting Data Apps*: After retrieving the fitting Data App from an App Store, it is installed at the Connector of the Data Consumer/User. The Connector also supports the installation of Data Apps, which do not come from an official App Store.

The sixth and last functional entity **Data Markets** is split into three aspects:

- *Clearing and Billing*: The owner of data is able to define the pricing model and the price of the data. The transaction of data is logged in the Clearing House. Based on the logged information on the transfer, the user of the data can be charged (e. g. pay per transfer, pay for access per day or month).
- *Domain-Specific Broker and Marketplaces*: The Broker can not only be used to search for data sources but it can also be understood as Marketplace for data, where data that can be bought by a participant of the IDS is presented via metadata.
- *Using Restrictions, Legal Aspects and Governance*: Governance is split into five aspects, which are data as economic good, data ownership, data sovereignty, data quality and data provenance.

3.1.3 Process Layer

The Process Layer of the IDS RAM specifies the interaction that takes place between the different components of the IDS. Therefore it is split into three main processes, which are:

1. Providing Data (cf. Figure 15)
2. Exchanging Data
3. Publishing and Using Data Apps.

In these three main processes, all roles described in the Business Layer (cf. section 3.1.1) are involved.

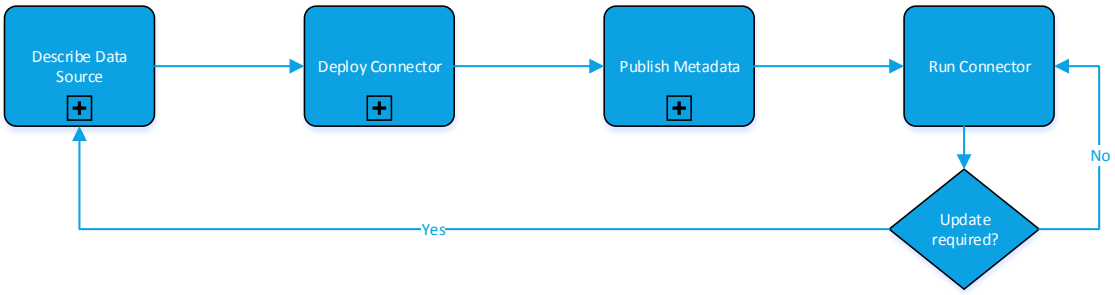


Figure 15: Overall Process of Providing Data

3.1.4 Information Layer

The Information Layer specifies the Information Model of the IDS and by doing so it defines the fundamental architecture of the IDS without being restricted to a specific domain. It primarily aims at describing Data and Applications in the context of IDS, which are the core components of the IDS (called resources). Besides that the Information Model also describes Participants of the IDS as well as their Interaction amongst each other (interchange of data) and basic Infrastructure Components that are necessary in order to operate an IDS ecosystem.

All the resources of the IDS and their dependencies are described on three different levels of abstraction that reach from generic and descriptive up to specific and executable representations (cf. Figure 16).

The **Conceptual Representation** is a very generic overview of the main resources of the Information Model. Pear group for this level of description are mainly the general public and management boards with an interest in the major context of IDS. It therefore mainly consist of textual documentations and graphical representations. The **Declarative Representation** adds a semantical level to the description of resources. Based on a stack of W3C technology standards (RDF, RDFS, OWL etc.) and standard modelling vocabularies (DCAT, ODRL etc.) it therefore provides a formal, machine-interpretable specification of the IDS resources. The **Programmatic Representation** is meant for developers as target group and therefore aims to seamlessly integrate the Information Model with common development infrastructures. It is the most specific description of the IDS and also executable.

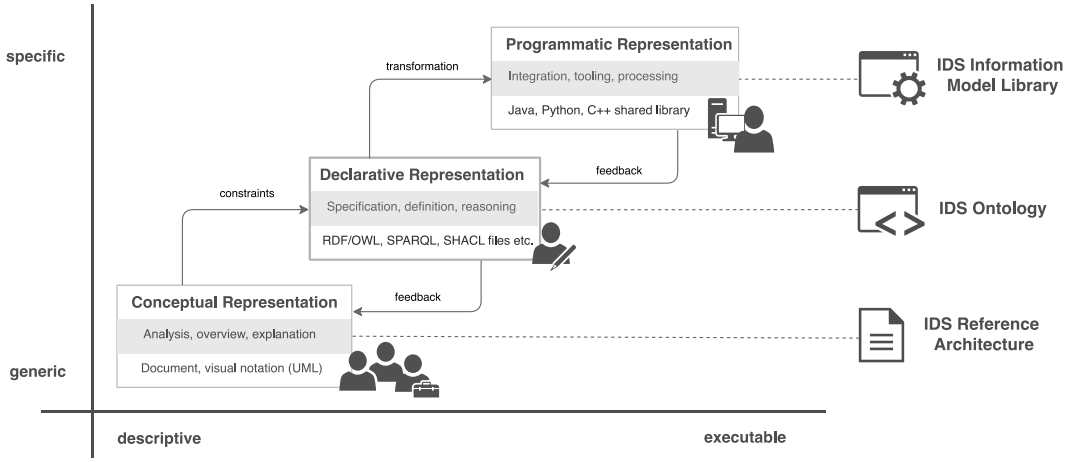


Figure 16: Representations of the Information Model

3.1.5 System Layer

The System layer maps the roles that are defined in the Business Layer with the data and service architecture on the Functional Layer in order to create the technical core of the IDS. There are three major technical components, which are required to realise an IDS on a fundamental level: The Connector, a Broker and an App Store. The interaction of these three components are depicted in Figure 17.

Besides that the System Layer also describes the architecture of the core components, e. g. for the connector.

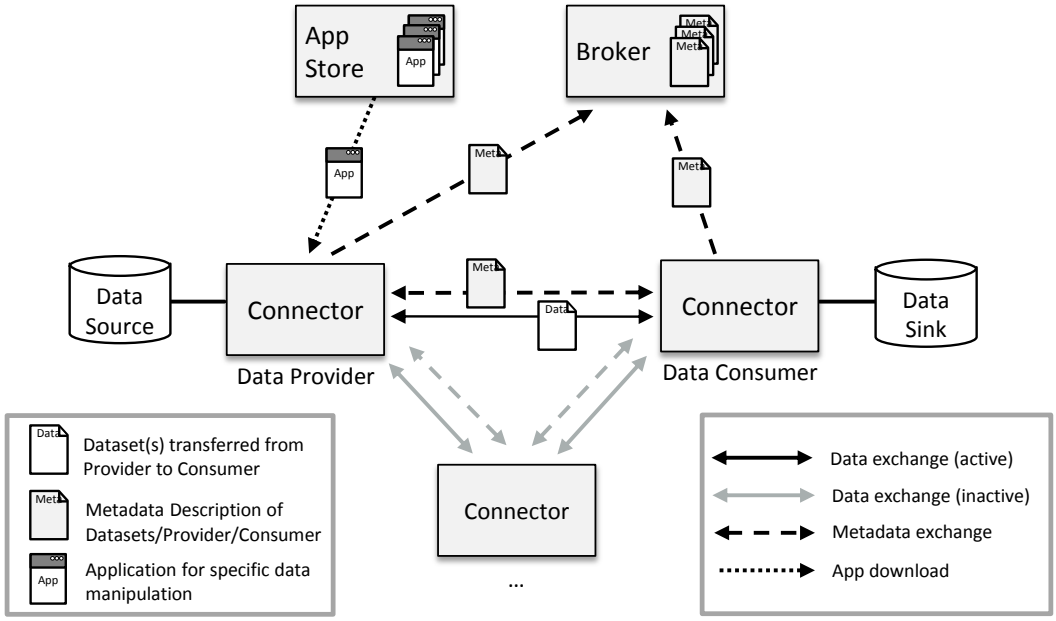


Figure 17: Interaction of Technical Core Components of the IDS

3.2 FIWARE Context Broker

The FIWARE[5] Context Broker is a component of the FIWARE ecosystems of technologies for simplifying the development of smart applications. The core functionality of the FIWARE Context Broker is that it integrates information coming from a variety of systems, ranging from robotic and other machines and sensors in industrial complexes, to information coming from systems in smart agricultures and cities. The Context Broker provides a holistic picture of what is happening with the data at any time instance (i. e. the "context" in a variety of use-case domains. This Context Information Management has been identified as a cornerstone in the design of smart solutions by relevant companies and organisations. As a result, the European Telecommunications Standards Institute (ETSI[6]) created an Industry Specification Group in January 2017, which has recently published a preliminary version of the ETSI NGSI-LD[7] open API specifications. These specifications are based on the FIWARE NGSIv2 API currently supported by the FIWARE Context Broker. The FIWARE Context Broker is envisioned to become an open source reference implementation of the ETSI NGSI-LD standard. This cornerstone component of the Reference Architecture is based on a European-rooted open standard.

Different Internet of Things (IoT) platforms can be integrated with the FIWARE Context Broker. These IoT platforms bring the conversion between a variety of contemporary IoT protocols (e. g. MQTT, CoAP, OMA-LWM2M) and the NGSI-LD. The FIWARE community maintains an open source implementation of several agents for most popular IoT protocols. The IoT platform to be integrated in the Context Broker may itself align with the ETSI OneM2M[8] specifications, as it is for example the case for the open source OpenMTC IoT platform developed by Fraunhofer FOKUS[9]. Following the advanced trends of cloud/fog/edge computing, technologies like EdgeX Foundry or FogFlow have either expressed their interest to become integrated or have already been integrated with FIWARE Context Broker.

Context information evolves over time, generating a continuous stream of data. Multiple processing engines can be connected to process these data streams. Processing engines include big data analysis (based on Apache Hadoop) for the extraction of insights, real-time stream processing platforms like Apache Flink, Spark or Storm, and engines supporting large-scale numerical computation based on data-flow graphs, for example

[5] https://www.fiware.org/
[6] https://www.etsi.org/
[7] https://docbox.etsi.org/ISG/CIM/Open/ISG_CIM_NGSI-LD_API_Draft_for_public_review.pdf
[8] http://www.onem2m.org/
[9] https://www.fokus.fraunhofer.de/en

TensorFlow. All these brings support to processing of complex events, as well as machine learning and other kinds of AI algorithms.

Current and historic context information can be visualised using operation dashboards, which rely on technologies suited for the visualization, as well as mashup technologies enabling users to design user interfaces tailored to their needs. More traditional business intelligence tools are also integrated as a part of the Reference Architecture bringing support to monitoring of KPIs, as well as reporting and analytic functionalities.

The features of FIWARE Context Broker make it a suitable candidate for the implementation of the IDS Reference Architecture Model defined by the International Data Spaces Association (cf. chapter 3.1), as will be in more details discussed in Section 3.2.2. The FIWARE Context Broker, the component enabling configuration of multiple data sink connectors and the proxy that implements API management functions like accounting, throttling or access control, together map to the core components of an IDS Connector, which is the core component of the IDS Reference Architecture. Using FIWARE tools designed for the deployment of FIWARE-based IDS Connectors, system adapters as well as processing, analysis, and visualization platform components can be deployed and properly configured to preserve a secure and trustworthy exchange of data between data providers and consumers in separate systems, where the provider has the full control over the data that is being shared with the consumer.

3.3 Block Chain

The following subsections explain how blockchain on a very general level works and how it is used in the context of the Hyperledger project.

3.3.1 Overview

In today's digitally networked world, no single institution works in isolation. Organizations aim at creating new value, optimize their business, and reduce risk within their business networks. Industry business networks are not an exception. These networks benefit from connectivity among customers, suppliers, manufacturers, partners, things, and cross geographic and regulatory boundaries. Wealth is generated by the flow of goods and services across these business networks in transactions and contracts. However, growth of wealth can be constrained if the networks are heavily silo'd or inefficient.

Ledgers are the key and **the** system of record for recording asset transfer in and out of a business. Ledgers are not new and have been in use since the 13th century for double-entry book keeping. *Assets* are anything owned or controlled to produce value. We distinct between two types of assets: tangible (e. g., a house) and intangible (e. g., a mortgage).

In today's manufacturing networks (Figure 18), each participant in a network keeps their own ledger(s) which are updated to represent business transactions as they occur. This is expensive due to duplication of effort and intermediaries adding margin for services. It is clearly inefficient, as the business conditions for transaction to occur – "the contract" – is duplicated by every network participant. It is also vulnerable because if a central system (e. g. Bank) is compromised due to an incident this affects the whole business network. Incidents can include fraud, cyber-attack or a simple mistake.

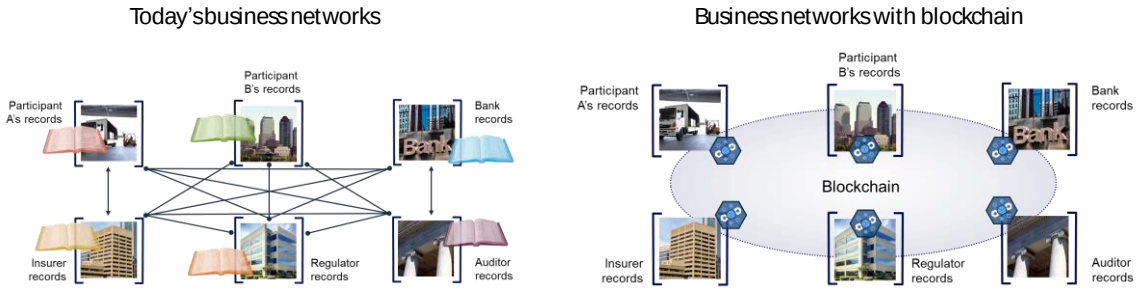


Figure 18: Business networks before and with blockchain

Blockchain is a digitally distributed ledger of database of records, transactions, or executed events that are shared across the participants/parties in the business network. Each transaction in the system is time stamped and verified by a consensus algorithm (see below).

Some fundamental characteristics of blockchain business networks:

- *Shared ledger* - There is one shared ledger instead of individual ledgers for each participant in the business network (Figure 18). This shared ledger is updated every time a transaction occurs through peer to peer replication. Businesses can have multiple ledgers for the multiple business networks in which they participate. A *transaction* is an asset transfer onto or off the ledger (e. g., John gives a car to Anthony) whereas a *smart contract* specifies the conditions for a transaction to occur (e. g., If Anthony pays John money, then car passes from John to Anthony). Blockchain allows the contract for asset transfer to be embedded in the transaction database. *Cryptography* is used to ensure that network participants see only the parts of the ledger that are relevant to them, and that transactions are secure, authenticated and verifiable.

- *Consensus* - Network participants agree how transactions are verified through *consensus* or similar mechanisms. Consensus means all participants agree that a transaction is valid. Government oversight, compliance and audit can be part of the same network.

- *Provenance* – means participants know where the asset came from and how it's ownership has changed over time
- *Immutability* – means no participant can tamper with a transaction once it's agreed. If a transaction was in error then a new transaction must be used to reverse the error, with both visible
- *Finality* – means that there is one place to determine the ownership of an asset or completion of a transaction. This is the role of the shared ledger.

Blockchain for business comprises the following four main blocks: shared ledger, smart contracts, privacy, and proof. Together they add *trust* to a business network. Shared ledger and smart contract are the 'things' that constitute blockchain; privacy and proof are the qualities of service.

Shared ledger – The shared system of records. It is an append-only distributed system of record shared across business network participants. Records all transactions across business networks. Participants have own copy through per to peer replication. It is permissioned, therefore participants see only appropriate transactions

Smart contract - Business rules associated with the transaction that express the business terms for execution. The smart contract is simply a piece of code that runs; the input parameters to the code (and a reference to the code itself) are stored as the transaction details. Smart contracts are encoded in programming language and verifiable.

Privacy - Transactions are secure with appropriate visibility. The ledger is shared, but participants require privacy and confidentiality. This means that not every member of business network can see the entire blockchain. Cryptography controls who can see what and is central for privacy. Privacy implies user registration process to build trust network and access permission via certification management. Transactions need to be authenticated whereas identity is not linked to a transaction.

Proof - Transactions are provably endorsed by relevant participants. Participants endorse transactions in the business network. Business network decides who will endorse transactions. Endorsed transactions are added to the ledger with appropriate confidentiality. In addition, the ledger is a trusted source of information as assets have a verifiable audit trail since transactions cannot be modified, inserted, or deleted. Proof is achieved through consensus, provenance, immutability, and finality

Blockchain technologies enable eliminating inefficiencies in business to business processes created by lack of trust and transparency, and create innovative business processes by streamlining the exchange of value along the business network. Blockchain

supports a new generation of transactional applications and streamlined business processes by establishing the trust, accountability, and transparency that are essential to smart connected factory 4.0 operations.

Manufacturing is one of the sectors to be most impacted by blockchain technology. According to a report from Frost and Sullivan[10], automotive ecosystem participants are expected to spend ~0.6% of their total IT spend on blockchain by 2025. Furthermore, smart manufacturing, supply chain logistics, retailing and leasing, connected living and IoT, mobility solutions, R&D, and aftermarket, are some of the automotive key functional areas to be based on blockchain technology.

3.3.2 Hyperledger Fabric

Hyperledger is a trademark of the Linux Foundation and is the umbrella for a group of projects designed to advance blockchain for business. Hyperledger was founded in February 2016; and today encompass five frameworks and four tool projects with more than 185 member organizations (partner IBM is a premier member). Hyperledger is the fastest growing project in Linux Foundation history and has an immense amount of industry support. The project promotes open source, open standards, and open governance. Open governance means that there is no one controlling organisation that governs the direction of the project, and no lock-in to one particular vendor.

Hyperledger Fabric (simple Fabric)[11] is one framework that is underneath the Hyperledger umbrella. Contributors to Fabric include IBM, DTCC, Fujitsu and others. Other frameworks include Iroha and Sawtooth. Fabric is an implementation of blockchain technology that is a foundation for developing blockchain applications, with emphasis on ledger, smart contracts, consensus, confidentiality, resiliency, and scalability. Fabric V1.0 was released in July 2017 and it involved 159 developers from 27 organizations. IBM is number one contributor of code, IP and development effort to Hyperledger Fabric.

Fabric provides an open source, industrial-grade implementation of a private blockchain used as the distributed ledger and smart contract engine for BOOST 4.0.

A *permissioned* or *private* blockchain restricts the actors who can contribute to the consensus of the system state. In a permissioned blockchain, only a restricted set of users have the rights to validate the block transactions. A permissioned blockchain may also restrict access to approved participants who can create smart contracts. Fabric is an example of a private blockchain targeted for business networks. On the other hand, in a

[10] Frost and Sullivan report K13A-18 "Blockchain Technology Revolutionizing Automotive Industry", 31 March 2017
[11] http://hyperledger-fabric.readthedocs.io/

public or *permissionless* blockchain anyone can join the network, participate in the process of block verification to create consensus and also create smart contracts. A good example of permissionless blockchain is the Bitcoin and Ethereum blockchains, where any user can join the network and start mining (for a comparison between permissioned and permissionless models refer to[12]).

According to a recent survey from Juniper from August 2017[13], IBM was perceived as the leading provider of Blockchain technology in the world with more than 400 engagements. In the BOOST project, partner IBM is responsible for the applicability of Blockchain technology to the project use cases.

3.4 Big Data Europe (BDE) Smart Big Data & Vocabularies

Combining data from various sources in different formats and modelling views requires both syntactic and semantic integration efforts. Whereas a syntactic integration can be established by relying on common data formats such as XML or JSON, the semantic variety needs an additional shared and formalized understanding of the meaning of the used terminology. In particular, searching and selecting data sources, exchanging datasets, and invoking and accessing software services require explicit information on the meaning of concepts.

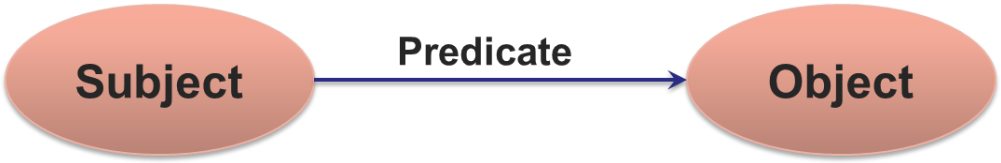


Figure 19: RDF structure. Each resource is identified by an URI

A common model to express shared vocabularies and to formalize relations between concepts is the Resource Description Framework (RDF[14]). RDF uses URIs to identify entities and relations among them, resulting in a triple-based structure being both human- and machine-interpretable (Figure 19). Subjects and objects can be regarded as nodes, being connected through predicates in the form of directed, labeled edges. In this way, RDF datasets form extensive knowledge graphs where concepts of any kind are clearly defined and interlinked. Furthermore, standardized axioms based on RDF Schema and OWL

[12] T. Swanson. Consensus-as-a-service: A brief report on the emergence of permissioned, distributed edger systems. Report, available online, Apr. 2015. URL: http://www.ofnumbers.com/wp-content/uploads/2015/04/Permissioned-distributed-ledgers.pdf.
[13] https://www.juniperresearch.com/resources/infographics/blockchain-enterprise-survey-august-2017
[14] https://www.w3.org/RDF/

definitions support detailed modelling of a domain of interest but also enable automated reasoning in order to explicitly provide implicit knowledge.

Portals such as the Linked Open Vocabularies[15] and the Fraunhofer VoCol instance[16] serve as starting points for discovering commonly used concepts. Additional domain specific vocabularies can be introduced in order to describe certain industry requirements or use cases. Commonly known examples are the Semantic Sensor Network Ontology (SSN)[17] or the Sensor, Observation, Sample, and Actuator (SOSA)[18] ontology. Widely used tools for creating and editing ontologies are the OWL IDE Protégé, the web-based WebVOWL and VoCol, the latter serving as a Reference implementation of the IDS Vocabulary Provider. All of those tools will be used in BOOST 4.0 in order to develop and maintain the shared vocabularies collaboratively.

3.5 Big Data Services & Marketplace.

The BOOST 4.0 Online Collaborative Analytics Service Marketplace will be the meeting point for a multi-level supply and demand data service business ecosystem. In this ecosystem will be available a wide variety of Big Data Services, Tools and Algorithms. The provided Marketplace services will be developed as IDS applications in order to support the collaborative and trust use of data in the Marketplace.

The use of IDS architecture, Vocabularies, IBM Hyperledger Fabric and FIWARE connectors will enable the creation of data-driven, secure and novel services for the Marketplace. These will enable apps injection to connectors to add the provided services to the top of the data exchange. The provided services will be related to data processing and analytics and will enable the remote execution of algorithms over the Marketplace. By using the aforementioned technologies and tools of the BOOST 4.0 platform, the developed data analytics services will rely on security, data sovereignty and standardized interoperability.

As the pilot partners of the project are related to industry domain, the available services to the Marketplace will be related to algorithms and data models for this domain and the supply chain domain which is considered as a supporting domain to the manufacturing one. The provided collaborative analytics services of the BOOST 4.0 Marketplace will be presented in details at the corresponding deliverable D3.3 Big Data Models and Analytics Platform v1 (M9) and of course in its final version D3.4 Big Data Models and Analytics Platform v2 on month 24. However, a short overview of the future available services at the Marketplace is provided in the following paragraphs as these will be the provided services

[15] http://lov.okfn.org
[16] https://vocol.iais.fraunhofer.de/
[17] https://www.w3.org/TR/vocab-ssn/
[18] http://www.w3.org/ns/sosa/

and algorithms running in the BOOST 4.0 Big Data Platform and they will contribute to the creation of a complete and competitive Big Data Value Space.

The main services related to predictive and prescriptive models, and data and visual analytics that will be available through the Marketplace to the pilot partners or other vendors from the same domain and even from different domains categorized as follow:

- Services, tools and algorithms related to prediction of production of defected products, based on the modelling of production assets' deterioration rate. Predictive modelling and early detection algorithms will be developed and used in order to add these types of services to the BOOST 4.0 Marketplace. Traditional techniques such as machine learning, data mining, information theory and statistics able to deal with this crucial problem will be offered. Furthermore, more innovative algorithms and new techniques such as Complex Event Processing (CEP) analysis techniques and trend analysis will be designed and be available to the Marketplace as well. The use of cognitive modeling for a cognitive manufacturing will also enable the creation of services that will predict unforeseen conditions during the production process and will boost the real-time decision-making.
- Services, tools and algorithms related to the detection of deterioration rate of production machines and their root cause. Model-based Fault diagnosis methods such as deterministic models and probabilistic models will be designed or adopted. Besides the Model-based Fault diagnosis methods Signal-based Fault diagnosis (Time-Domain, Frequency-Domain, Time-Frequency Domain) models will be offered as a services to the BOOST 4.0 Marketplace too.
- Services, tools and algorithms related to the advanced data visualization and visual analytics. New ways of interaction and display technologies to support analytical reasoning and collaborative decision making in the big data available to the Big Data Value Space of the BOOST 4.0 will be adopted by the project partners. Interactive Visual Analytics service in order to visualize, analyze and explore industrial data derived from multiple sources and enable the detection of events which are not clearly observed by providing multiple data views will be available as well.

The above described services and algorithms will extend the capabilities of the BOOST 4.0 Big Data Analytics platform and will be applied in the field of manufacturing. The analytic services will be running in the analytics platform and will be available through the Big Data Value space and the BOOST 4.0 Marketplace to everyone that will be eligible to participate in the Marketplace and set a capable connector. The input and the output of the services will be described by Vocabularies in order to provide interoperability and enable the use of the services from third parties that can provide and request data based on the terms of

these Vocabularies. The use of the Vocabularies and the provided Big Data Pipeline enables BOOST 4.0 project to offer a Marketplace in which the stakeholders will be able to use the platform and its services by setting their own IDS and FIWARE connectors.

3.6 Big Data Manufacturing Platforms

The role of Big Data Platforms is to integrate and extend a variety of existing commercial tools/platforms covering Decision Support, Data Analytics, Knowledge discovery, etc. Tool integration will focus on developing connectors capable of leveraging the Boost4.0 ecosystem which provides access to a wide variety of data but also Big Data Services, Tools and Algorithms, but also data governance and security features. Tool extension will be designed and planned based on the requirements of the Boost4.0 pilots (WP2) and the particular details of each manufacturing environment (WP4-8).

The details of the work carried out in context of T3.6 Big Data Analytics Platform is detailed in length in deliverable D3.3 Big Data Models and Analytics Platform v1 organized in paragraphs per tool/platform. Each paragraph provides an overview of the existing functionalities that are currently offered, describes existing data governance mechanisms and also needs that will be covered by leveraging Boost4.0 ecosystem features, and extensions planned in order to cover requirements coming from the manufacturing domain.

A short overview of existing tools is provided hereafter:

[1]. SAS analytics:

SAS Analytics provides a set of different tools and functionalities:

a. SAS Viya
b. SAS Visual Forecasting
c. SAS Visual Data Mining and Machine Learning
d. SAS Event Stream Process.

[2]. ATLANTIS DSS for smart maintenance:

A tool offering rule-based mechanism, which handles manually defined actions based on pre-specified static threshold violations. Such tools are used to assess the performance of the machines, to diagnose failures and overall to improve the maintainability and the operational efficiency of the production line.

[3]. Siemens MindSphere

The open, cloud-based IoT operating system from Siemens that lets you connect your machines and physical infrastructure to the digital world. It lets you harness big data from billions of intelligent devices, enabling you to uncover transformational insights across your entire business.

[4]. RISC framework data analytics extension

Includes the tools CALUMMA and Collibri mentioned in the application. Collibri focuses on the integration of large data volumes as well as continuous data streams. Further developments outside of the project Boost 4.0 can be read in the big data stack afterwards. CALUMMA is an ontology-based research infrastructure for domain expert driven knowledge discovery. The main principle behind the system is to put the domain experts with their knowledge and experience in the centre of the knowledge discovery process.

[5]. TRIMEK M3

The M3 platform which is poised to provide a structured solution for Metrology4.0, an edge-powered quality control analytics, monitoring and simulation system. This solution is used for the organization, analysis and reporting operations of the metrological information, taking advantage of the storage and computational capabilities of the cloud to carry out advanced operations and provide smart added value services

[6]. REAL TIME LOCATION SYSTEM AND NETWORK MONITORING

Time Asset and Network Monitoring framework will be based in two platforms: redborder provided by ENEO and i2track provided by i2CAT.

[7]. ESI cloud platform

ESI cloud platform delivers Virtual synchronized through the Internet of Things (IoT) and Big Data to receive real-life product feedback to maximize the longevity of a product through intelligent predictive maintenance.

[8]. CERTH IoT Platform

A cloud-based platform developed by CERTH/ITI. The platform:

- Enables Big Data Storage
- Supports real-time monitoring services
- Supports multiple continuous connected IoT devices
- Offers Real-time Predictive Analytics services

4 EIDS Enabling digital infrastructure

Besides the standardisation of the architecture, the EIDS has to be enabled on the digital infrastructure side. Specially for big data purposes, the EIDS has to match certain criteria. On one hand there has to be enough computing power to enable the BD algorithms at the connector of the EIDS participant which is processing the data. Therefore High Performance Computing (HPC) and edge technologies are of special interest. On the other hand there has to be high-performance networking technologies, which enable and ensure the ubiquitous/global communication and availability of huge amounts of data and/or data streams. Both aspects will be described in the following subsections.

4.1 High Performance Computing

Boost 4.0 partners are investigating how to use the HPC infrastructure at the university of Edinburg to run high-performance data analytics. Details about existing infrastructure are as follows:

- **Cirrus:** Housed at EPCC's Advanced Computing Facility, Cirrus is a flexible, state-of-the-art High Performance Computing system that provides an ideal platform for users to solve their computational, simulation, modelling, and data science challenges. Cirrus is based around the core compute resource: a 280 node (10,080 core) SGI ICE XA (link is external) high performance computing (HPC) cluster. The compute nodes each contain two 18-core Xeon "Broadwell" processors and have 256 GiB memory. The compute nodes are all linked using a low latency, high bandwidth FDR Infiniband network. [19]
- **Ultra:** Ultra is a large, shared memory machine. More specifically it is an SGI UV2000 (link is external) system with 64 Intel Xeon E5-4620 v2 processors each with 8 cores. The machine therefore has 512 cores which share 8Tb of memory. This large shared-memory capability makes the machine attractive for certain types of problem that are less well-suited to traditional distributed-memory machines (clusters), such as large scale data-analytics.[20]

[19] http://www.cirrus.ac.uk/about/hardware.html
[20] http://www.epcc.ed.ac.uk/facilities/other-facilities/ultra

4.2 5G Communication Networks and Support to European Industrial Data Space (EIDS)

Industry 4.0 requires a set of network communications in different stages of the production chain. It is very common to assume that communication is a commodity resource available all the time and with no degradation or problems. Indeed, from the point of view of 5G radio technology, industry 4.0 has been identified as one of the key services in by ITU in IMT-2020. This standardization institution defined that industry 4.0 (industry automation) should be achieve thanks to Ultra-reliable and Low Latency Communications (uRLLC). uRLLC defined restrictive requirements, that can be summarize in <1ms latency in user data plane communication, no interruption in mobility and reliability (successful packet rate delivery).

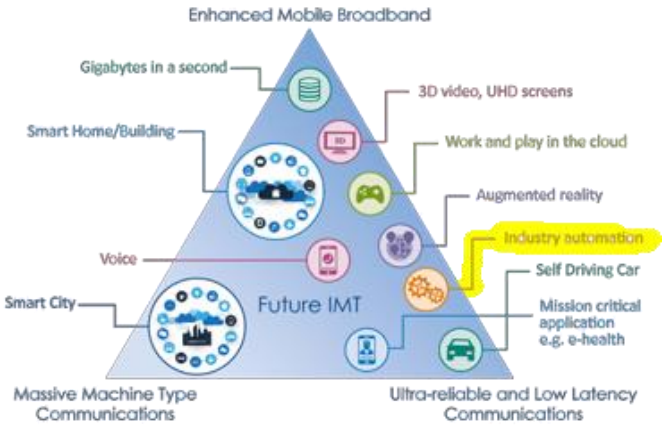


Figure 20: IMT-2020 requirements for 5G

It is clear that communication plays an important role in the Industry 4.0, and therefore network intelligence has been introduced in the map for BOOST 4.0.

4.2.1 Big data technologies for 5G and EIDS operation support

In order to support EIDS enabled big data services for Industry 4.0, the very same communication network must exploit big data and analytics/predictive technologies. To this end, Mouseworld is the 5G networking playground made available to build fit-for-purpose high performance networks that will support the advanced industry 4.0 digital capabilities envisioned in Boost 4.0. Mouseworld is an experimental Lab focused in Machine Learning applied to network communications. Designed as an agnostic to network transport technologies (OSI Layer1 and 2), it focusses in IP based communication. It allows to deploy and test different types of scenarios from Security to network management. The

Figure below shows a global perspective of the Mouseworld framework. This framework is composed by several types of modules: The traffic generators, the network infrastructure, the dataset collector, and the modules for labelling, training and validation. The framework allows to configure different types of applications and deploy customized network architectures to generate the desired traffic over the infrastructure. Some specific probes will collect the traffic packet by packet and group them to network flows summaries in different formats. These network flows will be converted in datasets suitable for ML application by adding a label representing the corresponding security threat to each flow. Finally, the obtained datasets are used for training the ML algorithms and for validating their accuracy and performance.

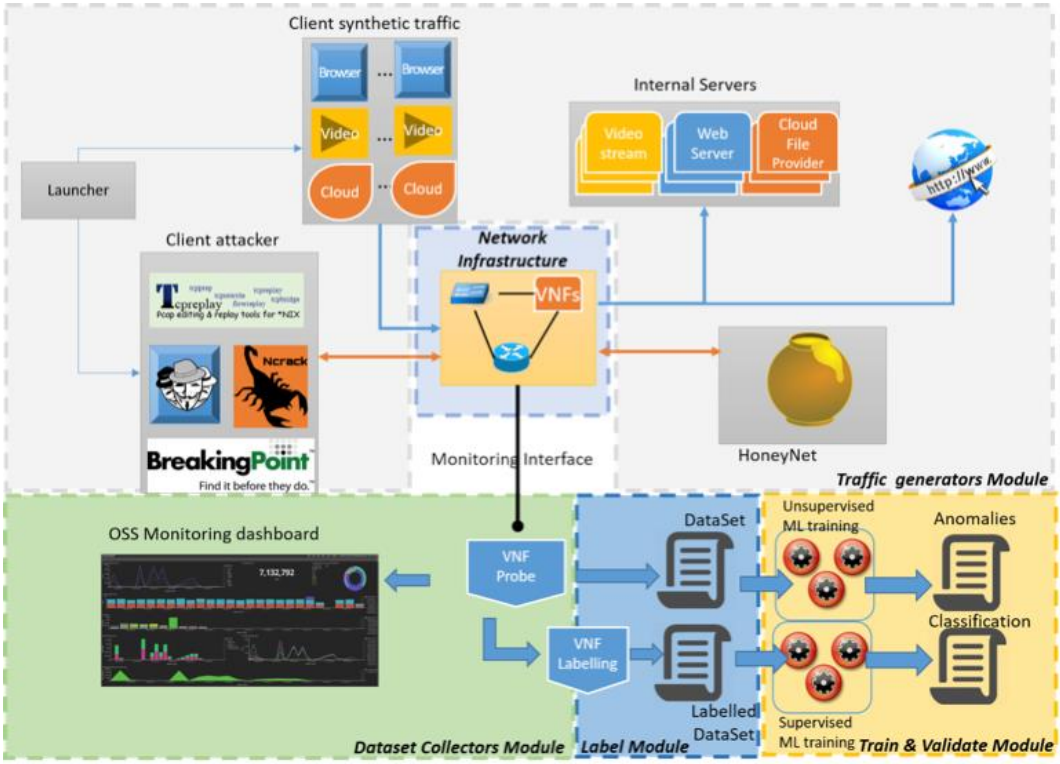


Figure 21: Mouseworld Lab

Mouseworld allows to work with ML applied to network management for improve network reliability. Some examples actually in use, related with network management problems are:

- Traffic classification. Identify different type of encrypted HTTPS flows, based on TCP/IP characteristics, without payload inspection. This is critical in signaling and management protocols of 5G networks where traffic is expected to be based on HTTPS REST API web services. Also, some relevant industrial 4.0 protocols such OPC_UA are based in the same protocols. Correct identification of nature of the flows will allow to guarantee better treatment and reliability.

- NFV infrastructure failures detection. Generating and collecting internal metrics in NFV Infrastructure in datacenters (servers, Virtual switches and virtual network functions), through orchestration software (Openstack), allows to identify and predict physical failures in the IT resources. Prediction of failures on the network infrastructure will again increase the reliability with early responses.
- Security attacks. Network flows are monitored with anomaly detection ML algorithms to detect attacks in the network, such ransomware, DDoS, etc. These kinds of attacks affect to Industry 4.0 through Industrial Control Systems, or IoT devices exposed. Detecting and mitigating them will reduce the impact in the Industry 4.0 services.

One of the main features from Mouseworld is the capacity of inject captured traffic from other environment to train and test ML algorithm. One possible applicability in BOOST4.0 can be collect the traffic generated by some industrial partner in the 5TONIC lab where 5G traffic will be used, and evaluate different algorithms for detection of network problems or for accurate classification of the traffic.

4.2.2 Machine Learning (ML) approaches to EIDS high performance networking infrastructure enhancement.

Telefónica Investigación y Desarrollo (TID), which is partner of the consortium is actively working in the evaluation and testing of different Machine Learning technologies applied to network communication and especially in the area of Network Management.

Traffic classification

Our aim is to research, design and develop a component for characterizing network traffic in real-time even if it is encrypted or some privacy limitations prohibit the access to packet payload. This component will be able to characterize network traffic flows in a similar way as current DPI technology does, but without breaking encryption or inspecting packet payloads and therefore obeying user privacy requirements. The component will be mainly based on supervised learning techniques such as Random forest and Deep Convolutional Neural Networks (CNN) using as source network flows in different formats:

- NetFlow V9/IPFIX[21]. Standard format that group network packets in flows based on a set of parameters (src.IP, dst. IP, src. Port, dst. Port, timestamp, etc.). It is widely used in the Telco industry and network managers to monitorised the health of the network

[21] RFC 7011, https://tools.ietf.org/html/rfc7011

- Tstat[22]. Alternative format, that enrich the data with statistical values, that suite better for ML algorithms.

Applying supervised machine learning techniques and using data sets of network traffic flows in the big data regime, we will train different models and architectures with the aim of classifying realistic set of network traffic flows in specific categories (e. g. video streaming, file download, web browsing, and cloud storage). In order to obtain the expected number of training examples, we will design and run complex and realistic experiments in a controlled environment using extensively the Mouseworld lab. In this context Launcher and Tagger components will be designed. The former will help to automate the deployment of experiments and the latter will add to each example the needed label for training supervised models in a totally automated way. For the traffic classification task we will design models based in CNNs, which is a class of deep, feed-forward artificial neural networks, most commonly applied to analysing visual imagery thanks to its shared-weights architecture and translation invariance characteristics. These characteristics encourage us to use CNNs as one of the corner stones of our traffic classification component.

Anomaly detection using unsupervised algorithms to detect anomaly patterns in the NetFlow v9 flows. There are different models available, such Isolation Forest, oneClass SVM, etc. This models allows to detect abnormal patterns, such network problems or security attacks.

Forecasting of network and infrastructure events.

Using different metrics collected from both network flows and IT infrastructure metrics (e. g. servers, virtual machines) we will design forecasting models in order to predict ahead of time network and infrastructure events that could impact in the management of the network and IT infrastructure. For example, it could be useful to detect in advance and avoid an upcoming degradation of the system performance or a network congestion problem. To this end we will design complex deep neural network architectures integrating recent proposals specifically tailored for time series prediction such as Recurrent Neural Networks (RNN) and in particular using extensively Long Short-Term Memory (LSTM) units. A recurrent neural network (RNN) is a class of artificial neural network where connections between nodes form a directed graph along a sequence. Additionally, and unlike feedforward neural networks, RNNs can use their internal state (memory) to process sequences of inputs. This architecture allows RNNs to exhibit temporal dynamic behaviour for a time sequence. Long short-term memory (LSTM) units are units of a RNN and are composed of a cell and several gates. The cell remembers values over arbitrary time

[22] http://tstat.polito.it/

intervals and the gates regulate the flow of information into and out of the cell. Compared with other approaches such as traditional RNNs and hidden Markov models, LSTM are perfectly-suited for making predictions based on time series data, since they are robust in their predictions when lags of unknown duration between important events appear in a time series.

5 EIDS Enabling Technologies Interaction

The following subsections introduce the interaction of the EIDS enabling technologies, their interdependencies, approaches for integration, and plans for joint operation.

5.1 FIWARE: An open source software (OSS) implementation of the IDS

The IDSA is an initiative to promote the adoption of existing technologies, standards, and governance models for facilitating data transaction in trusted business ecosystems, called International Data Spaces – IDS (cf. chapter 3.1). The IDS provides the keystone of new creative business processes as well as smart service scenarios. The IDS also ensures data sovereignty for the context information providers. To cover these functional requirements, the IDS provides a reference architecture model composed of business, functional, process, information, and system layers. This reference architecture comprises the following components, as depicted in Figure 22.

- **Connector:** the central component of the IDS architecture. The Connector is a dedicated data exchange server which will publish and notify data accordingly to a pre-defined Connector specification. Different types of connector servers will exist based on different aspects, such as the execution environment, trusted environment or developmental stage in terms of a data services or workflow being used. It is important to identify uniquely each of connector server type to facilitate the interaction between different Connectors instances. To do so, each Connector must send a detailed information about the metadata of its data source to the Broker instance.
- **Broker:** an intermediary that keeps and manages the metadata repository that provides syntax and semantic of Data Sources. Additionally, the Broker component offers clearing and settlement services for accounting and data exchange transactions. A Broker service will also provide the Data Source endpoint from which data is obtained.
- **App Store:** a secure platform for distributing and deploying Data Applications. Additionally, it includes some semantic representation of the Data Applications to facilitate the activity of the search engine to get the appropriate Data Application.

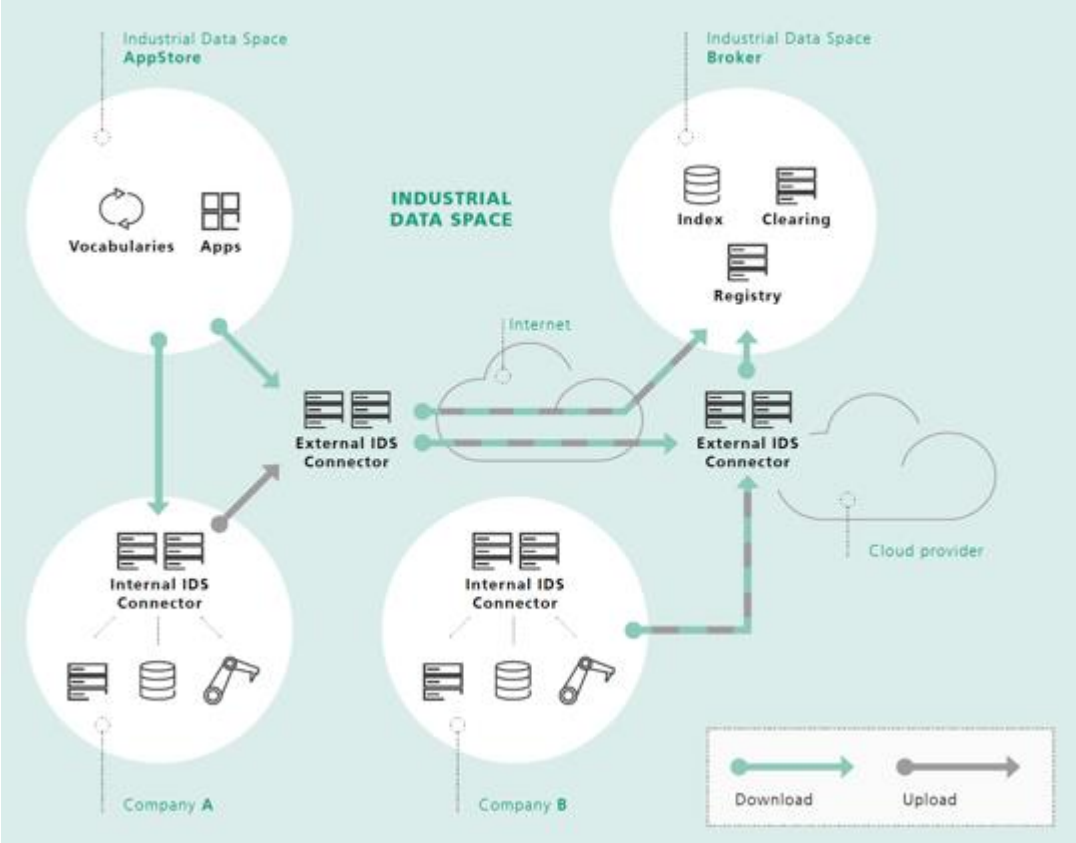


Figure 22: Overview of the main elements of the IDS architecture

5.1.1 IDS Connector

Figure 23 illustrates the IDS Connector architecture. The Data Hub brings access to and eventually stores data exported by between services or Connectors. The Data Hub provides the simplest method to exchange data between Connectors. The Data Router direct data from different sources to a Data Hubs the Data Hub can be replaced by alternative implementations in order to meet the requirements of the operator. The selection of an appropriate Data Hub may depend on various aspects (e. g., costs, level of support, throughput rate, quality of documentation, or availability of accessories).

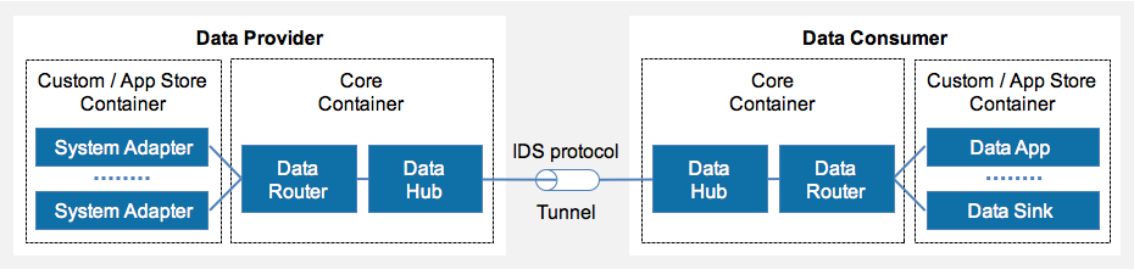


Figure 23: IDS connector model

Basic Access Control

To be able to make access control related decisions based on reliable identities and properties of participants, a concept for Identity and Access Management (IAM) is

mandatory in the IDS Architecture Model. This includes identification (i. e. claiming an identity), authentication (i. e. verifying the identity) and authorisation (i. e. making access decisions based on an identity). Every participant may possess attributes besides its own attributes and these attributes could vary dynamically. Connectors have to regulate access to data based on different criteria: the specific identity of Connectors, Connectors attributes or security profile requirements. Figure 24 shows the proposed architecture to achieve the Identity and Access Management IDS requirements using FIWARE components.

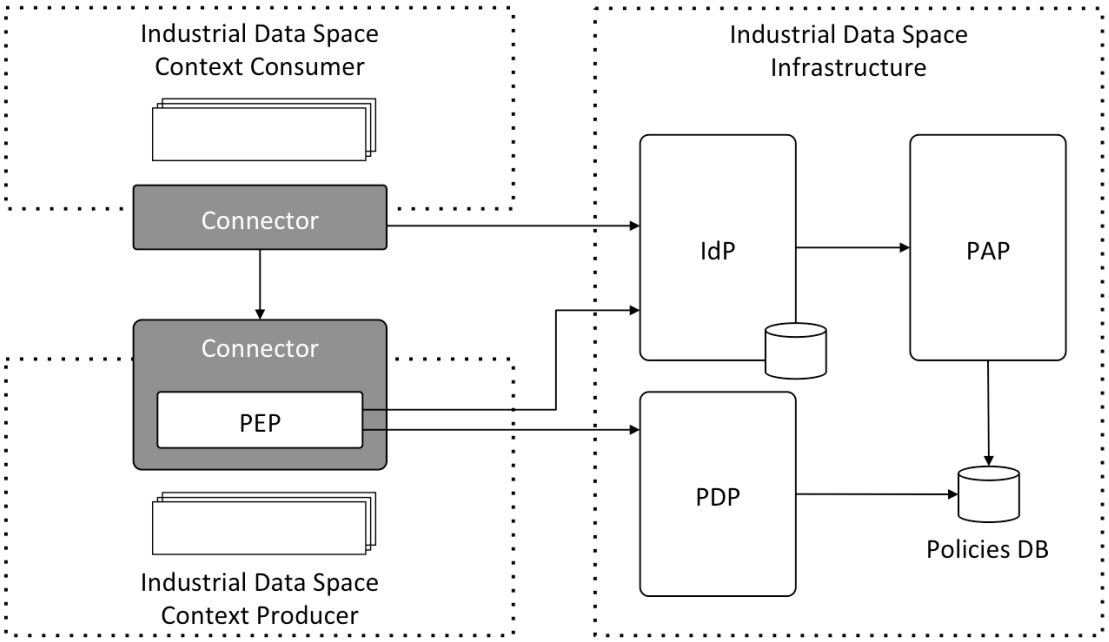


Figure 24: Implementation of the IDS Identity and Access Management using FIWARE components

The proposed model uses OAuth 2.0 protocol as authorisation framework. Using OAuth 2.0 delegates the authorisation process from IoT devices and so enables an application-scoped feature – Access Control policies can be defined in the scope of an application/service. By fostering this behaviour, a participant could have different permissions for different services, thus making it possible for the same participant to be reused among different services, being under different security conditions in each of them.

In Figure 24 we can see and IDS Context Consumer that accesses data provided by an IDS Context Producer. Of course, the communication between them is established through their respective IDS Connectors and using a secure channel. For protecting the access to the provided resources, every request is intercepted by a Policy Enforcement Point (PEP). On the other hand, in the IDS global infrastructure are deployed the rest of components that take part in the architecture. These components are deployed once and used by every IDS Connector in the environment. The Policy Administration Point (PAP) and the Policy Decision Point (PDP), together with the set of PEPs included in each Connector compose the widely-known Access Control architecture. The PAP stores the defined access control

policies in the Policies DB, where PDP checks them at decision time. Finally, the Identity Provider (IdP) is in charge of identification and authentication.

For the Access Control architecture to be used in all security contexts, the framework for describing authorisation policies should fit a level of granularity and flexibility. For this purpose, OASIS[23] (a global non-profit consortium that works on the standards definition for security, IoT and other areas) standardised the eXtensible Access Control Markup Language (XACML[24]), which allows the definition of fine-grained policies. XACML serves as a standard not only for the format of authorization policies and evaluation logic, but also for that of the request/response interactions that take place during an authorization decision.

Using XACML terminology, policies are composed by a set of rules, which are in turn made of a target (such as the resource), an effect (e. g., allow/deny) and a condition. In the approach used in , rules are used to implement permissions, in which the target is an action (e. g., an HTTP verb) plus a resource of the Context Producer. Of course, more complex policies may be also defined, provided that they are supported by the policy-description language. Roles are in turn sets of Permissions that serve as a container so that more than one permission can be assigned at once. Both Permissions, Roles and their relationships are defined in the PAP. Note that this approach allows both a simpler Role-Based Access Control (RBAC) scheme and a more complex, more flexible Attribute-Based Access Control (ABAC) scheme. It all depends on how the Permissions are defined when creating the XACML rules.

As introduced above, the Identity Provider provides identification and authentication following an Identity as a Service (IDaaS) approach. Thus, every IDS participant needs to be registered in the IdP, so that it gets a set of credentials (usually a username and a password), which it can use to authenticate and identify itself against the AC system. Groups of participants can be created in the IdP, to allow more complex authorization scenarios. Granting a permission to a group of participants, rather than to a single one, has the benefit of linking the given permission to the participants belonging to the group.

Once a participant is registered in the IdP, it can create an OAuth 2.0 access token for accessing data in a specific Context Producer. In OAuth 2.0 terminology that means creating a token in the scope of a consumer. This token represents the participant in the system and has to be included in every request sent to other Connector. As outlined before, these requests are intercepted by the PEP, that extracts the participant's access token and validates it with the IdP. This validation can be performed in three levels of security:

[23] https://www.oasis-open.org/
[24] https://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml

- **Authentication**: Using this level of security, the PEP just checks if the participant has been correctly authenticated against the IdP. Thus, at this level, every participant with an active account would be able to access the protected data. The check is performed by sending a validation request to the IdP.
- **Basic authorisation**: In this case, the PEP also checks if the participant has the required roles to perform the corresponding action (defined by an HTTP verb) in the corresponding data source (defined by an HTTP path). After the first check with the IdP, the PEP obtains the roles the participant has assigned in the scope of the Context Producer where the token was created. Once roles have been retrieved, the authorization check is sent to the PDP. PDP fetches the policies associated with the participant's roles from the Policies DB and decides whether or not access should be granted based on them.
- **Advanced authorisation**: This is the most complex, powerful case, because the authorization check is not only based on the HTTP verb and path, but also on other more advanced, customizable parameters, such as the request body or headers. To perform the check, a custom XACML policy request is sent to the PDP.

The following figure illustrates the interaction described above.

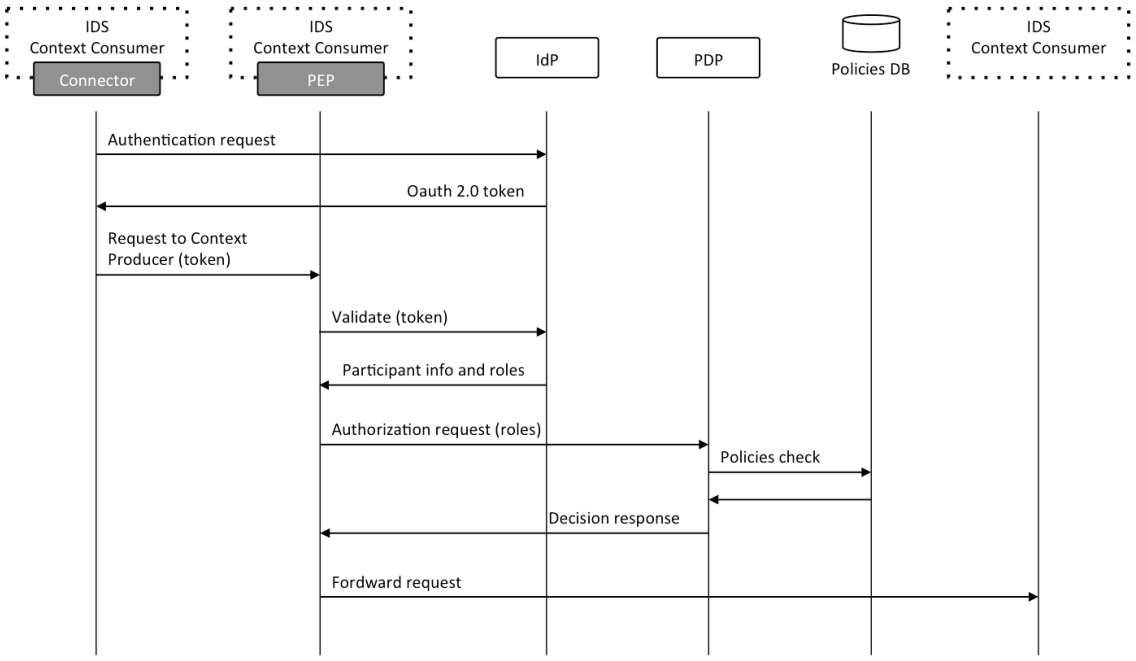


Figure 25:The overview of interaction between components at different security levels

Every component present in this architecture is implemented as a FIWARE Generic Enabler:

- **Identity Provider:** Identity Management GE is named KeyRock and has been implemented using Openstack technology as a starting point. KeyRock has two main components, named after their Openstack's counterparts: Python-based, back-end Keystone and Django-based, front-end Horizon. Both components have

been extended for supporting the specific features needed by the IDS Security Architecture.

- **Policy Administration and Decision Points**: this implementation integrates both PAP and PDP components and is called AuthZforce. It provides an API to get authorization decisions based on authorization policies and authorization requests from PEPs. The API follows the REST architecture style and complies with XACML v3.0.
- **Policy Enforcement Point**: the policy enforcement point GE is named Wilma and it is developed using Node.js. In the context of IDS FIWARE Architecture, Wilma is deployed as part of API Umbrella, an API Management system that enriches the PEP functionalities with features like accounting, API documentation or catching.

Management of Access Control for trusted applications

Establishing trust between participants in the Industrial Data Space is crucial when talking about Security IDS Architecture. It happens that some Context Providers offer data that is not directly managed by them (i. e. stored in their databases) but provided by a third-party Context Provider. In such cases, a participant could be authorized to get this data from the first Context Provider but not from the second one. The second provider, when receiving the delegated request, has to be able to authorise the request basing on the permissions the participant has in the scope of the original provider. For allowing this interaction, a trusted relationship has to be established between both Context Providers.

In the previous figure we can see an example of the explained scenario. The Context Consumer has the needed permissions to access a data entity provided by Context Producer 1 (ent-1). However, although Context Producer 1 is offering this entity, it is actually provided by Context producer 2 so it has to delegate the Context Consumer request to it.

When receiving the request, PEP of Context Producer 2 will check with the IAM infrastructure that the consumer has the needed permissions to access the entity in the scope of the Context Producer 1. Therefore, it should reject the request.

In FIWARE implementation we solve this issue with a mechanism that allows Context Producer to have a list of Trusted Context Producers. When receiving a request from a participant, the PEP will check if it has the required permissions to access the data in the scope of itself or in the scope of one of the Context Producers included in its Trusted Context Producers list. In the previous example, Context Producer 1 has to be included in the Context Producer 2 list. This way, when receiving the request, it will check that having the participant no permissions defined in its scope, it actually has the permission to access the entity in the scope of a trusted Context Producer.

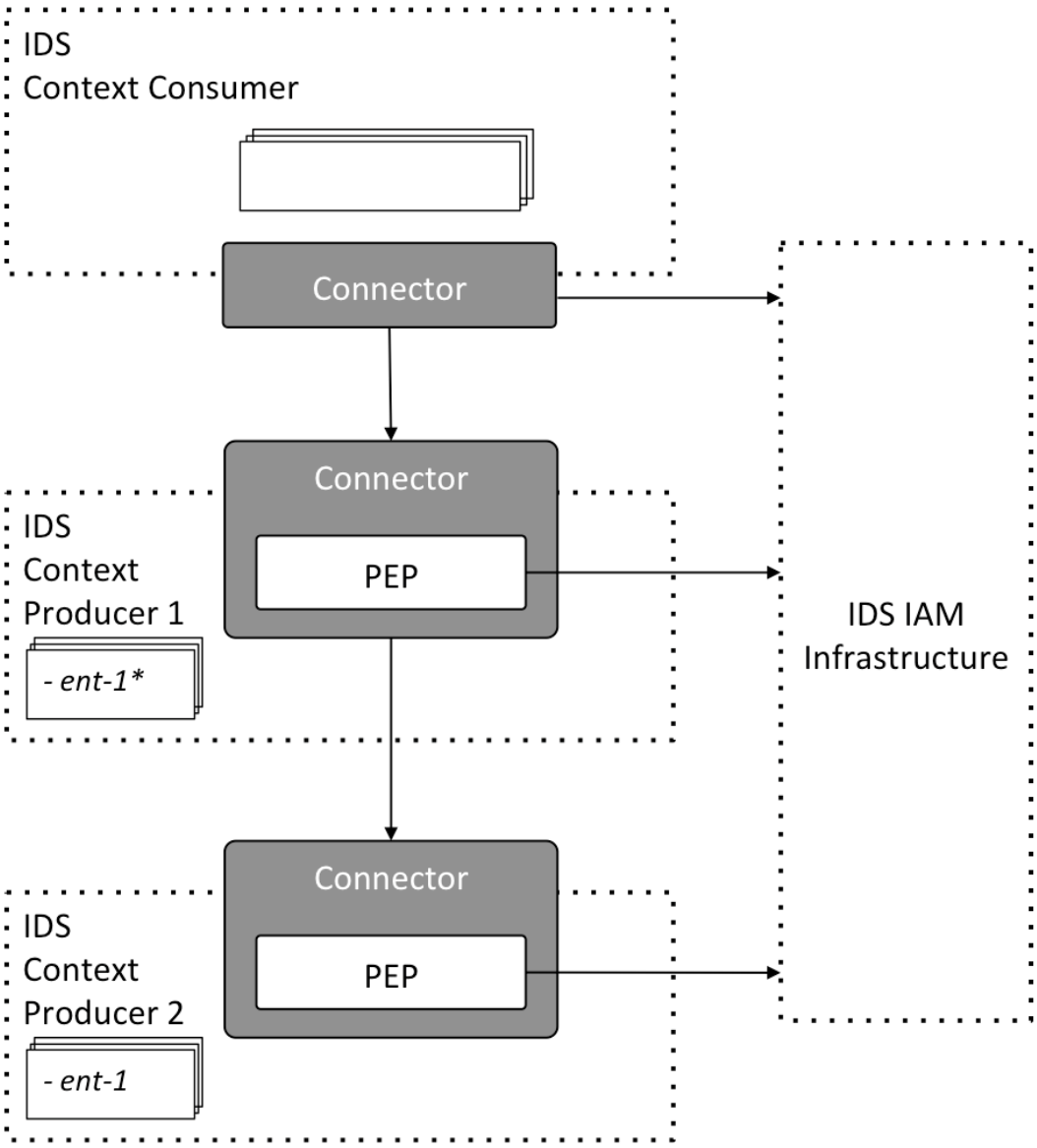


Figure 26: Establishment of a trusted relationship between context producers and consumers

5.1.2 IDS AppStore

In the FIWARE reference implementation of IDS, the AppStore is implemented using the FIWARE Business API Ecosystem GE. i. e. MasterMind will support the deployment of components (such as IDS Connectors) from the marketplace provided by the Business API Ecosystem GE. This solution provides some additional features such as the support for accepting term and conditions or enabling the monetization of such components.

The workflow to purchase a new component is from the AppStore is as follow:

1. For each component to be purchased in and used in MasterMind, the owner should follow the registration process in the Business API Ecosystem, including the component package (simpler solutions may be used in preliminary versions that does not require storing files in the App Store), the credentials to access a private

docker registry (if any), the license code (if any), the terms and conditions, characteristics, pricing etc.).

2. Once the offering is created, the Business API Ecosystem will automatically register the component in the MasterMind catalogue, using its API, providing all the included configuration files.
3. To ensure that only those users that purchased a particular component in the IDS AppStore can see and deploy it MasterMind, the security framework is used. In particular, the different components will be bound to Identity Manager roles, so when a user acquires a component in the AppStore, the BAE will grant the role bound to the component to the user. This way, when the user logs in MasterMind using the IDM, it will retrieve user information (including its roles in MasterMind), that can then be used internally to build create a custom catalog including only the components purchased by a user (or organizations she/he is member of).

5.1.3 IDS Broker

The Industrial Data Space architecture enables Data Providers to share their data with other participants (Data Consumers), who can discover available data sources and its characteristics using the IDS Broker. In particular, the IDS Broker stores metadata which describes the data source, including information about the Data Provider, the syntax and semantics of the data, or additional information, such as the pricing, the usage policies, etc.

In the FIWARE reference implementation of the IDS architecture, the data source metadata is described in NGSI using the NGSI context source format. This format enables to provide information about the endpoint, context data available, geographic availability, etc, as well as custom metadata which can be used to provide usage policies and pricing information.

To implement the IDS Broker supporting the NGSI context source format while providing a data catalogue for simplifying the registration and the discovery of data sources, two different FIWARE components are used jointly. On the one hand, a NGSI registry (in practice implemented by a context broker) stores the different NGSI context sources, providing a standard API that can be used for querying and subscribing to data availability. On the other hand, a FIWARE Extended CKAN provides a data publication platform that simplifies the registration, search and discovery of context data by end-users, while linking the IDS Broker implementation to the FIWARE monetization framework, realized by the integration of the extended CKAN, the Business API Ecosystem GE, and the Security Framework.

This implementation supports two flows for data source publication:

- If the metadata related to a data source is generated within the IDS connector and available before the data source publication, it can be directly feed to the API of the

NGSI Registry. Then, the FIWARE extended CKAN, subscribed to changes and new entries in the NGSI Registry, uses this metadata for automatically creating the new dataset entries in its catalog. Additionally, if the metadata includes usage policies or pricing information, the FIWARE Extended CKAN creates and configures all the needed products and offerings within the Business API Ecosystem.

- If the metadata is not automatically generated, it is possible to use the FIWARE extended CKAN portal for registering the NGSI context sources as dataset resources and creating its related products and offerings in the Business API Ecosystem GE. Once registered, the FIWARE Extended CKAN uses all the provided information in order to generate the data source metadata in NGSI format and feeds it to the NGSI Registry.

The resulting state after following any of the two flows for data source registration is the same, enabling to search and discover data sources, both, using the data publication portal offered by CKAN, or using the standard NGSI API provided by the NGSI registry.

The proposed architecture integrates the implementation of the IDS Broker with the data monetization framework already offered in FIWARE. In this way, those data sources which have some usage policies or pricing information included within its metadata, also have a related offering published in the Business API Ecosystem GE. Using this component, it is possible to provide support for the establishment of the agreement between the Data Provider and Data Customer as well as actual monetization of the data offered through a data source under different pricing models, including pay-per-use.

Note that the Business API Ecosystem delegates on the FIWARE Security Framework the access control, as described in the Basic Access Control section. In this regard, the different access policies of the data offered through a data source must have been established before the creation of the offering, being the Business API Ecosystem in charge of granting the needed permissions to those users acquiring access to a particular data source. Additionally, the Business API Ecosystem also delegates in the Security Framework (in the PEP) the accounting of the data usage which is used for processing the charging of usage-based pricing models.

5.1.4 Summary

This section has described the approach taken by FIWARE in terms of providing an open source implementation of the IDS business architecture. The figure and steps below, summaries how FIWARE open source components can be used to provide such implementation with the required extensions:

1. Docker-based tools relying on Docker Hub Services enabling automated deployment and configuration of Data Apps.

2. Standard vocabularies are being proposed at https://www.fiware.org/data-models
3. Data Apps map to NGSI adapters or Apps processing context information.
4. Both External and Internal IDS Connectors are implemented using FIWARE Context Broker components.
5. Extended CKAN Data Publication Platform.
6. FIWARE Context Broker components will be used as core component of IDS Connectors.
7. Interface between IDS connectors based on FIWARE NGSI.

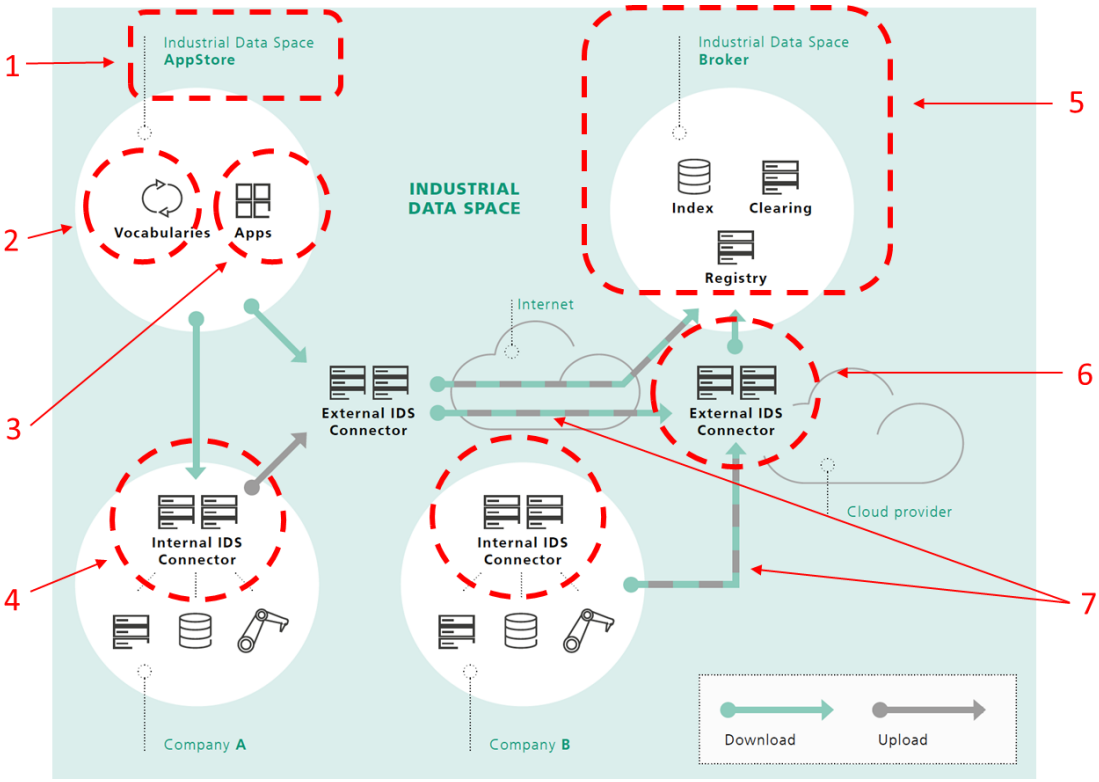


Figure 27: Mapping of FIWARE components to IDS business architecture

5.2 Approaches for integration of Hyperledger Fabric distributed ledger technology and IDS

In this section we discuss different alternatives for integration between blockchain and IDS, specifically between Hyperledger Fabric (simply Fabric) and IDS. Generally speaking, there are two levels of integration – the conceptual and technical levels.

Conceptually, we should first be able to identify the scenarios/use cases in which applying each technology by itself can fully meet the requirements, and scenarios in which we need a combination of these two technologies. At this stage, we believe that it's too early to answer this question and this requires a more use case oriented approach that will be answered after examining several use cases in the project. In this report we will address the second question, that is, knowing we need a combination of both technologies, what is the recommended way to integrate between the two.

Blockchain is an immutable, decentralized distributed ledger. IDS has been designed for secure data exchange in value chains. In both blockchain-oriented and IDS oriented architectures, integrating the two technologies can provide added value for different scenarios implementations. One of the most prominent scenarios in IDS is integrating blockchain to fulfil the role of clearing house and to serve as audit-proof ledger for recording of data exchanges (see IDS roles for explanation).

5.2.1 Blockchain-IDS integration design

We identified three integration options described below (inspired by the AMable project).

5.2.1.1 Option 1 - Customized blockchain IDS connector

This option is illustrated in Figure 28. A specialized IDS connector exposes access to blockchain ledger.

Each data transfer via a regular IDS connector is logged by the IDS connector's core module to blockchain IDS connector. Blockchain IDS connector implementation is based on a custom container enveloping a blockchain client application which in turn uses SDK to communicate with blockchain network.

The flow

1. Data consumer and data providers initiate data requests and response between them respectively
2. Once this happens, the core container of the IDS connector initiates communication with blockchain connector
3. The blockchain connector runs a custom container which communicates with blockchain infrastructure

Pros

4. The integration is transparent to the user as all functionality is "hidden" in the core container

5. The implementation is ledger-agnostic as long as the communication protocol between IDS connector and blockchain connector doesn't change.

Cons

- Implementation requires changes in the IDS connector core container.

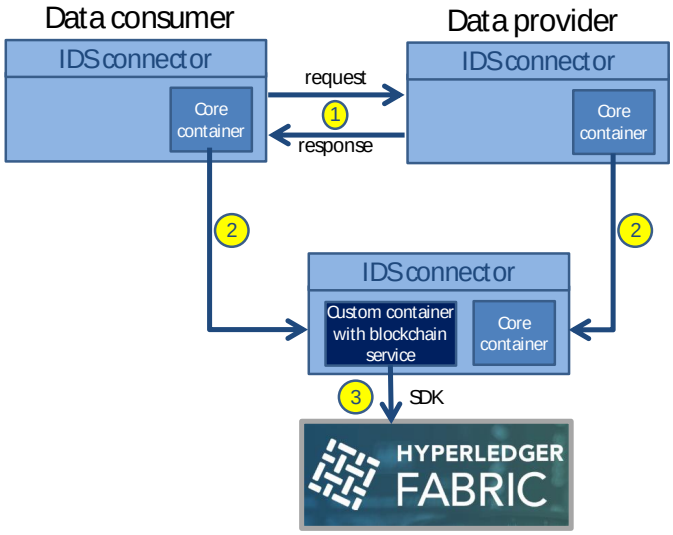


Figure 28: First integration option

5.2.1.2 Option 2 - Blockchain-adapted core container within IDS connector

A core container module within the IDS connector is adapted to serve as blockchain client. The service provided by the blockchain client is integrated as part of the workflow in the core container.

The flow

1. Data consumer and data providers initiate data requests and response between them respectively
2. Once this happens, the core container of the IDS connector invokes the blockchain client integrated in the core container
3. The blockchain client communicates with blockchain infrastructure

Pros

- The integration is transparent to the user

Cons

- Require changes in the core container
- If the blockchain technology changes will require redeployment

5.2.1.3 Option 3 - Blockchain client as a custom application container within IDS connector

This option is illustrated in Figure 29. In this alternative, the blockchain client is incorporated as a custom container within a regular IDS connector. The logging of the data transfers is not transparent but rather defined as workflow in the core container.

The flow

1. Data consumer and data providers initiate data requests and response between them respectively
2. When this happens the workflow manager within the IDS connector based on explicit workflow specification invokes the App Store/custom blockchain container which, in turn, communicates with the blockchain infrastructure via SDK

Pros

- No changes to core container is required
- The custom container can be published as application in App Store and downloaded as part of any IDS connector
- In case of changes to the blockchain implementation/provider it can be easily replaced

Cons

- Explicit workflow integration in the core container is required.

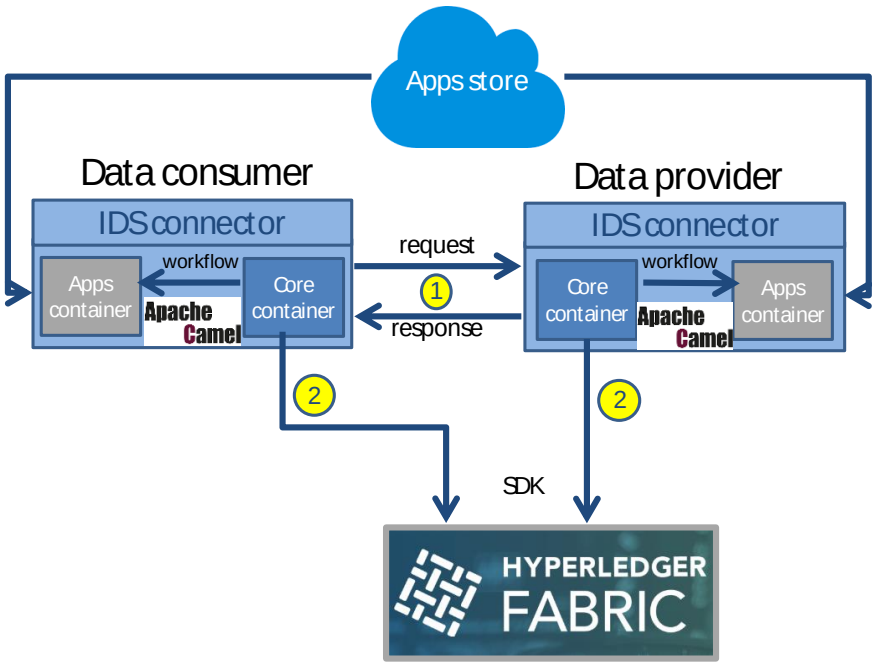


Figure 29 : Third integration option

5.2.1.4 Selected alternative

Option 3 has been chosen for BOOST project as cons outweigh pros and the ability to break implementation into several phases to minimize risk. More specifically, the implementation will be based on Hyperledger Fabric blockchain and will be carried out in two phases as described henceforth.

Phase 1: Implementation of a blockchain custom container.

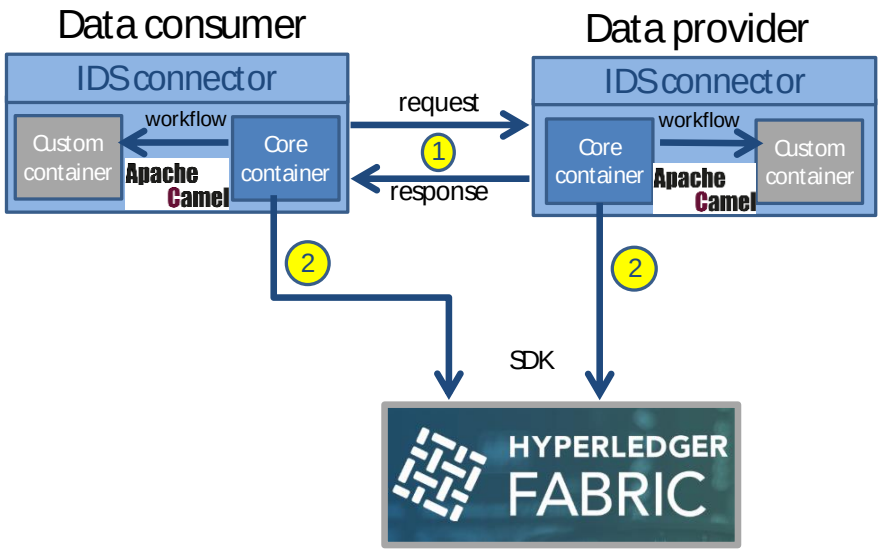


Figure 30: Third option phase I

In this way, integration is based on a custom container (Docker) which includes the Hyperledger Fabric SDK to communicate with Hyperledger Fabric. In addition, this extended IDS connector will have the following properties:

- The container can be run as a standalone container for interacting with blockchain network.
- The container will provide RESTful APIs for invoking chaincode/quering the state of the ledger
- In a second step, the custom container will be integrated in the IDS Connector with all Apache Camel needed configuration and model for workflows.

Phase 2: Implementation of an Apps container (Figure 29)

Once the App Store role provider is available in BOOST IDS implementation, the custom container will be provisioned as an App over the App Store.

5.2.1.5 Extended IDS connector for Fabric

The extended IDS connector will include the following components (Figure 31)

- Blockchain custom container
    i. Dockerized container
    ii. Blockchain nodeJS application serving as blockchain client using HFC SDK to communicate with blockchain network
    iii. RESTful APIs for invoking chaincode on the network/quering the network
    iv. The client will interact with matching chaincode which will be also defined and provided
- Workflow management definitions for explicit invocation of the blockchain custom container when data transfers occur or query of the state of the ledger is required.

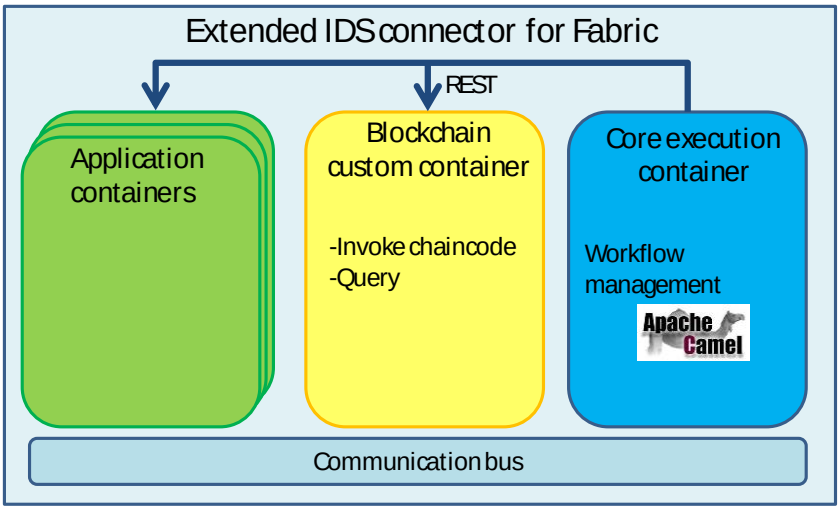


Figure 31: Extended IDS connector for Fabric

5.2.1.6 Our proposed approach for use case implementation

Figure 32 illustrates our general approach for the proposed blockchain-integrated solution. It includes participants in IDS network exchanging data. The data exchanges are logged via blockchain-integrated IDS connector on the ledger through invoking use case specific smart contract deployed on the blockchain network.

Validation of the approach will be carried out use case driven throughout the project.

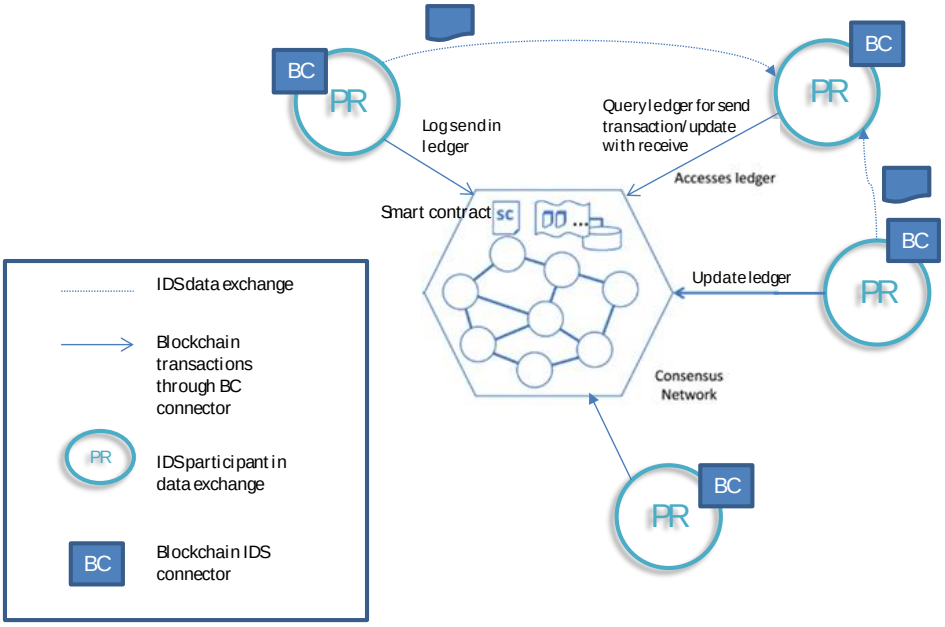


Figure 32: Proposed approach for use case implementation

5.2.2 Plans for integration of blockchain technologies and FIWARE IDS OSS implementation

The Industrial Data Space connectors, being the media of distribution of data from data owners to data consumers, should have full understanding in regard to the distribution of data (e. g. from and to whom the data is shared), which would allow monitoring and potential forensic in case of hurdles. Given a large amount of data, as well as potentially large numbers of data owners and consumers, the automated enforcement of consent and data usage purpose should be in place, which would also allow automated monitoring and potential forensics of materialized contracts. A technical solution for enforcing consent and data usage purposes comes in the form of smart contracts, while the blockchain technology enables automated monitoring and forensics of occurred contracts between data owners and data consumers.

Notwithstanding vivid activities around blockchain and smart contracts, both of these technologies are still in their infancies and therefore vulnerable. These vulnerabilities and challenges are predominantly related to the scalability and latency of both technologies. Another uncertainty comes from the fact that these contracts and not backed by governments, hence their enforcement is rather unclear and currently legally non-binding.

Given that the implementation of IDS connector using FIWARE technologies is an ongoing activity in the BOOST 4.0 project, as a part of the future work we will potentially aim at integrating the blockchain-based smart contracts with the FIWARE-based implementation of the IDS connector. We are at the moment doing research in both technical, i. e. for researching the scalability and latency capabilities of blockchain-based smart contracts,

and legal domains, i. e. learning about current regulations for aligning the enforcements of smart contracts with legally binding entities. We are doing that to decide if blockchain-based smart contracts could be a suitable addition to the ongoing FIWARE-based implementation of the Industrial Data Space connector.

5.3 Vocabularies ↔ IDS

The Information Model of the IDS primarily aims at detecting, describing, and publishing data products (Data Assets) and reusable data processing software (Data Apps) in the Industrial Data Space. Data Assets and Data Apps are the core resources of the Industrial Data Space, also referred to as IDS Resources. By means of a structured semantic annotation it is ensured that only relevant Resources are provided (i. e., Resources that are appropriate to meet the requirements of the Data Consumer). Once the Resources are identified, they can be exchanged and consumed via semantically defined service interfaces and protocol bindings in an automated way. Apart from those core commodities, the IDS Information Model describes essential properties of Industrial Data Space entities, its participants, its infrastructure components, usage control policies, and its processes in a central vocabulary. The vocabulary is implemented as an RDF-based ontology and published under an open license on GitHub[25].

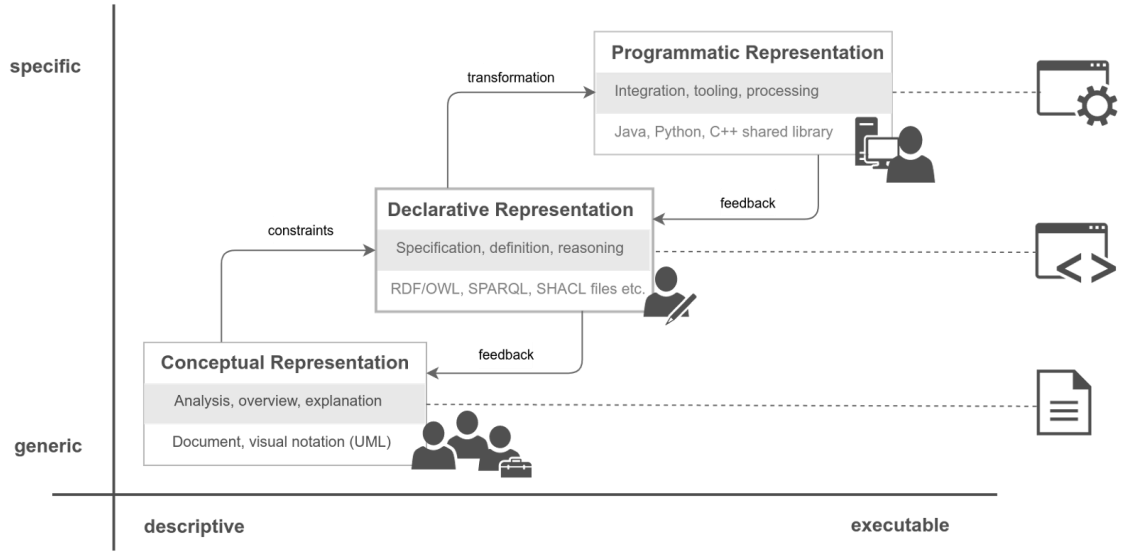


Figure 33: Representations of the IDS Information Model

In order to address the diverging needs, expectations, and usage scenarios of the different stakeholder groups, the Information Model allows for three different representation levels, with differing degrees of formalization: The Conceptual, the Declarative, and the Programmatic Representation. Each level corresponds to different forms of digital

[25] https://github.com/IndustrialDataSpace/InformationModel

representation, ranging from high-level, conceptual documents up to the level of operational code, as depicted in Figure 33. The Declarative Representation (IDS Ontology) is the only normative specification of the Information Model. A set of auxiliary resources, among them guidance documents, reference examples, validation, and editing tools are intended to support a competent, appropriate, and consistent usage of the IDS Ontology.

The Conceptual Representation of the Information Model provides an analysis and an overview of the main, largely invariant concepts of the Information Model, with no commitment to a particular technology or domain. It mainly targets the general public as well as management boards of organizations by means of textual document and understandable visual notations. The descriptions at this level are generic, provide basic information, allow comparative analyses, and promote a shared understanding of the concepts.

The Declarative Representation of the Information Model defines the normative ontology behind the Industrial Data Space in the form of the IDS Ontology. It has been developed along the analysis, findings, and requirements of the Conceptual Representation and depends on shared vocabularies. Based on the Semantic Web Stack ([RDF]; [RDFS], [OWL]) and standard modelling vocabularies ([DCAT], [ODRL] etc.), it provides a formal, machine-interpretable specification of concepts envisaged by the Conceptual Representation. Furthermore, it details out and formally defines entities of the Industrial Data Space in order to be able to share, search, and reason upon the structure of meta-data descriptions. As such, it comprises a complete referential model allowing to derive a number of Programmatic Representations. The IDS Ontology defines a reusable, domain-agnostic "core model". According to common best practices, existing domain vocabularies and standards are reused where possible fostering acceptance and interoperability.

The Conceptual Representation contributes to the Information Model by visualizing relations and formulizing constraints. The IDS Ontology as the key component of the Declarative Representation provides these essential concepts shared by all participants of the IDS. It serves as the foundation for the discovery, composition, and maintenance of any IDS related resource and for all types of data exchange. On the level of the Declarative Representation, additional vocabularies further extend the IDS core model where necessary in order to fulfil specific industry requirements. The thereby achieved shared understanding ensures a consistent modelling of both data and services and establish a solid basis for any interaction in the scope of the IDS.

5.4 Big Data Apps ↔ IDS

Big Data (BD) applications serve many different purposes. Nevertheless all BD applications have got one thing in common: The need for raw data in order to work properly and to

generate results. The IDS, as an ecosystem that allows to share data in a trustful environment, is enabler for BD applications and grants excess to all kinds of data sources (from public data like weather data up to data, which have the need for privacy). Due to the fact that there are no limitations by means of the kind of data sources, endless connectivity is enabled hereby. This comes with huge potentials for the providers of BD app, because in an IDS ecosystem they have access to all these sources and can start inventing all kinds of innovative functionalities.

The exchange of data is enabled were ever the usage policies grant access to the source of data for the user of data. The owner of data retains the full control over the data while taking advantage of the full functionality of the BD application. In such an environment data sovereignty is becoming reality.

The IDS can function as a mediator between BD applications with a specific need for data and the data endpoints that work as various sources for data. The functionality to enable BD application providers to search for data sources comes from the Broker. It on the one hand acts as search engine in order to find adequate data sources and on the other hand can be used by the data providers as market place in order to present their offering. The second mediator between BD application provider and other participants of the IDS ecosystem is the App Store. The IDS App Store can be understood as marketplace where the providers of BD application can sell their product to the customer. Therefore the IDS is enabler for Big Data on many levels.

6 Conclusions

This document provides a first version of the business reference architecture of the European Industrial Data Space. A shared model, like the EIDS, empowers the European industry to deal with industrial data as an economic good and is enabler for advanced analytic services while at the same time granting full sovereignty and control over ones data.

BOOST 4.0 and therefore the vision for the EIDS relies on the International Data Spaces (IDS) and other key European open initiatives and technologies. This report therefore provided a first introduction to the key technologies and initiatives that enable a European multi-homed open big data space for Industry 4.0 smart data applications. Besides the presentation of these corner stones of the EIDS, the report also provided a first vision for the synergies that will result in the combination of them.

The conclusion of this document is that there is a good alignment among the various open initiatives and that they can jointly contribute to the implementation of the European Industrial Data Space. Furthermore it is concluded that the first Proof of Concepts that were developed are of good value to the industry and that additional features in terms of better integration is necessary to ensure better usability of technologies. The approach is very promising regarding the full coverage of the 20% of scenarios, where data sharing is in focus, and that are requested 80% of the time by industries, willing to engage in inter-company or cross-company data sharing and exploitation.