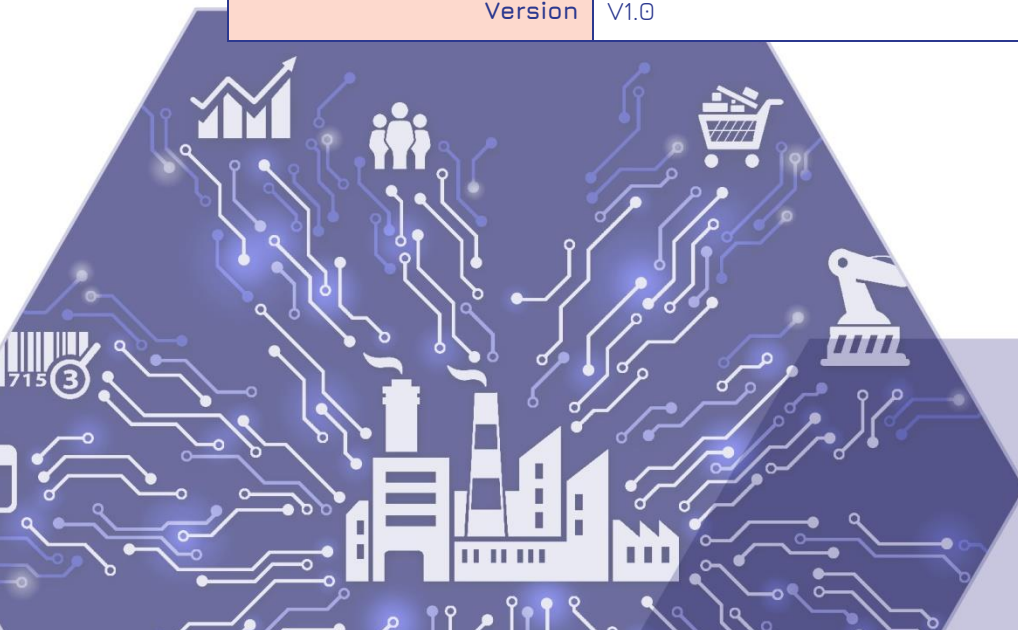




Big Data Value Spaces for Competitiveness of European Connected Smart Factories 4.0

Horizon 2020 EU Grant Agreement 780732

Title	D2.7 – Boost 4.0 standardization & certification v1
Document Owners	Dave Raggett – W3C/ERCIM
Contributors	Fernando UBIS – VIS
Dissemination	Public
Date	15/02/2019
Version	V1.0



Version history

05/02/2019	First version, ready for review
07/02/2019	Final version, reviewed by INTRA and INNO
15/02/2019	Extended by VIS

Document Fiche

Authors	Dave Raggett, W3C/ERCIM
Internal Reviewers	Konstantinos Sipsas, Intrasoftware Jesús Alonso-Rodríguez, Innovalia
Workpackage	WP2
Task	T2.5
Nature	Report
Dissemination	Public

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 Catalunya i2cat	I2CAT
TRIMEK S.A.	TRIMEK
CAPVIDIA N.V,	CAPVIDIA
Volvo Lastvagnar AB *	VOLVO
Chalmers Tekniska Högskola 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-Universität 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 document reports on the initial assessment of the role of standards and certification for the Boost 4.0. The focus is on the implications for the adoption of advanced digital technologies for vertical and horizontal integration, and the role of key technologies including graph data, microservices, distributed data storage and processing, shared ledgers and the move to enterprise wide data management and data governance. There are many relevant industry alliances and standards development organisations, this report focused in the most relevant for the project until its publication. The report discusses next steps, including plans for workshops on graph data, time-series, spatial and streaming data, along with the aim of developing a standards framework for interchange of data and schemas across database solutions from different vendors as a means to address integration across heterogeneous data silos. Preliminary plans for the use of standards in Boost 4.0 Pilots are reported.

Keywords: standardisation, standards, certification, compliance, standardisation gaps

Disclaimer

This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content. This document may contain material, which is the copyright of certain Boost 4.0 consortium parties, and may not be reproduced or copied without permission. All Boost 4.0 consortium parties have agreed to full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the Boost 4.0 consortium as a whole, nor a certain party of the Boost 4.0 consortium warrant that the information contained in this document is capable of use, nor that use of the information is free from risk, and does not accept any liability for loss or damage suffered by any person using this information.

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 N° 780732.

Table of Contents

<i>Executive Summary.....</i>	<i>5</i>
<i>Table of Contents.....</i>	<i>6</i>
<i>Abbreviations and Acronyms</i>	<i>8</i>
<i>Table of Figures.....</i>	<i>9</i>
<i>1 Introduction.....</i>	<i>10</i>
<i>2 The Role of Standardisation in Respect to Manufacturing</i>	<i>11</i>
2.1 Digitisation of Industry	12
2.2 Integration from shop floor to the office floor	14
2.3 Challenges associated with Graph Data	16
2.4 Data modelling frameworks	18
2.5 Moving away from monolithic data services	19
2.6 Distributed Data Storage and Processing	21
2.7 Data Sharing	22
2.8 Shared Ledgers using Blockchain	22
2.9 Data Governance	23
<i>3 Standardisation Initiatives.....</i>	<i>25</i>
3.1 The Digitising European Industry Initiative	25
3.2 Big Data Value Association	29
3.3 Industrial Internet Consortium	30
3.4 European Factories of the Future (EFFRA)	31
3.5 Alliance for Internet of Things Innovation (AIOTI)	31
3.6 US National Institute of Science and Technology (NIST).....	32
3.7 <i>De jure</i> organisations such as ISO, IEC, IEEE, CEN, CENELEC, ETSI	35
3.8 5G and Smart Factories.....	38

3.9	<i>De facto</i> organisations such as OPC, IETF, W3C, OMG and OASIS	41
3.9.1	OPC Foundation.....	41
3.9.2	IETF.....	41
3.9.3	W3C.....	41
3.10	Where next?	43
4	<i>Certification framework</i>	45
4.1	IDS architecture	45
4.1.1	International Data Space context	45
4.1.2	IDS Reference Architecture Model.....	49
4.1.3	IDS Connector Architecture.....	50
4.2	DATV certification methodology.....	56
4.2.1	DATV Certification Framework	58
4.3	IDS certification framework.....	60
4.3.1	IDS Connector Certification Framework	61
5	<i>The use of standards in the Boost 4.0 Pilots</i>	65
5.1	Gestamp	65
5.1.1	Experience with Standards Development.....	77
5.1.2	Data Management and Data Governance	78
5.1.3	Security, Safety, Privacy, Trust & Resilience	80
5.1.4	Testing and Certification.....	81
5.2	+GF+	81
5.2.1	Standards used	82
5.2.2	Identified Generic Standard for Further Development.....	82
5.2.3	Opportunities for new standards.....	83
5.2.4	Experience with standards development	83
5.2.5	Data Management and Governance	84
5.2.6	Security, Safety, Privacy, Trust and Resilience	84
5.2.7	Testing and Certification.....	85
5.3	WP8 – Benteler, Atlantis and Fraunhofer IEM	85
5.3.1	Data Management and Data Governance.....	87
5.3.2	Security, Safety, Privacy, Trust & Resilience	89
5.3.3	Testing and Certification.....	91
6	<i>Conclusions</i>	93

Abbreviations and Acronyms

Acronym	Meaning
A.I.	Artificial Intelligence
ANN	Artificial Neural Network
AML	Automation Markup Language
API	Application Programming Interface
CEP	Complex Event Processing
DSS	Decision Support System
ERP	Enterprise Resource Planning system
FDT	Fault Detection Tool
IoT	Internet of Things
KPIs	Key Performance Indicators
ML	Machine Learning
MES	Manufacturing Execution System
PaaS	Platform as a Service
PMT	Predictive Maintenance Tool
RDBMS	Relational Database Management Systems
SIEM	Security Information and Event Management
SSA	Singular Spectrum Analysis
SMS	Smart Manufacturing Systems
XML	eXtensible Markup Language

Table of Figures

Figure 1- Horizontal and Vertical Integration	13
Figure 2- WoT gateways hide IoT fragmentation to enable Web scale markets of services	20
Figure 3-Interoperability Layers.....	21
Figure 4 - DAMA Wheel	23
Figure 5 - RAMI 4.0 Architecture	25
Figure 6 - Standardisation and DEI – Jochen Friedrich, DEI-MSP WG, June 2018.....	27
Figure 7 - IoT SDOs and Alliances Landscape (vertical and horizontal domains)	32
Figure 8 - Data Science Sub-disciplines	33
Figure 9 - Standards aligned to the ISA95 model	35
Figure 10 - 5G applications	39
Figure 11 - Data Exchange Standards.....	47
Figure 12- Typical Enterprise architecture stack.....	48
Figure 13 - International Data Space and Cloud Platforms	48
Figure 14 - Interaction of technical components	50
Figure 15 - IDS Reference Architecture of Connector	52
Figure 16 - Industry 4.0 main HW and SW platforms in Digital Automation	56
Figure 17 - AUTOWARE Reference Architecture & SDA-SP	57
Figure 18 - Integrated approach for DATV Certification.....	59
Figure 19 - Integrated approach for DATV Certification.....	60
Figure 20 - Industry 4.0 migration path.....	60
Figure 21 - DATV Certification Approach for core components of the IDS	62
Figure 22 - Gestamp Big Data Pilot mapped in Boost4.0 architecture	66
Figure 23 - QIF Architecture.....	70
Figure 24 - QIF Model-Based Quality Workflow	71
Figure 25 - Simplified NetFlow Architecture	76
Figure 26 - Basic SNMP Communication.....	77
Figure 27 - Boost4.0 RA	86
Figure 28 - Automation Pyramid	87

1 Introduction

This document introduces the role of standardisation in respect to the challenges for big data in smart factories, and more generally, for realising the benefits of advanced digital technologies in industry as proposed by the EU's Digitising European Industry (DEI) initiative. Status reports are provided for the DEI/MSP, supporting organisations such as the BVDA and AIOTI, de jure and de facto standards development organisations, and some open source initiatives relevant to standardisation. Preliminary assessments of standards are provided on behalf of Boost 4.0 Pilots along with some recommendations. Plans are described for standardisation activities involving support from the Boost 4.0 project. Preliminary assessments are giving in respect to certification.

This Deliverable (D2.7) has been written as part of Work Package 2, Task 2.6 “Standardisation and Certification”, and should be read in conjunction with Deliverable D2.5 which describes the Boost 4.0 Reference Architecture. The analysis and information gathered for D2.7 is expected to assist with other Boost 4.0 Work Packages, e.g. WP3 for Task 3.1 – Industrial Data Space, Task 3.2 – Semantic Models, Vocabularies and Registry, Task 3.3 – Industry Data Space Connectors and Context Information Management, as well as for the Boost 4.0 Pilots covered by Work Packages 4-8

The report starts with a look at the role of standardisation in respect to manufacturing before turning to a survey of standardisation initiatives, covering the Digitising European Industry Initiative, industry alliances, de jure and de facto standards development organisations. This is followed by a look at opportunities for a certification framework, and then some preliminary information of the expected use of standards in Boost 4.0 pilots, and finally the report's conclusions.

2 The Role of Standardisation in Respect to Manufacturing

The need for agility in dealing with distributed collections of heterogeneous data silos with ever changing requirements. This becomes feasible by adopting an enterprise wide approach to data management and governance. This in turn relies on developing control over metadata and abstracting from the complexity inherent to data silos and the formats they use. Enterprise knowledge graphs can be built using W3C's approach to graph data (RDF/Linked Data). Information services can be simplified through abstractions that hide the underlying protocols and data formats – W3C's Web of Things

This section introduces the motivation for the introduction of advanced digital technologies in smart factories and the challenges that this brings, and the role that standards can play in addressing the challenges. This background will provide the context for subsequent sections that describe relevant work at a number of organisations including the DEI/MSP, pre-standardisation activities at organisations such as the BVDA and AIOTI, de jure and de facto standards development organisations.

Manufacturing has long been associated with the adoption of new technologies, for instance water and then steam power as the basis for the first industrial revolution. The introduction of electrical power and production lines for mass production. The use of electronic components for measurement and control. Computer controlled lathes and milling machines, the advent of information processing systems and the evolution of programmable robots for welding, painting, lifting and many other tasks. More recently smart sensors, advanced robotics, AI, big data lakes and cloud computing are helping to pave the way for gains in productivity, financial and operational performance, output, and market share as well as improved control and visibility throughout the supply chain.

Manufacturing agility for supplying highly customised products depends on rich information systems that control every aspect of operations, from product design, customer order handling, supply chain management, just in time and just in sequence assembly, production cells with symbiotic robots and human workers, distribution and post sales for servicing, customer relationship management and product improvements, through software and hardware upgrades.

This section continues with an introduction to what is being called the digitisation of industry and the associated aims for vertical and horizontal integration. This is followed by

subsections that explore the ideas this depends upon, including the move from tabular to graph data, the switch from monolithic applications to microservices, and related ideas for distributed storage and processing, shared ledgers, and the need for enterprise wide data management and data governance.

2.1 Digitisation of Industry

Digitalisation is a term used to express the increasing importance of digital technologies:

- Integration of digital technologies into everyday life by the digitization of everything that can be digitized. (Business Dictionary)
- The use of digital technologies to change a business model and provide new revenue and value-producing opportunities; it is the process of moving to a digital business. (Gartner)

According to McKinsey,¹ Senior management are under pressure to accelerate the digitisation of business processes. It is not sufficient to simply automate an existing process, and instead companies need to reinvent the entire business process, including cutting the number of steps required, reducing the number of documents, developing automated decision making, and dealing with regulatory and fraud issues. Operating models, skills, organisational structures, and roles need to be redesigned to match the reinvented processes. Data models should be adjusted and rebuilt to enable better decision making, performance tracking, and customer insights.

Some of the reasons² for the importance of digitisation include:

- Improve the efficiency of a business's process, consistency, and quality.
- Integrating conventional records into a digitised system removing
- redundancies and shortening the communications chain.
- Improve accessibility and facilitate better information exchange for staff and users.
- Improve response time and customer service anywhere in the world
- Reduce costs operating costs
- Ability to take advantage of analytics and real user data.
- Help with the flexibility of staff and reduced overheads

¹ Accelerating the digitization of business processes, Shahar Markovitch and Paul Willmott

² <https://www.talk-business.co.uk/2017/11/07/digitisation-important-business-users>

- Improvement plan for business continuity and growth

Companies are seeking to extract value from the vast amount of information generated by digitisation and the IoT. According to Forbes (May 2018) big data applications and analytics are projected to grow from \$5.3B in 2018 to \$19.4B in 2026.

Businesses want to establish control over the many data silos they have. This requires horizontal and vertical integration. Horizontally across organisational boundaries, including the value chains within an enterprise, the supply chain and post sales processes; and vertically from the shop floor to the office floor and upwards to the board room. Industry 4.0 thus builds upon ideas by Michael Porter on the role of value chains for competitive advantage³.

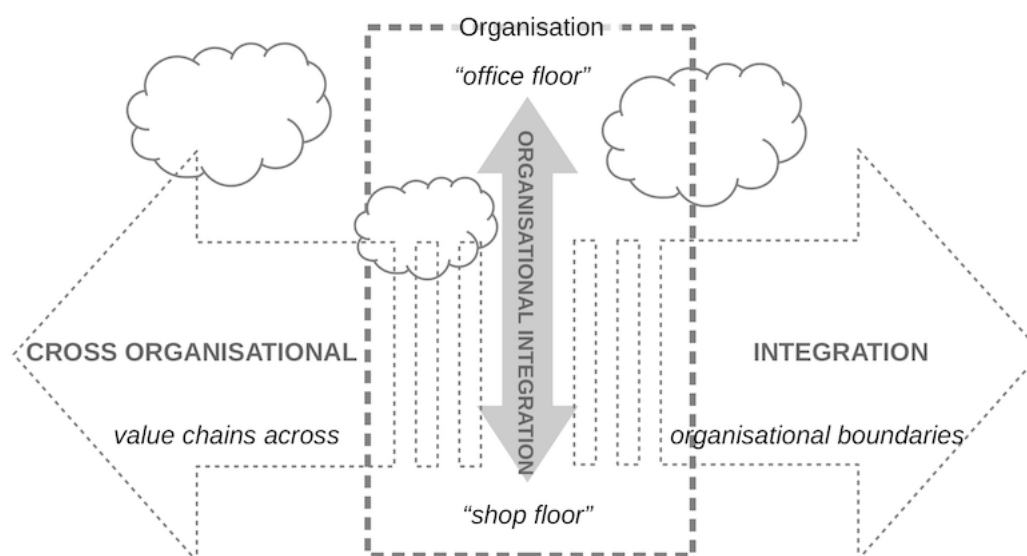


Figure 1- Horizontal and Vertical Integration

This level of integration is challenging to achieve in a way that provides agility in respect to continuously evolving requirements. Traditional technologies and methodologies are inadequate, and this is driving interest in graph data as compared to the tabular data used with relational database management systems (RDBMS). Graph data is better suited for combining information from heterogeneous data silos with ever changing requirements.

³ See Eric Porter's 1985 book "Competitive Advantage: Creating and Sustaining Superior Performance", published by Simon and Schuster, New York.

2.2 Integration from shop floor to the office floor

An attempt to review the different solutions for different requirements at different levels, e.g. real-time ethernet on the factory floor, 5G for wireless connections, LPWAN, MQTT, AMQP, etc. How to fit this into the above narrative? This needs to cover the traditional factory floor automation technologies with the very expensive connectors and how these can be replaced by much cheaper network connections.

An early example of factory automation is the Jacquard loom from the early 19th century, which used punched cards to automate the production of complex textiles. The Ford Motor Company introduced the assembly line for car production in 1913. Electrification greatly boosted factory productivity in the 1920's. Japanese companies developed the first micro-switch, protective relays and accurate timers in the 1930's, and in later years became the world leader in industrial automation.

Before the advent of solid-state electronics, control, sequencing and safety interlocks relied on relays, cam timers, drum sequencers and dedicated closed-loop controllers. Digital computers enabled the introduction of programmable logic controllers (PLC) in the 1960's. Early numerically controlled milling machines gave way to computer numerical control (CNC) for a broad range of applications.

Robotics, networking and advances in computers are paving the way for much greater flexibility in the assembly of highly customised solutions rather than mass production of standardised products. This requires sophisticated software to track materials and customer orders, with the need for integration along the supply chain, the value chain within an enterprise, and post sales for servicing and customer relationships, and feedback into product design.

The benefits of digitisation rely on a means to connect the different kinds of networks used in different parts of an enterprise. Some refer to this as the need to integrate industrial and enterprise networks. The benefits include greater connectivity and integration across plants, easier data sharing across the enterprise, and better visibility into real-time operations. The ISA/IEC-62443 Zones and Conduits Model developed by the ISA99 committee distinguishes three design areas:

- cell or area zone
- production site operations
- enterprise zone integration

On the shop floor (e.g. production cells), there may be tight constraints on latency for real-time operation. On the office floor (the enterprise zone), latency is less important than flexibility in handling different kinds of information transfers. In some contexts, it may be fine to drop events when needed, but in others, it may be critical to keep a continuous unbroken log for safety and audit purposes. Heavier weight protocols are needed when incomplete transactions need to be rolled back.

Many standards for factory automation have evolved over the years, e.g. Modbus, Profibus, the HART protocol, PROFINET, EtherNET/IP, SERCOS and EtherCAT. The connector assemblies involved are often complex and expensive. Newer standards seek to reduce the cost and complexity, and to increase the flexibility for reconfiguring factory floor machinery to fulfil rapidly evolving needs. One such example is Time Sensitive Networking (TSN) from the [IEEE 802.1 working group](#) which seeks to define mechanisms for the time-sensitive transmission of data over Ethernet networks.

According to Bill Lydon, Editor, Automation.com, TSN does not solve the problem of multivendor interoperability of controller to controller communications, which users have been complaining about for years. This is starting to be accomplished by progressive vendors, using OPC UA communications for [controller to controller](#), leveraging the joint [OPC Foundation/PLCopen](#) standards. Lydon adds: “with more powerful industrial controllers, process controllers, PLCs, and smart edge devices, the need for (specialised) industrial automation protocols may go away”.

Barcodes have proved an effective means to track parts as they progress through the supply chain, warehouse and factory floor. RFID provides an alternative that can be sensed electronically rather than optically along with the potential for scanning multiple devices at the same time. RFID can be combined with sensors, e.g. temperature for applications relating to cold chain, preventative maintenance, bulk material temperatures and electrical component monitoring. Barcodes and RFID are expected to play an important role in Industry 4.0 as the basis for tracking components under manufacture as they pass through production cells.

Wireless connectivity is very flexible, and includes technologies such as RFID, Bluetooth, ZigBee, LPWAN, WiFi and soon 5G. WiFi is used for local area networks, whilst LPWAN and 5G offer longer range. LPWAN allows for long lasting battery operation, but is restricted to low data rates.

Virtual LANs are a means to define a virtual OSI layer two network that runs over multiple physical networks, and allows administrators to separate traffic for different applications

for increased robustness and security. OSI layer three adds support for IP networking, routing and so forth. Smart routers can act as firewalls for added security.

2.3 Challenges associated with Graph Data

The strategic importance of data is driving the need for an enterprise wide approach to data management and governance. This in turn relies on developing control over metadata, and abstracting from the complexity inherent to data silos and the formats they use. “Knowledge Graph” is a popular term for expressing this metadata as a graph of nodes and links.

- Knowledge Graphs represent a collection of interlinked descriptions of entities – real-world objects, events, situations or abstract concepts, **Ontotext**;
- Knowledge Graphs are the only realistic way to manage enterprise data in full generality, at scale, in a world where connectedness is everything, **Kendall Clark**, 26 Jun 2017

There is now a wide choice of graph database solutions, but these suffer from a lack of portability unlike traditional relational database management systems and the maturity of the SQL standard for RDMS database query and update. W3C’s Resource Description Framework (RDF) is mature with two decades of experience, and has a standards suite including the OWL ontology language and the SPARQL query language. RDF supports globally unique identifiers in the form of URIs, enabling applications to use standard vocabularies for terms

Telemetry data from sensors in factory machinery provides time-series data. This can be processed to remove outliers, e.g. due to electrical interference. Data can be analysed to identify events and to look for evidence of the upcoming need for maintenance. Spatial data sets include three dimensional models and associated measurements, e.g. of temperatures or spatial discrepancies from the desired physical shape of a part under manufacturer.

The Boost 4.0 project is supporting W3C’s efforts to build bridges between the SQL/RDBMS, Property Graph, RDF/Linked Data and AI/ML communities. A W3C workshop on graph data is being organised for early March 2019 with a view to starting work on aligning graph query languages and enabling RDF to be used as an interchange framework between different graph databases. W3C is at an early stage in planning a follow-on workshop on time series data, spatial data and streaming data. The first workshop will attempt to bridge different communities including SQL/RDBMS, Property Graph, RDF/Linked Data and AI/ML. The aim is

to discuss the potential for new standards on aligning query languages for graph data, the role of RDF for interchange between different graph DB. In parallel work has started on extensions to RDF to make it easier to use by the vast majority of developers.

- The goal of the **Easier RDF** initiative⁴ is to make RDF, or some RDF-based successor, easy enough for *average* developers (middle 33%), who are new to RDF, to be consistently successful.
- Solutions may involve anything in the RDF ecosystem: standards, tools, guidance, etc. All options are on the table.
- Backward compatibility is highly desirable, but *less* important than ease of use.

Initial discussions have pointed to ideas for relaxing constraints on what RDF permits for the subject, predicate and object in respect to links between graph nodes. Perhaps the human language for a string literal can be treated as just another property? There are ideas for how to extend RDF serialisations such as Turtle and N3 to allow for links to and from a single link or a set of links. A higher level framework could directly support n-ary relationships as objects with properties, along with rules with conditions that are described in terms of graph traversal, and actions that offer flexibility in the construction and modification of graphs.

The potential for knowledge graphs is dependent on the development of vocabularies of terms. This presents plenty of opportunities for shared agreements. Sometimes this is just a matter for agreement between a handful of parties involved in sharing information, where it is necessary for them to agree on the meaning as well as the data types and formats. As the number of stakeholders involved goes up, then so does the need for a more rigorous approach to developing standards. It is unlikely that this could be met with a single standards organisation.

W3C is considering how to establish a shared vision with other standards development organisations on criteria for moving vocabulary standards along a dimension of increased maturity. One idea involves the means for trusted third parties to vet that work conforms to agreed criteria, which would include such things as the level of adoption, the level of review of design decisions, the availability and quality of documentation describing the vocabularies, their use cases and associated requirements, etc. We need to find sustainable approaches for vocabulary development that covers the cost for the infrastructure and the vetting against the agreed criteria. To encourage re-use, we need

⁴ See <https://github.com/w3c/EasierRDF>

an easy means to search for existing vocabularies, and preferably, to get in touch with people who worked on them when it comes to developing a deeper understanding.

Some vocabularies are essentially static and have a closed set of terms. Others need to continually expand to embrace new terms. This is analogous to natural language where nouns and verbs belong to a closed set of words, while proper names belong to an open set of words. If existing applications depend on particular terms, what happens when you want to evolve a vocabulary to meet new requirements? This is the challenge for versioning vocabularies when you need to support a heterogeneous mix of old and new applications.

A related challenge is how to relate vocabularies developed by different communities. They may have made different design choices. This could be due to differences in the use cases and requirements, but could also be down to differences in aesthetics, or more prosaically, due to communities using different human languages. One example is the vocabularies used for physical units of measure. The QUDT ontology takes a very detailed approach that relates units to the set of underlying physical dimensions (length, mass, time, electric current, temperature, amount of substance and luminous intensity). This level of detail is inappropriate in cases where all you want to know is the unit of measure and the scale factor, e.g. milliamps. We thus want to be able to use lightweight vocabularies and separately be able to relate them to other vocabularies as needed.

Sometimes it is possible to declare that a term in one vocabulary is the same as another term in a different vocabulary, but more generally, the mapping will be more complex. One approach is to describe mappings in terms of a more fundamental “upper” ontology that provide concepts which are common across domains. This can introduce considerable complexity when defining how terms are related to those in the shared upper ontology. It may be simpler to define direct peer to peer mappings between vocabularies. Such mappings may be context sensitive, and depend on the values of data. Work is needed to develop standards that simplify the effort needed to define mappings.

2.4 Data modelling frameworks

It is common to distinguish between conceptual, logical and physical levels. Conceptual models focus on the semantics involving entities and their relationships. Logical models express this in terms of tables, object-oriented classes or XML tags. The physical level describes how data is stored and accessed in terms of computer servers, disk partitions and so forth. Meta models are used to describe a model of a set of models, i.e. the vocabulary and relationships to be used when constructing a given class of models.

Developers have a choice in data modelling frameworks, including:

- OMG's model-driven architecture using QVT
- OMG's UML
- Chen's Entity-Relationship diagrams
- OPC Foundation's OPC-UA
- W3C's OWL ontology language
- AutomationML initiative

There is a risk that software is written in way that works good enough for the initial requirements, but is found to be hard to adapt to changes as the requirements evolve as experience is gained in everyday operations. A small change in how business is done may necessitate large changes in the underlying software and data models, increasing cost and risk of delays. This can be mitigated by implementing applications in terms of conceptual models and decoupled from the underlying details. This permits the underlying logical and physical levels to be restructured automatically, with limited changes needed to the software. A further challenge is that data cannot be shared with customers, suppliers and the value chain, because the structure and meaning of the data is not standardised or is described in a form that is not interoperable across different vendors tools.

2.5 Moving away from monolithic data services

Large monolithic applications tend to be complex and hard to maintain in respect to evolving requirements. This is motivating interest in moving to a more fluid approach based upon what have been called "microservices" in which applications are structured as a set of loosely coupled fine-grained services connected via lightweight protocols. Such services may be provided by third parties and form part of ecosystems with open markets of suppliers and consumers of services. This is being facilitated by high speed networking and cloud computing.

One approach to microservices involves direct use of RESTful APIs over HTTP and their description with e.g. OpenAPI (formerly known as Swagger), which allows you to declare an API in either JSON or YAML and then generate stub code in a variety of programming languages. This approach is being promoted by OpenAPI Initiative, an open source collaborative project of the Linux Foundation.

The Web of Things, by contrast, models services as software objects that stand as things on behalf of sensors, actuators and related information services. Applications interact with objects through the properties, actions and events that they expose. Things have URIs that can be used with RDF to describe the kinds of things, their capabilities, interrelationships and the context in which they reside. The URIs for things can be dereferenced to obtain RDF descriptions of their properties, actions and events.

Web Hubs are platforms that host applications that supply and consume things. The following figure illustrates how Web Hubs can act as gateways to isolate the details of the IoT technologies at the network edge, for improved security, and lower costs and risks for stakeholders. This approach decouples applications from the underlying protocols and data formats, and allows for rich descriptions of services as part of knowledge graphs.

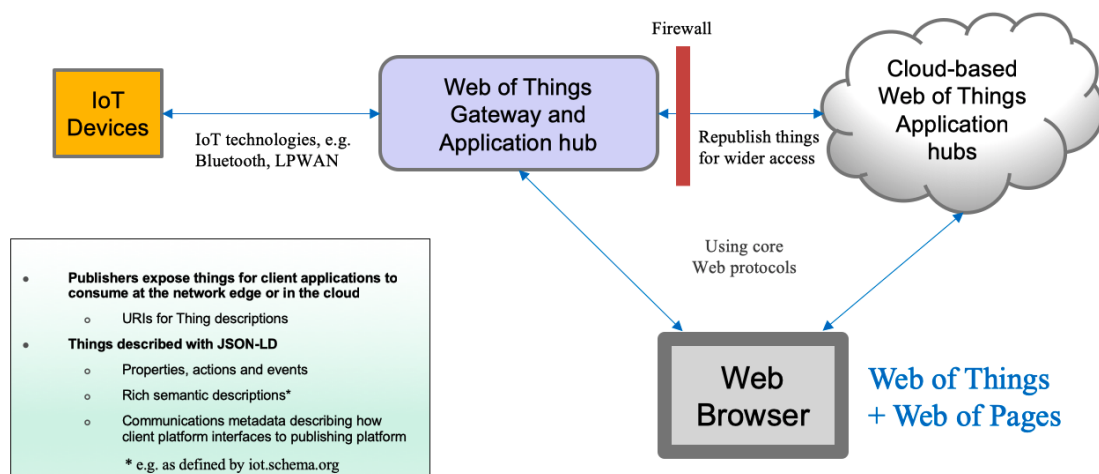


Figure 2- WoT gateways hide IoT fragmentation to enable Web scale markets of services

W3C's Web of Things seeks to overcome the challenge of fragmentation at the network edge through an abstraction layer that sits well above the details of the myriad IoT technologies and standards. As can be seen in the following figure, the Web of Things covers technical interoperability across multiple layers from knowledge awareness down to physical interoperability.

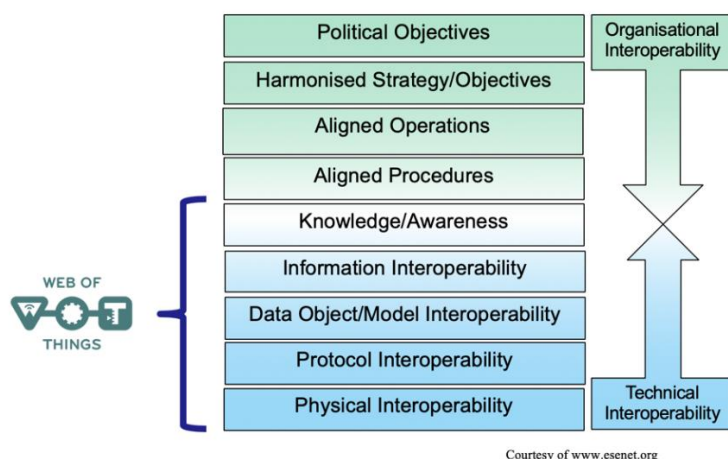


Figure 3-Interoperability Layers

W3C's work on data standards are valuable to smart factories in respect to a scalable framework for data and data models for data management and data governance. W3C's efforts on graph data etc. are described a later section.

With open marketplaces for microservices, we need standards for how suppliers can describe and publish their services, including metadata for security, terms & conditions and so forth. One approach suitable for distributed marketplaces would be via metadata embedded in, or linked from, web pages. Another approach for centralised marketplaces is via registration with a marketplace website. Search engines or marketplace operators need to be able to index services, and consumers need easy ways to install applications to local or cloud-based application platforms.

2.6 Distributed Data Storage and Processing

In practice, enterprises have to deal with multiple systems on multiple computers and connected by multiple networks. Operations on small amounts of data may be performed by downloading the data and carrying out the required operations locally. An example of this approach is where a Web application queries for some data and then operates on it as part of a web page script. The data may be transferred as comma separated value (CSV) files, JSON or XML. For larger amounts of data, the operations need to be performed close to where the data is stored. One example of this approach is where a complex SPARQL query is sent to a server hosting an RDF triple store.

For extremely large amounts of data the data and processing need to be distributed across a server farm, perhaps containing may thousands of servers. One example of this approach involves Apache Hadoop, a collection of open-source utilities designed for use with the Map-Reduce programming model. Data files are split into large blocks and distributed

across nodes in a cluster of servers. The operation to be performed is then distributed to each node to perform on that part of the data. Map-Reduce essentially divides the operation into a large number of tasks that can be performed in parallel and that feed into a network of nodes to combine the results into the desired output. The approach is designed to be fault tolerant against failures in some of the nodes.

2.7 Data Sharing

Industry 4.0 involves the need to share data and services across the supply chain. The challenge is how to do this securely, in an interoperable manner especially for really large amounts of data. One solution is to define a set of microservices that are exposed through servers using encryption and strong authentication. A refinement is to use a virtual network that offers improved control over the protocols and message exchange between participating parties. For semantic interoperability, all parties need to share the same semantic models. This means that enterprises need to distinguish between internal models used within the enterprise value chain and the models used for the services used with external parties. For really large amounts of data, it may be appropriate to hold the data in a distributed cloud storage farm that is shared in a secure way with the participating parties, and to move the processing to the cloud. Boost 4.0 seeks to make use of the International Data Space as a secure solution for data sharing. More details are given in a later section. There are expected to be opportunities for new standards work.

2.8 Shared Ledgers using Blockchain

Trust in a distributed system with multiple stakeholders can be established through the means for independent audit of operations. Traditionally this has involved each party maintaining its own ledger with complex processes for reconciling entries across different ledgers. Blockchain technologies allows a reduction in processing time and costs through the means to support a single distributed ledger that is shared across the participating parties. For practical reasons, the ledger is limited to recording transactions and is unsuitable for holding large amounts of data. This can be dealt with through a trusted distributed data store along with cryptographic checksums for verifying the integrity of the data. Boost 4.0 seeks to exploit the open-source Hyperledger project for shared ledgers. There are opportunities for standardisation in respect to a microservices architecture that supports the conceptual level operations (see above section on data modelling).

2.9 Data Governance

The need for enterprise wide oversight and management of data has given rise to the field of Data Governance. The Data Management Association⁵ (DAMA) provides extensive guidance, including the following diagram, which lists the many aspects involved:

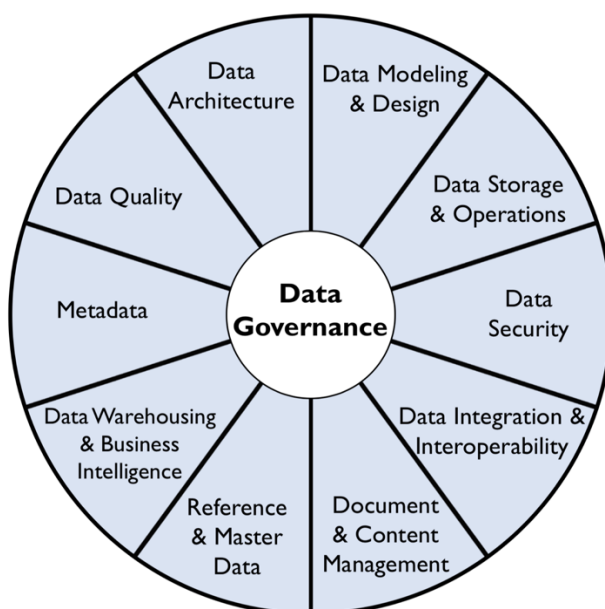


Figure 4 - DAMA Wheel

The DAMA Wheel describes eleven data management knowledge areas:

- **Data Governance** – planning, oversight, and control over management of data and the use of data and data-related resources. While we understand that governance covers ‘processes’, not ‘things’, the common term for Data Management Governance is Data Governance, and so we will use this term.
- **Data Architecture** – the overall structure of data and data-related resources as an integral part of the enterprise architecture.
- **Data Modelling & Design** – analysis, design, building, testing, and maintenance.
- **Data Storage & Operations** – structured physical data assets storage deployment and management.
- **Data Security** – ensuring privacy, confidentiality and appropriate access.
- **Data Integration & Interoperability** – acquisition, extraction, transformation, movement, delivery, replication, federation, virtualization and operational support.

⁵ <https://dama.org/>

- **Documents & Content** – storing, protecting, indexing, and enabling access to data found in unstructured sources (electronic files and physical records), and making this data available for integration and interoperability with structured (database) data.
- **Reference & Master Data** – Managing shared data to reduce redundancy and ensure better data quality through standardized definition and use of data values.
- **Data Warehousing & Business Intelligence** – managing analytical data processing and enabling access to decision support data for reporting and analysis • **Metadata** – collecting, categorizing, maintaining, integrating, controlling, managing, and delivering metadata.
- **Data Quality** – defining, monitoring, maintaining data integrity, and improving data quality.

This points to lots of opportunities for standards as a means to reduce the cost and complexity for data governance. Some of these have been analysed in earlier sections. There are considerable advantages to be had by adopting standard terminology and conceptual models, given that this eases communication between people within and between enterprises, as well as facilitating semantic interoperability between software applications. DAMA defines standard terminology and conceptual models for the data management and data, whilst the Industrial Ontologies Foundry⁶ (IOF) is defining this for the domain of digital manufacturing.

⁶ <https://sites.google.com/view/industrialontologies/home?authuser=0>

3 Standardisation Initiatives

This section surveys standardisation initiatives relevant to the aims of Boost 4.0. It encompasses an overview of the framework envisaged by the EU, which includes the rolling plan for ICT standardisation, the work underway by the MSP/DEI Working Group and the BVDA, and then looking at the work being done by IIC, EFFRA, AIOTI, and NIST as well as *de jure* standards development organisations such as ISO, IEEE, IEC, CEN, CENELEC, ETSI and *de facto* standards organisations such as OPC, IETF, W3C and AML. The deeper exploration of interoperability needs and the role of standards to tackle them, including the relationship to the Boost 4.0 pilots will be provided in deliverable D2.7 (Standardisation and Certification).

3.1 The Digitising European Industry Initiative

The European Commission launched the DEI with Communication COM(2016)176 in April 2016. The initiative aims to accompany the transformation of the global economy into a digital economy by trying to improve the framework for innovations across all sectors, including those covered by Boost4.0. Over many publications and public statements, the importance of Common standards to ensure interoperability is re-iterated. Those standards are seen to be crucial for an effective Digital Single Market. The EC is building on and complementing national initiatives such as Industrie 4.0, Smart Industry and l'industrie du future. RAMI 4.0 is German standard (DIN SPEC 91345). A three-dimensional structure describes the important aspects of Industrie 4.0. This allows us to better identify issues in their interrelations in a complex system.

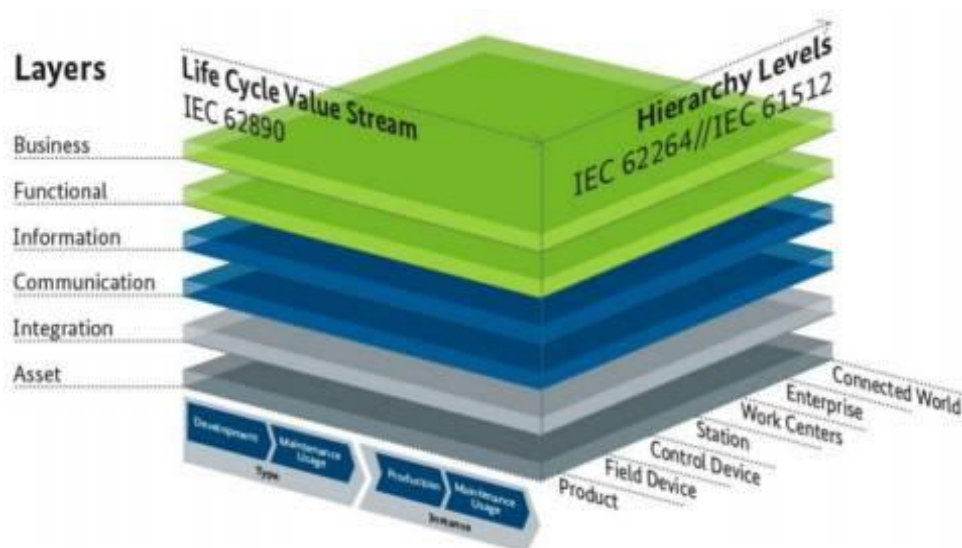


Figure 5 - RAMI 4.0 Architecture

The "**Hierarchy Levels**" axis represents the different functionalities within factories or facilities. The "**Life Cycle & Value Stream**" axis represents the life cycle of facilities and products, and is divided in two parts "Type" and "Instance" (a "Type" becomes an "Instance" when design and prototyping have been completed and the actual product is being manufactured). The "**Layers**" axis describes the decomposition of a machine into its properties structured layer by layer.

The mapping allows objects such as machines to be classified according to the model. The exercise of describing concepts using this model helps to better understand them in order to provide a common understanding for standards and use cases. It may help the understanding of use cases within Boost 4.0.

The reference model alone may not bring interoperability without an additional layer of syntactical and semantical interoperability. For instance, W3C uses linked data to achieve it. Linked data also allow tackling the issue of a large variety of data types without losing the ability to process them.

DEI - Actions are planned in terms of policy instruments, financial support, coordination and legislative powers with the aim of triggering further public and private investments in all industrial sectors. Some of the challenges that need to be overcome include: differences in the level of digitisation across industry sectors, Member States and regions; only 1 in 5 companies across the EU are highly digitised; around 60% of large companies and 90% of SMEs are lagging behind in digital innovation; Europe is lagging behind in respect to online platforms, and has a major short fall in people with appropriate digital skills.

One of the instruments to accompany the standardisation is the Rolling Plan on ICT standardisation of the European Commission with input from the European Multi-Stakeholder Platform (MSP). The plan lists topics identified as EU policy priorities where standardisation, standards or ICT technical specifications ought to play a key role in the implementation of that policy. It also contains a list of ongoing or notable actions in the areas of interest, including DEI. Of note is the June 2018 workshop in which the European Commission presented the interim results of the MSP-DEI Working Group. The MSP-DEI WG is working on:

1. Identifying standardisation needs for manufacturing sector
2. Landscaping ongoing standardisation activities, fora & consortia, Large Scale Pilots, Public-Private Partnerships (PPPs), DE/IT/FR trilateral cooperation, and other research projects, etc. that are relevant to the digitalisation of European industry

3. Developing a model for synchronisation of standardisation activities
4. Proposing a roadmap, taking national standardisation roadmaps into account and specifying concrete actions for inclusion in the Rolling Plan on ICT standardisation

The standardisation ecosystem is split into formally recognised standards bodies such as IEC, ISO, ITU, CENELEC, CEN, ETSI, 3GPP and oneM2M; and standards bodies that play a key role, but which are not formally recognised, such as W3C, OASIS, OMG, IETF, IEEE, AML and the OPC Foundation. This distinction is made by WTO, EU Regulation 1025/2012, national government contracts and rules. Other distinctions include:

- The distinction between standards for meeting regulatory requirements, and standards defining technical specifications that enable interoperability
- The difference between horizontal standards and technical specifications
- Between normative standards and technical specifications that define the basis for compliance and informative publications which may serve many different purposes

The following figure illustrates the different categories of standards considered by the DEI-MSP WG

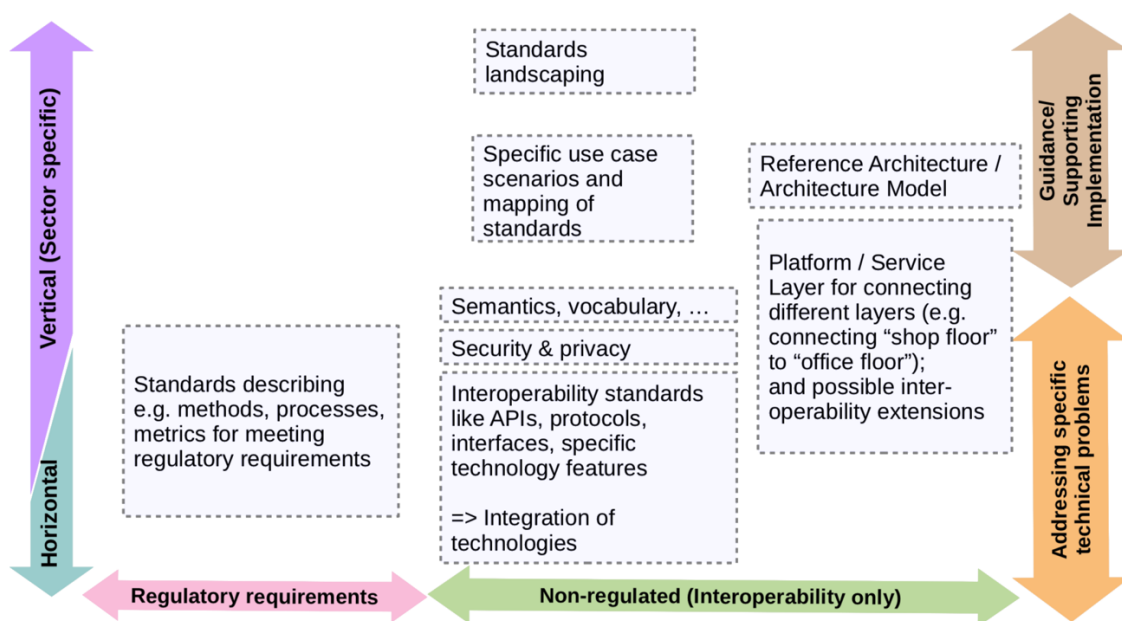


Figure 6 - Standardisation and DEI – Jochen Friedrich, DEI-MSP WG, June 2018

The final report of the MSP DEI Working Group was issued end of November 2018. Task 1 & 2 of the Working Group was to create Recommendations around landscape and gaps in standardisation. They gave a good overview of the landscape of current initiatives including a list of Specification Developing Organisations (SDOs), Reference models, Standardisation initiatives such as oneM2M and political initiatives like the trilateral

activities around the topic in Germany, France and Italy. Both tasks ended with the recommendation of nine actions:

1. Common communications standards and a reference architecture for connections between machines (M2M) and with sensors and actuators in a supply chain environment are a basic need and a priority.
2. Check whether the Skills – Agenda of the EC contains the appropriate topics and standards
3. Conduct a study to identify and analyse opportunities for revisions of existing standards
4. Improve interoperability and reduce overlap, redundancy and fragmentation.
5. Interoperable and integrated security - SDOs should work on interoperability standards for security and for linking communication protocols in order to provide end-to-end security for complex manufacturing systems including the span of virtual actors (from devices and sensors to enterprise systems).
6. Create a hierarchical catalogue of technical and social measures for assuring privacy protection and task all SDOs impacting the DEI domain in general and the advanced manufacturing domain in particular to comment on and prioritize the elements in the catalogue.
7. Standards should be developed to define the main characteristics for all levels of the interaction from mechanical to electrical to protocol to semantic levels between robot and tool to ensure the exchangeability and to enable the design of generic tooling (plug-and-play).
8. Start the discussion about the possible development of harmonised standards in the area of additive manufacturing.
9. Develop standards for ensuring long-term traceability of material to enable re-use and recycling.

Many of those suggestions concern metadata and align well with the IDS model used by Boost4.0 and explained further down.

Task 3 takes into consideration the specificity of the European standardisation landscape. While it is of common use that there are several competing standards on the same topic in

the US system, the EU has a tradition of trying to avoid competing standards. There are numerous ways to harmonize existing work and Task 3 provided recommendations on how to synchronise the large variety of standardisation activities in the area of DEI and industry 4.0. They suggest to create an additional networking platform for the variety of organisations and initiatives active in the area. Task 4 had to create a first roadmap and reiterated the recommendation of Task 3 to ask the European Standardisation Organisations (ESOs, CEN, CENELEC and ETSI) to host such a networking platform. Given its size, Boost4.0 is certainly qualified to participate in such a platform.

3.2 Big Data Value Association

The Big Data Value Association (BVDA) is the European contractual Public-Private Partnership on Big Data and seeks to build a data drive economy across Europe. BVDA has an official liaison with ISO/JTC1/WG9 with the aim of channelling European input into global standards, and has organized several workshops to engage with standardisation bodies such as ETSI, CEN / CENELEC, W3C, OneM2M, and RAMI 4.0. Efforts are underway to further develop the BVDA Reference Model to align with others such as oneM2M, BDE Platform, AIOTI, RAMI 4.0, etc.

The Big Data Value Chain starts with data acquisition, then data processing and analysis, then data storage and curation, and ends with data visualisation and services. Horizontal concerns include data protection and management, and the location for data processing, e.g. at the edge, near to the edge (fog computing) or in high performance cloud-based systems. Vertical concerns include data types and semantics, standards, communication and connectivity, cybersecurity, engineering and DevOps, marketplaces, industrial platforms, personal data platforms, and ecosystems for data sharing and innovation.

Standardisation is needed to address the lack of an existing standard platform, limiting stakeholders from participating in the European Digital Single Market, and the lack of clear governance models (reference models, guidelines and best practices) regulating the secure and trusted exchange of proprietary data. The aim is to enable industrial participants to safely and within a clear legal framework, to monetise and exchange data assets.

The BVDA SRIA notes NOSQL databases as an example of where the lack of standards is a major obstacle. As a result, application code is tied to specific storage solutions and query mechanisms. Similar concerns apply to complex event processing for real-time data, a lack of portability across graph databases, and a lack of agreed standards for query languages, data storage and management.

3.3 Industrial Internet Consortium

The Industrial Internet Consortium (IIC) was founded in 2014 and is now incorporated in the US. Initially IIC has set its own scope to best practices thus excluding hard core standardisation. In the meantime, this isn't very clear anymore. The most prominent work from IIC is the IIC Industrial Internet Reference Architecture (IIRA). The architectural template and methodology use existing standards and assists people in designing systems in the area of IoT in order to achieve interoperable IoT systems in the industrial area.

The Technology Group within IIC has several Task Groups⁷:

- The Architecture Task Group has authored the mentioned Industrial Internet Reference Architecture (IIRA).
- The Connectivity Task Group is responsible for connectivity aspects of the Industrial Internet Reference Architecture and produces the Industrial Internet Connectivity Framework (IICF) technical report.
- The Digital Twin Interoperability Task Group defines digital twin characteristics and their interoperability.
- The Distributed Data Interoperability and Management (DDIM) Task Group defines the properties of a data service framework for the Industrial Internet. Its purpose is to provide a ubiquitous data-sharing integration framework for all architecture elements.
- The Edge Computing Task Group identifies and evaluates standards, practices, deployment models and characteristics best suited for addressing the IIoT space from a holistic perspective and highlighting gaps where needed. This Task Group recently published an Introduction to Edge Computing in IIoT white paper and is now developing the Industrial Internet Edge Framework (IIEF).
- The Industrial Artificial Intelligence Task Group defines the properties of realizable, comprehensive analytical techniques.
- The Industrial Distributed Ledger Task Group is addressing industrial distributed ledger technologies.
- The Innovation Task Group offers a stage for research communities and start-ups to present or demo innovations

⁷ <https://www.iiconsortium.org/wc-technology.htm>

- The IT/OT Task Group develops best practices related to technology deployed at the IT/OT boundary.
- The Networking Task Group is identifying and analysing requirements, trends and technologies for various usage scenarios across industrial verticals for OSI layers 1, 2 and 3 in the context of IIoT. This Task Group recently published an Industrial Networking Enabling IIoT Communication white paper and is now developing the Industrial Internet Networking Framework (IINF).
- The Vocabulary Task Group recently published a new version of the Industrial Internet Vocabulary technical report which is a common and reusable vocabulary of terms as they apply to specific IIC outputs. This vocabulary will be used in all published IIC deliverables to ensure consistent terminology.

This way, IIC tries to cover all aspects given in their reference architecture. But nowhere it mentions how it relates to other frameworks like RAMI 4.0. It is also unclear how the effort relates to SDOs developing similar components like networking in the IETF or vocabulary development in W3C.

3.4 European Factories of the Future (EFFRA)

The European Factories of the Future (EFFRA) is a non-for-profit, industry-driven association promoting the development of new and innovative production technologies. It is the official representative of the private side in the 'Factories of the Future' public-private partnership with the European Commission, and aims to bring together private and public resources to create an industry-led programme in research and innovation with the aim of launching hundreds of market oriented cross-border projects throughout the European Union.

3.5 Alliance for Internet of Things Innovation (AIOTI)

The AIOTI is the private partner in the public-private partnership with the European Commission in respect to the Internet of Things, and H2020 funded projects such as the IoT Large Scale Pilots. AIOTI has many working groups. Of particular note is Working Group 3 (WG03) on Standardisation. This group is working on the IoT standardisation landscape, see below, semantic interoperability and a standardisation gap analysis and associated recommendations.

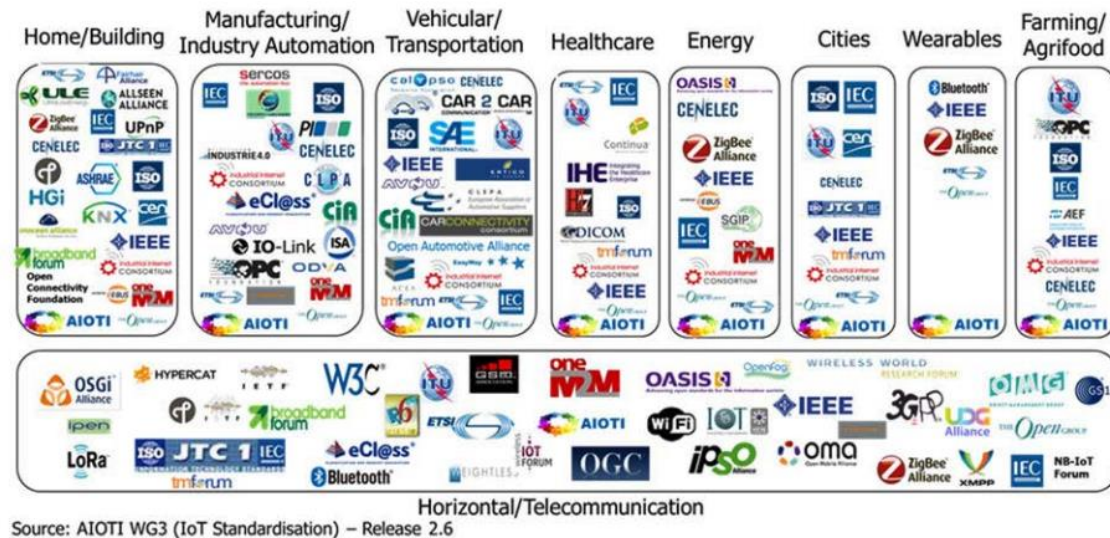


Figure 7 - IoT SDOs and Alliances Landscape (vertical and horizontal domains)

3.6 US National Institute of Science and Technology (NIST)

NIST is the U.S. National Institute of Standards and Technology. The NIST Big Data Public Working Group has published a report on definitions and taxonomies⁸. This introduces the primary characteristics of Big Data and associated frameworks, and its relationship to other technological innovations including high performance computing, cloud computing, the Internet of Things, cyber-physical systems and blockchains, as well as the recently coined discipline of “Data Science”.

NIST define Data Science as the extraction of useful knowledge directly from data through a process of discovery, or of hypothesis formulation and hypothesis testing. Data Science is this tightly coupled to the analysis of Big Data, statistics and data mining. The following figure is courtesy of Nancy Grady Drew Conway, and illustrates sub-disciplines of Data Science.

⁸ <https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-1r1.pdf>

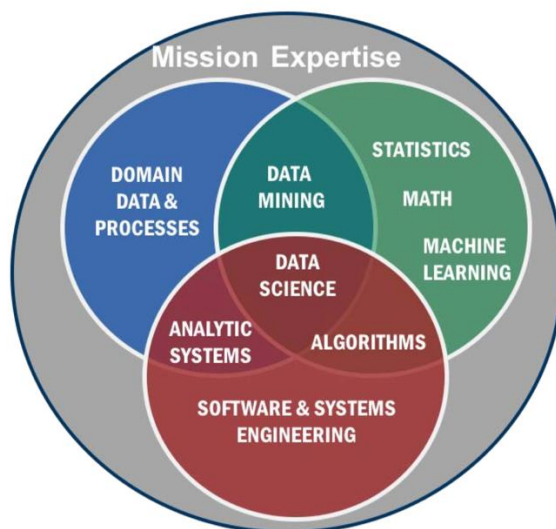


Figure 8 - Data Science Sub-disciplines

One challenge for Data Science is how to handle bias in the data coverage, especially for machine learning techniques based upon Deep Learning. This can in part be addressed through metadata that supports analysis of the training set. Newer work has looked at how to compensate by getting the training phase to focus more heavily on under-represented examples.

Another challenge is dealing with variations in data quality either within a given dataset or variations in quality between different datasets. A “data cleaning” step when importing data can help. This looks for values that are either outside of the expected limits, or outside statistical predictions based upon historical data. There could be metadata that indicates that a particular sensor is broken, and in some cases, it may be possible to replace a suspect value with one computed on the basis of preceding and following values, or readings from adjacent sensors.

Anomalous values may indicate that something significant has happened rather than say a faulty sensor or electrical interference etc. This points to opportunities for stream processing for event detection, and to successive layers of interpretation that progressively reduces the volume and velocity of information that feeds into centralised cloud-based systems.

NIST’s 2016 report NISTIR 8107⁹ covers the current standards landscape for smart manufacturing systems. It describes the need to replace the traditional hierarchical control model, as used by the classical manufacturing system, by a new paradigm based

⁹ <https://nvlpubs.nist.gov/nistpubs/ir/2016/NIST.IR.8107.pdf>

upon distributed manufacturing services. This involves smart networked devices, embedded intelligence at every level, predictive analytics and cloud computing and makes the point that all these technologies depend on standards.

Smart manufacturing characteristics:

- Digitisation of every part of a manufacturing enterprise with interoperability and enhanced productivity
- Connected devices and distributed intelligence for real time control and flexible production of small batch products
- Collaborative supply chain management with fast responsiveness to market changes and supplying chain disruption
- Integrated and optimal decision making for energy and resource efficiency
- Advanced sensors and big data analytics throughout the product lifecycle to achieve a fast innovation cycle

The report identifies the following priority areas for standards advancement:

- Smart manufacturing systems reference model and architecture
- Internet of Things reference architecture for manufacturing
- Manufacturing service models
- Machine to machine communication
- PLM/MES/ERP/SCM/CRM integration
- Cloud manufacturing
- Manufacturing sustainability
- Manufacturing cybersecurity

The report further includes tables of relevant standards for modelling practices, product model and data exchange, manufacturing model data, production system modelling and practice, production system engineering, production lifecycle data management, production system O&M, general standards for modelling and executing business processes, enterprise level standards, MOM level standards, SCADA and device level, and finally, cross-level standards

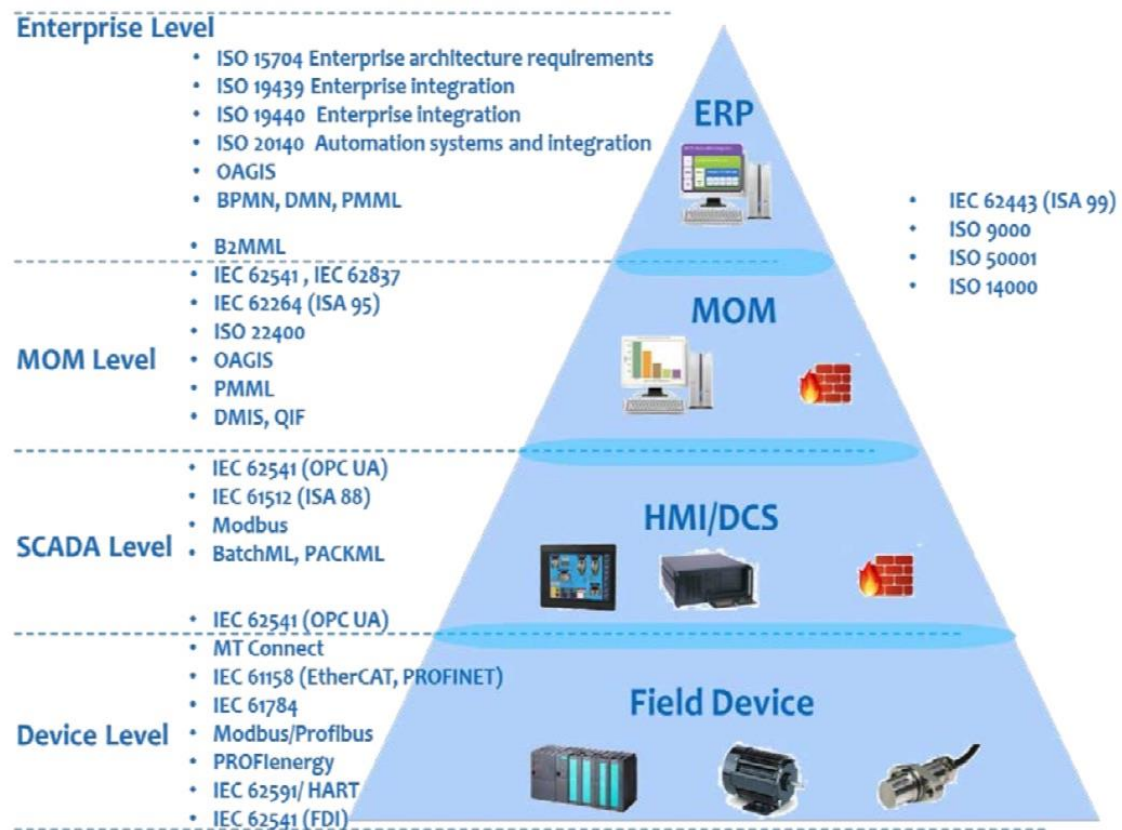


Figure 9 - Standards aligned to the ISA95 model

NIST note that traditional standards development efforts, primarily focused on incremental improvements of existing standards, are unable to keep pace with the speed of evolving technology. Instead, new requirements for smart manufacturing are being identified by a collaboration across standards development organisations, national manufacturing initiatives and industrial consortia.

3.7 *De jure* organisations such as ISO, IEC, IEEE, CEN, CENELEC, ETSI

De jure standards development organisations are those whose standards are officially recognised by governments and mandated by legal requirements.

Here are some examples of ISO and IEC standards relevant to smart manufacturing:

- ISO/TC 108/SC 5 – Condition monitoring and diagnostics of machine systems
- ISO/TC 184/SC 4 – Industrial Data
- ISO/TC 184/SC 5 – Interoperability, integration and architectures for enterprise systems and automation applications

- ISO/TC 299 – Robotics
- ISO/TC 261 – Additive manufacturing
- IEC/TC 65 – Industrial process management, control and automation
- IEC/TC 3 – Information structures and elements, identification and marking principles, documentation and graphical symbols
- IEC/SC 3D – Product properties and classes, and their identification
- ISO/IEC JTC1 – Information technology
- ISO/IEC JTC1/SC 7 – Software and systems engineering
- ISO/IEC JTC 1/SC 38 – Cloud computing and distributed platforms
- ISO/IEC JTC1/SC 41 – Internet of Things and related technologies
- ISO/IEC JTC1/SC 42 – Artificial Intelligence

See footnote¹⁰ for a link to the list of IEC technical committees and subcommittees.

The ISO/IEC JTF 1 is a joint ISO/IEC task force focusing on the smart manufacturing standards map. It plans work in three phases. The first phase is creating an initial compilation of terms and definitions for smart manufacturing. The second phase will classify the standards map according to existing reference models. The third phase will maintain this map together with external bodies.

The ITU-T is currently addressing smart manufacturing and digitisation of industry from the perspective of technologies and applications for the Internet of Things. ITU-T SG 20 “IoT and Smart Cities and Communities is working on requirement, capabilities and use cases across verticals, smart manufacturing in the context of industry internet of things, fundamental characteristics and high-level requirements for manufacturing systems, reference model for product life cycles for smart manufacturing, and smart manufacturing related use cases.

CEN/CENELEC have a number of tasks and working groups relevant to smart manufacturing:

- Safety of machinery
- Industrial process measurement, control and automation
- Printed electronics
- Cybersecurity and data protection
- Advanced automation technologies and their applications
- Additive manufacturing
- Electromagnetic compatibility (EMC)

¹⁰ See <https://www.iec.ch/dyn/www/f?p=103:6:0##ref=menu>

- Maintenance
- Electrical energy measurement and control
- Home and building electronic systems
- Electrotechnical aspects of telecommunication equipment
- System aspects of electric energy supply

See footnote¹¹ for the CEN/CENELEC work program related to machinery.

The IEEE has published a standards landscape for smart manufacturing systems (SMS)¹². A separate paper covers the Industry 4.0 standards landscape from a semantic integration perspective¹³.

ETSI focuses on telecommunication standards and works closely with oneM2M on machine to machine standards as well as 3GPP and the 5G Alliance. Here is a list of some relevant ETSI technical committees:

- Human factors – ETSI TC HF Human factors
- Information management, semantic interoperability and information display
 - ETSI ISG ARF Augmented reality
 - ETSI ISG CIM – Context information management
 - ETSI TC SmartM2M
 - ETSI STF 534 – SAREF Extensions in smart city, industry + manufacturing and agri-food domains
 - ETSI STF (RP2017) – SAREF extensions in automotive, wearable, e-health/aging well and watering domains
 - ETSI STF 547 (RP2017) – Coordinated approach for security/privacy and (semantic interoperability of standardised IoT platforms (supporting AIOTI))
- Cybersecurity, identity and privacy
 - ETSI TC Cyber
 - ETSI ISG ISI Information security indicators
 - ETSI TC ESI
- Communications, radio spectrum
 - ETSI TC MSG
 - ETSI TC NTECH
 - ETSI TC ERM

¹¹ <https://www.cenelec.eu/standards/Sectors/Machinery/Pages/WorkProgramme.aspx>

¹² See <https://ieeexplore.ieee.org/document/7294229>

¹³ See <https://ieeexplore.ieee.org/document/8247584>

- o ETSI TC DECT

In addition, there are national standardisation bodies and smart manufacturing initiatives. The latter includes Germany's Industrie 4.0, France's Industrie du future, Manufacturing USA, Korea's Manufacturing Innovation 3.0, and China's Made in China 2025. This is not a complete list as many other countries have launched similar initiatives.

3.8 5G and Smart Factories

5G offers much faster speeds and lower latency for data transfers compared to previous generations of mobile networks. This section focuses in summarizing the standardization organizations and initiatives that represent 5G networking, in particular those related with Industry 4.0.

3GPP (3rd Generation Partnership Project) is the main organization defining and creating 5G cellular/mobile communication standards including the radio access, core transport and service capabilities. Currently, 3GPP has the mission to accomplish in its 5G standards the requirements defined by another standardization organization: ITU (International Telecommunication Union). ITU has defined the KPIs (Key Performance Indicators) for mobile networks by 2020 and beyond, under the umbrella of the IMT-2020 framework, summarized in:

- Enhanced Mobile Broadband (eMBB) to deal with hugely increased data volumes, overall data capacity and user density
- Massive Machine-type Communications (mMTC) for IoT, requiring low power consumption and low data rates for very large numbers of connected devices
- Ultra-reliable and Low Latency Communications (URLLC) to cater for safety-critical and mission critical applications.

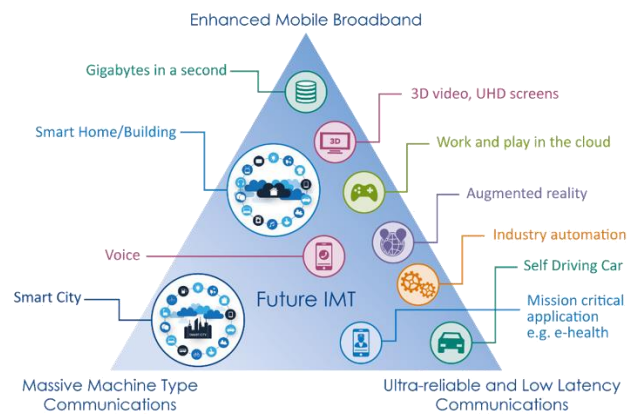


Figure 10 - 5G applications

Industry 4.0, such as connected factories, relies heavily in IMT-2020 KPIs or expected benchmarks offered by the network to accomplish their needs.

Currently, 3GPP has split 5G standardization work in several phases:

- Release 15 (5G phase 1)¹⁴. Published in June 2018, allows 4G and earlier Core network with NR (New Radio) for 5G to live together.
- Release 16 (5G phase 2)¹⁵. Expected by end of 2019, will defined 5G system to its completion.

One of the key technologies of 5G with high interest to Industry 4.0 is the **network slicing**, which enables the operator to create logically partitioned networks at a given time, customized to provide optimized services for different market scenarios. This includes:

- Network virtualization
- Isolation and security
- Elasticity and quality of service

Some aspects of these technologies, however, are being standardized by other organizations:

ETSI¹⁶ (European Telecommunications Standards Institute) works in several topics related with 5G networks. The more relevant ones are 5G infrastructure aspects related with

¹⁴ <http://www.3gpp.org/release-15>

¹⁵ <http://www.3gpp.org/release-16>

¹⁶ <https://www.etsi.org/>

Network Virtualization and slicing (ETSI ISG NFV), security (ETSI TC cyber), edge computing (ETSI MEC) and Artificial Intelligence (ETSI ENI).

IETF¹⁷ (The Internet Engineering Task Force) specifies the relevant communication protocols used in the Internet. Some topics such as service chaining and slicing, as well as most of the basic network transport protocols used by 5G (HTTP, TLS, TCP, SCTP, etc...) are specified in this standardization body.

5GPPP¹⁸ (5G Infrastructure Public Private Partnership) is an initiative between the European Commission and European ICT industry (ICT manufacturers, telecommunications operators, service providers, SMEs and researcher Institutions) that contributes in other standards and trials in 5G.

As part of 5GPPP standardization role, this Initiative published a white paper¹⁹ with common recommendations for **Factories in Industry 4.0**. Most relevant aspects proposed are: focus on deterministic wireless communication for zero defect manufacturing, security and high availability mechanisms, research network capabilities to manage heterogeneity to reduce TCO or focus on networked data management to create new data-driven business.

5G-ACACIA²⁰ (5G Alliance for Connected Industries and Automation) searches for the best **applicability of 5G Technology for connected network industries**. Also, they are searching for correct standardization and regulation to cope with industry needs.

5GSA²¹ (5G Slicing Association) is an industry Initiative that aims at leveraging the benefits of **5G network slicing to provide value to various industries** and the society as a whole.

Finally, **5GAA**²² (5G Automotive Association) is a global, cross-industry organisation of companies from the **automotive, technology, and telecommunications industries (ICT)**, working together to develop end-to-end solutions for future mobility and transportation services.

¹⁷ <https://www.ietf.org/>

¹⁸ <https://5g-ppp.eu/>

¹⁹ <https://5g-ppp.eu/wp-content/uploads/2014/02/5G-PPP-White-Paper-on-Factories-of-the-Future-Vertical-Sector.pdf>

²⁰ <https://www.5g-acia.org>

²¹ <http://www.5gnsa.org/>

²² <http://5gaa.org>

3.9 *De facto* organisations such as OPC, IETF, W3C, OMG, AML and OASIS

This section surveys some of the industry alliances and related standards development organisations. These are not formally recognised by the European Commission and national governments, but are nonetheless very important for many aspects of industry.

3.9.1 OPC Foundation

The OPC Foundation is an industry consortium focused on widely used standards for industrial automation, including industrial control systems and process control generally. OPC originated in 1994 with work on applying Microsoft's Object Linking and Embedding (OLE) technology to process control. The OPC Unified Architecture (OPC-UA) defines a service oriented cross platform architecture for machine to machine communication, featuring an object-oriented information model inspired by UML. Key target industries include pharmaceutical, oil and gas, building automation, industrial robotics, security, manufacturing and process control. One of the yet to be investigated challenges is how to map OPC-UA models with ontologies and models based upon RDF, such as those under development by the Industrial Ontologies Foundry.

3.9.2 IETF

The Internet Engineering Task Force (IETF) is responsible for standards relating to the packet-based Internet Protocol (IP), e.g. TCP/IP, HTTP, CoAP, WebSockets, SMTP and DNS. The IETF has vastly improved the address space through the transition from IPv4 to IPv6. IP protocols operate at ISO layer 4 and rely on layer 3 protocols for data transport using layer 3 device and port identifiers, e.g. Ethernet which uses 48-bit network interface identifiers. Layer 4 identifiers and protocols, by contrast, support packet routing across a network of networks, enabling the Internet to grow exponentially over successive decades in respect to the number of devices, bandwidth, and data carried. IETF standards are essential for big data in smart factories and includes an emphasis on security, e.g. transport level security (TLS) for datagram and session based transport protocols.

3.9.3 W3C

The World Wide Web Consortium is an international community dedicated to developing standards for web technologies, e.g. in respect to the Web of Pages, such as HTML, CSS, SVG and associated JavaScript APIs for use by Web browsers, and in respect to the Web of Data, e.g. RDF, OWL, SPARQL and SHACL. Web browsers are available across a range of

devices including desktop, tablets, and smart phones, and offer a flexible solution for user interfaces to networked applications for smart factories.

W3C's Web of Things seeks to overcome the fragmentation of the IoT through an abstraction layer that sits well above the details of the myriad IoT technologies and standards. The Web of Things associates URIs with things that stand for sensors, actuators or related information services. These URIs can be used for RDF/Linked Data descriptions of the kinds of things, their relationships and the context in which they reside. Things are exposed to applications as software objects with properties, actions and events, described in a platform neutral way using JSON-LD and JSON Schema. This decouples developers from the details of the underlying networking technologies.

This points to an opportunity to more closely connect RDF with object-oriented descriptions, n-ary terms, and Property Graphs

Some observations:

- Many developers are familiar with JSON but not with RDF and Linked Data
- Property Graphs are graphs with sets of objects whose properties are literals or other objects. Properties can be associated with sub-properties.
- Property names are locally scoped to the thing as you would expect in object-oriented programming languages
- In the Web of Things, properties can have sub-properties and so forth, just as in Property Graphs
- The Web of Things can be considered as a superset of Property Graphs with the addition of actions and events
- You can associate metadata with things, properties, actions and events, e.g. data types, data constraints, units of measure and so forth
- You can distinguish between properties and other kinds of metadata
- You can describe how to interact with things via updating property values, invoking actions and listening for events
- JSON Schema isn't ideal, e.g. it doesn't directly allow you to specify things as first class types that can be used for property values, passed to and from actions, or

passed with events, and there isn't a direct way to distinguish a value as a link except as a property that has a string value and is typed as a link.

- Likewise, there isn't a direct way to annotate the link except as an annotation on the property rather than the value – link annotations are valuable and can be used for provenance, temporal constraints (e.g. relationships that hold during a given time interval), spatial constraints, data quality, and so forth.
- As yet there are no proposals for rule languages for the Web of Things, although it is possible to use RDF shape rules in SHACL etc. to express constraints on things.

Further work could address these limitations in a way that is easy to use by the vast majority of developers without the need to be aware of the details of the RDF core that underpins it. Investigations are needed to explore the potential for a higher level framework that builds on top of the RDF core, and can be used as an interchange framework between Property Graphs.

3.9.4 AutomationML Association

The AutomationML association is an industrial initiative focused in the development of a data exchange format for production systems. The format characterized to be neutral, vendor independent, open and freely accessible standard. The language under development, AML (Automation Markup Language) is a neutral data format XML for the storage and exchange of plant engineering information and standardized within the International Electrotechnical Commission (IEC).

Data flow in the life cycle of production systems has always played an important role for enhancing productivity, reducing errors and maximising efficiency. The AML provides an effective solution in the context of the engineering of production systems to serialize and handle data through the different steps (component and technology development, production system engineering, commissioning, maintenance and reconfiguration and decommissioning).

The advent of the Industry 4.0 and similar initiatives has pushed the standard, being adopted and extended by different frameworks, such as RAMI 4.0.

3.10 Where next?

This section considers opportunities for further work on standardisation in the second and third year of the Boost 4.0 project. The W3C Workshop on Graph Data in March 2019 will provide a timely opportunity for an exchange of perspectives across the SQL/RDBMS, Property Graph, RDF/Linked Data and AI/ML communities, and is expected to recommend next steps, for instance, launching W3C Groups to incubate ideas for alignment on query languages for graph data, and a framework for interchange across databases that will facilitate horizontal and vertical integration across data silos.

We want to look at the feasibility of launching a W3C Business Group to foster discussion on use cases and requirements for standards for integration across heterogeneous information systems, along with the role of knowledge graphs for data governance. This would seek to exploit existing work in industrial alliances and SDOs, including the public-private partnership organisations associated with the EU research and innovation projects on manufacturing, big data and the Internet of Things.

A further workshop is at an early stage of consideration for late 2019 on time-series data, spatial data and streaming data, and is expected to be very relevant to Boost 4.0's work on Big Data for smart factories. By that time, it should be much easier to engage with the requirements emerging from the Boost 4.0 pilot projects.

Other areas for investigation include the role of standards for distributed storage and information sharing across organisational boundaries. This can build upon the requirements identified by the International Data Space (IDS) in respect to the Boost 4.0 pilots. A similar opportunity is presented by work on applying blockchains to shared ledgers in the context of smart manufacturing.

4 Certification framework

The International Data Space IDS is a virtual data space leveraging existing standards and technologies, as well as accepted governance models for the data economy, to facilitate the secure and standardized exchange and easy linkage of data in a trusted business ecosystem. It thereby provides a basis for smart service scenarios and innovative cross-company business processes, while at the same time making sure data sovereignty is guaranteed for the participating data owners. The Connector Architecture uses Application Container Management technology to ensure an isolated and secure environment for individual data services. Based on DATV methodology, IDS connector certification framework is presented, ensuring the certification of Big Data driven solutions for connected Smart Factories through the validation of involved components, developed systems and deployed solutions.

4.1 IDS architecture

The International Data Space IDS is a virtual data space leveraging existing standards and technologies, as well as accepted governance models for the data economy, to facilitate the secure and standardized exchange and easy linkage of data in a trusted business ecosystem. It thereby provides a basis for smart service scenarios and innovative cross-company business processes, while at the same time making sure data sovereignty is guaranteed for the participating data owners.

Data sovereignty is a central aspect of the International Data Space. It can be defined as a natural person's or corporate entity's capability of being entirely self-determined with regard to its data. The IDS initiative proposes a Reference Architecture Model for this particular capability and related aspects, including requirements for secure and trusted data exchange in business ecosystems.

After an initial analysis on IDS Architecture model, the IDS connector concept is presented focusing on its capability to offer an isolated context where data is processed guaranteeing the ownership of data and restricted access from service owners to provide end-users the knowledge/results defined.

4.1.1 International Data Space context

Novel digital products and services often emerge in business ecosystems, which companies enter to jointly fulfil the needs of customers better than they can do on their own. In such ecosystems, which emerge and dissolve much faster than traditional value

creating networks, the participating companies have a clear focus on end-to-end customer processes in order to jointly develop innovative products and services. Examples of business ecosystems are numerous and can be found across all industries; many of them have been analysed and documented by the Smart Service Welt working group. Key to all these scenarios is the sharing of data within ecosystems. End-to-end customer process support can only be achieved if ecosystem partners team up and jointly utilize their data resource.

From these two developments – data turning into a strategic resource, and companies increasingly collaborating with each other in business ecosystems – results a fundamental conflict of goals as a main characteristic of the digital economy: on the one hand, companies increasingly need to exchange data in business ecosystems; on the other hand, they feel they need to protect their data more than ever before, since the importance of data has grown so much. This conflict of goals is all the more intensified, the more a company is engaged in one or more business ecosystems, and the higher the value contributed by data to the overall success of the collaborative effort.

Data sovereignty is about finding a balance between the need for protecting one's data and the need for sharing one's data with others. It can be considered a key capability for companies to develop in order to be successful in the data economy. To find that balance, it is important to take a close view at the data itself, as not all data requires the same level of protection, and as the value contribution of data varies, depending on what class or category the data can be subsumed under.

Cross-company data exchange and inter-organizational information systems are not a new topic, but have been around for decades. With the proliferation of Electronic Data Interchange (EDI) in the 1980s, many different data exchange scenarios emerged over time being accompanied by the development of respective standards. Figure 11 shows the evolution of different classes of data exchange standards and identifies a need for standardization. Data sovereignty materializes in "terms and conditions" that are linked to the data upon its exchange and sharing. However, these terms and conditions (such as time to live, forwarding rights, price information etc.) have not been standardized yet. In order to foster the emergence of data sovereignty in the exchange of data within ecosystems, standardization activities are needed. This does not mean that existing standards will become obsolete. Contrary to that, the overall set of standards companies need to comply with when exchanging and sharing data is extended.

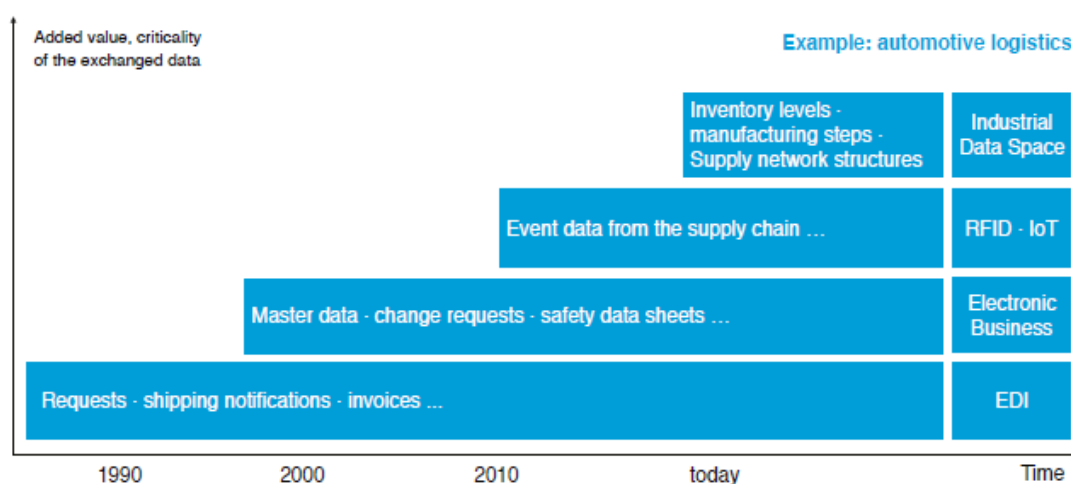


Figure 11 – Data Exchange Standards

The growing number of industrial cloud platforms will also drive the need towards a standard for data sovereignty. With the large number of different platforms emerging – driven by technology providers, software companies, system integrators, but also existing intermediaries – it is very much likely that the platform landscape will be heterogeneous – at least for a significant amount of time. Platform providers will increasingly have to provide capabilities for secure and trusted data exchange and sharing between their own platform and other platforms in the ecosystem. Furthermore, the cloud platform landscape is likely to be characterized by a “plurality” of architectural patterns ranging from central approaches, such as so-called “data lakes”, to completely distributed architectures, such as applications of blockchain technology.

Data owners and data providers will choose the platform depending on the business criticality and the economic value of the data goods they want to exchange and share via the respective platform. As the entire data resource of a company consists of data of different criticality and value, many companies will use different platforms for different needs.

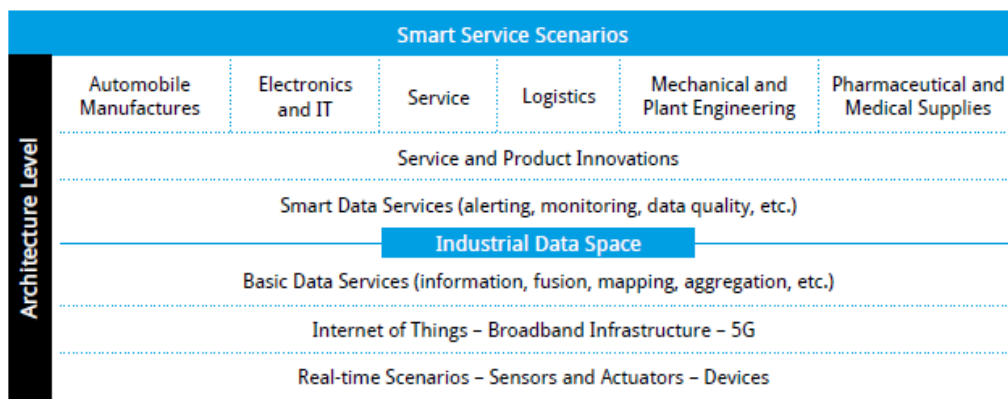


Figure 12- Typical Enterprise architecture stack

By proposing an architecture for secure data exchange and trusted data sharing, the International Data Space contributes to the design of enterprise architectures in commercial and industrial digitization scenarios. It does so by bridging the gaps between research, industrial stakeholders, political stakeholders, and standards bodies. The architecture is designed with the objective that the differences between top-down approaches and bottom-up approaches can be overcome. Figure 12 shows a typical architecture stack of the digital industrial enterprise. The International Data Space connects the lower-level architectures for communication and basic data services with more abstract architectures for smart data services. It therefore supports the establishment of secure data supply chains from data source to data use, while at the same time making sure data sovereignty is guaranteed for data owners.

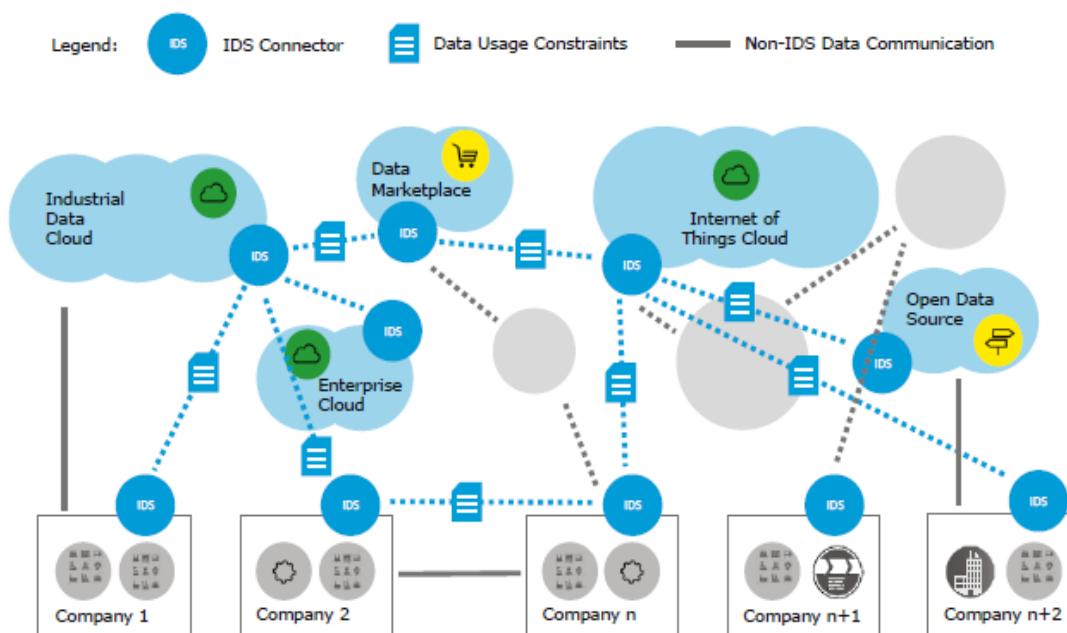


Figure 13 - International Data Space and Cloud Platforms

When broadening the perspective from an individual use case scenario to a platform landscape view, the International Data Space positions itself as an architecture to link different cloud platforms through secure exchange and trusted sharing of data, short: through data sovereignty. By proposing a specific software component, the International Data Space Connector, industrial data clouds can be connected, as well as individual enterprise clouds and on-premises applications and individual connected devices (see Figure 13).

4.1.2 IDS Reference Architecture Model

On the System Layer, the roles specified on the Business Layer are mapped onto a concrete data and service architecture in order to meet the requirements specified on the Functional Layer, resulting in what is the technical core of the International Data Space. Resulting from the requirements identified are three major technical components: Connector, Broker and App Store. How these components interact is depicted in Figure 14. The Connector, the Broker, and the App Store are supported by four additional components (which are not specific to the International Data Space):

- Identity Provider,
- Vocabulary Hub,
- Update Repository (source for updates of deployed Connectors), and
- Trust Repository (source for trustworthy software stacks and fingerprints as well as remote attestation checks).

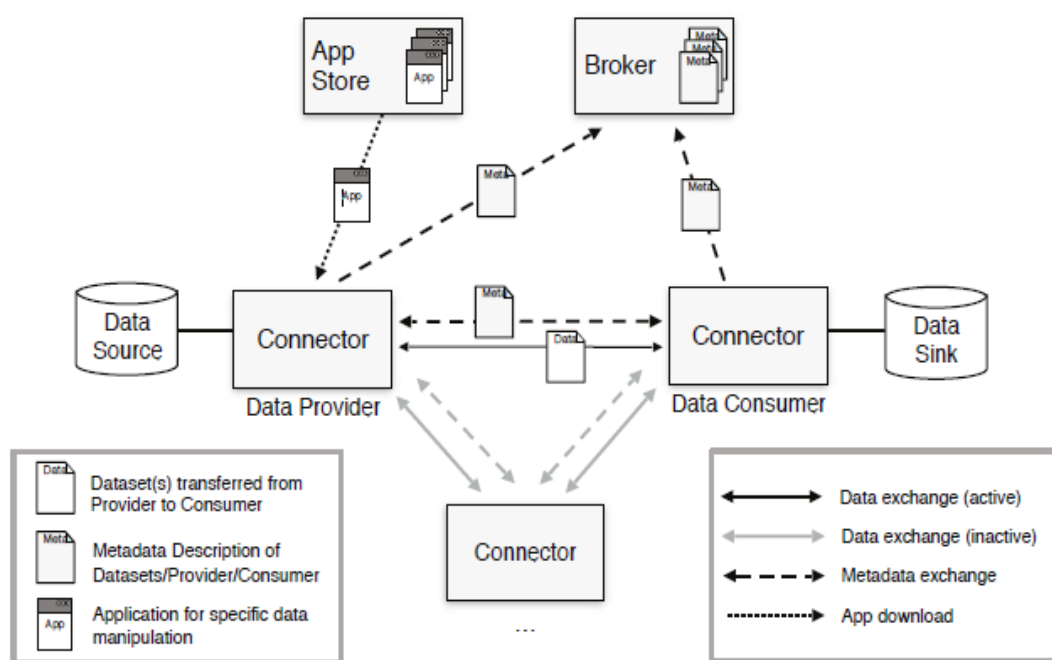


Figure 14 - Interaction of technical components

A distributed network like the International Data Space relies on the connection of different member nodes (here: Data Endpoints). The Connector is responsible for the exchange of data, as it executes the complete data exchange process. The Connector thus works at the interface between the internal data sources and enterprise systems of the participating organization and the International Data Space. It provides metadata to the Broker, including a technical interface description, an authentication mechanism, exposed data sources, and associated data usage policies. It is important to note that the data is transferred between the Connectors of the Data Provider and the Data Consumer (peer-to-peer network concept). There may be different types of implementations of the Connector, based on different technologies and depending on what specific functionality is required. Two basic versions are the Base Connector and the Trusted Connector.

Connectors can be distinguished into External Connectors and Internal Connectors. An External Connector executes the exchange of data between participants of the International Data Space. Each External Connector provides data via the Data Endpoints it exposes. The International Data Space network is constituted by the total of its External Connectors. This design avoids the need for a central data storage instance. An External Connector is typically operated behind a firewall in a specially secured network segment of a participant (so-called “Demilitarized Zone”, DMZ). From a DMZ, direct access to internal systems is not possible. It should be possible to reach an External Connector using the standard Internet Protocol (IP), and to operate it in any appropriate environment. A participant may operate multiple External Connectors (e.g., to meet load balancing or data partitioning requirements). External Connectors can be operated on-premises or in a cloud environment. An Internal Connector is typically operated in an internal company network (i.e., which is not accessible from outside). Implementations of Internal Connectors and External Connectors may be identical, as only the purpose and configuration differ. The main task of an Internal Connector is to facilitate access to internal data sources in order to provide data for External Connectors.

4.1.3 IDS Connector Architecture

The Connector Architecture uses Application Container Management technology to ensure an isolated and secure environment for individual data services. To ensure privacy of sensitive data, data processing should take place as close as possible to the data source. Any data pre-processing (e.g., filtering, anonymization, or analysis) should be performed by Internal Connectors. Only data intended for being made available to other participants should be transferred to External Connectors. Data Apps are services

encapsulating data processing and/or transformation functionality bundled as container images for simple installation by Application Container Management.

Three types of data apps can be distinguished:

- **self-developed Data Apps**, which are used by the Data Provider's own Connector (usually requiring no certification from the Certification Body),
- **third-party Data Apps**, which are retrieved from the App Store (and which may require certification), and
- **Data Apps provided by the Connector of the Data Consumer**, which allow the Data Provider to use certain functions before data is exchanged (e.g., filtering or aggregation of data) (and which may also require certification).

In addition, data apps can be divided into two more categories:

- **System Adapters** are Data Apps on the Data Provider side, establishing interfaces to external enterprise information systems. The main task of a Data App belonging to this category (in addition to wrapping the enterprise information system and perhaps transforming from an internal data model to a data model recommended or standard for a given application domain) is to add metadata to data.
- **Smart Data Apps** (or Data Sink Connectors) are Data Apps on the Data Consumer side, executing any kind of data processing, transformation, or storage functionality. Normally, the data provided from, or sent to, a Smart Data App is already annotated with metadata (as described in the Information Layer section).

Using an integrated index service, the Broker manages the data sources available in the International Data Space and supports publication and maintenance of associated metadata. Furthermore, the Broker Index Service supports the search for data sources. Both the App Store and the Broker are based on the Connector Architecture (which is described in detail in the following paragraphs). Figure 15 illustrates the internal structure of the Connector. A concrete installation of a Connector may differ from this structure, as existing components can be modified and optional components added. The components shown in Figure 15 can be assigned to two phases: Execution and Configuration.

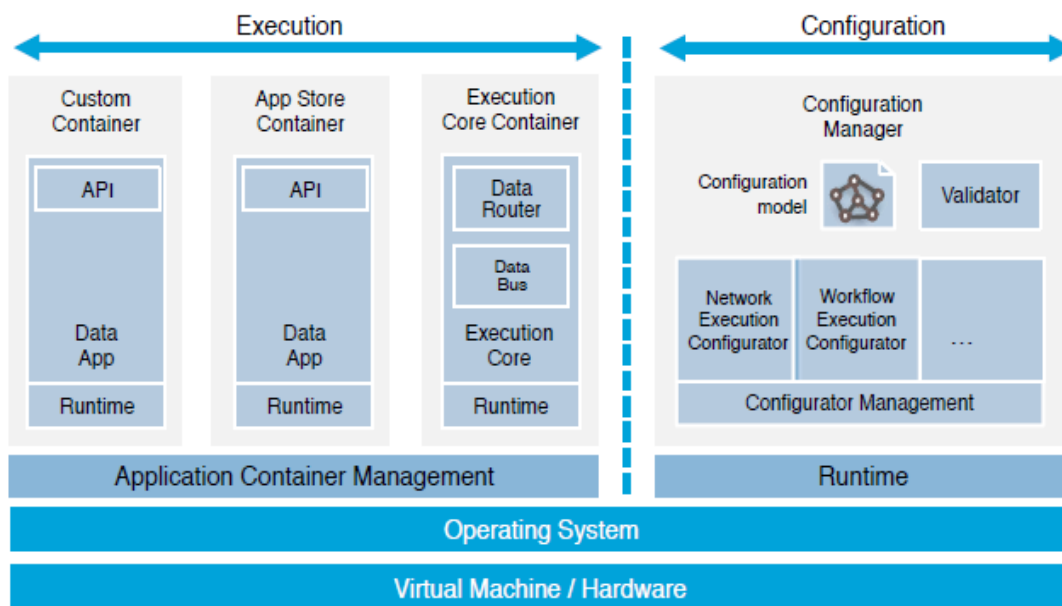


Figure 15 - IDS Reference Architecture of Connector

The execution phase of a connector involves the following components:

- **Application Container Management:** In most cases, the deployment of an Execution Core Container and selected Data Services is based on application containers. Data Services are isolated from each other by containers in order to prevent unintended interdependencies. Using Application Container Management, extended control of Data Services and containers can be enforced. During development, and in case of systems with limited resources, Application Container Management can be omitted. Difficulties in container deployment can be handled by special Execution Configurators (see below).
- **An Execution Core Container** provides components for interfacing with Data Services and supporting communication (e.g., Data Router or Data Bus to a Connector).
- **A Data Router** helps configure Data Services to be invoked according to predefined configuration parameters. In this respect, it is responsible of how data is sent (and received) to (and from) the Data Bus from (and to) Data Services. Participants have the option to replace the Data Router component by alternative implementations of various vendors. Differences in configuration can be handled by specialized Execution Configurator plug-ins. If a Connector in a limited or embedded platform consists of a single Data Service or a fixed connection configuration (e.g., on a sensor device), the Data Router can be replaced by a hard-coded software, or the Data Service can be exposed directly.

- The Data Bus exchanges data with Data Services and Data Bus components of other Connectors. It may also store data within a Connector. Usually, the Data Bus provides the method to exchange data between Connectors. Like the Data Router, the Data Bus can be replaced by alternative implementations in order to meet the requirements of the operator. The selection of an appropriate Data Bus may depend on various aspects (e.g., costs, level of support, throughput rate, quality of documentation, or availability of accessories).
- An App Store Container is a certified container downloaded from the App Store, providing a specific Data Service to the Connector.
- A Custom Container provides a self-developed Data Service. Custom containers usually require no certification.
- A Data Service defines a public API, which is invoked from a Data Router. This API is formally specified in a meta-description that is imported into the configuration model. The tasks to be executed by Data Services may vary. Data Services can be implemented in any programming language and target different runtime environments. Existing components can be reused to simplify migration from other integration platforms.
- The Runtime of a Data Service depends on the selected technology and programming language. The Runtime together with the Data Service constitutes the main part of a container. Different containers may use different runtimes. What runtimes are available depends only on the base operating system of the host computer. From the runtimes available, a service architect may select the one deemed most suitable.

The configuration phase of a connector involves the following components:

- The Configuration Manager constitutes the administrative part of a Connector. Its main task is the management and validation of the Configuration Model, followed by deployment of the Connector. Deployment is delegated to a collection of Execution Configurators by the Configurator Management.
- The Configuration Model is an extendable domain model for describing the configuration of a Connector. It consists of technology-independent, inter-connected configuration aspects.
- Configurator Management loads and manages an exchangeable set of Execution Configurators. When a Connector is deployed, the Configurator Management delegates each task to a special Execution Configurator.
- Execution Configurators are exchangeable plug-ins which execute or translate single aspects of the Configuration Model to a specific technology. The procedure

of executing a configuration depends on the technology used. Common examples would be the generation of configuration files or the usage of a configuration API. Using different Execution Configurators, it is possible to adopt new or alternative technologies and integrate them into a Connector.

- The Validator checks if the Configuration Model complies with self-defined rules and with general rules specified by the International Data Space, respectively. Violation of rules can be treated as warnings or errors. If such warnings or errors occur, deployment may fail or be rejected.

As the Configuration phase and the Execution phase are separated from each other, it is possible to develop, and later on operate, these components independently of each other. Different Connector implementations may use various kinds of communication and encryption technologies, depending on the requirements given.

4.1.3.1 Configuration Model

The Configuration Model describes the configuration of a Connector, which is exported during deployment. This description is technology-independent and can be deployed to different environments (e.g., development, test, or live systems). The following aspects of the Configuration Model are translated with the help of special Execution Configurators:

- The Dataflow defines the configuration of connections established by the Data Router between the Data Services and the Data Bus (for multiple data pipelines).
- Metadata describes the data types for input and output used by different Connector components. Data Services can provide metadata descriptions, which can be imported into the Configuration Model.
- Networking means to define network parameters (ports, IPs, etc.) for being used inside the Connector as well as for connections to external Connectors.
- Service Configuration defines how configuration parameters for Data Services or other Connector components have to be set.
- Identity Management defines the Identity Provider, which is closely integrated with the Connector. To be able to connect to Identity Providers, Data Services may need additional libraries.
- Publishing defines which Dataflows or Data Services are provided to external participants. This information is submitted to Brokers.
- The Lifecycle summarizes information on single Dataflows and Data Services. In addition to the lifecycle information of the Connector, information on the service configuration is stored here.

- For Accounting of the data exchange between participants, it is necessary to record additional information, such as contract specifications, pricing models, or billing details.
- Clearing describes which Clearing House should be informed regarding a certain data transaction.
- Compliance Rules can be specified to be checked by the Validator before Connector deployment. If warnings or errors occur, deployment may be cancelled.
- The Security settings contain information about e.g. which SSL certificates should be used for connections or which public key infrastructure should be used.

4.1.3.2 *Special Connector Implementations*

What type of Connector is to be implemented may depend on various aspects, such as the execution environment given or the current developmental stage regarding Data Services or Dataflows used. In the following, three exemplary scenarios are outlined:

- **DEVELOPER CONNECTOR** As is the case for the development of any software, developing Data Services or configuring Dataflows comprises several phases (specification, implementation, debugging, testing, profiling, etc.). For reasons of simplification, it may be useful to run Connectors without Application Container Management. In doing so, the development process can be accelerated, as packing and starting the container can be omitted, and debugging can be done in the development environment. After successfully passing all tests, the configuration model used for the developer Connector can be used to deploy a productive (live) Connector. Upon deployment in the live environment, the Connector is ready for being used.
- **MOBILE CONNECTOR** Mobile operating systems (e.g., Android, iOS, or Windows Mobile) use platforms with limited hardware resources. In such environments, Application Container Management is not necessarily required. The same applies for operating systems which do not support application containers (e.g., Windows). In such environments, Data Services (and the execution core) can be started directly on the host system, without requiring any virtualization. The differences between Connectors with containers and Connectors without containers can be met by different Execution Configurator modules.
- **EMBEDDED CONNECTOR** Another way of Connector miniaturization offers the Embedded Connector. Embedded Connectors have the same design as mobile Connectors, and do not necessarily require Application Container Management either. However, unlike mobile or developer Connectors, the Configuration Manager is not part of the Connector hardware platform here, which is why remote

configuration capabilities of the platform are required (e.g., using an API or configuration files).

Additional steps for Connector miniaturization may include the use of a common runtime for all components, or simplified versions of the Data Router and the Data Bus. If data is to be sent to a fixed recipient only, a simple Data Bus client library may be sufficient. Similarly, it may be sufficient to hard-code a single, fixed connection to the Data Bus instead of using a configurable component. To save communication overhead, simple API calls inside the common runtime could be used.

4.2 DATV certification methodology

DATV Digital Automation Technology Validation methodology is centred in AUTOWARE²³ Reference Architecture (RA) and aligned with main open HW and SW Platforms groups and initiatives in Digital Automation area for Industry 4.0 as can be seen in Figure 16. AUTOWARE has defined a complete open Framework including a novel modular, scalable and responsive Reference Architecture for the factory automation, defining methods and models for the synchronization of the digital and real world based on standards and certified components. AUTOWARE Reference Architecture aligns several cognitive manufacturing technical enablers, which are complemented by usability enablers, making it easy to access and operate by manufacturing SMEs. The third key element is the DATV certification framework for the fast integration and customization of digital automation solutions.



Figure 16 - Industry 4.0 main HW and SW platforms in Digital Automation

AUTOWARE RA targets all relevant layers for the modelling of CPPS automation solutions: Enterprise, Factory, Workcell and Field devices. To uphold the concept of Industry 4.0 and to move from the old-fashioned automation pyramid, the communication pillar enables

²³ H2020 AUTOWARE project website. Available at: <http://www.autoware-eu.org/>

direct communication between the different layers by using Fog/Cloud concepts. Finally, the Software Defined Autonomous Service Platform (SDA-SP) broadens the overall AUTOWARE RA with the mapping of main technologies and CPPS services (see Figure 17).

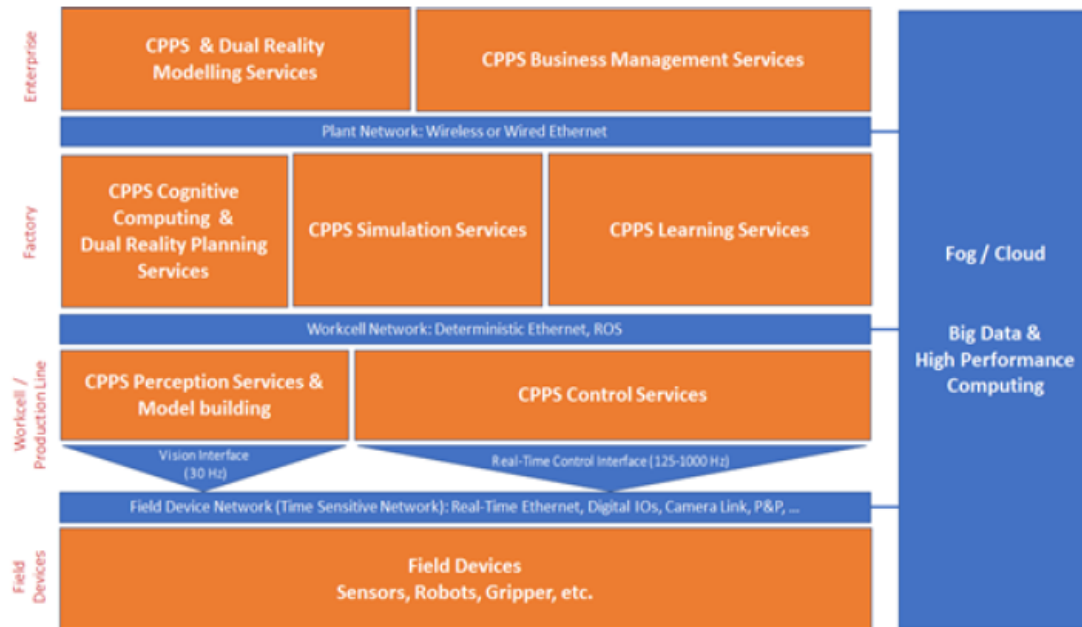


Figure 17 - AUTOWARE Reference Architecture & SDA-SP

In order to ease the digital transformation process for manufacturing companies, the digital automation technology-based solutions are defined based in the following structure:

- Technological components (from well-known technology providers and aligned to open HW, SW and platforms)
- Core Products (architectural, functional, non-functional, normative and S&S compliant, validated for a purpose VPP)
- Certified solutions (safety compliant: certified Components and Core Products validated for a specific application/service)
- Validated deployments, developed by trained professional integrators, for SME's customised automation solutions

Where the Digital Automation Technology Validation (DATV) framework aims for the different technologies, components, tools and services validation for a specific use under certain conditions, normatives, standards... based on the use of AUTOWARE V&V enablers. This approach offers both top-down and bottom-up vision to safely implement and certificate digital transformation strategies and secure I4.0 digital automation systems in manufacturing area and smart factories.

4.2.1 DATV Certification Framework

The planning and control of production systems has become increasingly complex regarding flexibility and productivity as well as regarding the decreasing predictability of processes. Thus, validation and certification processes offer easy adoption, secure environment and greater credibility to smart factories. It is well accepted that every production system should pursue the following three main objectives:

- Providing capability for rapid responsiveness.
- Enhancement of product quality.
- Production at low cost.

These requirements can be satisfied through highly stable and repeatable processes. However, they can also be achieved by creating short response times to deviations in the production system, the production process or the configuration of the product in coherence to overall performance targets. In order to obtain short response times, a high process transparency and the reliable provisioning of the required information to the point of need at the correct time and without human intervention is essential. As a result, variable and adaptable systems are needed resulting in complex, long and expensive engineering processes. Although Big Data driven solutions are defined to correctly work under several environment conditions, in practice, it is enough if it properly works under specific conditions. In this context, certification processes help to guarantee the correct operation under certain conditions easing the engineering process for smart factories that want to include Big Data driven solutions in their businesses.

In addition, certification can increase the credibility and visibility of Big Data driven solutions, as it guarantees its correct operation even following specific standards. If a Big Data driven solution is certified to follow some international or European standards or regulation, it is not necessary to be certified in each country, so the integration complexity, cost and duration highly reduce. Nowadays, security and privacy are one of the major concerns for every business. Connected smart factories need to be able to quickly assess if an item provides confidence or if required security and privacy is provided. For example, a minimal required barrier may need to be set to deter, detect and respond to distribution and use of insecure interconnected items throughout Europe and beyond.

The DATV certification framework offers a complete description of the Core Products including the achieved classification in the different technology levels (visualization, security, connectivity, open standards...), its set of components, main features, RA mapping,

component providers, estimated investment cost & deployment time table depending on complexity level.

DATV compliant components are the base for the development of Core Products designed and validated for a purpose (e.g. predictive maintenance, zero-defect manufacturing, energy efficient manufacturing...). Each Core Product (as shown in Figure 18) should be composed by a set of DATV compliant components with their matching datasheets (features and performance specifications), configuration & programming profiles and validation for purpose profiles (VPP), as guidance to ensure DATV validation when integrated in future solutions.

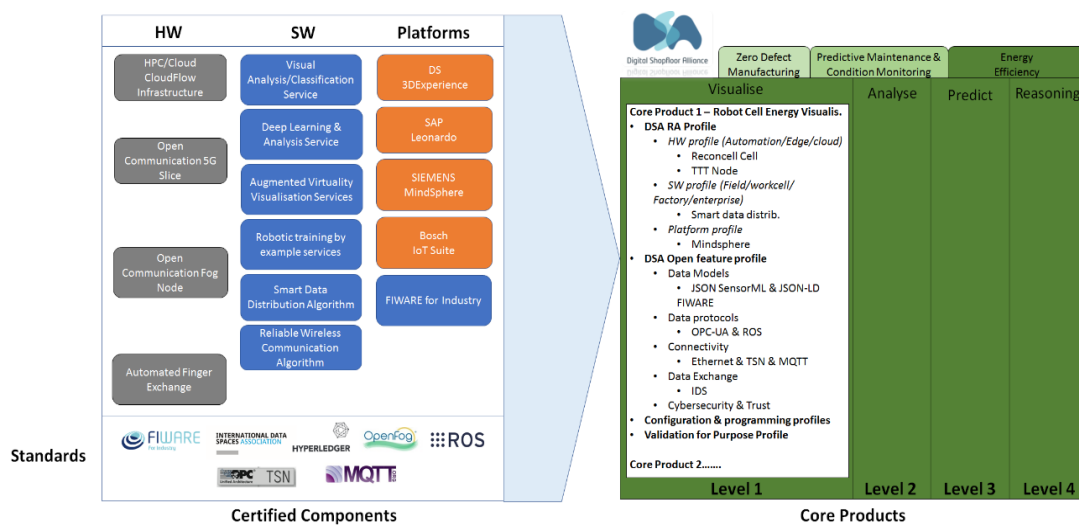


Figure 18 - Integrated approach for DATV Certification

Individual components should support relevant open standards, APIs and specifications to become DATV compliant. However, DATV does not promote the simple certification of individual components but moreover the availability of **core products** (HW infrastructure and software services and digital platforms) that are constructed following the RA architecture; **built for a purpose** (visualisation, analysis, prediction, reasoning) **in the context of specific digital services** (energy efficiency, zero defect manufacturing, predictive maintenance...) **for manufacturing lines** (collaborative workspaces, robots, reconfigurable cells, modular manufacturing) as can be seen in Figure 19.

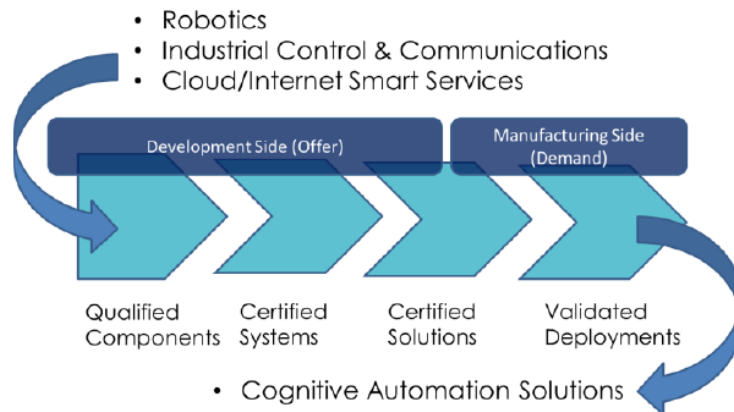


Figure 19 - Integrated approach for DATV Certification

The DATV approach will reduce considerably the integration and customization costs of validated deployments, maximising Industry 4.0 ROI and ensuring future scalability of digital shopfloor in smart factories. Industry 4.0 migration path (shown in Figure 20) guides smart factories strategy in order to leverage their automation solutions visibility, analytic, predictability and autonomy.

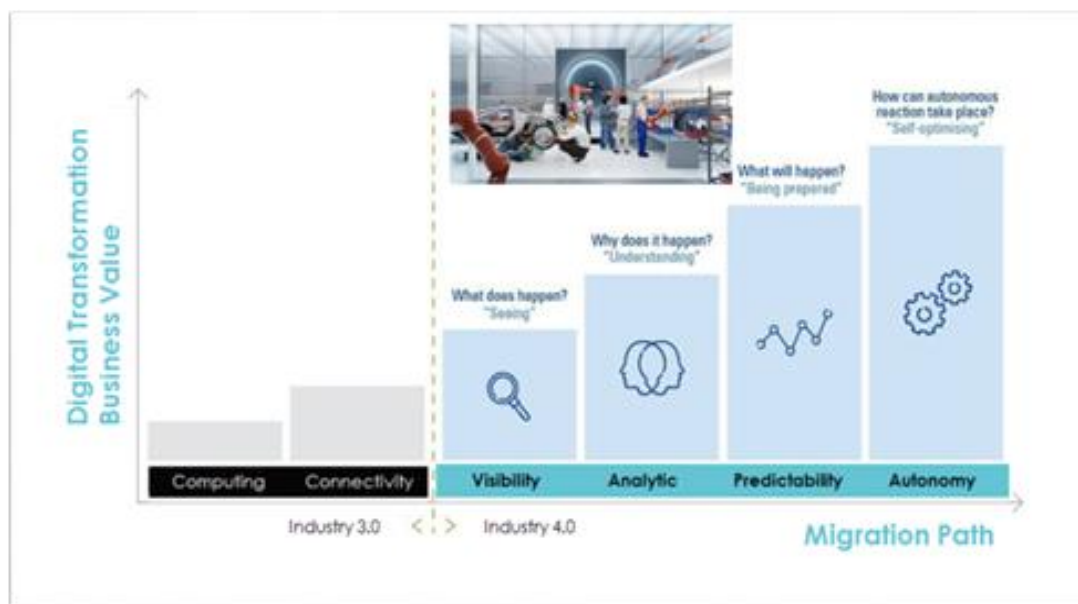


Figure 20 - Industry 4.0 migration path

4.3 IDS certification framework

Data security and data sovereignty are the fundamental characteristics of the International Data Space. Data sovereignty is a natural person's or legal entity's capability of exclusive self-determination with regard to their data goods. Participants within the

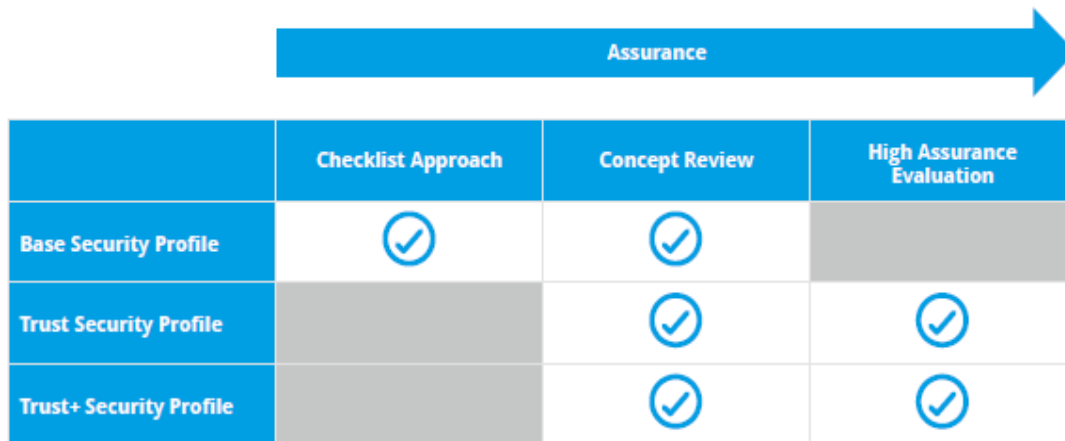
International Data Space must therefore use certified software (e.g., IDS Connector) in order to securely exchange data in a sovereign way. Furthermore, data is only exchanged if the exchange takes place between trustworthy and certified participants.

The International Data Space certification scheme encompasses all processes, rules and standards governing the certification of participants and core components within the International Data Space. Being the point of access to the International Data Space, the Connector provides a controlled environment for processing and exchanging data, ensuring secure data exchange between the Data Provider and the Data Consumer. Trust in the correct and complete implementation of the functionality required by the Reference Architecture Model can only be ensured by independent evaluation and certification of Connectors from an approved Evaluation Facility and the Certification Body of the International Data Space.

Participants and core components shall provide a sufficiently high degree of security regarding the integrity, confidentiality and availability of information exchanged in the International Data Space. Therefore, an evaluation and certification of the core components as well as of the technical and organizational security measures is mandatory for participating in the International Data Space. This requirement for compliance necessitates the definition of a framework in order to ensure a consistent and comparable evaluation and certification process amongst all International Data Space participants and core components. Hence, a certification scheme has been defined following best practices from other internationally accredited certifications.

4.3.1 IDS Connector Certification Framework

To secure the intended cross-industrial and cross-company information exchange, IDS core components must provide the required functionality and an appropriate level of security. As such, the core component certification is interoperability- and security-focused, while aiming to strengthen the development and maintenance process of these components. A matrix certification approach as shown in Figure 21 was defined for the core components of the International Data Space. This ensures on the one hand a low entry barrier and on the other hand a scalable certification to meet high information security requirements.



	Checklist Approach	Concept Review	High Assurance Evaluation
Base Security Profile	✓	✓	
Trust Security Profile		✓	✓
Trust+ Security Profile		✓	✓

Figure 21 – DATV Certification Approach for core components of the IDS

4.3.1.1 Assurance Levels

The depth and rigor of an evaluation consists of the following three assurance levels as defined by the IDS certification scheme:

- **Checklist Approach** The core component must fulfil security features (security requirements, security properties, security functions) as defined by the corresponding checklist. The vendor of the component validates the claims made about the implementation. Additionally, an automated test suite will be used to verify the component's security features.
- **Concept Review** Instead of the checklist approach, an in-depth review by an International Data Space evaluation facility is necessary. The review includes an evaluation of the provided concept as well as practical functional and security tests.
- **High Assurance Evaluation** For the third level, in addition to the functional and security tests, the vendor must provide the source code of all security relevant components and an in-depth source code review will be performed by an evaluation facility. Furthermore, the development process will be evaluated, including an audit of the development site.

4.3.1.2 Security Profiles

Whenever two components establish a communication channel, it's up to them to decide which information they will send to the communication partner. Therefore, the identity and certification level (for both the participant and the component) must be provided by each component in the form of a digital certificate containing this information. As with the participant certification, this approach enables the data owner and data consumer to specify the security profile required for the core components used during data exchange.

For this purpose, the International Data Space certification scheme defines three security profiles for the core components.

- **Base Security Profile** This profile includes basic security requirements: limited isolation of software components, secure communication including encryption and integrity protection, mutual authentication between components, as well as basic access control and logging. However, neither the protection of security related data (key material, certificates) nor trust verification are required. Persistent data is not encrypted and integrity protection for containers is not provided. This security profile is therefore meant for communication inside of a single security domain.
- **Trust Security Profile** This profile includes strict isolation of software components (apps/services), secure storage of cryptographic keys in an isolated environment, secure communication including encryption, authentication and integrity protection, access and resource control, usage control and trusted update mechanisms. All data stored on persistent media or transmitted via networks must be encrypted.
- **Trust+ Security Profile** This profile requires hardware-based trust anchors (in the form of a TPM or a hardware-backed isolation environment) and supports remote integrity verification (i.e., remote attestation). All key material is stored in dedicated hardware isolated areas.

4.3.1.3 *Certification Criteria Catalogue*

The catalogue of certification criteria for the IDS core components [CRIT-C] was defined as part of the Fraunhofer research project »Industrial Data Space« and fine-tuned with the members of the WG Certification. The catalogue is split into three thematic sections, i.e. IDS-specific requirements, functional requirements taken from the industry standard ISA/IEC 62443-4-2 [62443-4-2] and best practice requirements for secure software development. Each criteria section targets a set of evaluation goals:

- The IDS-specific requirements aim to evaluate the Core Component's conformity to the IDS Reference Architecture Model, both in regard to functionality (e.g. support of the IDS information model) as well as security (e.g. conformance to the IDS security architecture).
- The requirements taken from ISA/IEC 62443-4-2 target the implemented functionality and security measures in relation to industry-wide accepted requirements for industrial automation and control systems, e.g. the capability to obscure feedback of authentication information during the authentication process.

- To round off the catalogue, the best practice requirements for secure software development aim to evaluate the security of the processes during the development of the component, e.g. design documentation, physical security measures and test processes.

To reduce the financial entry barrier not only for IDS participants but also for the developers of core components, the component certification approach is designed to use existing certification schemes whenever reasonable. Where such certification schemes do not exist or are not widely recognized, e.g., for IDS-specific aspects, criteria defined within the International Data Space certification scheme will be employed.

The functional and security requirements of the core components to be evaluated will be defined based on the IDS Reference Architecture Model, specific component specifications like the Connector Specification as well as widely recognized requirement catalogues like ISA/IEC 62443-4-2 (e.g. for functional requirements such as data confidentiality and system integrity).

The evaluation at the various assurance levels can also be supported and facilitated by requiring appropriate measures used throughout the lifecycle of the component as defined in ISA/IEC 62443-4-2, such as using the approach for thorough elicitation of the Security Requirements, enforcing those Security Requirements at the Architecture level (e.g., using Security-by-Design) and tracing them to the Secure Implementation level, supported by relevant Guidance Documents, Verification & Validation approaches, as well as a Secure Defect Management & Secure Update Management.²⁴

²⁴ US Cert: Build Security In: Modeling Tools, 2013

5 The use of standards in the Boost 4.0 Pilots

Information was sought from the Boost 4.0 pilot projects via a template chapter in the trial handbooks that focused on standardisation. This template attempted to provide the background and context for questions on the following areas:

- Standardisation
 - What standards are important for each pilot?
 - Which part of the Boost 4.0 architecture do these apply to?
 - Which Boost 4.0 partners are involved with or knowledgeable about SDOs?
 - What are the known standardisation gaps for each pilot?
- Testing and Certification
 - Which Boost 4.0 partners have relevant experience they can share with the project?
- Data Management and Governance
 - What are the requirements for each pilot and how are these to be addressed?
- Security, Safety, Privacy, Trust and Resilience
 - What are the requirements for each pilot and how are these to be addressed?

5.1 Gestamp

Gestamp general pilot is aimed at developing new initiatives focused on the improvement of the overall efficiency of the plants. In spirit of this, Gestamp business scenarios focus on two of the main pillars, logistics and quality in order to test different solutions to optimize the logistics of the plant through the localization of assets and the inspection and quality control workflow, improving product lifecycle management. A new high-density and virtualised metrology approach will be developed to ensure traceability, interoperability and to gain knowledge and understanding about products and processes. In this context, this data will be accessible so it will be able to be used to prevent major failures, avoid defects and waste of material.

At the highest level, the overall objective of the trial will be to introduce new protocols in how factory data is used. Once the trial has been completed successfully the factory will have a fully integrated data collection and analysis system which provides information in real time to decision makers.

According to Boost4.0 Reference Architecture, Gestamp pilot has identified main layers and components that are essential for the correct implementation of the pilot.

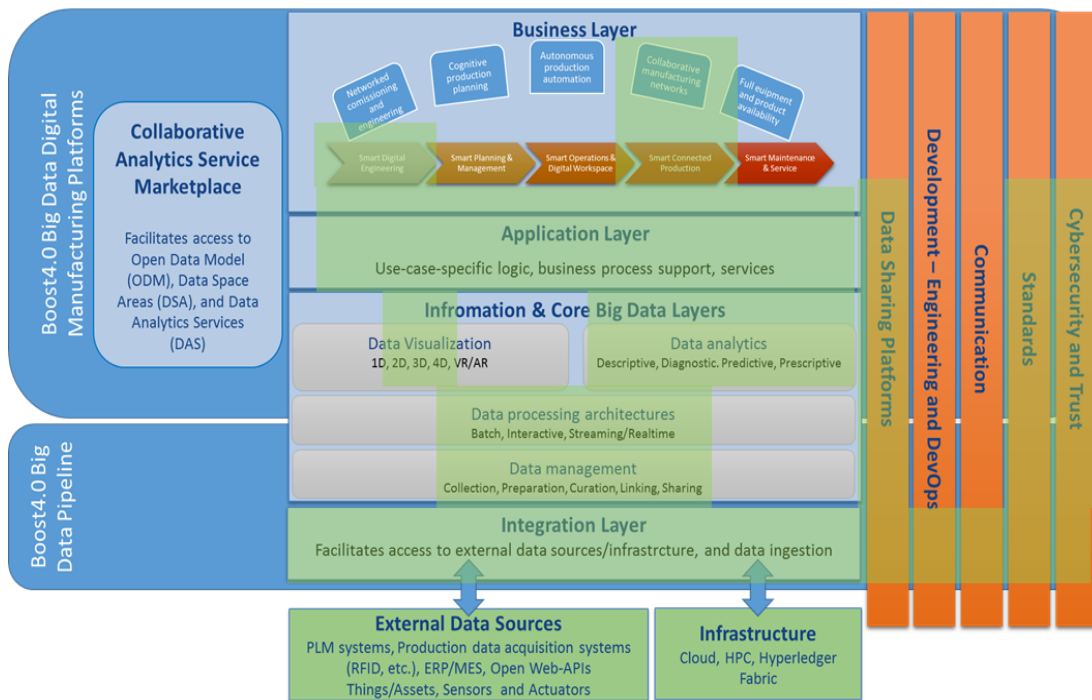


Figure 22 – Gestamp Big Data Pilot mapped in Boost4.0 architecture

In this context, Zero Defect Manufacturing powered by massive metrology is one of the objectives of Gestamp pilot. The aim of this business scenario is to implement a high-density metrology workflow and improved massive point clouds visualization in order to generate benefit in terms of quality efficiency and optimization. To gain knowledge about products and production process, to optimize quality control management and to implement predictive actions, it is vital to enhance the product lifecycle to ensure interoperability between the different actors and steps, and guarantee traceability of data and products.

It is important to define the data models and tools for cognitive storage structures and semantically mediated data environments which will facilitate the interoperability. Considering ZDM massive metrology, some standards have been selected in order to ensure interoperability and traceability of products and quality data. These standards are: **Quality Information Framework (QIF) 2.1 (Dimensional Metrology)**, **STEP (data model)** and **STL (data model)**. All of them are related to data processing, data management and data visualization.

Manufacturing quality systems can be generally categorized into five sub-systems, namely product definition, measurement process planning, creation of the inspection

system, measurement process execution, and measurement results reporting. Several past standards efforts addressing manufacturing quality data interoperability have focused only on pieces of the total manufacturing quality system.²⁵

In the **product definition process**, the most accepted standards are: The Initial Graphics Exchange Specification (IGES)²⁶ and the Standard for the Exchange of Product model data (STEP)²⁷. Within STEP, various Application Protocols (APs) were developed to describe product data for different sections of manufacturing processes. STEP AP 203²⁸ (ISO 2007a) models 3D product design information. The first edition of STEP AP 203 does not have sufficient geometric dimensioning and tolerancing (GD&T) information to support automated processing of information by downstream quality processes. A newer version of STEP AP 203 – AP 203 edition 2²⁹, which models both annotated and semantic GD&T information in 3D product design is now available to address this. The GD&T definition from AP 214³⁰ was harmonized with AP 203 edition 2. These GD&T definitions are mainly for annotation purposes; therefore, they are not sufficient for automatic generation of dimensional measurement process plans.

DMIS is the only standard that combines measurement features and operation instruction information within the same measurement process definition. It is a language for **controlling dimensional measuring equipment** and includes an input and an output language. Part of the DMIS input language defines features, tolerances, sensors, etc. The output language serves both as a log of action commands and settings and a report of results, with actual and nominal point data, features, and tolerances. However, it does not define complete measuring equipment resources. Measuring equipment resource data is necessary to complete the effectiveness of DMIS.

²⁵ Psymbiosys Project

²⁶ IGES 1980. Initial Graphics Exchange Specification (IGES <http://ts.nist.gov/standards/iges/>)

²⁷ ISO 10303-1: Industrial automation systems and integration - Product data representation and exchange - Part 1: Industrial Automation System and Integration - Product Data Representation and Exchange Part 1: Overview and Fundamental Principles.

²⁸ ISO 2007a. ISO 10303-203: Industrial automation systems and integration - Product data representation and exchange - Part 203: Application Protocols: Configuration controlled 3D design.

²⁹ ISO 10303-203:2009: Industrial automation systems and integration - Product data representation and exchange - Part 203: Application protocol: Configuration controlled 3D design of mechanical parts and assemblies.

³⁰ ISO 10303-214: Industrial automation systems and integration - Product data representation and exchange - Part 214: Application Protocol: Core data for automotive mechanical design processes.

STEP AP 219³¹ specifies an application protocol for the **exchange of information** resulting from the dimensional inspection of solid parts, which includes administering, planning, and executing dimensional inspection as well as analysing and archiving the results. AP 219 is inadequate in providing complete definitions of dimensional measurement features, dimensional measurement results collections, and analysis methods. There are many entities in AP 219 that were left empty for further development.

As for the interface between **measurement process execution and equipment control**, there are two publicly available specifications/standards, one of which is formalized as an official ANSI and ISO standard – the equipment module of DMIS Part 2³² (ANSI 2003). The other is the I++DME Interface Specification (I++DME, 2005) which is a specification for dimensional measuring equipment information exchange developed by several European automakers and measuring equipment vendors. There are no known product implementations of DMIS Part 2. There are many software implementations of I++DME worldwide, but it is not yet ubiquitous for either coordinate measuring machine (CMM) software or CMM systems to offer I++DME in their published product offerings. Several vendors have I++DME simulators available to enable quick and accurate development of I++DME implementations within measurement process execution software.

Dimensional Mark-up Language (DML)³³, DMIS Output, AP 219, and Quality Measurement Data (QMD)³⁴ are **specifications/standards for measurement results reporting**. DML is having moderate usage largely in North America. A format for CMM measurement results is defined within DMIS, and has enjoyed some usage, wherever DMIS is used. Within the STEP effort, AP 219 was defined to cover all important metrology information, including, but not limited to, measurement results. As mentioned earlier, the latest ISO standard version of AP 219 only defines measurement results information. The QMD Data Model describes a non-proprietary and open standard XML schema for variable, attribute, and binary quality measurements, including non-dimensional measurements and gage measurements. QMD targets quality measurements from measurement devices other than CMMs. The standard is unidirectional – it defines the measurement export only. There are multiple standards and/or specifications that define traceability data such as DMIS, DML, and ISO 10303 AP 238 (ISO 2004). However, the link between traceability and measurement data is insufficient.

³¹ ISO 10303-219, Industrial automation systems and integration – Product data representation and exchange – Part 219: Application protocol: Dimensional inspection information exchange.

³² Dimensional Measuring Interface Standard Part 2: Object Interface Specification, Consortium for Advanced Manufacturing – International.

³³ DML 2009. Dimensional Markup Language (DML) <http://www.aiag.org/>

³⁴ Quality Measurement Data (QMD) <http://www.aiag.org>

One of the goals of the pilot is to demonstrate the complete metrology workflow starting from the product definition up to the advanced analysis of the metrological data, ensuring interoperability and traceability. In this sense, it is key the collaboration between different departments: engineering (product design) and the quality control.

The Quality Information Framework (QIF) standard³⁵ defines an integrated set of information models which enable the effective exchange of metrology data throughout the entire manufacturing quality measurement process – from product design to inspection planning to execution to analysis and reporting.

The goal of the QIF specification is to facilitate interoperability of manufacturing quality data between system software components. Solving the metrology interoperability problem will benefit manufacturers by avoiding wasted resources spent on non-value-added costs of translating data between the different components of manufacturing quality systems. Users should gain flexibility in configuring quality systems and in choosing commercial components, and achieve effortless and accurate flow of data within their factory walls as well as with suppliers and customers. Solution providers should be able to eliminate their efforts previously spent in data translations, and there should be increased opportunities to sell their products and to improve and expand the features of their solutions.

Figure 23 shows a high-level view of the QIF information model.

³⁵ Quality Information Framework – An Integrated Model for Manufacturing Quality Information. Part 1: Overview and fundamental Principles Version 2.0, Dimensional Metrology Standards consortium, 2014.



Figure 23 - QIF Architecture.³⁶

At the core of the QIF architecture is the reusable QIF library which contains definitions and components that are referenced by the application areas, thereby ensuring interoperability and extensibility. Around the QIF library core, there are seven QIF application area information models, MBD, Plans, Resources, Rules, DMIS, Results and Statistics.

The flow of QIF data starts with generation of CAD + PMI data exported as QIF MBD application data. Quality planning systems import the MBD and generate Plans (whats), then import Resources and Rules information and export Plans (whats and hows). Programming systems import Plans to generate DME-specific programs, or general instructions to guide inspection. Dimensional measurement equipment executes programs and evaluates characteristics of a single manufactured part or assembly and export the measurements as Results. Analysis systems, typically performing statistical process control, import single parts Results and generate analysis of multiple part batches as QIF Statistics data.

Users of the QIF information model are not required to implement the entire model. Any of the application models may be used singly for exchange of quality data between software systems. Further, other data models and exchange formats can coexist in an enterprise with QIF data.

³⁶ J. Herron, C. Brown, D Campbell, QIF: Quality Information Framework, Model-Based Enterprise Summit 2018, April 2018.

Figure 24 shows a model-based quality workflow activity diagram and the use of QIF formats to convey information between computer-aided quality processes. It includes all manufacturing design and quality information required to assess product quality, and also to improve manufacturing processes and product design. The work flow model shows five major activities of a quality metrology enterprise:

- Define Product
- Determine Measurement Requirements
- Define Measurement Process
- Execute Measurement Process
- Analyse & Report Quality Data

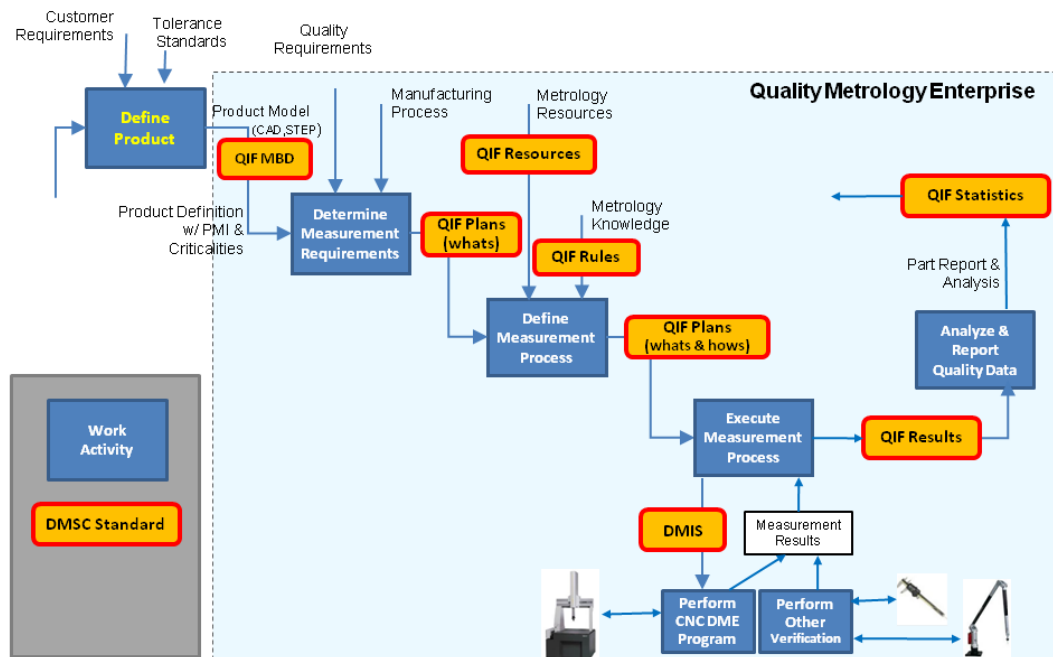


Figure 24 - QIF Model-Based Quality Workflow

Activities may export and/or import quality information that can be formatted according to the QIF information model and XML encoding rules. The diagram does not show activities that generate manufacturing process information or that implement a manufacturing execution system. QIF information is conveyed in a product-neutral format and is modularized into six application areas plus the QIF Library. These features of QIF facilitate efficient flow of enterprise quality data in a way that does not specify or constrain a user's system architecture.

The Define Product activity generates a model-based definition of a part that can support an enterprises digital product verification. The product definition contains geometry information plus semantically linked product manufacturing information (PMI). PMI commonly includes geometric dimensioning and tolerancing (GD&T) information, key characteristic criticalities, material, surface texture, roughness, colour and hardness. GD&T includes associations between geometric elements of the product and dimensions, tolerances, and datums. The product definition is expressed using the QIF MBD information model.

The Determine Measurement Requirements activity imports QIF MBD information or its equivalent expressed in another format or formats. Based upon enterprise quality requirements and/or manufacturing process knowledge, measurement requirements for a part are generated as a set of measurement criteria also known as a bill of characteristic instances (BOCI). A characteristic instance is typically a tolerance or specification applied to a feature or product that needs verification. The plan may also specify the measuring sequence and resources to be used, but is not required to do so. This BOCI constitutes a high level quality plan of “what” needs to be inspected or verified, expressed as a QIF Plans (whats) information packet.

The activity that generates QIF Resources information is not shown, but can be called Define Measurement Resources. This activity defines enterprise hardware and software resources available, either serial-number specific, or generic, that can be harnessed to meet inspection requirements for individual part features. The QIF Resources data format can be used to specify resources required in an inspection plan, or the resources actually used in an inspection. The scope of the QIF Resources application model includes:

- definition of resources both hardware and software,
- DMEs,
- application software,
- fixtures,
- go-no-go gages,
- manual instruments.

The activity that generates QIF Rules is not shown, but can be called Define Measurement Rules. This activity generates inspection practices required by an enterprise to be used in-house or by contractors. QIF Rules data defines, for each possible feature type on a part,

the information elements required to fully specify and constrain the measurement on that feature type. The information elements include things like measurement point density, measurement point pattern, and feature fitting algorithm. The QIF standard defines the generic format to express enterprise Rules, but does not contain specific rules. Information areas that are in scope include:

- product measurement point number or density and sampling method for each product feature type
- feature fitting algorithm for each feature type
- rule ID and corporate ownership.

The Define Measurement Process activity inputs resource and metrology knowledge, and the QIF Plans (what), and generates additional instructions on “how” to inspect or verify the bill of characteristic instances. The completed inspection plan is output as QIF Plans (whats and hows). The scope of the QIF Plans 2.0 application model includes:

- dimensional product information, e.g., geometric features, measurement features, nominal dimensions, measurement features, and tolerance values
- non-dimensional product information, e.g., product IDs, customer information, key contact, temperature, and roughness
- product characteristics
- traceability values and pointers
- work instructions
- CAD entity relationships.

The downstream activity Execute Measurement Process activity imports the QIF Plan, and if needed generates a detailed resource specific inspection program. The programs are machine-level measurement programs, formatted according to DMIS or some other measurement programming language, that provide equipment level commands to specific coordinate measuring machine (CMM) control units, to collect point data, fit features to data, and output feature and characteristic data. The workflow shows the export of non-QIF format subsequently translated according to the QIF Results information model. Measurement processes that adopt QIF will likely export results directly without translation. Measurement Execution can also include software solutions that issue instructions to human operators using callipers, go/no-go gauges, and specialized

inspection equipment, and generate results data. Actual measurement values may be numerical or non-numerical. Measurement results may include not only raw measurement values, but also summary statistical or derived results (e.g., cylinder radius with standard deviation). Measurement results may also include description of the algorithmic means (e.g., least squares) by which the derived results are calculated. All necessary nominal (as designed) target values may also be included to allow reanalysis. Any other information relevant to the measurements is also in scope. This includes information called inspection traceability, which includes the shift, the equipment operators name, a description of the item measured, the date and time of the measurement, etc.

Finally, the measurement results for two or more parts are collected, analysed, and reported by the activity Analyse & Report Quality Data. The output, expressed using the QIF Statistics model, is generally an analysis of a multi-part batch. QIF Statistics is designed to carry information to transport statistical quality control plans, corrective action plans and detailed summary quality statistics. It builds on the QIF Results framework through supporting multi-part measurement results that can apply to a number of quality study types beyond single or first article inspection. It is designed to haul information in an unambiguous form for pre-production, capability, and production quality studies. In addition, it supports the full extent of measurement systems analysis studies including Gage R&R.

Quality information generated in QIF format can be used as input by many other quality and manufacturing management components not shown in Figure 2, including, but not limited to, first article inspection plan and report generation, statistical process control (SPC), materials resource planning (MRP), measurement systems analysis (MSA), manufacturing execution systems (MES), and computer aided manufacturing (CAM).

The digital interface between Execute Measurement Process and the DME (dimensional measurement equipment) has been satisfied by the Dimensional Measuring Interface Standard (DMIS), ANSI/DMIS 105.2 Part 1-2009. DMIS can also be used as a numerical control part program for DMEs such as coordinate measuring machine (CMM).

QIF 2.1 enhances the previous ANSI Standard, QIF V2.0 containing quality planning and measurement results, by providing a complete and accurate 3D product definition with semantic geometric and dimensional tolerances, definitions for measurement resources, template for measurement rules, and statistical functionality. All of this to satisfy the digital interoperability needs for a wide variety of use cases including feature-based dimensional metrology, quality measurement planning, first article inspection, and discrete quality measurement.

Network Traffic Monitoring and Analysis: Although there is no standard approach to monitoring and analysing network traffic all, any solution devised to work with existing infrastructure must work with existing industry standards. The GESTAMP pilot will need to use hardware which is capable of providing network activity information using industry recognized protocols. The most common protocols and those which will be used in during the trial include the following:

- **NetFlow** – This feature was first introduced on Cisco routers in the 1990s. It was designed to provide the ability to collect IP network traffic as it enters or exits an interface. NetFlow data analysis can provide network administrators information on traffic sources and destinations, service classes, and possible causes of network congestion, among others. At the most basic level, a flow monitoring setup using NetFlow is comprised of three primary components:
 - **Exporter:** This component is responsible for aggregating packets into flows and exporting them to the collectors.
 - **Collector:** This element is responsible for the reception, storage and pre-processing of flow data received from a flow exporter.
 - **Analysis application:** The application analyses received flow data to provide the necessary information to the network administrator. In the most typical use cases this consisting of intrusion detection or traffic profiling.
- **IPFIX** – The Internet Protocol Flow Information Export (IPFIX) is a network flow standard which is currently run by the Internet Engineering Task Force (IETF). The goal behind the IPFIX protocol was the creation of a common, universal standard of export for flow information from routers, switches, firewalls, and other infrastructure devices. In practice, IPFIX provides definitions on how flow information should be formatted and transferred from an exporter to a collector. At the standardization level, IPFIX is documented in IETF's RFC 7011 through RFC 7015 as well as RFC 5103³⁷. Cisco NetFlow Version 9 is the basis and main point of reference for IPFIX. IPFIX changes some of the terminologies of NetFlow, but in essence they are the same principles of NetFlow v9. The main advantages of IPFIX versus NetFlow are IPFIX's ability to integrate information that would normally be sent to [Syslog](#) or SNMP information directly in the IPFIX packet, thus eliminating the need for these additional services collecting data from each network device. Furthermore, IPFIX allows fields that are "Variable" length, which means that there is no fixed length an ID has to conform to while NetFlow does not.

³⁷ RFCs (Request for Comments) are proposals for internet standards presented to the IETF.

- **Sflow** - This sampling technology is embedded within switches and routers provided by multiple vendors. sFlow enables the continuous monitoring of application level traffic flows at wire speed on all interfaces simultaneously. The sFlow Agent is a software process that runs as part of the network management software within a device. It combines interface counters and flow samples into sFlow datagrams that are sent across the network to a sFlow Collector. Packet sampling is typically performed by the switching/routing ASICs, providing wire-speed performance. The state of the forwarding/routing table entries associated with each sampled packet is also recorded. The sFlow Agent does very little processing. It simply packages data into sFlow Datagrams that are immediately sent on the network. Immediate forwarding of data minimizes memory and CPU requirements associated with the sFlow Agent.

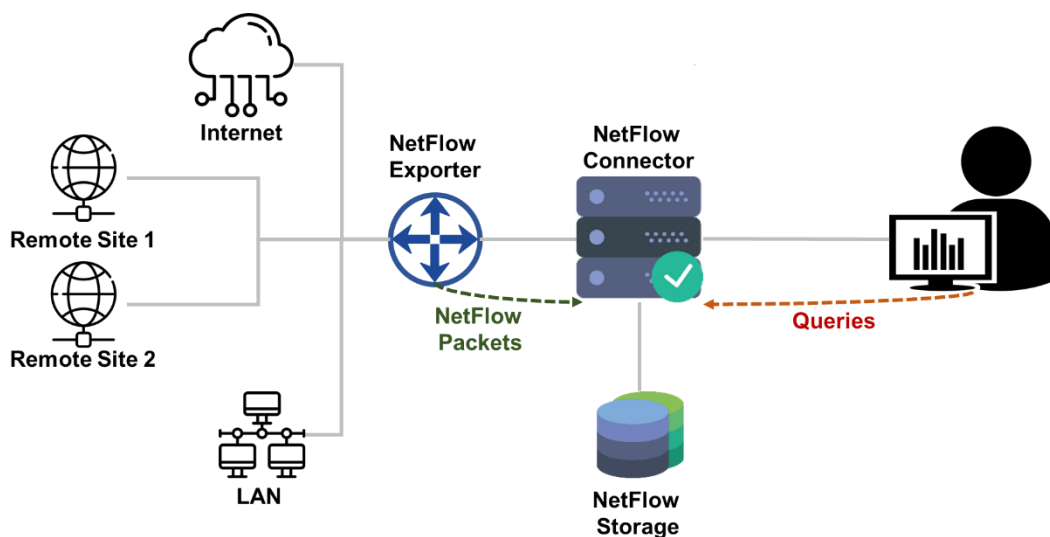


Figure 25 - Simplified NetFlow Architecture

Infrastructure monitoring: The most widely used standard for collecting and organizing information on managed devices on IP networks is SNMPv3 (Simple Network Management Protocol). Simple Network Management Protocol version 3 (SNMPv3) is an interoperable, standards-based protocol that is defined in RFCs 3413 to 3415. This module discusses the security features provided in SNMPv3 and describes how to configure the security mechanism to handle SNMP packets. The protocol enables both infrastructure monitoring as well as the ability to modify information to change device behaviour. The most common devices which support this protocol are cable modems, routers, switches, servers, workstations and printers, among others.

SNMP is a commonly used protocol in network management and monitoring operations as it exposes management data as a set of variables on the managed systems organized in

a management information base (MIB) which describe the system status and configuration. The protocol allows for these variables to be queried and manipulated remotely through management applications.

The protocol is currently on its third major version which has greatly improved the performance, flexibility and features of previous iterations. One of the most important improvements in version 3 are the security enhancements which include the following specific features:

- **Message integrity** – This feature provides mechanisms which are able to ensure that a given packet has not been tampered with or altered during transit.
- **Authentication** – The authentication functionality is used to determine whether a message has been sent from a valid source.
- **Encryption** – This functionality encrypts the contents of packets to prevent them from being intercepted and analysed by an unauthorized source.

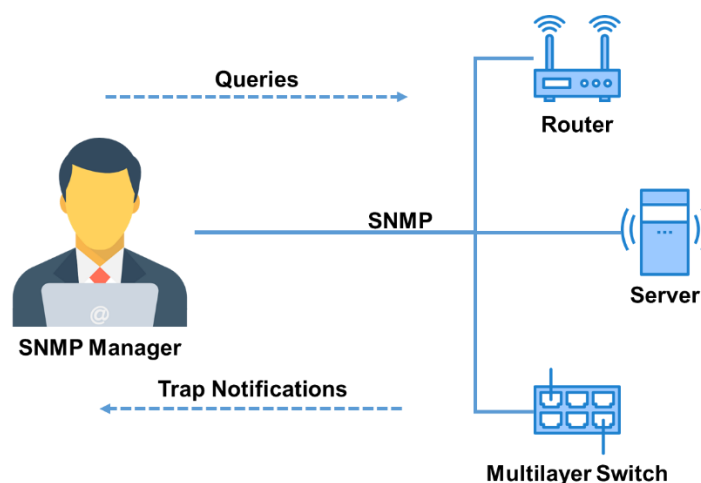


Figure 26 - Basic SNMP Communication

5.1.1 Experience with Standards Development

Innovalia Metrology and Capvidia are members of the Dimensional Metrology Standards Consortium (DMSC) –which is a non-for-profit, cooperative sponsorship organization focused on or relating to digital dimensional metrology. It is dedicated to identifying, promoting, fostering, and encouraging the development and interoperability of standards that benefit the dimensional metrology community such as QIF. It is an accredited national standard-making organization with international presence.

The use of Open Source technologies is one of the most important means by which the consortium can participate in the development of new standards. In the specific case of network traffic monitoring and analysis, the redborder platform is built on Open Source components such as Hadoop, Snort, and Druid OLAP among others. ENEO is an active member of the Open Source community of developers and is committed to furthering the development and implementation of the Open Source model. The Open Source community is at the forefront of standards development and ENEO will strive to include the results of the Boost 4.0 trials in its broader efforts to promote more effective standards.

There are also current initiatives which the consortium will be well-positioned to assist in their development of new standards, such as the European Cyber-Security Organization's Working Group 3.1 on "Cybersecurity for Industry 4.0" and OW2's Open Source cyber-range, among others. There are also a number of organizations which are continuously introducing security guidelines, rules and regulations, and standards for the security of industrial control systems including the Information Sharing and Analysis Center (ISACs), Industrial Control Systems Cyber Emergency Response Team (ICS-CERT), and National Institute of Standards and Technology (NIST), to name a few.

5.1.2 Data Management and Data Governance

Data Management and Data Governance are key points in Gestamp pilot. The use of TRIMEK Big Data M3 platform and the implementation of a complete QIF metrology workflow ensure data integration and interoperability as well as high quality data, data storage, management, modelling and other operations. Moreover, the M3 platform allows data sharing/publishing to be used by M3 modules and authorized external applications through the implementation of IDS Connectors.

Finally, data security will be addressed through the use of ENEO's redborder platform. One of the core efforts of the trial with regards to the use of the redborder platform is to provide a more efficient and actionable data management and security tool. This will be achieved by making improvements to all stages of the data acquisition, processing and analysis tasks. The overall goal is to achieve a seamless integration of heterogeneous data sources which will enable more complex analysis on the data in both real time and on historical registers as well. The overall process improvement will involve developments in each of the following areas:

- **Data capture:** A complete analysis of the data sources to be included in the integration with the analytics platform will be used to determine the type, format and specific requirements. The ability to capture data from multiple sources will be the

cornerstone of the platform and this analysis will be used to identify potential development needs to increase compatibility.

- **Data enrichment:** One of the most important advancements in terms of data management is a more useful enrichment process. With the goal of providing a more robust set of analytical tools the collected data will need to be enriched with data from additional sources including metadata. By combining multiple sources of data, a more complete analysis will be possible. To accomplish this goal the development stage will need to identify the logical associations between data sources and types, including the identification of metrics and business problems to be solved and what information is needed for each. The result will be the definition of new enrichment protocols which correlate data from multiple sources in a manner which adds value to the analytics process.
- **Indexing:** Without proper indexation the improved data management strategy will not provide the necessary increases in efficiency. The collected data will need to be given a structure in order to optimize data retrieval operations. The search keys (set of attributes which can look up the data points) and pointers (address of the data stored in memory) to be created. Although this is a standardised process in many ways, during the development stage the decision of how to approach the indexing process, including whether to use an ordered or hash index, will need to be made to optimise the process.
- **Record:** The use of historical data along with real-time analysis will represent one of the most important advancements in terms of data management. However, further optimization of the data is required to before historical data can be analysed. The structure of the record, prioritisation of data, and overall approach to data storage are aspects which will need to be considered in the development phase of the pilot. Once implemented the process will be constrained by the size of the overall historical record which is to be analysed at any given moment.
- **Service broker:** In terms of data management tools and strategies, the service broker will be a fundamental part as it will control all data queries. Its design and optimisation will permit queries to be performed on data in near real-time as well as to compare those results with historical data.
- **Data visualisation:** One of the core developments related to data management strategies and the use of novel analytical tools will be the implementation of customised dashboards. These tools will allow the technicians to observe the results of the data collection and analysis processes in real time and in a manner, which greatly facilitates interpreting and taking decisions based on the data collected. The visualization tool will be designed to provide each user with the precise information necessary at the correct moment to allow them to make more informed decisions.

5.1.3 Security, Safety, Privacy, Trust & Resilience

Security and data protection are the core goal of ENEO's redborder platform which includes the following specific modules which will be deployed during the trial and are aimed at providing industry-leading real-time NTA and cybersecurity:

- **IPS/IDS system (Intrusion):** The intrusion module is a rule-based approach to network security. The system matches network activity to a set of predefined rules which it then uses to alert network managers or directly prevent traffic which meets the established criteria. The main advantage against the current approach to data security is that the module provides the ability to define specific, customized rules based on the conditions and use cases which are specific to each industrial environment. Furthermore, the module is designed to adapt and grow as the needs of the organization evolve through the development of additional rules in the future. In terms of the evolution of industrial approaches to security this presents an important means to fully customized solutions which are uniquely adapted to the specific needs of a single customer, sector or vertical. Furthermore, a rule-based approach provides the framework necessary to implement AI-based threat detection and neutralization strategies.
- **Network Traffic Analysis (NTA):** This module allows network technicians to monitor the status and activity on the network in real time. This approach is an ideal complement to the rule-based IPS/IDS module as it allows technicians to focus on identifying potential threat activity which is not currently classified. The module allows technicians to immediately take action when they identify potential threats. This module also provides a general overview of network activity and health, aiding in the identification of potential technical problems not directly related to cybersecurity. By providing a visual representation of network activity, technicians will be able to better understand precisely what is occurring on the network at any given moment. This is valuable not only for security reasons but also in terms of data protection, privacy and trust as it will provide the industrial partners with the means to verify all network activity, not only known threats.
- **SIEM system (Vault):** This module can analyse and manage a large number of logs, offering the possibility of detecting more complex threats which the IPS system cannot detect. This approach works by analysing historical network data against security events to identify how the event occurred. The vault module offers capabilities such as metadata extraction to normalize data from numerous hardware vendors, data enrichment, correlation, and storage. This module represents the fusion of new approaches to data management and security.

During the trial the goal will be ensure that these modules are able to provide the necessary level of security in the industrial environment. Furthermore, these elements are widely-used approaches to network and data security. Therefore, the trial will present an excellent opportunity to produce a series of recommendations on improving their application in industrial environments as well as defining best practices.

5.1.4 Testing and Certification

Gestamp pilot's partners expect to test and certify:

- A unique QIF workflow, enable the effective exchange of metrology data throughout the entire manufacturing quality measurement process – from product design to inspection planning to execution to analysis and reporting.
- Validate new technologies such as colour mapping with textures for massive point clouds as part of the QIF workflow.
- Integration and compatibility between QIF and EIDS.
- Security data transfer
- The compliance of all network monitoring and security mechanisms with industrial standards and certifications

5.2 +GF+

Each Trial leader is expected to insert details for each of the standards they plan to use along with the role they play in the system architecture (by reference to the above) and why each standard has been selected. We are also interested in learning when several standards were considered, which ones were considered, and the reasons why particular standards were rejected.

- Data Management:
 - The database type used will be a time-series based database such as InfluxDB. InfluxDB is a high-performance data store written specifically for time series data. It allows for high throughput ingest, compression and real-time querying of that same data. It is ideal for machine process data.
 - The other way would be to use csv file as a first step for a matter of simplicity and availability of file format. It will be the main data format in transition phase.

- Communication Protocol: The communication protocol used will be OPC-UA for its open-source, cross-platform and security aspect. This multi-layered approach accomplishes the original design specification goals of:
 - Functional equivalence: all COM OPC Classic specifications are mapped to UA
 - Platform independence: from an embedded micro-controller to cloud-based infrastructure
 - Secure: encryption, authentication, and auditing
 - Extensible: ability to add new features without affecting existing applications
 - Comprehensive information modelling: for defining complex information
- Development: The development will be done in Python for its number of libraries in the field of machine learning provided, as well as for its open-source aspect.

5.2.1 Standards used

- In the Boost 4.0 context, and for the GF use case, INENDI Inspector is used as a standalone software for visualization of data sets, with the aim to identify useful subsets of data for building predictive models. The only standard involved in that scope is the one regarding the input and output of data sets. For the transition phase, we planned to use the csv file format which respect the RFC4180 (according to Internet Engineering Task Force) for a matter of universality/simplicity/availability of the file format.
- In that same context, Scilab software will be used to access/display data and potentially run machine learning algorithms. For the transition phase, this will be possible exchanging data trough csv file and prototyping algorithms in Scilab/Python/Matlab languages.

For both ESI's software, OPC UA communication protocol should be available through VISUAL Environment interface.

5.2.2 Identified Generic Standard for Further Development

- In the Boost 4.0 context, and for the GF use case, INENDI Inspector is used as a standalone software for visualization of data sets, with the aim to identify useful subsets of data for building predictive models. The only standard involved in that scope is the one regarding the input and output of data sets. For the transition phase, we planned to use the csv file format which respect the RFC4180

(according to Internet Engineering Task Force) for a matter of universality/simplicity/availability of the file format.

- In that same context, Scilab software will be used to access/display data and potentially run machine learning algorithms. For the transition phase, this will be possible exchanging data through csv file and prototyping algorithms in Scilab/Python/Matlab languages.

For both ESI's software, OPC UA communication protocol should be available through VISUAL Environment interface.

5.2.3 Opportunities for new standards

People working on the Boost 4.0 trials may be in a position to spot opportunities for new standards where they can see a gap in existing standards, e.g. because of fresh insights for how to look at the requirements, or when working on new areas for which no standards exist as yet.

- The standards that need to be investigated deeper is the data sharing platform. We will investigate platform such as Renku, platform developed by EPFL and Swiss Data Science Center
- EPFL and +GF+ milling machines take part in activities related to standard ontology development, in cooperation with the Industrial Ontologies Foundry (IOF), which aims at creating a set of core and opening ontologies that spans the entire domain of digital manufacturing. The +GF+ pilot was submitted to the IOF Product Planning & Scheduling (PPS) subgroup, and it will provide the testbed for standard-ontology implementation. Meanwhile, the +GF+ domain ontology will be used together with other business scenarios to capture common industry entities.

5.2.4 Experience with standards development

To help Boost 4.0 succeed in meeting its aims, it will be helpful to gather information on which partners have experience with particular standards development organisations and industry alliances. We would like to know which such organisations you are involved with and in what role. We are further looking for information summarising their main standards (relevant to smart manufacturing) and the work in progress – their standardisation pipeline!

- Python: GFMS has relevant experience in Python and in machine learning thanks to data and process specialist.

- Information Modelling: The Software Development team has a strong knowledge in OPC information modelling and in building OPC-UA framework.
- Time-Series DB: GFMS has run a test bench phase in order to test, explore and choose the best type of database. During this phase, our specialists have acquired a good understanding of this type of database

5.2.5 Data Management and Governance

The data platform might play an important role and bring added-value here. As explained, data sharing platform such as Renku are designed to connect independently administered platforms and positions itself as a unique one-stop shop for high quality data by allowing a federated access across institutions, giving each the freedom to enforce its own access controls over resources.

From GFMS point of view, those questions are still under investigation and our Software Team strongly involved and we hope that Boost 4.0 might help us better define those critical points.

5.2.6 Security, Safety, Privacy, Trust and Resilience

The use of OPC UA is firewall-friendly while addressing security concerns by providing a suite of controls:

- Transport: numerous protocols are defined providing options such as the ultra-fast OPC-binary transport or the more universally compatible SOAP-HTTPS, for example
- Session Encryption: messages are transmitted securely at 128 or 256 bit encryption levels
- Message Signing: messages are received exactly as they were sent
- Sequenced Packets: exposure to message replay attacks is eliminated with sequencing
- Authentication: each UA client and server is identified through OpenSSL certificates providing control over which applications and systems are permitted to connect with each other
- User Control: applications can require users to authenticate (login credentials, certificate, etc.) and can further restrict and enhance their capabilities with access rights and address-space “views”
- Auditing: activities by user and/or system are logged providing an access audit trail

The data platform that will be picked will include a high level of security, privacy, trust and resilience. For example, Renku makes use of state of the art security and privacy preserving technologies and best practices. It will give fine grained control over who accesses any data, from where and how. GFMS will have a strong focus and pay particular attention to security and confidentiality of the data when choosing its data sharing platform.

5.2.7 Testing and Certification

By drawing upon the experience gained in each of the Boost 4.0 trials, we are hoping to develop recommendations on architecture, standards, tooling and testing. We therefore are seeking to gather information on the approaches that trials expect to take in respect to testing, and certification against the standards.

GFMS expects to test and validate the following points:

- The integration of different data type (from process, measurement to assembly lines) in one platform and the results in term of analytic. Numbers of KPI will be defined in order to measure the improvement in the different steps of the production (machines, assembly lines and quality).
- Validate the performance of the different technologies in an industrial state of the art production plant, incl. Time Series Database, security of data transfer,
- Benchmark the different sharing platform and learn how to share efficiently big amount of critical data and results with industrial and academic partner.

5.3 WP8 – Benteler, Atlantis and Fraunhofer IEM

Each Trial leader is expected to insert details for each of the standards they plan to use along with the role they play in the system architecture (by reference to the above) and why each standard has been selected. We are also interested in learning when several standards were considered, which ones were considered, and the reasons why particular standards were rejected.

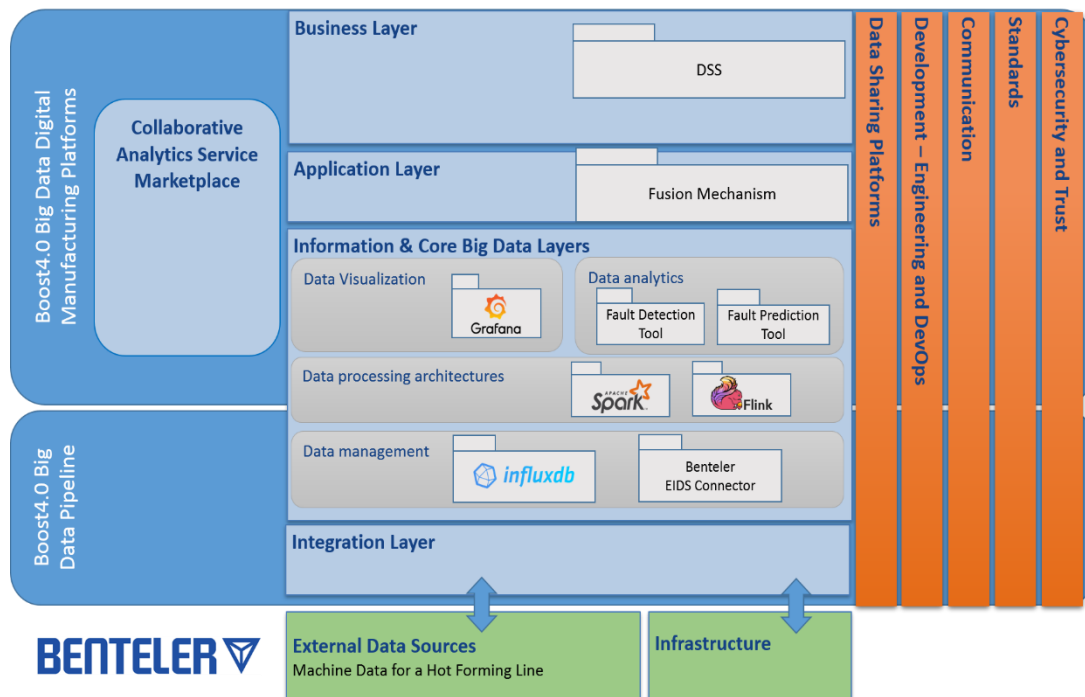


Figure 27 - Boost4.0 RA

Figure 27 presents the mapping of the different components used in the Benteler pilot to the Boost4.0 Reference Architecture (RA). The rest of this Section, will be based on that Figure to present the relation of the selected standards to the different layers of the Boost4.0 RA.

Benteler uses state of the art techniques in all steps of the manufacturing process, from design to commissioning. This is result of a systematic and strategic planning process. One of the key aspects is the standardization of the production process. It is important to follow Standard Operating Procedures (SOPs), which cover not only the manufacturing part, but also quality, data handling etc. Benteler applies multiple standards for data gathering, communication and transfer and is thus an excellent pilot partner for Boost4.0, as the standardization aspect is strong in the project and it affects all architectural levels.

Starting from the bottom layer of the RA, the Integration, which communicates with the External Data Sources and the Infrastructure, process control standards like Programmable Logic Controllers (PLC) are used for the automation of the production line (i.e. the control of the machinery) and the monitoring of the process. A PLC is a modular industrial computer control system that continuously monitors the state of input devices and makes decisions based upon a custom program to control the state of output devices. They are key to consistent replication of processes while also allowing collection and communication of vital information.

The communication between the Integration and the Data management layers, for the population and storage of the data collected through the monitoring process, is based on the OPC Unified Architecture (**OPC-UA**) open standard. More specifically, OPC-UA is used for the communication of devices within machines, between machines in the production line and from machines to systems in a convergence of Information Technology (IT) or Operation Technology (OT). Access control, authentication and encryption are embedded in the OPC-UA standards.

The communication between the Data management layer and the other layers of the Information & Core Big Data layers (i.e. Data processing architectures, Data analytics, Data visualization) is encrypted. The Transport Layer Security (**TLS**) standard is used for the validation of the Secure Sockets Layer (**SSL**) certificates for the data exchange between the server in the Data management layer and the clients in the other layers. Hypertext Transfer Protocol Secure (**HTTPS**) protocol is used for the encrypted (HTTP over TLS) communication orchestration between the server and the clients.

For the data exchange between the tools of the Data analytics, the Application and the Business layer, formats like JavaScript Object Notation (**JSON**) and Extensible Mark-up Language (**XML**) are used. Both the formats are easily interpretable by both humans and machines.

In the Business layer the DSS communicates with an **SAP** software solution (i.e. Production Module), which is used for the production and maintenance planning.

5.3.1 Data Management and Data Governance

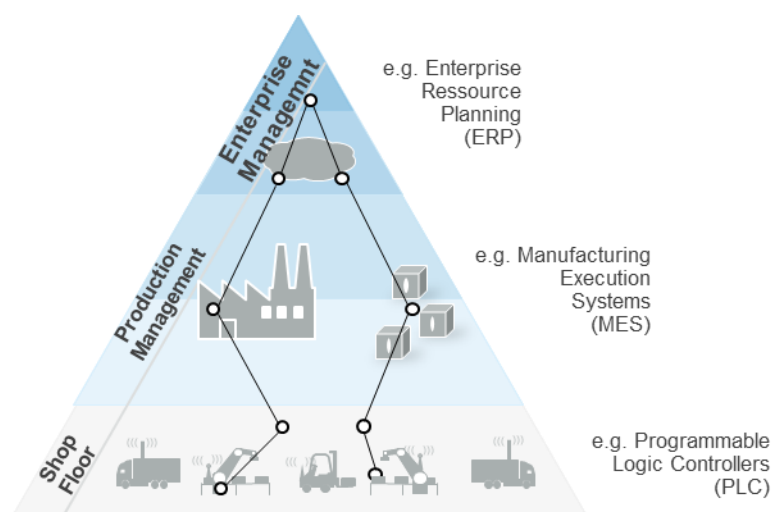


Figure 28 - Automation Pyramid

In context of smart maintenance within the manufacturing domain, the **automation pyramid** is a suitable way to structure and organize data sources as well as data sinks and data collecting entities. From bottom to top, different technical levels are addressed. At shop floor level, sensor data and machine data are available from PLCs. Real time data is made available via industry communication standards as e.g. OPC-UA or MQTT. This data is essential to identify the condition of machines and thus basis for any prediction of future machine state.

At production management level, context information of the process state is available through operation data. This includes information about production planning, e.g. part and stock numbers, manufacturing order numbers with production start/stop times and volumes, but also manual input from the shop floor, e.g. scrap numbers, machine faults or maintenance incidents and diagnosis. This data is typically held within MES systems.

At enterprise level, process data is linked to company-wide data with more static characteristics. This are e.g. part lists, bill of materials, stock levels, order status, but also overall maintenance information and production and machine effectiveness information. In order to make use of data for smart maintenance applications, **data management and governance has to be aligned to all levels of the automation pyramid.**

Integration of data sources among all technical levels has to be ensured. For ***data integration and interoperability***, this means integration of heterogeneous data sources. Concerning ***documents and contents***, these structured data as machine data or enterprise resource information from SAP, as well as unstructured data from shift books. While MES and ERP systems are wide-spread in production, machine data is often not provided in a way suitable for big data processing. This has to be addressed in the design of the ***data architecture***. This includes the ability to quickly request and process historical data from a given data source and time span. For ***Data storage and operations***, time series data bases as e.g. InfluxDB are a suitable solution. Considering ***reference and master data***, it should be considered to select a centralized data base as a single source of truth. ***Metadata*** is a challenge in manufacturing context, since there is a wide-spread range of machines available, which are not standardized. Thus, basic information as time stamps, production line, machine name is available. However, a more precise, standardized description of data sources would be useful in order to get information about e.g. physical properties and characteristics of the signal. This is essential in order to extract meaningful information from the data. ***Data quality*** has to be ensured, it is thus of importance for the company to have specific control over large parts of the data pipeline, starting from the data fetching mechanisms. Care has to be taken when inputting fetched data into the time series database, since timing information may be easily obscured, though it is of

importance for the correct subsequent processing. For *Data security*, industry standards and procedures are available, e.g. SSL certification and encryption for secure data transfer, and user authentication for access control. Concerning *Data Warehousing and Business Intelligence* in the Benteler pilot, the decision support system (DSS) processes the outcomes of the data analytics algorithms (i.e. fault detection and prediction). It then enables access to decision support on shop-floor level, considering the maintenance organization as well as management requirements.

5.3.2 Security, Safety, Privacy, Trust & Resilience

1) **Security:**

Benteler policies pose strict restrictions and requirements to security of data transfer due to the confidentiality of data. Secure communication between algorithms and machine databases is ensured through the application of security standards, in order to prevent unauthorized access to Benteler data and thus data from end-customers and OEMs. In order to meet these restrictions and requirements, the systems being developed within Boost4.0 will offer access control and user authentication. Additionally, encryption in the communication and Sandboxed IT development and network environments are some of the security measures that will be applied. The goal and logic behind the aforementioned planned actions is to setup a system with build in security which can be used in a competitive and confidential industrial environment. It should be noted that the aspect of physical security does not apply in the Boost4.0 scenarios in the Benteler pilot, in the sense of equipment or material malicious transfer and/or removal. However, it is a general issue that Benteler addresses already and, thus, there was no need to consider it in the smart maintenance scenarios.

2) **Safety:**

The transition from corrective or even planned maintenance to smart or predictive maintenance has a great potential to positively affect worker's safety at shopfloor level. The personnel do not have to intervene in a rush, in potentially hazardous conditions with hazardous materials, spills etc., as is many times the case in corrective maintenance, which takes place when a fault occurs. Even in the case of preventive maintenance, the risk of fault or failure is not identified. The ambient may still hide dangers for the personnel. Moreover, it can be a stressful task, especially when preventive maintenance activities are squeezed in short time periods, in order to affect production and equipment availability as less as possible. Smart/Predictive maintenance strategies and tools are key factors for ensuring a

less or non-stressful environment for the personnel, safer working conditions for all, lower risk of failures and faults.

3) **Privacy**

The protection of personal data is ensured since May 25th 2018 by the Regulation (EU) 2016/679, also known as GDPR (General Data Protection Regulation) which is repealing Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data. The objectives and principles of Directive 95/46/EC remain sound, but GDPR puts more stress in privacy, security, control over data, right to access etc. Within the Boost4.0 project, GDPR is anticipated to influence the handling of data primarily in issues related to privacy by design, data portability and access rights. In the case of the Benteler pilot, it is not expected to involve personal data, rather than machine data. However, the principles of GDPR are of relevance and will be considered to the extent that they are applicable.

Within the Benteler pilot and the smart maintenance Boost4.0 tools, there is the concept and evaluation for integrated resource planning of equipment for maintenance process optimization. This is to be done by interconnecting manufacturing, maintenance and diagnostic/predictive analytics data. It is noted that no personal data is anticipated to be needed nor used.

4) **Resilience**

The resilience of the production line is of critical importance to the industry. To cope with machinery faults, physical redundancy is a trivial but effective measure, where critical parts of the production line are duplicated. Functional redundancy relies on one machine but with a specification of multiple configurations per machine.

For cyber-attacks resilience, Benteler incorporates physical, logical, and other cybersecurity controls to prevent, detect, and mitigate cyber-attacks. Preventative controls are implemented, like firewalls, antiviruses, antimalware, etc. to minimize the impact and likelihood of successful attacks, detective controls to identify attacks in early stages, and corrective controls to mitigate the impact (like network redundancy, etc.).

5) **Trust**

The quality of the gathered machine data directly affects the business decisions

extracted from data analytics. Hence, the trust in the data quality is important. In the Benteler case, the data flow is exclusively under Benteler's control, as no other external business partner is involved in the data gathering and storing process. The quality of the raw machine data, that are fed to the data analysis tools, is reassured through a thorough iterative process of validation and optimization, in order to obtain reliable and near real time data measurement reporting. Other than the data gathering process, the trust in the data processing and data exchange procedures of the higher layers of the RA, is reassured through the adoption of the IDS technology and its trusted ecosystem. Inside the IDS ecosystem, all participants, data sources, and data services are certified against commonly defined rules.

5.3.3 Testing and Certification

Gather your inputs in respect to testing and certification. Do you have a plan for testing your implementation and monitoring its operation? If not, can you explain why? Otherwise we would like you to describe the details. In respect to certification against the standards you plan to use, you may be able rely on third party libraries that have been certified for the standards the libraries implement. How these vendors describe their approach to certification? Otherwise, please describe what you expect to do to ensure that your implementation conforms to the standards it uses.

By drawing upon the experience gained in each of the Boost 4.0 trials, we are hoping to develop recommendations on architecture, standards, tooling and testing. We therefore are seeking to gather information on the approaches that trials expect to take in respect to testing, and certification against the standards.

Testing and validation of algorithmic results is a challenging task in case of predictive maintenance. Often times, very little reference data is available for maintenance cases due rare occurrence of incidents and incomplete or imprecise documentation. Automated data-driven testing is not possible. Moreover, a human-based cross checking is necessary, where detected incidents are manually analysed in conjunction with reference data e.g. from SAP or shift-books. This process is done iteratively, in order to test for a wide range of incidents with varying conditions. However, this leads to testing and validation with long lead times and high research/development overhead.

Another form of validation is the transfer of algorithms to other plants with similar parameters. This increases the variance of the available data and thus the confidence of the validity measure. However, care must be taken when choosing the right machinery and

evaluating similarity of scenarios, since systematic deviations may give misleading conclusions.

For widely-used standardized components (e.g. OPC-UA, SSL, ...), testing and certification is more straight forward and based on industry standards.

6 Conclusions

Standardisation has a major role to play in enabling the adoption of advanced digital technologies across the enterprise, and this applies to the opportunities for the digitalisation of smart factories. NIST describe this in terms of a transformation from a hierarchical control model to one based upon distributed services. Industrie 4.0, by contrast, proposes a generic model featuring three axes. One addresses the lifecycle for product design and manufacture, another addresses the elements needed for manufacturing control systems, and the third addresses the different functional abstractions (assets, integration, communication, information, functional and business). These are related to IEC standards such as IEC62890, and IEC 62264/IEC61512 (derived from ANSI/ISA-95 and ANSI/ISA-88 respectively).

Whilst there are many applicable technologies on the shop floor, the trend is to replace older standards with new ones based upon modern networking standards. The networking requirements vary across the functional abstractions, e.g. real-time requirements are important for shop floor control systems. This can be realised through gateways between the different networks and a compartmentalised approach to security and resilience. There is inevitably a heterogeneous mix of technologies. Integration, as envisaged by Michael Porter, thus necessitates the use of abstraction layers that decouple the considerations appropriate to each layer. This includes the means to integrate older systems that would be costly to replace with newer designs.

Traditional tabular databases (SQL/RDBMS) are ill-suited to the needs of agile business models with rapidly evolving needs in response to changes in business conditions. This is driving the adoption of graph data as this facilitates integration of diverse data sources. The push for enterprise-wide integration along with data management and data governance, is driving interest in knowledge graphs and a Web inspired approach for access across federated systems. At this time, there is a lack of interoperability across graph databases from different vendors, and this is ripe for standards work on an interchange framework and query language. W3C's RDF is in a good position to feed into this work in conjunction with a higher-level framework aligned to property graphs and the Web of Things. The framework will need to address mappings between different classes of identifiers and schema languages, e.g. mapping OPC-UA models to industrial ontologies.

Smart manufacturing benefits from work on big data and digital twins³⁸. Big data focuses on how to benefit from the vast amounts of data that can now be collected during the design, manufacturing and operation phases for products. The sheer amount of information poses strong challenges to efficient processing, but at the same time is a good fit for machine learning algorithms hungry for data. There are opportunities around standards in relation to distributed storage and stream processing.

Digital twin is a term that refers to the idea of providing a virtual copy of a device for monitoring and control purposes. This is essentially the same idea as “things” in the Web of Things, where software objects acting as digital twins expose software interfaces to applications, decoupling applications from the underlying networks and protocols. The Web of Things embeds digital twins within a rich framework for describing the kinds of things, their capabilities and relationship to the context in which they reside. This enables devices to be integrated into an enterprise-wide knowledge graph.

Boost 4.0 can constructively support work on the missing standards through gathering the use cases and requirements from the experience gained with the Boost 4.0 pilot projects. The need for greater agility is a challenge for traditional standards development practices. This motivates the need for work on lighter weight processes for agreements that don’t need the full weight of international standards. Industry alliances such as the Industrial Ontologies Foundry³⁹ look promising in that regard. Boost 4.0 could help by supporting a dialogue across standards development organisations on a common vision for standards work at various stages of maturity, and with different requirements for agility versus stability.

Boost 4.0 can help to drive standards work through support from Boost 4.0 consortium partners for standardisation workshops such as the W3C Workshop on Graph Data scheduled for March 2019. Such workshops facilitate bringing people together from different backgrounds for fresh insights as to what new work is needed. This can create the momentum to launch new groups that can incubate ideas to the point where they are ready for entry into a formal standardisation process. There is a rich landscape of industry alliances and standards development organisations, and Boost 4.0 has the potential to help with breaking down barriers between different organisations and maximising opportunities for complementary roles for these organisations, with ensuing benefits for European manufacturers.

³⁸ See “Digital Twin and Big Data Towards Smart Manufacturing and Industry 4.0: 360 Degree Comparison”, Quinlin Qi and Fei Tao, available from <https://ieeexplore.ieee.org/document/8258937>

³⁹ See <https://sites.google.com/view/industrialontologies/home>