Fraunhofer

INTERNATIONAL DATA SPACES ASSOCIATION

IDS REFERENCE ARCHITECTURE MODEL

INDUSTRIAL DATA SPACE

VERSION 2.0

AUTHORS & CONTRIBUTORS

- Prof. Dr.-Ing. Boris Otto, Fraunhofer ISST
- Dr.-Ing. Steffen Lohmann, Fraunhofer IAIS
- Sebastian Steinbuß, International Data Spaces Association
- Andreas Teuscher, SICK

- Prof. Dr. Sören Auer, L3S Research Center
- Martin Böhmer, Fraunhofer IML
- Dr. Jürgen Bohn, Schaeffler
- Gerd Brost, Fraunhofer AISEC
- Dr.-Ing. Jan Cirullies, Fraunhofer ISST
- Constantin Ciureanu, T-Systems
- Eva Corsi, Boehringer Ingelheim
- Søren Danielsen, GateHouse Logistics
- Andreas Eitel, Fraunhofer IESE
- Thilo Ernst, Fraunhofer FOKUS
- Dr. Sandra Geisler, Fraunhofer FIT
- Joshua Gelhaar, Fraunhofer ISST
- Roland Gude, Fraunhofer IAIS
- Dr.-Ing. Christian Haas, Fraunhofer IOSB
- Jürgen Heiles, Siemens
- Juanjo Hierro, FIWARE
- Joachim Hoernle, ATOS
- Manuel Huber, Fraunhofer AISEC
- Christian Jung, Fraunhofer IESE
- Prof. Dr. Jan Jürjens, Fraunhofer ISST
- Dr. Anna Kasprzik, L3S Research Center
- Dr. Markus Ketterl, msg systems
- Judith Koetzsch, Rittal
- Jacob Köhler, Deloitte
- Dr. Christoph Lange, Fraunhofer IAIS
- Dorothea Langer, Deloitte
- Jörg Langkau, nicos
- Dominik Lis, Fraunhofer ISST
- Sven Löffler, T-Systems
- Dr. Ulrich Löwen, Siemens
- Dr. Christian Mader, Fraunhofer IAIS
- Bernhard Müller, SICK
- Nadja Menz, Fraunhofer FOKUS
- Andreas Müller, Schaeffler
- Lars Nagel, International Data Spaces Association
- Dr. Ralf Nagel, Fraunhofer ISST
- Harri Nieminen, Fastems
- Thomas Reitelbach, Bosch
- Aleksei Resetko, PricewaterhouseCoopers
- Daniel Pakkala, VTT Technical Research Centre of Finland
- Florian Patzer, Fraunhofer IOSB
- Heinrich Pettenpohl, Fraunhofer ISST
- René Pietzsch, eccenca
- Jaroslav Pullmann, Fraunhofer FIT
- Matthijs Punter, TNO
- Dr. Christoph Quix, Fraunhofer FIT
- Dr. Dominik Rohrmus, Siemens
- Lena Romer, Boehringer Ingelheim
- Jörg Sandlöhken, REWE Systems
- Patrick Schöwe, agma data

- Daniel Schulz, Fraunhofer IAIS
- Dr. Julian Schütte, Fraunhofer AISEC
- Dr. Karsten Schweichhart, Deutsche Telekom
- Prof. Egbert-Jan Sol, TNO
- Peter Sorowka, Cybus
- Prof. Dr.-Ing. Gernot Spiegelberg, Siemens
- Markus Spiekermann, Fraunhofer ISST
- Christian Spohn, ATOS
- Gerrit Stöhr, GESIS
- Dr. Michael Theß, Signal Cruncher
- Dr. Sebastian Tramp, eccenca
- Dr. Mona Wappler, thyssenkrupp
- Ann-Christin Weiergräber, Uniklinik RWTH Aachen
- Dr. Sven Wenzel, Fraunhofer ISST
- Oliver Wolff, Advaneo
- Heike Wörner, DB Schenker

PUBLISHERS

International Data Spaces Association
Anna-Louisa-Karsch-Str. 2
10178 Berlin
Germany

Fraunhofer-Gesellschaft zur Förderung
der angewandten Forschung e.V.
Hansastr. 27 c
80686 München
Germany

COPYEDITING

Tom Fleckstein, Text-World

Federal Ministry
of Education
and Research

DLR Project Management Agency

Grant ID 01IS15054

TABLE OF CONTENTS

01

INTRODUCTION

THE INDUSTRIAL DATA SPACE 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.

1.1 GOALS OF THE INDUSTRIAL DATA SPACE

Data sovereignty is a central aspect of the Industrial 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 Industrial Data Space initiative proposes a Reference Architecture Model for this particular capability and related aspects, including requirements for secure and trusted data exchange in business ecosystems.

The Industrial Data Space is an initiative that is institutionalized by two main activities: a strategic Fraunhofer research initiative entitled "Industrial Data Space", and the "International Data Spaces Association" (see Figure 1.1). While the strategic research initiative is concerned with the design and prototype implementation of the Reference Architecture Model, the association aims at setting an international standard. To reach this goal, it pools the requirements from various industries and provides use cases to test the results gained from its implementation. The standard materializes in the Reference Architecture Model itself and defined methods for secure data exchange between the various Industrial Data Space connectors.

Numerous actors in the market may then take up the Industrial Data Space standard and provide software services and technology. All offerings must comply with the Industrial Data Space standard and, thus, undergo a certification process.

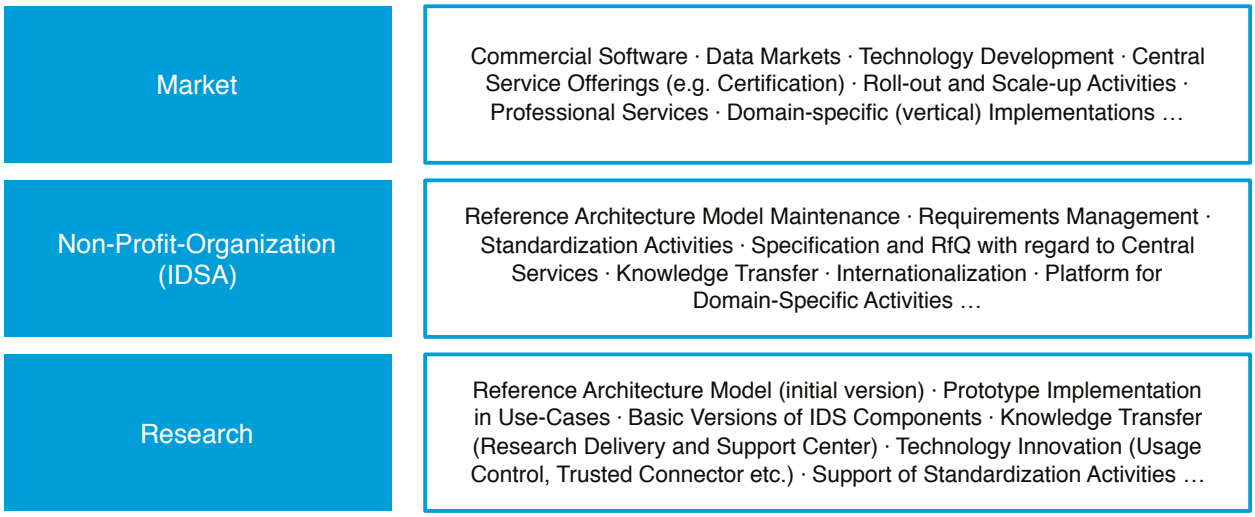


Figure 1.1: Industrial Data Space Activities

THE INDUSTRIAL DATA SPACE AIMS AT MEETING THE FOLLOWING STRATEGIC REQUIREMENTS:

» TRUST: Trust is the basis of the Industrial Data Space. It is supported by a comprehensive identity management focusing on the identification of participants and providing information about the participant based on the organizational evaluation and certification of all participants.

» SECURITY AND DATA SOVEREIGNTY: Components of the Industrial Data Space rely on current security measures. Next to architectural specifications, this is realized by the evaluation and certification of the components. In line with the central aspect of ensuring data sovereignty, a data owner in the Industrial Data Space attaches usage restriction information to its data before it is transferred to a data consumer. The data consumer may use this data only if it fully accepts the data owner's usage policy.

» ECOSYSTEM OF DATA: The architecture of the Industrial Data Space does not require central data storage capabilities. Instead, it pursues the idea of decentralization of data storage, which means that data physically remains with the respective data owner until it is transferred to a trusted party. This approach requires a holistic description of the data source and data as an asset combined with the ability to integrate domain-specific vocabularies for data. Brokers in the ecosystem enable comprehensive real-time search for data.

» STANDARDIZED INTEROPERABILITY: The Industrial Data Space Connector, being a central component of the architecture, is implemented in different variants and from different vendors. Nevertheless, each connector is able to communicate with every other connector or component in the ecosystem of the Industrial Data Space.

» VALUE ADDING APPS: The Industrial Data Space enables app injection to connectors to add services on top of the pure data exchange. This includes services for data processing as well as the alignment of data formats and data exchange protocols, but also enables analytics on data by the remote execution of algorithms.

» DATA MARKETS: The Industrial Data Space enables the creation of novel, data-driven services that make use of data apps. It also fosters new business models for those

Being the central deliverable of the research project, the Reference Architecture Model of the Industrial Data Space (IDS-RAM) constitutes the basis for a variety of software implementations, and thus for a variety of commercial software and service offerings.

THE RESEARCH AND DEVELOPMENT ACTIVITIES, AS WELL AS THE STANDARDIZATION EFFORTS, ARE DRIVEN BY THE FOLLOWING GUIDELINES:

» OPEN DEVELOPMENT PROCESS: The International Data Spaces Association is a non-profit organization institutionalized under the German law of associations. Every organization is invited to participate, as long as it adheres to the common principles of work.

» RE-USE OF EXISTING TECHNOLOGIES: Inter-organizational information systems, data interoperability, and information security are well-established fields of research and development, with plenty of technologies available in the market. The work of the Industrial Data Space initiative is guided by the idea not to "reinvent the wheel", but to use existing technologies (e.g., from the open-source domain) and standards (e.g., semantic standards of the W3C) to the extent possible.

» CONTRIBUTION TO STANDARDIZATION: Aiming at establishing an international standard itself, the Industrial Data Space initiative supports the idea of standardized architecture stacks.

1.2 PURPOSE AND STRUCTURE OF THE DOCUMENT

The purpose of this document is to introduce the Reference Architecture Model of the Industrial Data Space. Focusing on the generalization of concepts, functionality, and overall processes involved in the creation of a secure "network of trusted data", it resides at a higher abstraction level than common architecture models of concrete software solutions do. The document provides an overview supplemented by dedicated architecture specifications defining the individual components of the Industrial Data Space (Connector, Broker, App Store, etc.) in detail.

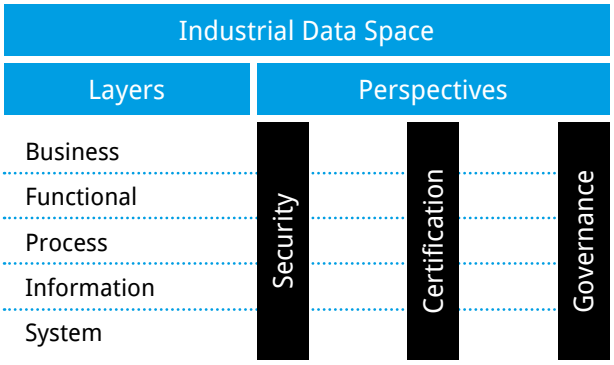


Figure 1.2: General Structure of Reference Architecture Model

In compliance with common system architecture models and standards (e.g., ISO 42010, 4+1 view model), the Reference Architecture Model uses a five-layer structure expressing various stakeholders' concerns and viewpoints at different levels of granularity.

The general structure of the Reference Architecture Model is illustrated in Figure 1.2. The model is made up of five layers: The Business Layer specifies and categorizes the different roles which the participants of the Industrial Data Space may assume, and it specifies the main activities and interactions connected with each of these roles. The Functional Layer defines the functional requirements of the Industrial Data Space, plus the concrete features to be derived from these. The Process Layer specifies the interactions taking place between the different components of the Industrial Data Space; using the BPMN notation, it provides a dynamic view of the Reference Architecture Model. The Information Layer defines a conceptual model which makes use of Linked Data principles for describing both the static and the dynamic aspects of the Industrial Data Space's constituents. The System Layer is concerned with the decomposition of the logical software components, considering aspects such as integration, configuration, deployment, and extensibility of these components. Directly related to the five layers of the Reference Architecture Model are three cross-sectional Perspectives: Security, Certification, and Governance. These are an integral part of the Reference Architecture Model in order to make sure three major core concepts of the Industrial Data Space are implemented across all five layers.

02

CONTEXT OF THE INDUSTRIAL DATA SPACE

2.1 DATA IN THE SMART SERVICE WELT

Novel digital products and services often emerge in business ecosystems, which companies enter to jointly fulfill 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 analyzed 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 (as shown by a number of examples in Figure 2.1).

2.2 DATA SOVEREIGNTY AS A KEY CAPABILITY

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.

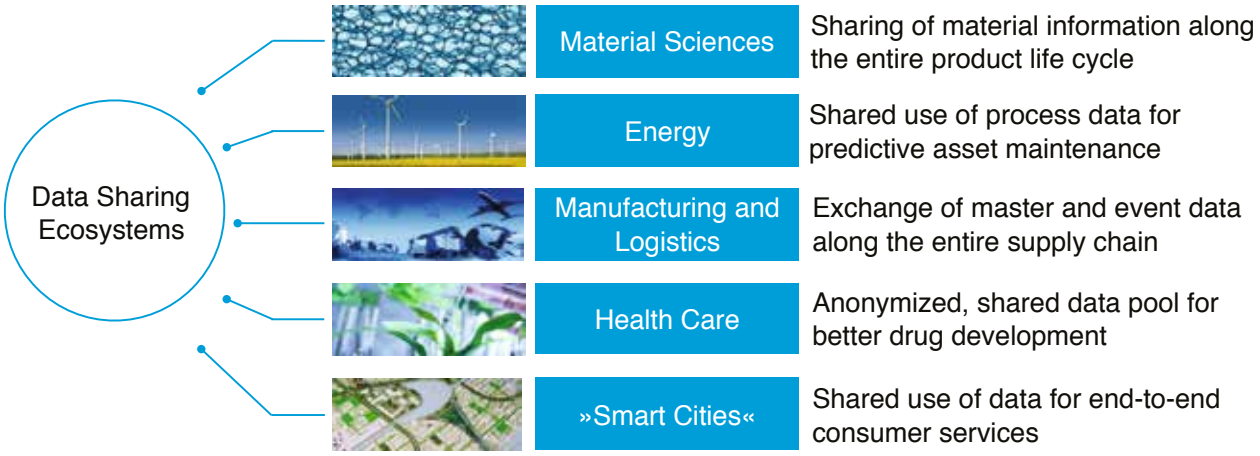


Figure 2.1: Data Sharing in Ecosystems

2.3 DATA AS AN ECONOMIC GOOD

It is indisputable that data has a value and that data management produces costs. Today, data is traded in the market, it has a price, and many companies monitor the costs incurred for data management. However, data as an intangible good differs from tangible goods with regard to a number of properties, among which the fact that data is non-rival is considered the most important one. The value of data increases as it is being used (and, in many cases, as the number of users increases). These differences hinder the transfer and application of legal provision to the management and use of data. It does not prevent the fact, though, that data is in fact an economic good.

As the value data contributes to the development of innovative products and services varies (depending on what category the data can be assigned to), the need for protection of data is not the same across all categories. Public data, for example, which can be accessed by any company, requires a lower level of protection than private data or club data.

Because of these differences and distinctions made with regard to data, a generally accepted understanding of the value of data has not been established so far. Nevertheless, there is a growing need to determine the value of data, given the rapid developments taking place in the

2.4 DATA EXCHANGE AND DATA SHARING

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 2.2 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.

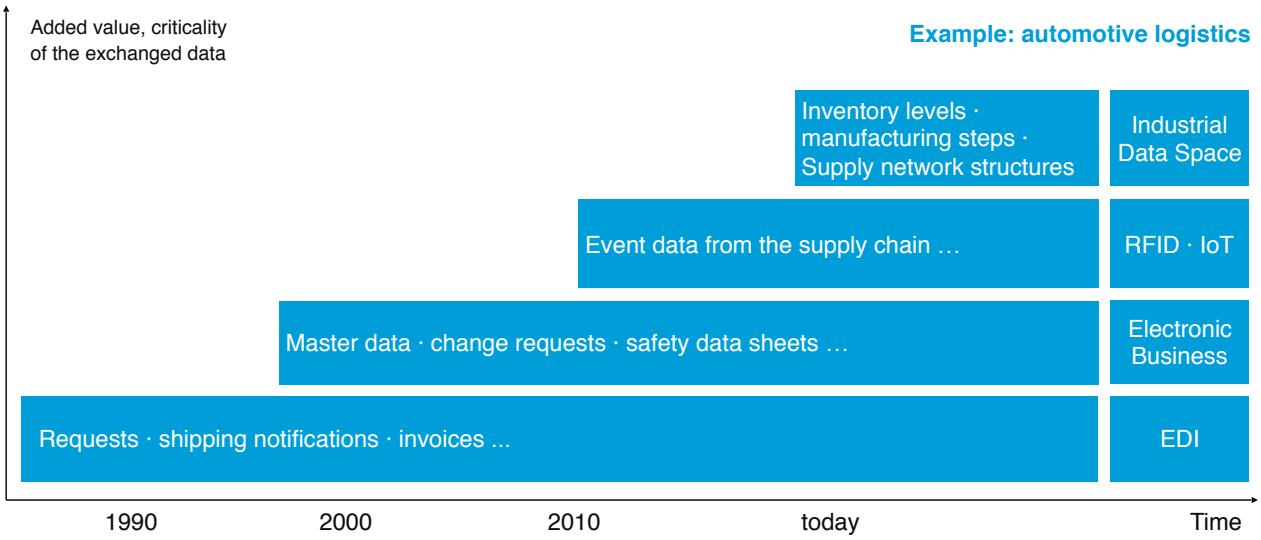


Figure 2.2: Data Exchange Standards

2.5 INDUSTRIAL CLOUD PLATFORMS

The growing number of industrial cloud platforms will also drive the need towards a standard for data sovereignty. With the large amount 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.

2.6 CONTRIBUTION OF THE INDUSTRIAL DATA SPACE

By proposing an architecture for secure data exchange and trusted data sharing, the Industrial 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 2.3 shows a typical architecture stack of the digital industrial enterprise. The Industrial 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.

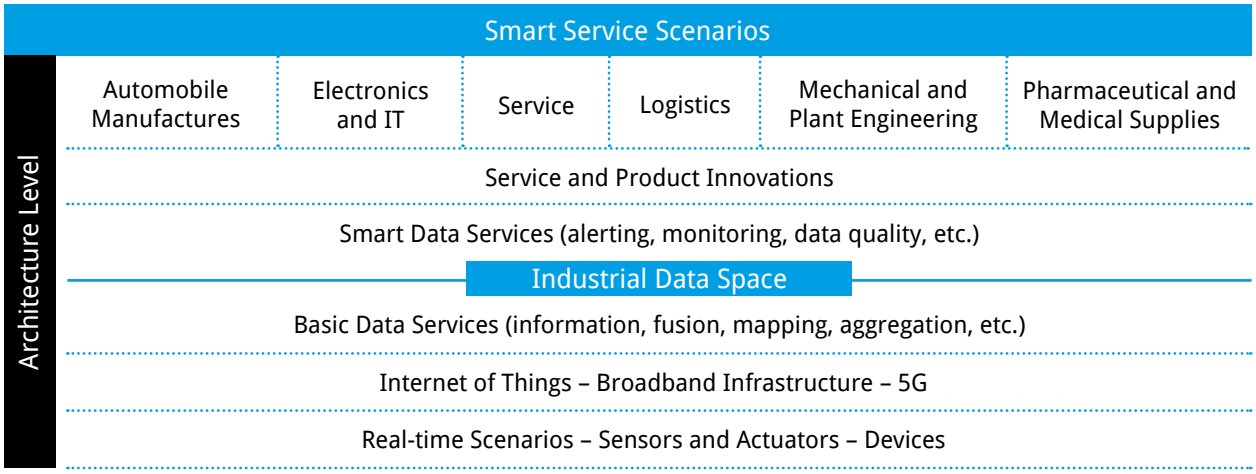


Figure 2.3 Typical enterprise architecture stack

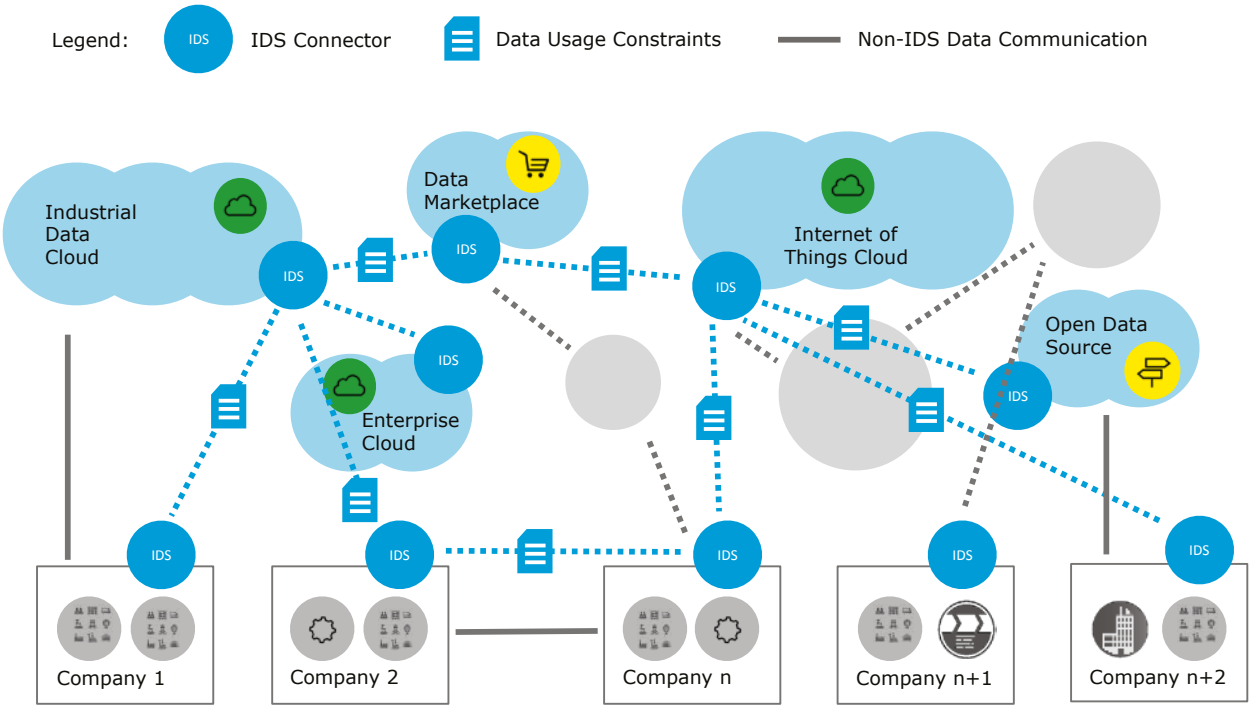


Figure 2.4: Industrial Data Space and Cloud Platforms

When broadening the perspective from an individual use case scenario to a platform landscape view, the Industrial 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 Industrial 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 2.4).

With this integrating ambition, the Industrial Data Space initiative positions itself in the context of cognate initiatives on both national and international level. Founded in Germany, the activities of the Industrial Data Space are closely aligned with Plattform Industrie 4.0. It is important to note that Plattform Industrie 4.0 addresses all relevant architectural layers, whereas the Industrial Data Space initiative focuses on the data layer and economy (see Figure 2.5). On the other hand, the Industrial Data Space initiative has a broader scope than Plattform Industrie 4.0 does, as it includes also smart-service scenarios from all domains and is not limited to industrial scenarios only.

The Industrial Data Space initiative has established, and aims to establish, liaisons with other initiatives, among them

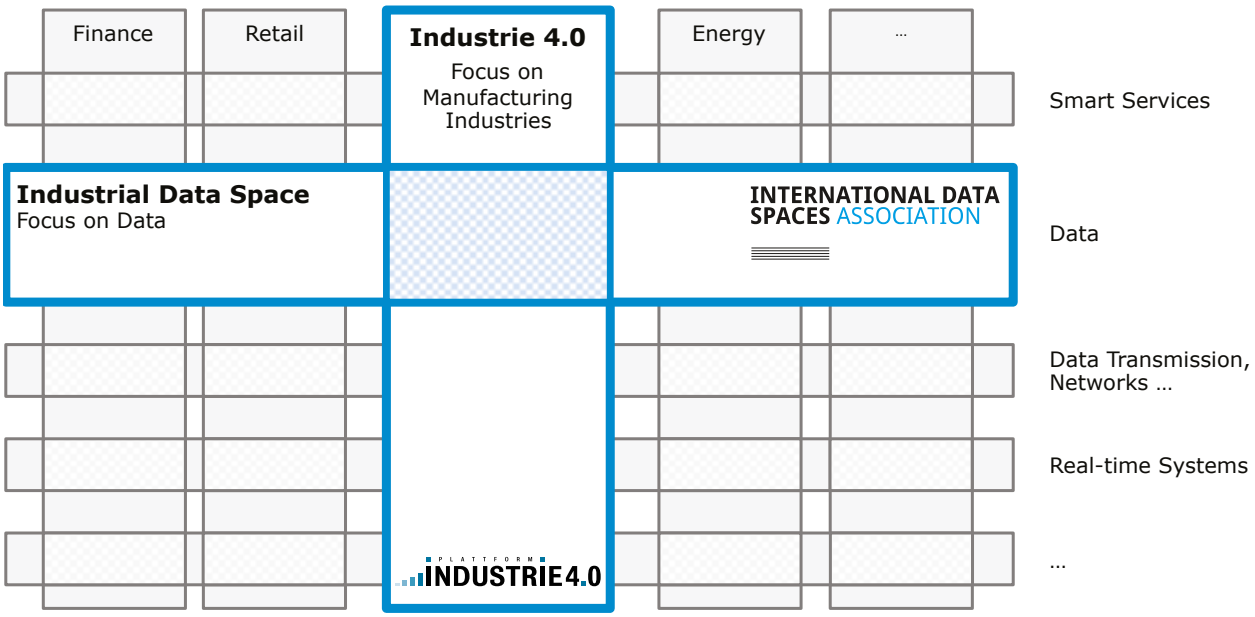


Figure 2.5: Relations with Platform Industrie 4.0

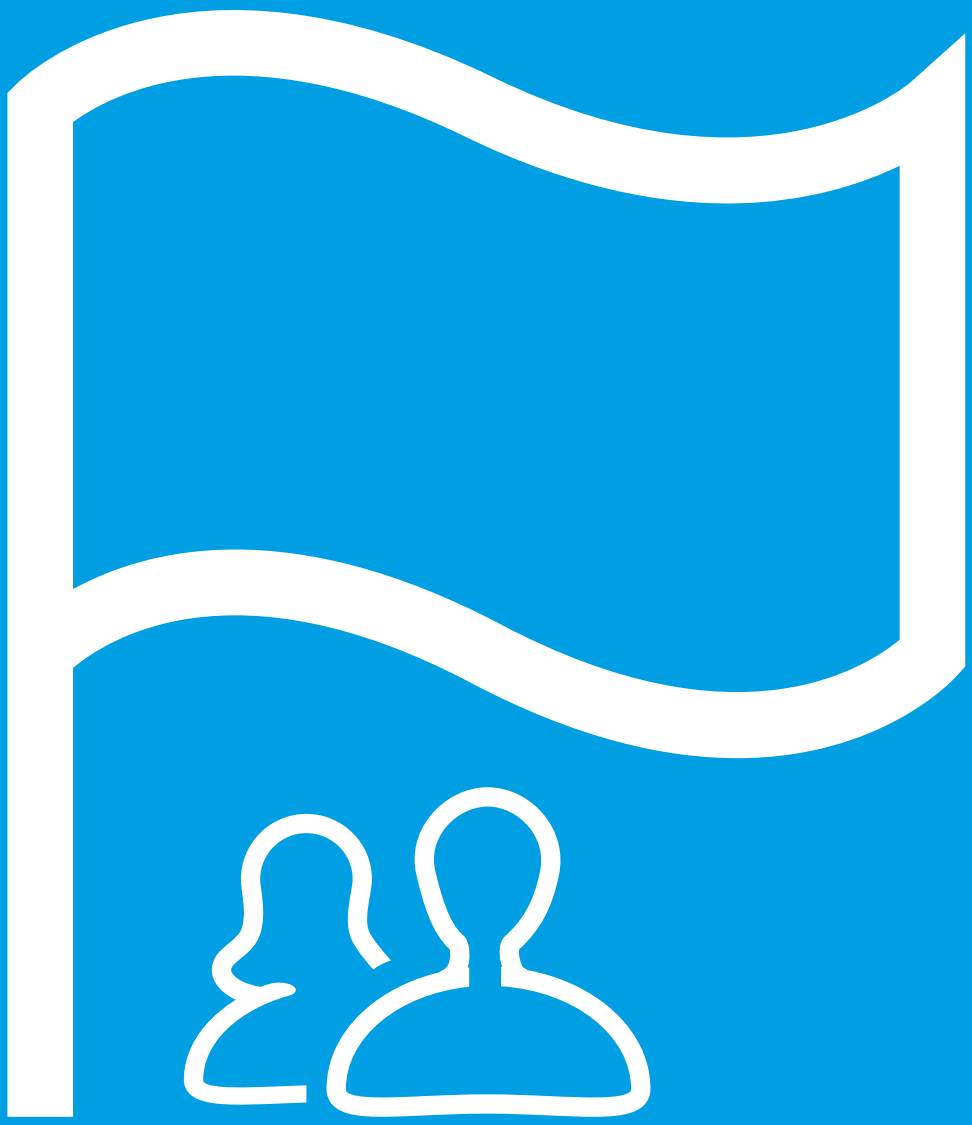
- Alliance for Internet of Things Innovation,
- Big Data Value Association,
- Data Market Austria,
- eCl@ss,
- FIWARE Foundation,
- Industrial Internet Consortium,
- OPC Foundation,
- Plattform Industrie 4.0,
- Standardization Council Industrie 4.0, and
- World Wide Web Consortium.

Furthermore, the Industrial Data Space initiative seeks collaboration and exchange of ideas with existing research and standardization initiatives.

By functioning as a mediator between top-down and bottom-up approaches, bridging the gaps between research, industry, politics, and standards bodies, aligning the requirements of the economy and society, and fostering ties with other initiatives, the Industrial Data Space can be considered a unique initiative in the landscape of Industry 4.0, not only affecting manufacturing industries but also other industrial and service sectors, such as health, finance, retail or energy (see Figure 2.5).

03

LAYERS OF THE REFERENCE ARCHITECTURE MODEL



THE FIVE LAYERS OF THE REFERENCE ARCHITECTURE MODEL ARE PRESENTED IN DETAIL IN THE FOLLOWING SUBSECTIONS.

3.1 BUSINESS LAYER

The Business Layer specifies and categorizes the different roles which the participants in the Industrial Data Space may assume. It thereby contributes to the development of business models that can be applied by the participants in the Industrial Data Space. In addition, the Business Layer specifies the main activities and interactions connected with each of these roles, which is important in the subsequent sections to identify the components of the architecture.

While the Business Layer provides an abstract description of the roles in the Industrial Data Space, it can be considered a blueprint for the other, more technical Layers. The Business Layer can therefore be used to verify the technical architecture of the Industrial Data Space (e.g., to check whether all interfaces required between the Industrial Data Space components have been specified, or whether all information required for running the business process is available for the Industrial Data Space components).

3.1.1 ROLES IN THE INDUSTRIAL DATA SPACE

In the following, the roles of the participants, together with the basic activities assigned to these roles, are described in detail. The majority of roles require certification of the organization that wants to assume that role, including certification of the technical, physical, and organizational security mechanisms the organization employs. Certification of organizations that want to participate in the Industrial Data Space is considered a measure to establish trust among all participants (especially with regard to roles that are crucial for the functioning of the Industrial Data Space, such as the Broker Service Provider, the App Store, the Identity Provider, or the Clearing House). The Certification Scheme applied in the participant evaluation process is described in detail in Section 4.2.

DATA OWNER
The Data Owner holds all legal rights of, and has complete control over, its data. Usually, a participant acting as a Data Owner automatically assumes the role of the Data Provider as well. However, there may be cases in which the Data Provider is not the Data Owner (e.g., if the data is technically managed by a different entity than the Data Owner, such as in the case of a company using an external IT service provider for data management).

In cases in which the Data Owner does not act as Data Provider, the only activity of the Data Owner is to authorize a Data Provider to make its data available to be used by a Data Consumer. Any such authorization should be documented by a contract, which should include data usage policy information for the data provided (cf. Section 4.1.10). The contract needs not necessarily be a paper document, but may be an electronic file as well.

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

Model of the Industrial Data Space. Providing a Data Consumer with data from a Data Owner is the main activity of the Data Provider. To facilitate a data request from a Data Consumer, the Data Provider should provide a Broker Service Provider (see below) with proper metadata about the data. However, a Broker Service Provider is not necessarily required for a Data Consumer and a Data Provider to establish a connection.

Exchanging data with a Data Consumer needs not necessarily be the only activity of the Data Provider. At the end of a data exchange transaction completely or partially executed, for example, the Data Provider may log the details of the successful (or unsuccessful) completion of the transaction at a Clearing House (see below) to facilitate billing or resolve a conflict. Furthermore, the Data Provider can use Data Apps to enrich or transform the data in some way, or to improve its quality (Data Apps are specific applications that can be integrated into the data exchange workflow between two or more participants in the Industrial Data Space).

If the technical infrastructure for participating in the Industrial Data Space is not deployed by the Data Consumer, a Data Provider may use a Service Provider (see below) to connect to the Industrial Data Space.

DATA CONSUMER

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

Before the connection to a Data Provider can be established, the Data Consumer can search for existing datasets by making an inquiry at a Broker Service Provider. The Broker Service Provider then provides the required metadata for the Data Consumer to connect to a Data Provider. Alternatively, the Data Consumer can establish a connection with a Data Provider directly (i.e., without involving a Broker Service Provider). In cases in which the information to connect with the Data Provider is already known to the Data Consumer, the Data Consumer may request the data (and the corresponding metadata) directly from the Data Provider. Like a Data Provider, the Data Consumer may log the details of a successful (or unsuccessful) data exchange transaction at a Clearing House, use Data Apps to enrich, transform, etc. the data received, or use a Service Provider to connect to the Industrial Data Space (if it does not deploy the technical infrastructure for participation itself).

DATA USER

Similar to the Data Owner being the legal entity that has the legal control over its data, the Data User is the legal entity that has the legal right to use the data of a Data Owner as specified by the usage policy. In most cases, the Data User is identical with the Data Consumer. However, there may be scenarios in which these roles are assumed by different participants. For example, a patient could use a web-based software system to manage their personal health data and grant access to this data to a health coach. The data could be received from a hospital. In this case, the health coach would be the Data User and the provider of the web-based software system would be the Data Consumer.

BROKER SERVICE PROVIDER

The Broker Service Provider is an intermediary that stores and manages information about the data sources available in the Industrial Data Space. As the role of the Broker Service Provider is central but non-exclusive, multiple Broker Service Providers may be around at the same time (e.g., for different application domains).

An organization offering broker services in the Industrial Data Space may assume other intermediary roles at the same time (e.g., Clearing House or Identity Provider, see below). Nevertheless, it is important to distinguish organizations and roles (e.g., assuming the role of a Broker Service Provider means that an organization deals only with metadata management; at the same time, the same organization may assume the role of a Clearing House, for which completely different tasks are defined).

The activities of the Broker Service Provider mainly focus on receiving and providing metadata. The Broker Service Provider must provide an interface for Data Providers to send their metadata. The metadata should be stored in an internal repository for being queried by Data Consumers in a structured manner. While the core of the metadata model must be specified by the Industrial Data Space (i.e., by the Information Model, see Section 3.4), a Broker Service Provider may extend the metadata model to manage additional metadata elements.

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

CLEARING HOUSE
The Clearing House is an intermediary that provides clearing and settlement services for all financial and data exchange transactions. In the Industrial Data Space, clearing activities are separated from broker services, since these activities are technically different from maintaining a metadata repository. As already stated above, it might still be possible that the two roles "Clearing House" and "Broker Service Provider" are assumed by the same organization, as both roles require acting as a trusted intermediary between the Data Provider and the Data Consumer.

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

IDENTITY PROVIDER
The Identity Provider should offer a service to create, maintain, manage and validate identity information of and for participants in the Industrial Data Space. This is imperative for secure operation of the Industrial Data Space and to avoid unauthorized access to data. More details about identity management can be found in Section 4.1.

APP STORE
The App Store provides Data Apps, i.e., applications that can be deployed in the Industrial Data Space to facilitate data processing workflows. Data Apps might be certified by a Certification Body, following the certification procedures defined in Section 4.2.

The App Store is responsible for managing information about Data Apps offered by App Providers (see below). The App Store should provide interfaces for publishing and retrieving Data Apps plus corresponding metadata.

APP PROVIDER
App Providers develop Data Apps to be used in the Industrial Data Space. To be deployable, a Data App has to be compliant with the system architecture of the Industrial Data Space (see Section 3.5). In addition, Data Apps can be certified by a Certification Body in order to increase trust in these applications (especially with regard to Data Apps processing sensitive information). Each Data App must be published in the App Store for being accessed and used by Data Consumers and Data Providers. App Providers should describe each Data App using metadata (in compliance with a metadata model) with regard to its semantics, functionality, interfaces, etc.).

VOCABULARY PROVIDER
The Vocabulary Provider manages and offers vocabularies (i.e., ontologies, reference data models, or metadata elements) that can be used to annotate and describe datasets. In particular, the Vocabulary Provider provides the Information Model of the Industrial Data Space, which is the basis for the description of data sources (see Section 3.4). In addition, other domain specific vocabularies can be provided.

SOFTWARE PROVIDER
A Software Provider provides software for implementing the functionality required by the Industrial Data Space (i.e., through software components, as described in Section 3.5). Unlike Data Apps, software is not provided by the App Store, but delivered over the Software Providers' usual distribution channels, and used on the basis of individual agreements between the Software Provider and the user (e.g., a Data Consumer, a Data Provider, or a Broker Service Provider). This procedure implies that the agreements between Software Providers and Data Consumers, Data Providers, etc. remain outside the scope of the Industrial Data Space.

SERVICE PROVIDER
If a participant does not deploy the technical infrastructure required for participation in the Industrial Data Space itself, it may transfer the data to be made available in the Industrial Data Space to a Service Provider hosting the required infrastructure for other organizations. This role includes also providers offering additional data services (e.g., for data analysis, data integration, data cleansing, or semantic enrichment) to improve the quality of the data exchanged in the Industrial Data Space. From a technical point of view,

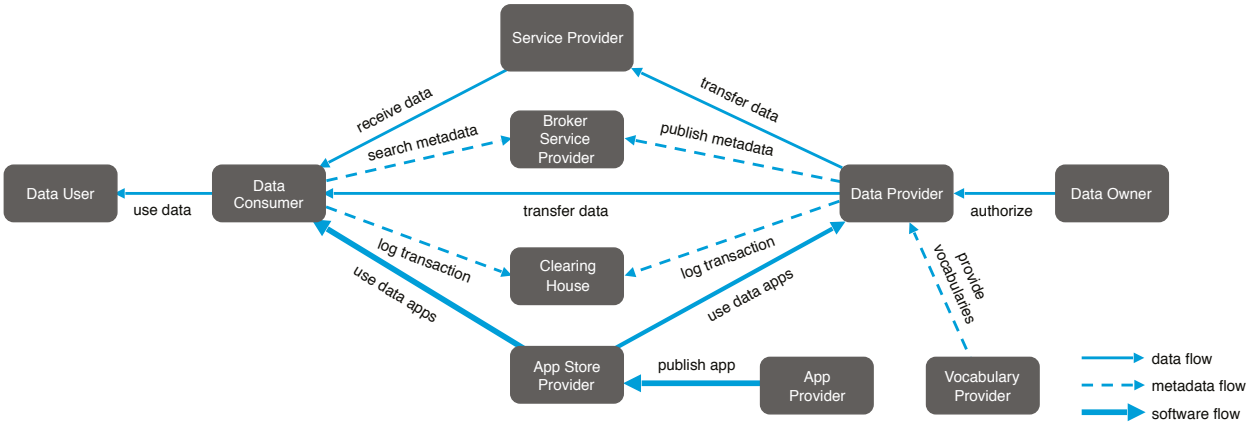


Figure 3.1: Roles and interactions in the Industrial Data Space

such a Service Provider can be considered a Data Provider and a Data Consumer at the same time (e.g., as a Data Consumer, it receives data from a Data Provider, then provides its specific service, and then turns into a Data Provider itself and offers the data in the Industrial Data Space). Unlike the services provided by a Service Provider, Data Apps can be installed in the IT environment of a Data Consumer or Data Provider for implementing additional data processing functionality. To use the functionality of a Data App, the data therefore does not have to be transferred to an external Service Provider.

CERTIFICATION BODY AND EVALUATION FACILITY
The Certification Body and the Evaluation Facility are in charge of the certification of the participants and the technical core components in the Industrial Data Space. The Certification Scheme applied is described in Section 4.2.

3.1.2 ROLE INTERACTION AND CATEGORIZATION

Figure 3.1 gives an overview of the roles and the interactions taking place between them. As some of the roles (Certification Body and Evaluation Facility) are not actively involved in the everyday operations of the Industrial Data Space, they are omitted from the illustration. Also, the figure does not include Software Providers and Identity Providers, because of the redundant connection of those roles with all other roles. The Software Provider would be connected to all other roles with the relation “provides software”. Likewise, the Identity Provider would be connected to all other roles with the relation “provides identity”. Based on this overview and the previous descriptions, each role can be assigned to one of four categories.

CATEGORY 1: CORE PARTICIPANT
Core Participants are involved and required every time data is exchanged in the Industrial Data Space. Roles assigned to this category are Data Owner, Data Provider, Data Consumer, and Data User. The role of a Core Participant can be assumed by any organization that owns, wants to provide, and/or wants to consume/use data.

Benefit for participants in the Industrial Data Space is created by these roles by providing or consuming/using data. Data Providers and Data Consumers may apply business models (including pricing models) as deemed appropriate.

CATEGORY 2: INTERMEDIARY
Intermediaries act as trusted entities. Roles assigned to this category are Broker Service Provider, Clearing House, App Store, Vocabulary Provider, and Identity Provider. Only trusted organizations should assume these roles.

Benefit for participants in the Industrial Data Space is created by these roles by establishing trust and providing metadata, creating a business model around their services.

CATEGORY 3: SOFTWARE AND SERVICES
This category comprises IT companies providing software and/or services (e.g., in a software-as-a-service model) to the participants of the Industrial Data Space. Roles subsumed under this category are App Provider, Service Provider, and Software Provider.

Benefit is created by these roles by providing applications, software, and services to the participants of the Industrial Data Space. As far as Data Apps are concerned, the value chain is part of the processes managed by the Industrial Data Space. The same applies to services that are provided by Service Providers. The process of providing software used for establishing the endpoints of a data exchange is not part of the Industrial Data Space, however, as it takes place before an organization joins the Industrial Data Space.

CATEGORY 4: GOVERNANCE BODY
The Industrial Data Space is governed by the Certification Body. These two bodies make sure that only compliant organizations may participate in this trusted business ecosystem. Benefit for participants in the Industrial Data Space is created by these roles by taking care of the certification process and issuing certificates (both with regard to organizations that want to participate and with regard to software components that are to be used).

3.2 FUNCTIONAL LAYER

The Functional Layer defines, irrespective of existing technologies and applications, the functional requirements of the Industrial Data Space, and the features to be implemented resulting thereof. The Industrial Data Space initiative has drawn up a document entitled “Functional Overview”, containing all functional requirements identified. Figure 3.2 shows the overall functional architecture, grouping the individual requirements into six functional entities to be provided by the Industrial Data Space (in accordance with the strategic requirements given in Section 1.1).

Each of these functional entities is characterized by different requirements. The full list of functional requirements can be found in the “Functional Overview” document referred to above. In the following, a brief summary of the Functional Overview is given.

3.2.1 TRUST

Although requirements related to trust are usually non-functional, they are addressed by the Functional Layer, since they represent fundamental features of the Industrial Data Space. The trust entity can be split into three main aspects: roles in the Industrial Data Space, identity management and user certification – complemented by respective governance aspects (see Section 4.3).

ROLES
The roles in the Industrial Data Space have different tasks. For example, the identity provider has the task to verify the participants. More information about the roles is given in Section 3.1.

IDENTITY MANAGEMENT
Every Connector participating in the Industrial Data Space must have a unique identifier and a valid certificate. Each Connector must be able to verify the identity of other Connectors (with special conditions being applied here; e.g., security profiles).

USER CERTIFICATION
The organizations participating in the Industrial Data Space require certification in order to establish trust among all participants. More information about the certification process is given in Section 4.2.

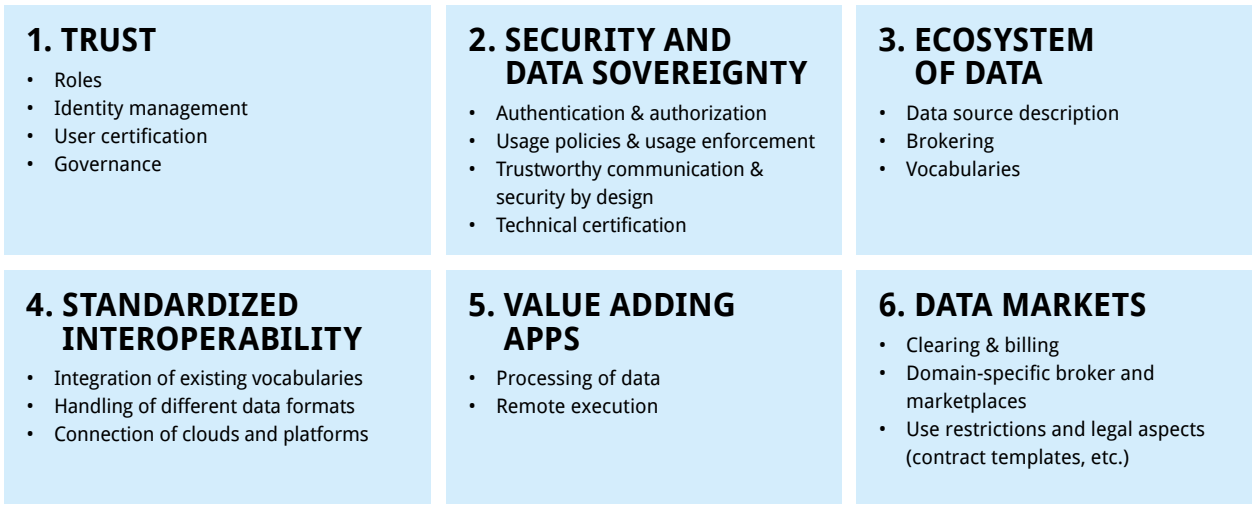


Figure 3.2: Functional architecture of the Industrial Data Space

3.2.2 SECURITY

Although requirements related to security are also usually non-functional, they are addressed by the Functional Layer, since they represent fundamental features of the Industrial Data Space. The Security entity contains different aspects: authentication & authorization, usage policies & usage enforcement, trustworthy communication & security by design and technical certification.

AUTHENTICATION & AUTHORIZATION
Each Connector must have a valid X.509 certificate. Therefore, each participant of the Industrial Data Space (operating an endpoint) is able to verify the identity of other participants by the X.509 certificate. Certain conditions, e.g. security profiles, may also apply here. More information about authentication is given in Section 4.1. The Connector serving as data source must be able to verify the receiving Connectors capabilities and security features as well as his identity. More information about authorization is given in Section 4.1.

USAGE POLICIES & USAGE ENFORCEMENT
Data Providers can be sure that their data is treated according to the specified usage policies by the IDS Connector of the Data Consumer. Each participant is able to define usage control polices that are attached to outbound data. Policies might include restrictions, e.g. disallowing persistence of data or transfer of data to other parties. More information about authorization is given in Section 4.1.

TRUSTWORTHY COMMUNICATION & SECURITY BY DESIGN
Connectors, App Stores, and Brokers can check if the Connector of the connecting party is running a trusted (certified) software stack. Any communication between (external) Connectors can be encrypted and integrity protected. Each Data Provider must be able to ensure that its data is handled by the Connector of the Data Consumer according to the usage policies specified, or the data will not be sent. To reduce the impact of compromised applications, appropriate technical measures must be applied to isolate Data Apps from each other and from the Connector. Data Providers and Data Consumers can decide about the level of security of their respective Connectors by deploying Connectors supporting the selected security profile. More information about security is given in Section 4.1.

TECHNICAL-CERTIFICATION
The core components of the Industrial Data Space, and especially the Connector, require certification from the Certification Body in order to establish trust among all participants. More information about the certification process is given in Section 4.2.

3.2.3 ECOSYSTEM OF DATA

Being able to explain, find and understand data is another key aspect of the Industrial Data Space. Therefore, every data source in the Industrial Data Space is described based on the Industrial Data Space vocabulary.

DATA SOURCE DESCRIPTION
Therefore, participants must have the opportunity to describe, publish, maintain and manage different versions of metadata. Metadata should describe the syntax and serialization as well as the semantics of data sources. Furthermore, metadata should describe the application domain of the data source. The operator of the Connector must be able to define the price, the price model and the usage policies regarding certain data. More information about data source description is given in Section 3.4.

BROKERING
The operator of a Connector must be able to provide an interface for data and metadata access. Each Connector must be able to transmit metadata of its data sources to one or more brokers. Every participant must be able to browse and search metadata in the metadata repository, provided the participant has the right to access the metadata. Every participant must be able to browse the list of participants registered at a broker.

VOCABULARY
To create metadata, the operator may use vocabularies, which help structure metadata. The operator can use standard vocabularies, create own vocabularies, or work collaboratively with others on new vocabularies provided by vocabulary hubs. Vocabulary hubs are central servers that store vocabularies and enable collaboration. Collaboration may comprise search, selection, matching, updating, suggestion of vocabulary changes by users, version management, deletion, duplicate identification, and unused vocabularies. Vocabulary hubs need to be managed. More information about vocabulary is given in Section 3.4.

3.2.4 STANDARDIZED INTEROPERABILITY

The standardized data exchange between the participants is the fundamental aspect of the Industrial Data Space. The Industrial Data Space Connector is the main component in this case.

OPERATION
Participants should be able to run the Connector software in their own IT environment. Alternatively, they may run a Connector on mobile or embedded devices. The operator of the Connector must be able to define the data workflow inside the Connector. Users of the Connector must be identifiable and manageable. Passwords and key storage must be protected. Every action, data access, data transmission,

incident, etc. should be logged. Using this logging data, it should be possible to draw up statistical evaluations on data usage etc. Notifications about incidents should be sent automatically.

DATA EXCHANGE
The Connector must receive data from an enterprise backend system, either through a push mechanism or a pull mechanism. The data can be provided via an interface or pushed directly to other participants. To do so, each Connector must be uniquely identifiable. Other Connectors may subscribe to data sources, or pull data from these sources. Data can be written into the backend system of other participants.

3.2.5 VALUE ADDING APPS

Before or after the actual data exchange, data may be processed or transformed. For this purpose, the Industrial Data Space offers Data Apps. Each Data App have a lifecycle, spanning its implementation, provision in the App Store, and installation and support. The App Store should therefore be clearly visible and recognizable to every participant.

DATA PROCESSING AND TRANSFORMATION
A data processing app (subtype of a Data App) should provide a single, clearly defined processing functionality to be applied on input data for producing an expected output. A data transformation app (subtype of a Data App) should be able to transform data from an input format into a different output format in order to comply with the requirements of the Data Consumer (without any substantial change made to the information contained in the data; i.e., loss-less transformation).

DATA APP IMPLEMENTATION
The developers of Data Apps should be able to annotate the software with metadata (about exposed functionality and interfaces, pricing model, license, etc.). Data Apps must explicitly define their interfaces, dependencies, and access requirements.

PROVIDING DATA APPS
Any authorized Data App developer may initiate a software provision process (App Store publication). Prior to publication in the App Store, Data Apps must pass an optional evaluation and certification process controlled by the Certification Body. The App Store should support authorized users in their search for a matching application in an adequate fashion. Access of privileged users (e.g., administrators or operators) should require strong authentication (e.g., 2-factor authentication).

INSTALLING AND SUPPORTING DATA APPS
A dedicated Connector service should support authorized users in (un-)installing Apps not originating from an official App Store. A dedicated Connector service should support authorized users in searching, installing, and managing (e.g., removal or automated updates) Apps retrieved from an App Store.

3.2.6 DATA MARKETS

Some of the data which will be exchanged trough the Industrial Data Space also have a financial value. Therefore, the Industrial Data Space also have to integrate Data Market concepts like clearing and billing, but also governance.

CLEARING & BILLING
The data owner is able to define the pricing model and price. For example: Pay per transfer, pay for access per day/month/year etc. Any transaction of participants can be logged. There must be a simple & standardized clearing process included in the data exchange.

USING RESTRICTIONS, LEGAL ASPECTS AND GOVERNANCE
Governance in the Industrial Data Space splits into five aspects: data as economic good, data ownership, data sovereignty, data quality and data provenance. More information about data governance is given in Section 4.3.

3.3 PROCESS LAYER

The Process Layer specifies the interactions taking place between the different components of the Industrial Data Space. It thereby provides a dynamic view of the Reference Architecture Model. In the following, three major processes of the Industrial Data Space are described, involving all roles introduced in the Business Layer section:

1. providing data,
2. exchanging data, and
3. publishing and using Data Apps.

The processes are illustrated using the Business Process Modeling Notation (BPMN).

3.3.1 PROVIDING DATA

The overall process of providing data consists of four main steps, as illustrated in Figure 3.3. To provide data in the Industrial Data Space, a Data Provider first must describe the data source (e.g., a backend system of the enterprise) in accordance with the Information Model (see Section 3.4), using optionally generic and/or domain-specific vocabularies offered by a Vocabulary Provider (see "Describe Data Source" sub-process illustrated in Figure 3.4).

The result of this sub-process is a metadata object describing the data source and including data usage policy information.

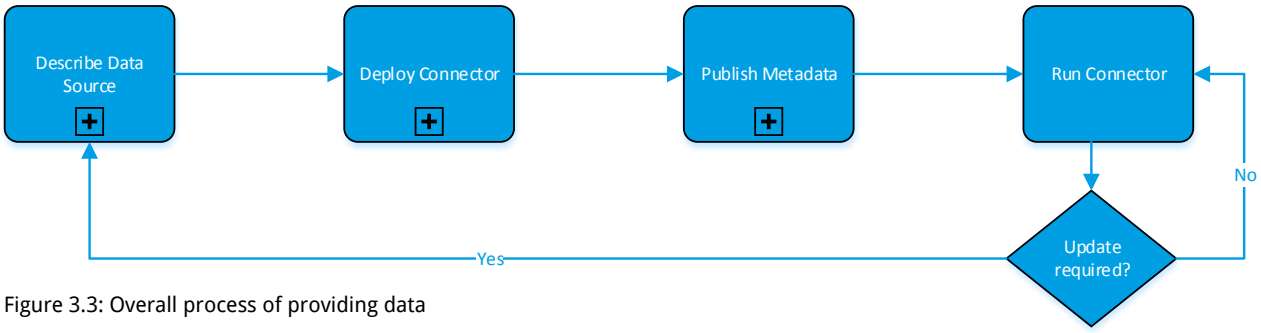


Figure 3.3: Overall process of providing data

The metadata describing the data source is part of the configuration of the Connector, and mandatory for the Connector to be deployable. This might include activities such as defining a connection to the data source, deploying a System Adapter inside the Connector, or configuring and using data processing and transformation Apps.

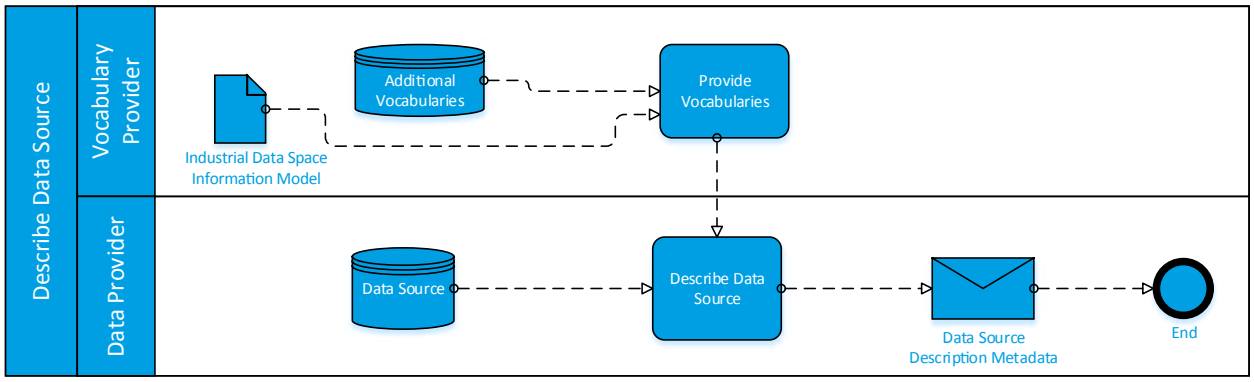


Figure 3.4: „Describe Data Source“ sub-process

All these activities result in a configuration model, which constitutes the basis for the deployment of the Connector.

The “Deploy Connector” sub-process is shown in Figure 3.5

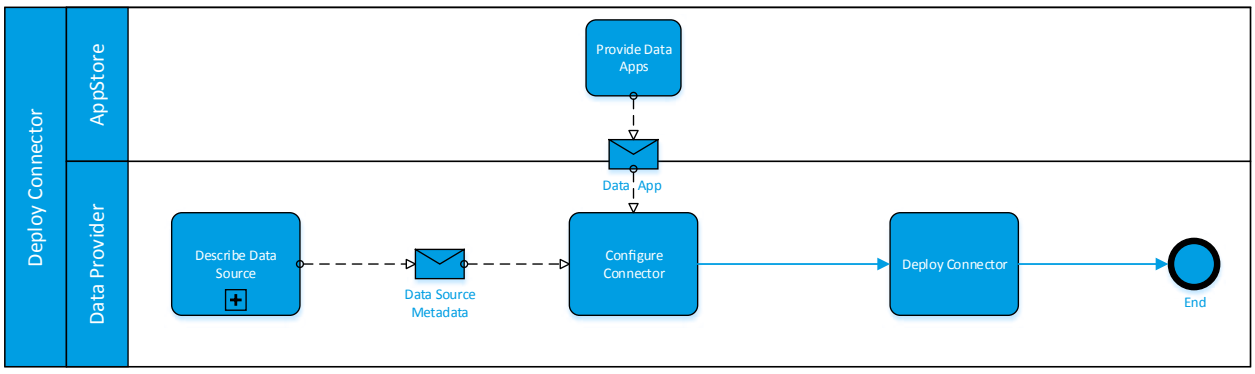


Figure 3.5: „Deploy Connector“ sub-process

After deployment, the Connector sends metadata about the data source to the Broker Service Provider.

The Broker Service Provider indexes the metadata and returns an acknowledgement of receipt to the Connector. This acknowledgement of receipt may include an identifier generated by the Broker Service Provider for unambiguous identification of the data source. After the Connector has been successfully deployed, the Data Provider must run and maintain the Connector in order to make sure it is able to handle data requests. The "Publish Metadata" sub-process is illustrated in Figure 3.6.

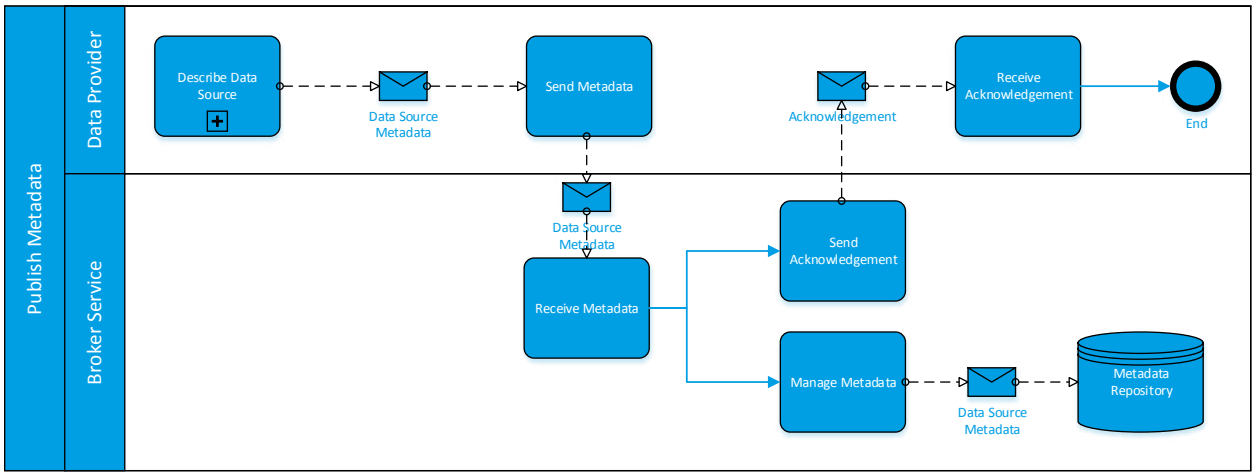


Figure 3.6: „Publish Metadata" sub-process

3.3.2 EXCHANGING DATA

The overall process of exchanging data consists of three sub-processes, as illustrated in Figure 3.7. It starts with a Data Consumer searching for a Data Provider. If the provider is found (or already known beforehand), the Data Consumer and the Data Provider can start to exchange data.

The final sub-process is the logging of the transaction details at a clearing house.

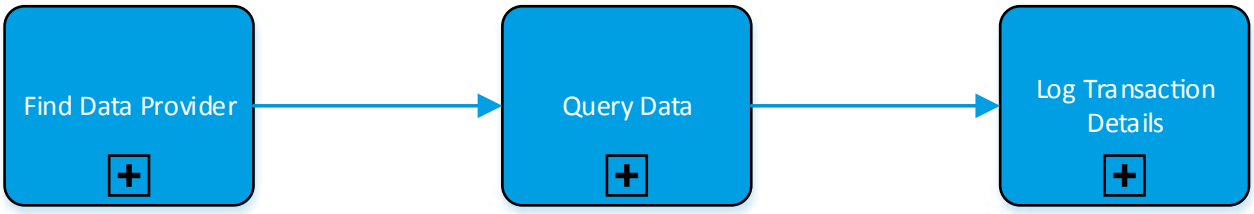


Figure 3.7: Overall process of exchanging data

To find a Data Provider, the Data Consumer must send a query to a Broker Service Provider. The Broker Service Provider then compiles a list of metadata describing different data sources in the Industrial Data Space, and sends this information back to the Data Consumer. From this list, the Data Consumer selects the Data Provider deemed most suitable. If the Data Provider is already known to the Data Consumer, the Data Consumer can configure its Connector to directly connect to the corresponding Connector of the Data Provider. The "Find Data Provider" sub-process is shown in Figure 3.8. A sub-process that is not shown here is the establishment of a legal agreement between the Data Provider and the Data Consumer. This sub-process is omitted because it lies beyond the scope of the current version of the Reference Architecture Model (upcoming versions may include functions to establish legally binding contracts between Data Consumers and Data Providers; e.g., in the form of one-click agreements). Also omitted here is the process of orchestration of the data flow inside the Connector, as it can be very complicated (the data provided by the external partner may have to be integrated with data from other external or internal sources; part of this step may be the use of Data Apps for data transformation or processing; this sub-process is described in the following).

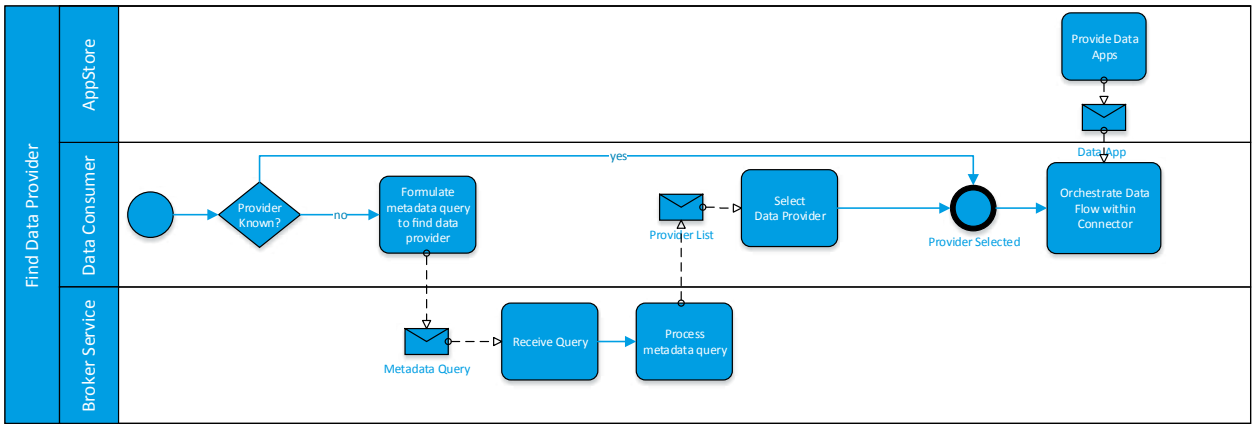


Figure 3.8: „Find Data Provider" sub-process

Data usage policy information is an important element of legal agreements and is therefore modeled as first-class objects on the Information Layer (see Section 3.4). The handling of data usage policy information is shown in detail in the "Query Data" sub-process (Figure 3.9). Usage policies are extracted from the query result, i.e., the data to be provided to the Data Consumer. In an automated negotiation process performed by the usage control frameworks of the participating Connectors, the Data Consumer and the Data Provider need to agree on a data usage policy. If an agreement has been reached, this policy is instantiated and deployed inside both Connectors. The Data Provider then sends the result of the query (i.e., the payload) to the Data Consumer.

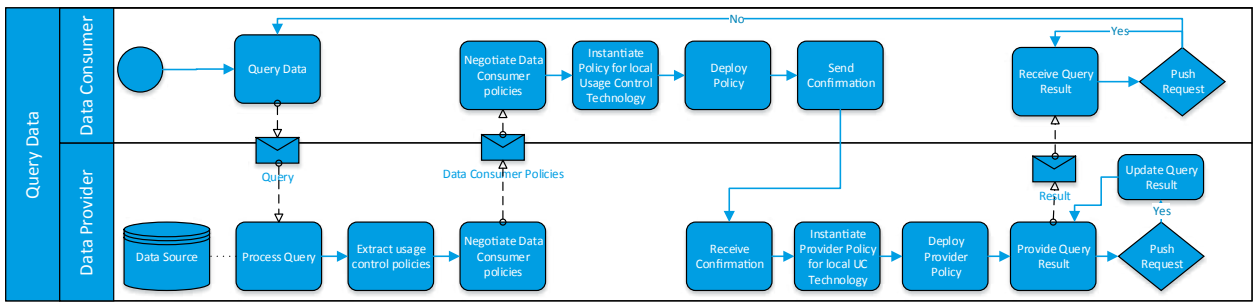


Figure 3.9: „Query Data" sub-process

Communication between the Connectors can be asynchronous (i.e., the Data Consumer does not have to wait in idle mode for the result to arrive, but will be notified by the Data Provider as soon as the result is available). Instead of a pull request, a push request can be sent, which means that the Data Consumer asks for updates regarding the requested data. The updated query results can be provided either after certain events (e.g., after the data mhas been updated by the Data Provider) or within certain time intervals (e.g., every five minutes). If a push-request is made, the Data Consumer repeatedly receives updated query results from the Data Provider.

In case of a pull-request, the Data Consumer can repeat the last part of the process to query data again (using the same or a different query).

The final sub-process of the total process of exchanging data is the logging of the transaction details at the Clearing House (see Figure 3.10). To do this, both the Data Consumer and the Data Provider must send a message to the Clearing House, confirming the transaction was successfully completed. To keep track of what kind of data was requested and what result was sent, the query information and the result (or metadata about it) are also logged by the Clearing House.

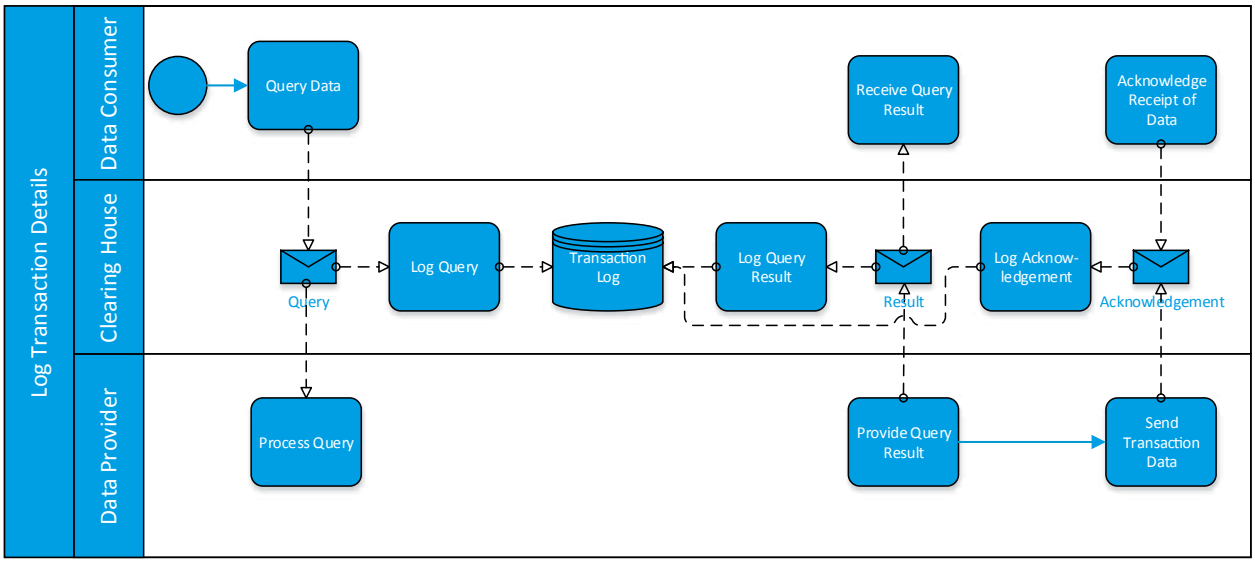


Figure 3.10: „Log Transaction Details“ sub-process

3.3.3 PUBLISHING AND USING DATA APPS

Data Apps can be used by Connectors for specific data processing or data transformation tasks. They can perform tasks of different complexity, ranging from simple data transformation to complex data analytics. An example of data transformation may be a Data App parsing a single string field with address information and producing a data structure consisting of street name and number, zip code, name of the city, and name of the country. Another example may be map matching (i.e., matching of geographical coordinates consisting of latitude and longitude of an address or a street section).

With regard to Data Apps, two sub-processes are relevant. First, a Data Apps needs to be published by an App Provider so that it can be offered in the App Store. Some Data Apps may require certification from a Certification Body in order to be published (see Figure 3.11).

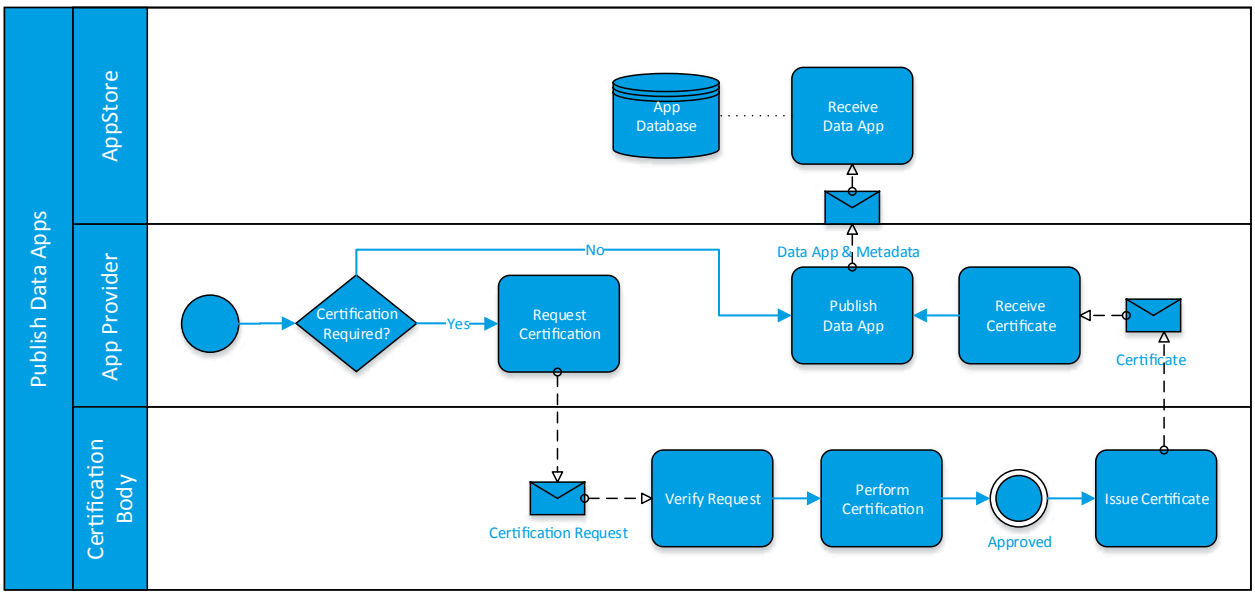


Figure 3.11: „Publish Data App“ sub-process

For each Data App that was successfully certified, the corresponding metadata is stored in the App Store for being retrieved by users (e.g., Data Consumers or Data Providers) via a search interface. Searching for a Data App is part of the second sub-process, "Use Data App" (Figure 3.12). If a user finds a suitable Data App in the App Store, the App can be requested. The App Store then offers the user a contract based on the metadata defined by the App Provider. This contract includes a pricing model, but also license information, data usage policy information, and information about resources required (this process is very similar to the process of granting access permissions when downloading an app to a mobile phone).

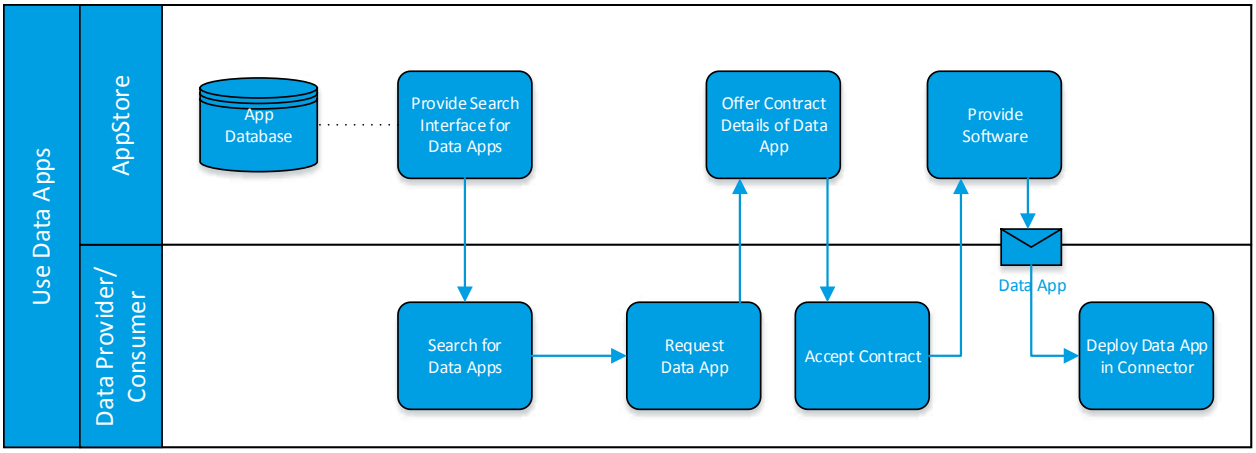


Figure 3.12: „Use Data App" sub-process

The user then has two options: to accept the contract or to reject it. If the user accepts the contract, the App Store provides the user with the selected App (i.e., the App is deployed inside the user's Connector).

3.4 INFORMATION LAYER

The Information Layer specifies the Information Model, defining the domain-agnostic lingua franca of the Industrial Data Space. The Information Model constitutes a central agreement shared by its participants and components, facilitating compatibility and interoperability. As part of the Reference Architecture, it provides a framework enabling the implementation of a variety of concrete architectures compliant with the reference architecture.

3.4.1 SCOPE

The Information Model primarily aims at describing, publishing and detecting data products (Data Assets) and reusable data processing software (Data Apps) in the Industrial Data Space. Data Assets and Data Apps are the core resources of the Industrial Data Space, and are hereinafter referred to as resources. By means of a structured semantic annotation it is ensured only relevant resources are provided (i.e., resources appropriate to meet the requirements of the Data Consumer). Once the resources are identified, they can be exchanged and consumed via semantically defined service interfaces and protocol bindings in an automated way. Apart from those core commodities, the Information Model describes essential properties of Industrial Data Space entities, its participants, its infrastructure components, and its processes.

The Information Model is a generic model, with no commitment to any particular domain. Domain modeling is delegated to shared vocabularies and data schemata, as provided e.g. by domain-specific communities of the Industrial Data Space. The Information Model does not provide a meta-model for defining custom structured datatypes comparable to the OData or OPC-UA standards. Considerations beyond the scope of modeling digital assets and their interchange are considered out-of-scope. The Information Model does not deal with the side effects of data exchange (on Data Consumer's side), for example in scenarios where data is used for real-time machine control. RPC (remote procedure call) semantics of data messages is also not covered by the Information Model.

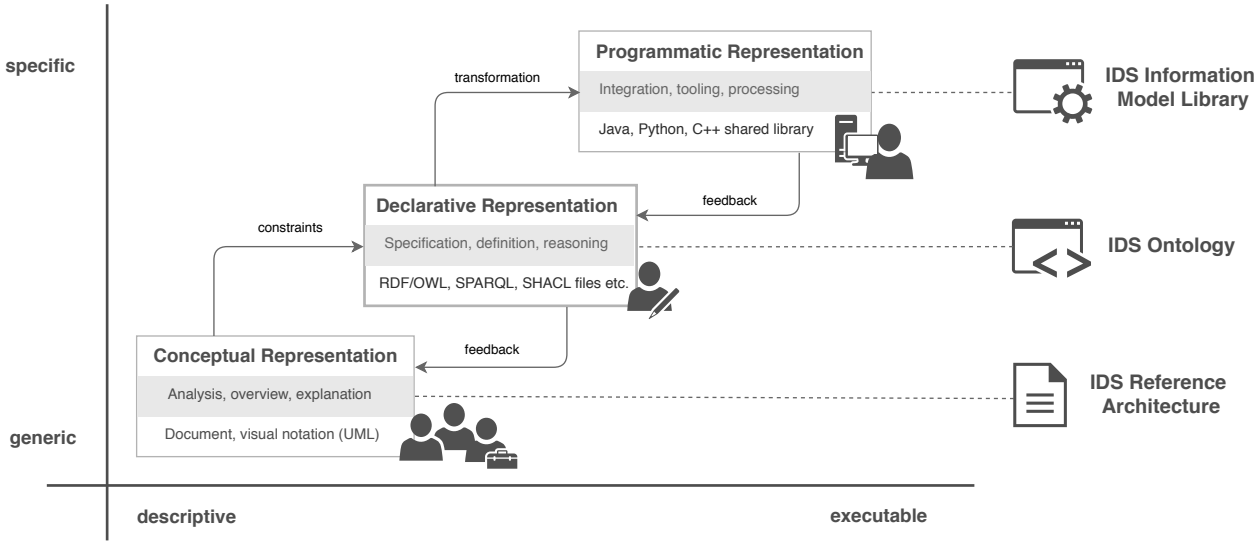


Figure 3.13 Representations of the Information Model

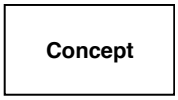
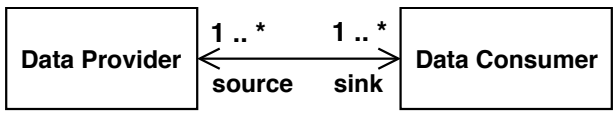
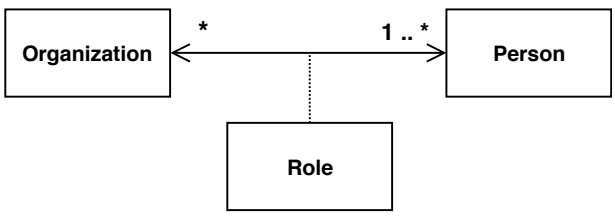
3.4.2 REPRESENTATIONS

The Information Model has been specified at three levels of formalization. Each level corresponds to a digital representation, ranging from a high-level, conceptual document up to the level of operational code, as depicted in Figure 3.13.

The Declarative Representation (IDS Vocabulary) is the only normative specification of the Information Model. A set of auxiliary resources, among others, guidance documents, reference examples, validation, and editing tools is intended to support a competent, appropriate, and consistent usage of the IDS Vocabulary.

CONCEPTUAL REPRESENTATION
The Conceptual Representation of the Information Model (CRIM) provides an analysis and a high-level overview of the main, largely invariant concepts of the Information Model, with no commitment to a particular technology or domain. It mainly targets the general public as well as management boards of organizations by means of a textual document and an understandable visual notation. The descriptions at this level are generic, providing basic information, allowing comparative analysis, and promoting a shared understanding of the concepts. References to the Declarative Representation and the Programmatic Representation are provided, encouraging the reader to take a look into these representations as well in order to learn more details.

VISUAL NOTATION
Alongside with figurative, explanatory images, a simplified version of UML class diagrams is used throughout this section, as given in Table 1.

Element	Description
	**CLASS DIAGRAM** Class diagrams represent in the context of this document concepts with no immediate correspondence to data types defined by the concrete representations of the Information Model. Class attributes are designed as external, associated entities for readability purposes.
	**ASSOCIATION** Association lines represent relationships among concepts. Optional arrow heads indicate the orientation (navigability) of the relation, when appropriate. Optionally, the role and cardinality (multiplicity) of the involved concepts is depicted at the respective end. In the given example, the Data Consumer class represents a data sink of a Data Provider, while there is at least one instance of each concept present.
	**ASSOCIATION CLASS** Association class is a standard UML class with the particular purpose of providing additional information about a relationship between two classes. The association class in the given example describes the nature of the Role a Person plays within an Organization.

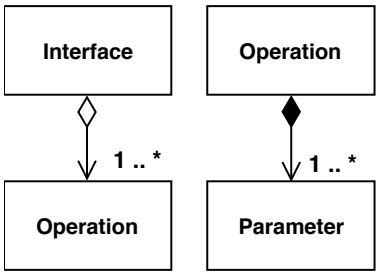
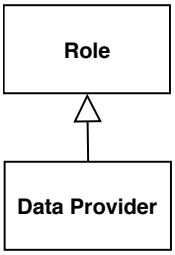
Element	Description
	**AGGREGATION AND COMPOSITION** Aggregation indicates a containment relation among a part (arrow end) and the whole (unfilled diamond end). The parts exist independently of an aggregate. The composition is a stronger type of association. The composite governs the life-cycle of its components, which may not exist independently of the whole (filled diamond end). Paraphrasing the shown example, an Interface contains Operations and both may exist (as concepts) independently. On the contrary, Parameters make only sense as part of an Operation.
	**GENERALIZATION** Relationship indicating inheritance, i.e., the hierarchy of concepts. The closed head arrow points from the sub-class to its super-class (generalization).

Table 1: Elements of the visual notation

DECLARATIVE REPRESENTATION
The Declarative Representation of the Information Model (DRIM) defines the normative Information Model of the Industrial Data Space. It has been developed along the analysis, findings and requirements of the Conceptual Representation. Based on a stack of W3C technology standards (RDF, RDFS, OWL, etc.) and standard modeling vocabularies (DCAT, ODRL, etc.), it provides a formal, machine-interpretable specification of concepts envisaged by the Conceptual Representation. Furthermore, it details out and formally defines entities of the Industrial Data Space in order to be able to share, search for, and reason upon their structured meta-data descriptions. As such, it comprises a complete referential model allowing to derive a number of Programmatic Representations. The ontology is typically used and instantiated by Knowledge Engineers, and Ontology Experts. The Declarative Representation defines a reusable, domain-agnostic "core model". It relies on third party standard and custom vocabularies in order to express domain-specific facts. According to the common practice, existing domain vocabularies and standards are reused where possible fostering acceptance, and interoperability.

PROGRAMMATIC REPRESENTATION
The Programmatic Representation of the Information Model (PRIM) targets Software Providers (developers) by supporting a seamless integration of the Information Model with a development infrastructure they are familiar with. This representation comprises a programming language data model (e.g., Java, Python, C++ classes) shipped as a set of documented software libraries (e.g., JAR files). The Programmatic Representation provides best-effort mapping of the Declarative Model onto native structures of a target programming language. This approach supports type-safe development, well-established unit testing, and quality assurance processes. It allows developers to easily create instances of the Information Model that are compliant with the Declarative Representation, relieving them from intricacies of the RDF graph model and ontology processing.

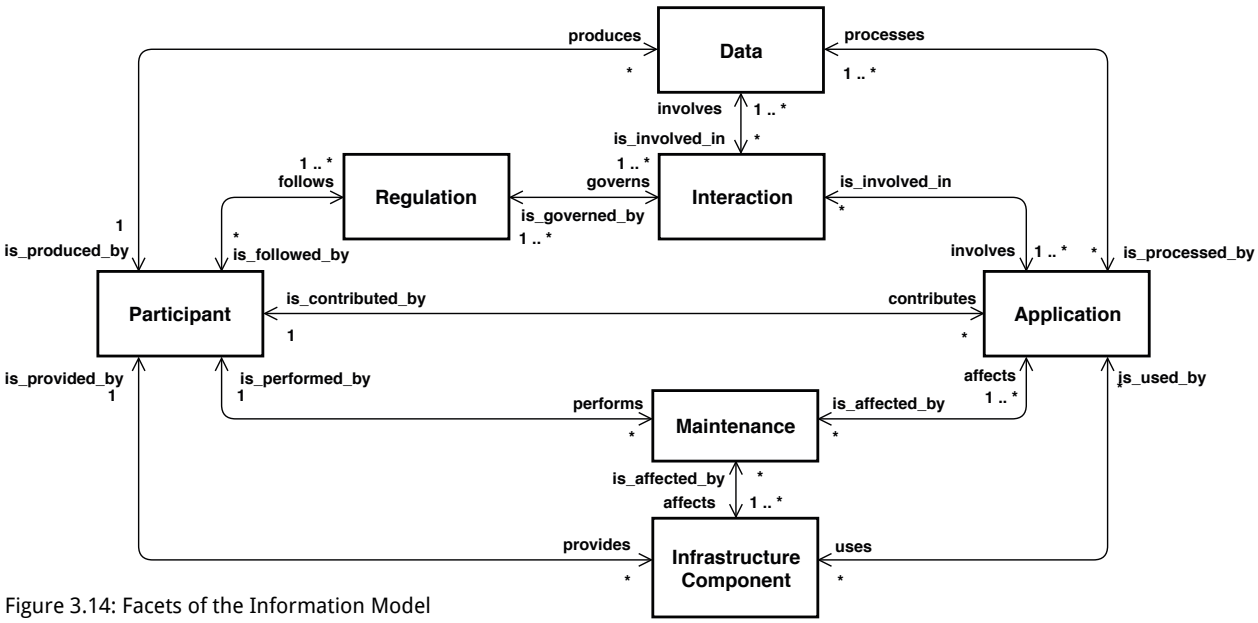


Figure 3.14: Facets of the Information Model

3.4.3 FACETS

Each of the three representations outlined above expresses the complete Information Model in a particular way. For the purpose of this document, the overall view was divided into logical groups of concepts (hereinafter called facets), addressing different static and dynamic aspects of the Industrial Data Space. Most of the facets were named after the key concept they contain. Unqualified citations refer to the concept, unless explicitly addressing the facet. Facets do not necessarily correspond to a physical organization of the model (e.g., modules or namespaces), but rather identify the core assets and the different modeling concerns:

- **Resource:** Concepts related to the description, provision, commoditization, and usage of resources, i.e., Data Assets and Data Apps, exchanged as digital commodities by participants of the Industrial Data Space.
- **Data:** Concepts particular to Data Assets, beyond the scope of general resources
- **Service:** Concepts particular to Data Apps, beyond the scope of general resources that are installed within the infrastructure in order to communicate or process data on behalf of participants of the Industrial Data Space.

The following Information Model facets deal with the description of entities constituting the Industrial Data Space:

- **Infrastructure:** Concepts related to description and verification of certified components used by participants in the Industrial Data Space in order to perform business interactions, or be managed as part of maintenance processes.
- **Participant:** Concepts related to the description, and verification of legal or natural persons that interact using the infrastructure of the Industrial Data Space, assuming certain roles and adhering to formal regulations.
- **Regulation:** Concepts related to the description, formal definition, and enforcement of contracts and usage policies governing the interactions of participants and their use of resources.

The remaining facets deal with the description of dynamic scenarios, i.e. the value generating interactions and the maintenance of internal resources and the IDS infrastructure:

- **Interaction:** Concepts related to description, instantiation, and evolution of business interactions between participants of the Industrial Data Space, leading to the exchange and consumption of resources in compliance with defined regulations.
- **Maintenance:** Concepts related to the description, execution, monitoring, and clearing of the operational processes within the infrastructure of the Industrial Data Space and the life-cycle management of resources.

Figure 3.14 illustrates the facets, depicted as high-level concepts, involved in various relationships. Being a mere abstraction, the resource facet was omitted from the figure. A set of illustrative examples will be introduced per facet in order to motivate and demonstrate its application.

These examples are reused as a reference across the representations of the Information Model and expressed as ontology instances (DRIM) or Java objects (PRIM).

FACET 1: RESOURCE

The resource concept is the root of a simple taxonomy of Industrial Data Space assets, comprising the Data Asset and Data App concepts (see Figure 3.15). A resource, as defined here, is an identifiable, valuable, digital (non-physical) commodity traded and exchanged among participants by means of the infrastructure of the Industrial Data Space. Examples of Data Assets are, among others, textual documents, time series of sensor values, communication messages, archives of image files, or media streams. Data Assets are subject to forwarding, processing, or consumption, with a particular demand for the modeling of related aspects (i.e., context and provenance, structure and usage control). On the contrary, the usage of Data Apps is rather straightforward and largely determined by their functionality. The Data App concept therefore emphasizes a formal description of the function, deployment prerequisites, and maintenance life-cycle (updates).

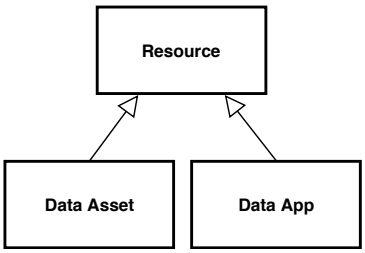
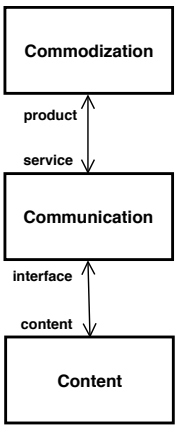


Figure 3.15: Taxonomy of the resource concept



Despite these differences, both resource types, the Data Asset and Data App, may uniformly be modeled in their capacity as a shared, digital commodity. As depicted in Figure 3.16, a stratified approach was chosen in order to disaggregate the spectrum of concerns related to their interchange. It resulted in the definition of dedicated views looking at the Content, Communication, and Commodization of resources (here termed as “3C Principle”).

Figure 3.16: Views of the resource (3C principle)

The Content View describes the inherent substance of a resource. The Communication View defines the means to communicate that content in terms of service operations. Legal, contractual, and commercial aspects complementing the resource concept are described by the Commodization View. Each view introduces a particular, new perspective on the resource. In order to cope with its complexity, a view may be refined into complementary layers, each one providing level of detail that build upon another, as illustrated in Figure 3.17.

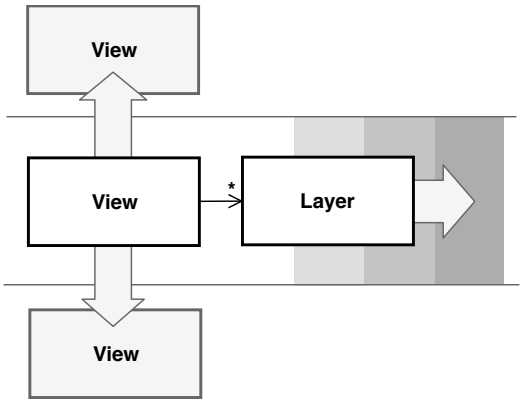


Figure 3.17: Relation of Views and Layers

CONTENT VIEW

The Content View considers the resource per se, regardless of its distribution, at three distinct layers (see Figure 3.18). The kind layer addresses the abstract content structure, e.g. „image“, „table“, „data record“, „service“, or collection of above, independently of a physical representation.

The Representation Layer concretizes a related content kind by introducing further dimensions and constraints unique to its particular serialization, e.g. JPEG image, Excel sheet, SenML XML document or Debian software package. Both layers represent prototypical „blueprints“ of content, i.e., a set of virtual instances that may comply with those models. The Artifact layer concentrates on individuals (deliverable artifacts), and such it allows to express aspects that are specific to a concrete resource instance, e.g., a particular document, image or service.

Figure 3.18: Layers of the Content View

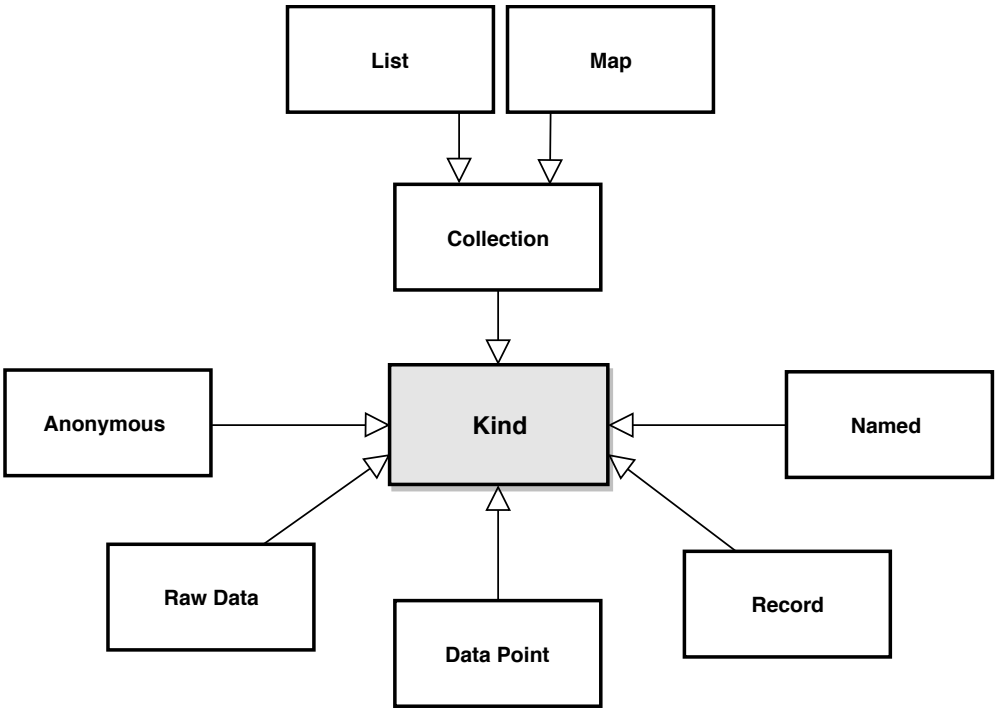


Figure 3.19: Partial taxonomy of content kinds

KIND LAYER

For modeling purposes, different, generic kinds of content are assumed. Named kind of content has a permanent identifier that is unique within a context or collection. There is no identifier (name) for anonymous content. Both kinds of content are disjoint, i.e., there is no single entity that is simultaneously an instance of both concepts.

Raw Data is an opaque sequence of bytes which is either bounded (e.g., a binary file) or unbounded (e.g., a media stream). No assumptions are made about its internal nature. Data Point consists of a single, primitive value which is an instance of a simple, basic data type. Record corresponds to a complex data type composed of nested structures and terminal primitives.

Collections are a utility kind of content used to internally organize and enable access to groups of the aforementioned content kinds without interfering with the definition of the included melements. Lists are collections ordered according to a sort criterion allowing for a position/index-based access to elements, their sorting and grouping. Lists of resources may be ordered according to one or more dimensions. Data Points are usually ordered by time(stamps) or the element values. Records further allow for ordering by the attributes of embedded structures. Standalone elements (files) allow for ordering by file properties. Maps are collections that support a random, key-based access relying on a persistent identifier given to a resource. Whereas the concepts of Raw Data, Data Point and Data Record distinguish different levels of structuring – which often coincide with various stages of processing that data has undergone – collections are generic containers for bundling those kinds of content. Standardized serializations of the collection concept should be defined to comply with the respective representation of content.

The content kind of a resource and the type of collection determine the strategies to address the resource, or to select a range of (one or more) elements out of the collection. Table 2 summarizes some envisaged reference strategies.

Referencing strategy	Description
Reference by ID	A standalone resource, or an element of a Map, is referred to by its unique name (identifier)
Reference by index	An element of a List is referred to by its absolute numeric position (index)
Selection by volume	A range of an ordered data continuum (List or stream of Raw Data) is selected by data volume (e.g., every 5 MB)
Selection by time	A range of a time-ordered data continuum is selected by a time instant (index) or time range
Selection by count	A range of ordered data items is selected by counting (e.g., every 10,000 items)

Table 2: Summary of referencing strategies

The following table summarizes the relation of the content kind, Collection type, referencing strategies, and operations available.

Content kind	Properties
Raw data	Opaque sequence of bytes (e.g. binary file or media stream) • Access by ID, if named • Access by time (range, instant) or volume, if unbounded • Operations: No filtering, no grouping, no sorting
Value collection	Collection of transient, anonymous Data Points or Records (e.g. sensor readings) • Access by index, volume and count, if ordered • Access by time, if time-ordered (time series) • Operations: - Listing (values) - Pagination, if ordered - Filtering, grouping, sorting, if ordered and structured
Resource collection	Collection of persistent resources, e.g. files • Access by ID • Access by index, volume, and count, if ordered • Access by time, if time-ordered • Operations: - Listing (IDs, values) - Pagination, if ordered - Filtering, grouping, sorting, if ordered and structured an or on file-property level

Table 3: Summary of referencing strategies per content kind

REPRESENTATION LAYER

The Representation Layer defines serializations, i.e. physical representations of a related content kind. For example, the „image" kind of content might be provided as a raster (JPEG, PNG, GIF) or a vector graphics Representation (SVG). Developers of a „Data App for image anonymization" might provide alternative software Representations (Windows EXE, Debian DEB, or Java JAR) supporting different software environments and operating systems. A Representation of a content kind is defined, among others, by a Data Type specified in terms of a Schema (i.e., formal description of the structure of data), and a Media Type, optionally augmented by Profiles (i.e., additional informal specifications and constraints that may apply). A Reference to a standard specifying that type of information should be provided, when existent.

The Representation might specify a Mapping to an equivalent, but syntactically incompatible serialization. The Representation may further indicate available Packaging options to combine contents into a single Archive (tar), apply Compression (gzip) and Encryption algorithms (AES).

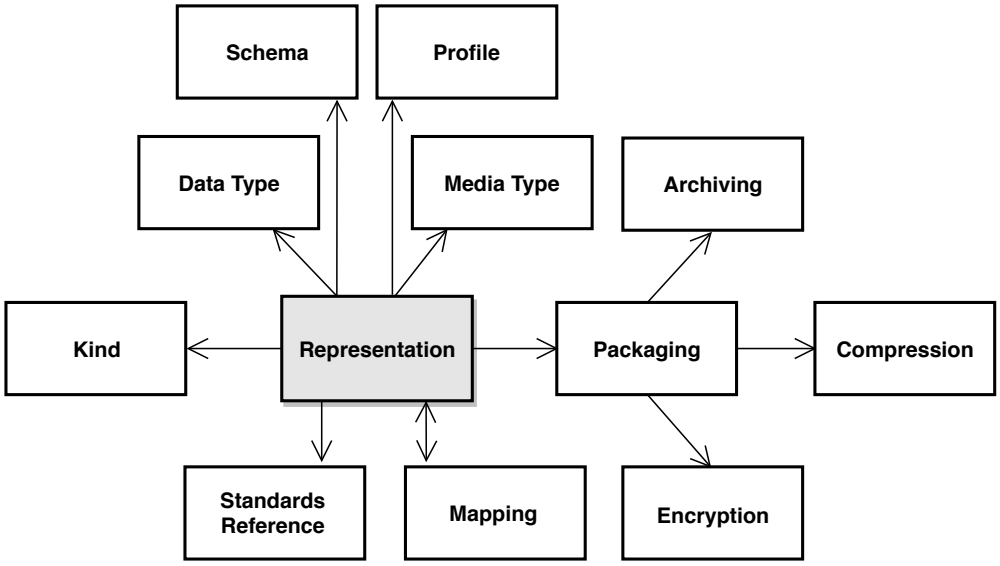


Figure 3.20: Outline of the Representation concept

ARTIFACT LAYER

The Artifact layer focuses on the description of deliverable resource instances. Going beyond the prototypical kind and representation models, it captures properties that are unique to individual materializations of the resource. Such, for example, a particular assembly of data might be individually referenced and associated with a custom Commodization model. The Artifact view of a Data App would, for example, define its inherent characteristics, the distribution size, configuration options or software dependencies etc. The previous sections introduced the content layers of a resource. Aspects that apply to a description of content in general are presented in the following. They will be augmented later on by aspects of Data Asset and Data App that apply only to the respective subclass of the resource concept.

PROVENANCE

Provenance is concerned with the origin of the content, as well as with the traceability of the processing steps the content has undergone, and finally, also with the Agents that are responsible for those Activities. The main goal of provenance tracking is to ensure the reliability of the content, so that modifications are made explicit, comprehensible and may be analyzed for defects.

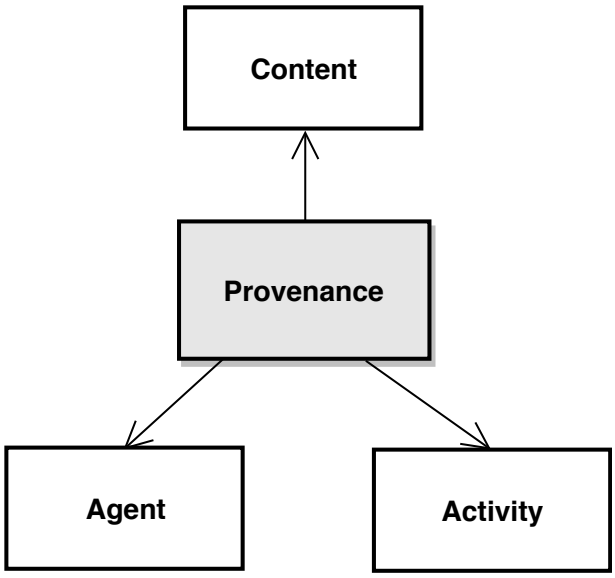


Figure 3.21: Outline of the Provenance concept

COMMUNICATION VIEW

The Communication View deals with the (dynamic) communication of resource content. Similarly to the Content View, it is defined at multiple levels of detail. The Interface Layer conceptualizes the interchange of digital artifacts as a set of uniform operations (interactions primitives). The Service Layer defines bindings of such generic operations to concrete communication protocols turning them into operable resource endpoints.

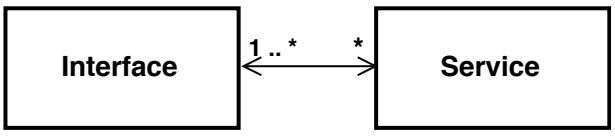


Figure 3.22: Layers of the Communication view

INTERFACE LAYER

Following the Service-oriented Architecture paradigm (SOA) this Layer defines the Interface concept comprising a set of Operations. There are multiple reasons motivating the definition of such an abstract service contract:

- Separating a service interface from its implementation is a common practice and mandated by standards like WSDL.
- A high-level description of a service interface (with a focus on functionality) allows Data Consumers to easily identify and interpret the interaction logic (i.e., operational capabilities).
- Protocol-specific interface definition languages may either not exist (e.g., MQTT), or require reverse engineering in order to infer such information (e.g., Open API).
- Conventions and best practices in resource interchange have been informally established within several technological communities (e.g., REST-architecture paradigm). The concept of an abstract, technology-agnostic interaction interface may help to formalize those implicit patterns and foster their re-usability beyond the scope of protocols originally designed for this task (HTTP).

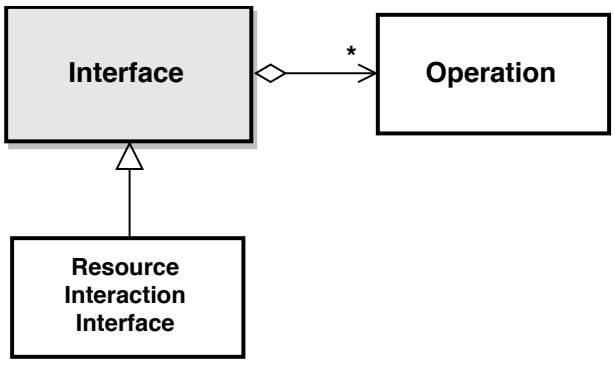


Figure 3.23: Outline of the Interface concept

Inspired by the REST-architecture paradigm the set of operations available in resource interactions has been restricted to a selection of generic, reusable interaction primitives. The expressiveness of the resultant Resource Interaction Interfaces (RII) has been purposefully limited in favor of designing simple, uniform interfaces that could be easily interpreted by generic, automated clients.

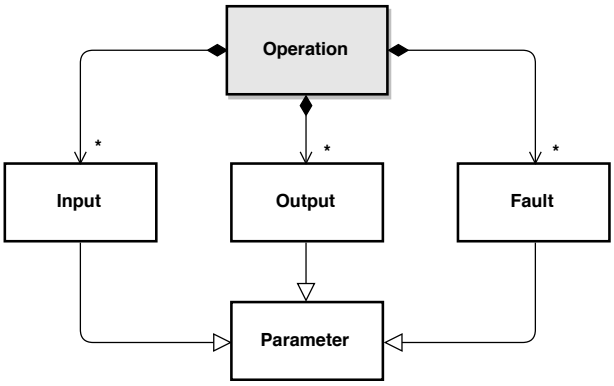


Figure 3.24: Outline of the Operation concept

OPERATIONS

Operations are the building blocks of an Interface. The operation signature lays down the expected input, its content promise (output parameters) and functional errors that might occur during the invocation (fault parameters).

Depending on operations, the interface may support various ways of data provision (Data Source), data reception (Data Sink), and meta-queries allowing the Data Consumer to introspect the interface as depicted by Table 4. Some descriptions refer to Parameter types subsequently defined in Table 5.

Parameter type	Description
Resource	Parameter used to mediate the resource content, contrasted to parameters conveying auxiliary information.
Identifier	Parameter for passing identifiers of data elements as defined by a data collection. The resource identifier is unique and valid regardless of the actual extent, ordering, and view (filtering) of the Collection.
Index	Parameter conveying the transient, positional identifier of a resource in the context of an ordered Collection. The resource index is temporarily unique and valid only with respect to the actual extent, ordering, and view (filtering) of the Collection.
Order	Parameter indicating the order of data elements when retrieving or providing a collection of data elements; either implicit, following the natural order of the collection, or based on the “Sort key” parameter; valid values are formal equivalents of “none”, “ascending”, and “descending”.
Sort key	Parameter of the “Path” type indicating the key values underlying the order of data elements in a collection; to be applied to collections of “structured” data only.
Offset	Parameter indicating the absolute offset (number of data elements to skip) within an ordered data Collection.
Limit	Parameter indicating the number of data elements retrieved or provided at once within a paginated subset (page).
Filter	Parameter holding a filter expression used to retrieve a matching subset of a collection’s data elements.
Path	Extension of “Filter” parameter supporting hierarchical, nested data structures; examples are XPath and JSONPath.
Selector	Parameter holding a selector expression used to retrieve a matching subset of a collection’s data elements.

Table 4: Resource Interaction Interface – overview of parameter types

Operation type	Description
Query parameter range	Meta-query operation used by a (potential) Data Consumer to retrieve (a dynamically generated) enumeration of input parameter values (input options); suitable for use cases in which the complete parameter range cannot be specified beforehand.
List identifiers	Extension of "Query parameter range" operation; to be used by a (potential) Data Consumer to retrieve an enumeration of available values for an input parameter of the "Identifier" type; to be applied to collections of data elements; depending on the type of collection, the identifier may be a unique name (map) or a numeric index (list); the Data Consumer may use the identifiers for a subsequent call to "Provide data" operation.
Provide data	Operation for providing data via the operation's output parameter(s) from the Data Provider to the Data Consumer; (optional) input parameters do not convey significant content and merely configure the operation's invocation; the description focuses on the Data Provider's interface: depending on its implementation, the data is either provided for retrieval upon the Data Consumer's request (PULL) or on a subscription basis (PUSH).
List data	Extension of "Provide data" operation; it is used by a Data Consumer to retrieve an enumeration of values for an input parameter of type „resource". Optional parameters of type „Order", „Sort Key", „Offset" and „Limit" may be used to create and navigate page-like groupings of data (pagination).
Filter data	Extension of "List data" operation; requires a mandatory input parameter of the "Filter" type (for example, an LDAP filter); the filter is used to provide the Data Consumer with a filtered, custom subset of the original data elements compliant with the operation's output definition; operation is to be applied to "structured" data elements or file properties of binary data elements (such as file extension, file name, file type, etc.).
Select data	Extension of "Filter data" operation; requires a mandatory input parameter of the "Selector" type (for example, a SPARQL CONSTRUCT query or a partial data template); the selector is used to provide the Data Consumer with a selective, custom view of the original data elements compliant with the expression's statement; operation is to be applied to "structured" data elements only.
Consume data	Operation for receiving data via the operation's input parameter(s) from the Data Provider to the Data Consumer; (optional) output parameters do not convey significant content and merely indicate the status of the operation's invocation; the description focuses on the Data Consumer's interface: depending on its implementation, the data is either retrieved via the Data Consumer's request (PULL) or received on a subscription basis (PUSH).

Table 5: Operation types of the Resource Interaction Interface

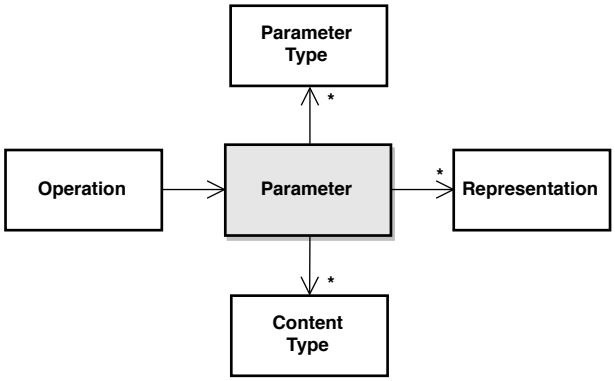


Figure 3.25: Outline of the Parameter concept

PARAMETERS

Parameters are named slots of data exchange via operations of the resource Interaction Interface. They are defined in terms of a content type, Parameter type, and a Representation (serialization). The content type designates the semantics of the data passed through (not to be confused with the homonymic HTTP header). Parameters might refer to structures of the resource content, that are mediated by the Parameter, (e.g. a table column) in order to re-use their semantics definition.

Parameters share the Representation definition provided above. This is useful when mediating transient data which is not modeled as part of the resource content. The Parameter type provides hints to interface clients about the purpose and intended usage of the Parameter, and may e.g. support e.g. a query generation process. Table 6 provides a listing of currently envisaged, standard Parameter types.

SERVICE LAYER

The resource Interaction Interface can be turned into an executable service by binding it to a concrete communication protocol. A protocol binding provides a vocabulary to map the abstract operation signatures onto the concrete structures (e.g., HTTP headers or query parameters), configuration parameters (e.g., MQTT broker), and interaction patterns (e.g., WSDL, Message Exchange Patterns) of a protocol. Each instance of a Protocol Binding defines a resource Endpoint, an addressable and operable point of resource exchange which communicates Representations of a resource in compliance with the definitions of underlying Resource Interaction Interface. The Information Model does not constrain Data Providers in the way they configure the individual protocol bindings but it should provide a guidance and example instances demonstrating a recommended practice.

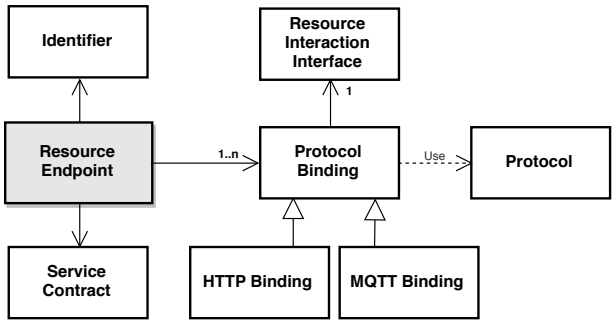


Figure 3.26: Outline of the resource endpoint concept

COMMODIZATION VIEW

The Commodization view focuses on the “commodity” aspects of a resource, amongst others its price, licensing model, and usage restrictions. It optionally lists the available Quality of Service options (per resource endpoint). Once published, the static dimensions of the Product concept are augmented by dynamic statistics and community feedback (rating, comments, etc.) represented by the Feedback concept. The Product information allows a potential Data Consumer to estimate the expenses and commercial exploitability of a resource.

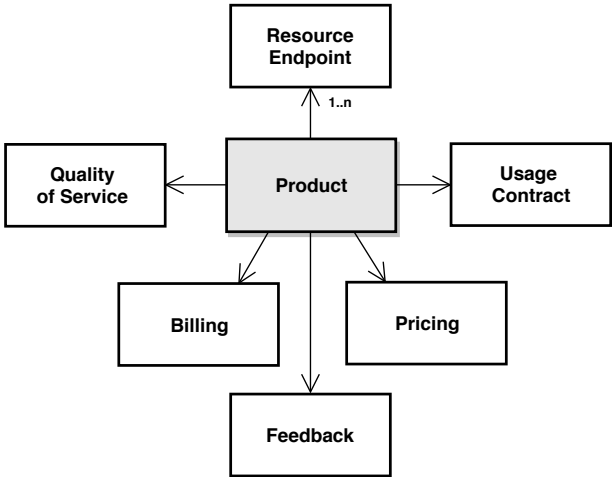


Figure 3.27: Outline of the Product concept

PRICING

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

REGULATIONS

The regulatory aspects of the Commodization view are discussed in a separate section (see above), because of their key role in implementing the data sovereignty of Data Owners and App Providers.

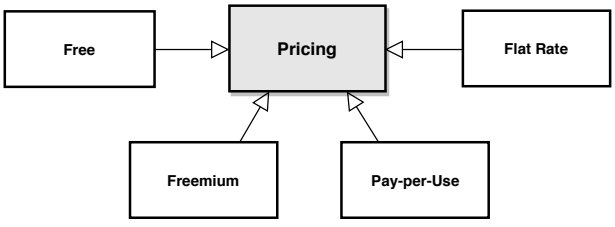


Figure 3.28: Taxonomy of Product Pricing concepts

SUMMARY

This section introduced the concept of an Industrial Data Space resource, a generalization of the core asset concepts, the Data Asset and Data App. The resource is an identifiable, valuable, digital (non-physical) commodity traded and exchanged between participants by means of infrastructure components of the Industrial Data Space. The specification of resource concept was given in terms of the Content, Communication, and Commodization views (3C-Principle). A refinement of the views by orthogonal layers lead to a complete description matrix as summarized in Figure 3.29.

FACET 2: DATA

Data is the central asset of the Industrial Data Space. This section elaborates upon the concept of a Data Asset, an identifiable, non-physical entity that comprises data, or a service interface to data. The Data Asset concept is described only in the extent going beyond the description of the parent resource concept given above. Reference examples are presented to demonstrate the concept of a Data Asset. They demonstrate differences in the provision of static data versus dynamic data, different usage policies applied, different interaction patterns chosen, and different transfer protocols used.

EXAMPLES

The reference data stems from a hypothetical scenario of measuring traffic conditions at defined locations of the highway E37 for purposes of traffic control, predictive road maintenance, toll fee optimization and so on.

Example DAT1: Off-line, free data download

The example DAT1 showcases an easy, non-interactive access to free, historical data. Monthly reports on traffic statistics collected during a year are provided for download at a fixed web address *(.../trafficreport/)*. File names (e.g., *E37_up_2018_01.csv.zip*) consist of the (underscore separated) identifier of the highway (e.g., "E37"), the direction of travel ("up" or "down", relative to highway mileage), year (e.g., "2018") and month (e.g., "01"), and (optionally) the file (csv) and compression extension (zip). HTTP content negotiation or default settings may supplement missing values for file type (Accept-header) and compression (Accept-Encoding-header). The reports comprise tabular data with a fixed number of labeled columns. Each row corresponds to an individual value tuple collected in a certain sampling area within a certain sampling period. The sampling

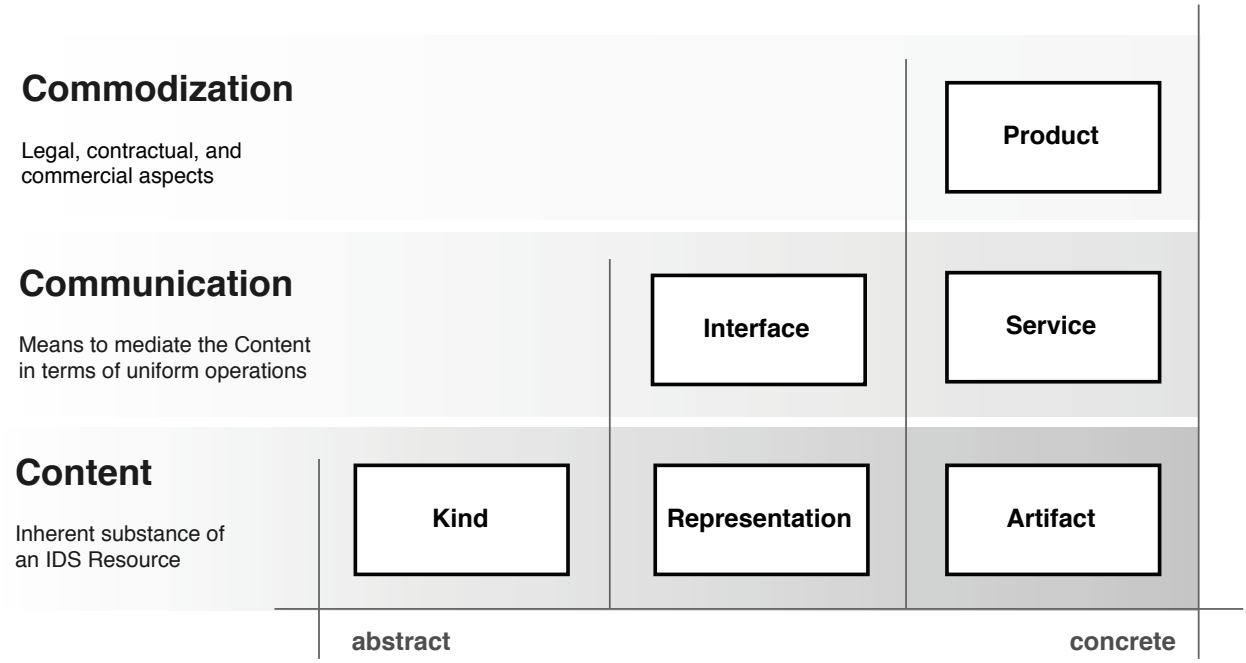


Figure 3.29: Description matrix of Industrial Data Space resources

area is identified by a readable name (String), a distance marker (double, km), and the geo-location (according to a predefined geo-spatial reference system). The remaining columns contain the measurement details, i.e. the time stamp of the sampling period (ISO 8601 period format, *YYYY-MM-DDThh:mmPnYnMnDTnHnM*), the average velocity (double, km/h), and the number of vehicles passing (integer). The data may be used free of charge, but the policy requires a credits citation.

Example DAT2: On-line, commercial data query

Example DAT2 introduces interactive features going beyond the retrieval of alternative representations of static content, allowing the Data Consumer to probe and accordingly operate services providing access to extended, growing datasets. In order to explore the dataset, the Data Consumer may request the value range of enumerable parameters (*trafficreport/column/areaId*), define valid filter conditions, and limit the report coverage to fit consumers' informational needs (*trafficreport?filter=in(areaId,[id1,id2,id3]*) in a fully automated manner. Elaborating upon the report structure of Example DAT1, the Data Consumer may learn about the available properties/columns (*trafficreport/columns*) and configure the report layout accordingly (*trafficreport?column=areaId,timestamp,avgSpeed&orderBy=areaId,timestamp&order=asc*). For some properties to be elicited, investments into dedicated sensory infrastructure may be required (e.g., weighbridge, vehicle type detection), making such values only commercially available (*avgWeight, countVehicleTypeTruck*). Pricing models may allow for discounts when combining payed properties. Depending on consumer's request behavior, various payment models may be applied (pay-per-use, volume or time-based subscription, etc.). The usage policies of this sample prohibit resale of the commercial data parts.

Example DAT3: Preprocessed, live data subscription

While data exchange in the two previous samples was driven by the Data Consumer (pull-pattern), Example DAT3 showcases a data-driven delivery, for which the Data Consumer is provided with content on the basis of a previously made subscription (push-pattern). In the context of the traffic monitoring scenario, a Data Consumer subscribes to traffic parameters, which values match a particular complex event pattern deployed on Data Provider premises as part of the subscription (see "Facet 3: Data AppsData Apps" for details on examples of such rules). The following sections summarize aspects that are considered specific to Data Assets.

DYNAMICITY
Data can differ significantly in terms of dynamicity (i.e., the way data expands and can be updated). As far as frequency is concerned, data may change spontaneously (i.e., on an irregular basis) or regularly (e.g., at a certain sampling rate). A change may represent an extension, i.e., an insertion in the middle of, or an addition at the head of, an ordered collection, a partial or complete update (replacement), or deletion of a collection item. (Continuously) extended, live collections (sensor measurements, log entries, message queues, etc.) differ from static collections. The time variance of data needs to be explicitly modeled and considered when selecting the appropriate interaction and communication protocol.

CONTEXT
The context is defined by the temporal, spatial, and socio-economical (or world) coverage of the data, i.e. the range of time, space, or real world entities referred to by the data. Accurate and meaningful context modeling gives answers to questions like “when”, “where”, and “what”, and is seen as a prerequisite for the assessment of data‘s relevance and business value with respect to the needs of Data Consumers. In the traffic scenario introduced above, the temporal context is the overall time period the data was collected in; its upper bound (end time) is undefined here because of the continuously extended live data). The spatial context of the examples may be defined by the geographical extent (union of bounding boxes) enclosing the sampling area. The world context may comprise the enumeration of the highways as a real-world objects of interest. An overly broad and excessive context description might impede the discoverability and value assessment of the Data Asset.

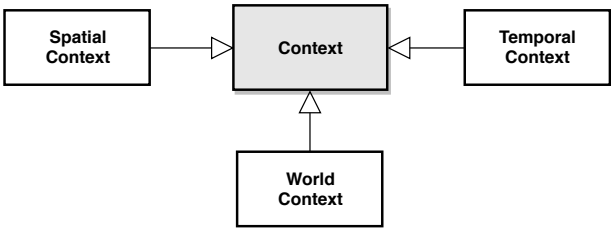


Figure 3.30: Taxonomy of the Data Asset Context

TOPIC
The topic of a Data Assets emphasizes the essential statement of the data, its purpose, or interpretation. It might express the relation of data to the world context. Topics appropriate in a given traffic scenario are, for example „monitoring“, „statistics“, etc.

FACET 3: DATA APPS
The Data Apps facet focuses on the description of reusable software and auxiliary artifacts delivering a data-specific functionality. Data Apps are self-contained and self-descriptive software packages (e.g. Linux Containers) extending the functionality of the generic Connector with custom capabilities. In addition, there are Data App Plug-ins and Data App Assets. A Data App Plug-in is an add-on of a Data App, adding new capabilities to it.

The extension management process for selection, installation, and maintenance of such plugins has to be implemented by the respective Data App in accordance with the security policies of the Connector. A Data App Asset is a machine-interpretable Data Asset, such as a script file, algorithm, rule set, or another type of code, which execution relies on a particular runtime environment.

EXAMPLES
The following reference examples demonstrate the provision, extension, and configuration of Data App logic in context of the traffic scenario.

Example DAP1: Data App for image anonymization

The photographs taken by the surveillance camera have to be anonymized before being forwarded to a Data Consumer. This sample accepts images of standard traffic scenarios in various file formats (e.g. PNG, JPG) recorded in compliance with the international norm EN 50132-7. It is trained to locate particular personal information (e.g., the license plate of a car) and to apply image processing techniques to irreversibly obfuscate this information.

Example DAP2: Data App Plugin for advanced image processing

There may be scenarios that impose advanced privacy requirements and require a dedicated plug-in to augment the aforementioned sample with a capability of advanced image processing (e.g., face anonymization).

Example DAP3: Data App Asset as interpreted CEP rule (DAP3)

The Data Consumer in the traffic scenario might define complex event processing (CEP) rules as part of a data subscription in order to shift the task of processing and monitoring live data at the edge of the network (edge computing). One such rule may request a notification sent every time the average speed in a critical area dropped below 10 km/h within the last 5 minutes (risk of congestion). Likewise, a notification is sent every time a truck weighing more than 20t heads towards a bridge that has only limited load carrying capacity (limited access). The content of the notification message, the communication protocol (MQTT), the quality of the service parameters (at-least-once delivery), and other details are defined by the rule as part of the subscription.

DIMENSIONS

In course of their life-cycle Data App may be considered according to various dimensions, as illustrated by Figure 3.31.

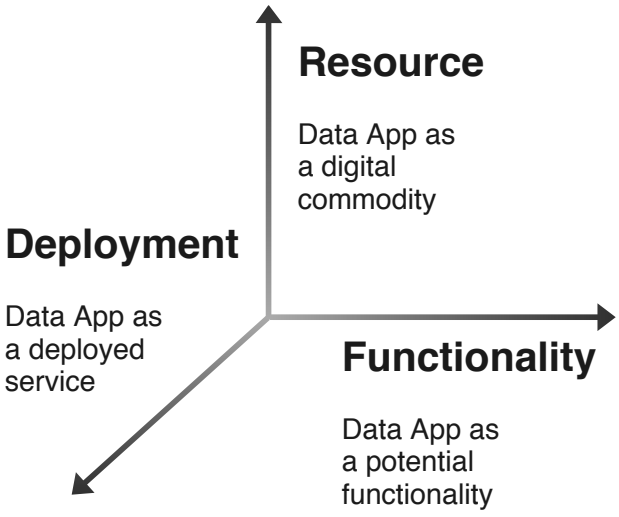


Figure 3.31: Dimensions of Data Apps

The Resource dimension, shared by Data Assets and Data Apps, specifies their quality as a tradable digital commodity according to the 3C-Principle. The Functionality dimension expresses the functional potential, i.e. data handling capabilities, of a Data App published via the App Store component. The Deployment dimension deals with the runtime aspects of a concrete Data App deployment (security updates, quality of service and usage control enforcement etc.).

RESOURCE

The following sections focus on the Resource dimension of the Data Apps. The views and layers of the 3C-Principle are instantiated according to characteristics of Data Apps.

CONTENT VIEW

The Content View considers the static, structural aspects of the Data App Resource. Its general kind is expressed by a reference to a shared taxonomy of Data App Categories, while a detailed modeling of the functionality is delegated to the Functionality dimension. The Representation Layer defines the distributions available as a combination of available software file formats and general properties of the target system (hardware architecture, operating system). Optionally, the Artifact Layer may elaborate about the structure, dependencies, configuration and requirements of a particular Data App.

The Structure concept discloses the internal software components the Data Apps uses or is based on. It allows to estimate their technical maturity, potential technical and security risks, e.g. once defects or security vulnerabilities of those components were reported. The Dependencies concept deals with the reliance on external software artifacts. The Environment concept encompasses the requirements on the execution context of the Data App, among others the runtime environment (J2EE, Linux-Container runtime), its configuration, and resources made available to the Data App (storage volume, network ports, memory, CPU). Finally, the Configuration concept describes the configuration options and default settings etc. The Signature concept covers the verifiable identity, integrity and formal IDS certification of the Artifact.

COMMUNICATION VIEW

The Communication view deals with the physical distribution of the Data App resource. Depending on the distribution strategy, a signed Data App might be provided in a decentralized manner by the App Provider, similarly to a Data Asset, or retrieved from a central App Store repository. In the former case, the App Provider has to define a resource Endpoint within a local Connector and publish it to the App Store Registry. Its resource Interaction Interface should enable the prospective Data App user to select, customize and download an appropriate Data App resource. In the latter, default case, these tasks are handled by a generic resource Endpoint exposed by the App Store Repository.

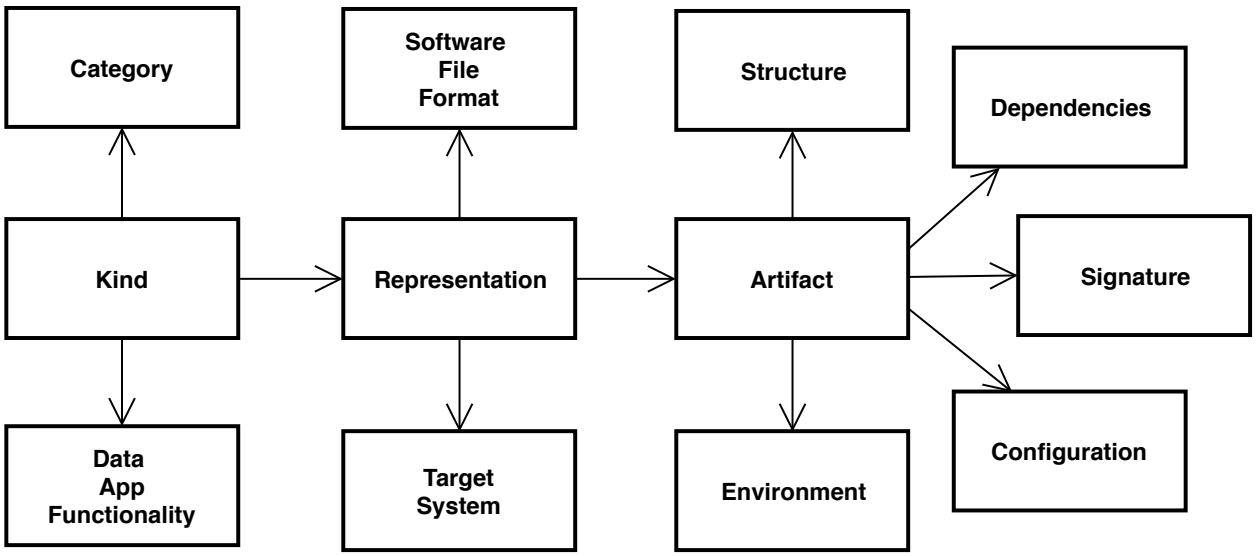


Figure 3.32: Content view of the Data App resource

COMMODIZATION VIEW

In addition to the general considerations of the Resource Commodization View, specific aspects apply for Data Apps. An obvious example are the various deployment options, as listed in Table 6. Both on-premises deployment options impose additional agreements with regard to maintenance, upgrades, and usage policy enforcement.

FUNCTIONALITY

The Functionality dimension expresses the capabilities of a Data App to handle a type of data in a particular way. The content view details out the kind and syntactic Representation of the data in question. At the definition time there are no concrete data instances to be handled, therefore the Artifact layer of the content view is omitted. Please refer to Section Content View for details. The Communication view defines a custom Data Interface in terms of Operations exposed by the Data App. Figure 3.33 summarizes the main aspects of the Functionality dimension.

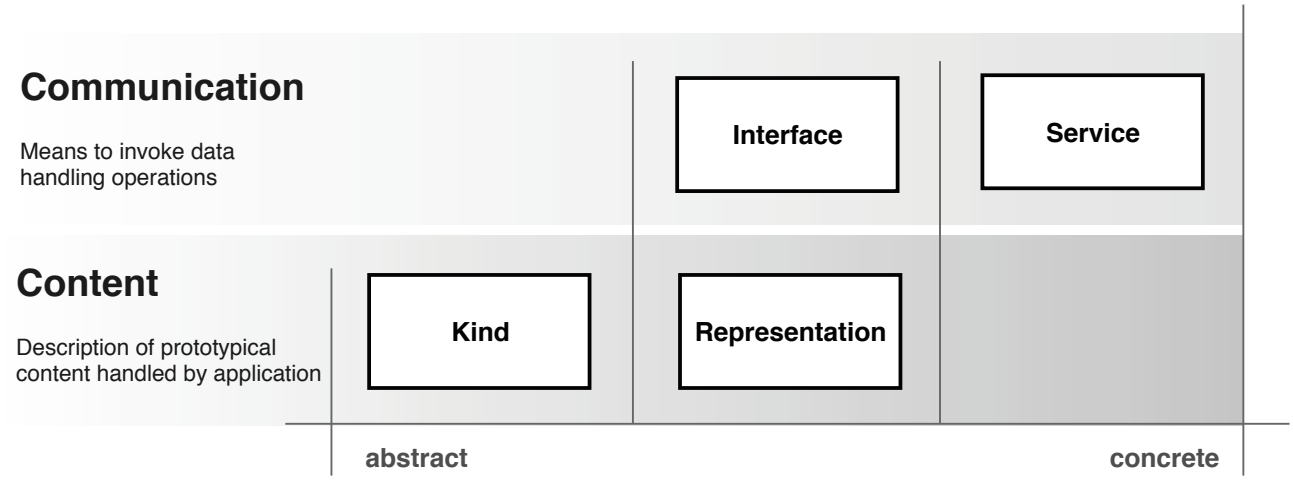


Figure 3.33: Description matrix of the Data App Functionality dimension

Deployment option	Description
On-premises installation	A Service Provider deploys the Data App inside of an on-premises IDS Connector on behalf of the Data Provider. This is assumed to be the default case.
On-premises injection	A Service Provider deploys the Data App inside of an on-premises IDS Connector on behalf of the Data Consumer (asking for customized data preprocessing, according to contract specifications; e.g., edge computing).
Remote integration	A Service Provider integrates a remote Data App service on behalf of the Data Provider. In this scenario, the Data App is hosted by different participants and used remotely.

Table 6: Deployment options of Data Apps

COMMUNICATION VIEW

The Communication view considers in this context the abstract Data Interface (Interface Layer) of a Data App and its materialization as a Data Service (Service Layer).

The Data Interface models the effective functionality of a Data App. It encapsulates a range of Operations upon data passed via the Parameters of the Operation. The semantic type of an Operation indicates the processing of and effect on input data in an interoperable way. The set of available Operation types includes the subset of Resource Interaction Interface Operation types and is deliberately not restricted. Data App Developers are free to specify custom Operation types in accordance to the Information Model governance rules. Depending on the data flow and interactions supported by the individual Operations, a Data App may act as a Data Providing App, Data Processing App or a Data Consuming App. These concepts are not disjoint, a single Data App may simultaneously implement any combination of these roles.

A Data Providing App exposes data by means of at least one Provide data Operation, as illustrated by Figure 3.35. Equally a Data Consuming App exposes at least one Consume data Operation in order to receive (and store) data. Please refer to Table 5 for a definition of those Operation types.

Data Processing Apps expose custom functionality via at least one Process Data Operation. The range of such Operation types is rather infinite, Table 7 provides some examples of possible subclasses.

At the Service layer Data Apps may require bindings to further, e.g. native protocols (IPC socket) in addition to „remote“, web-based protocols involved in exchange of resources. The corresponding requirements and examples are being collected and will be included in the next document iteration. The Service description in context of the Functionality dimension is inevitably incomplete, the Data Service model remains a template with no references to a real Deployment.

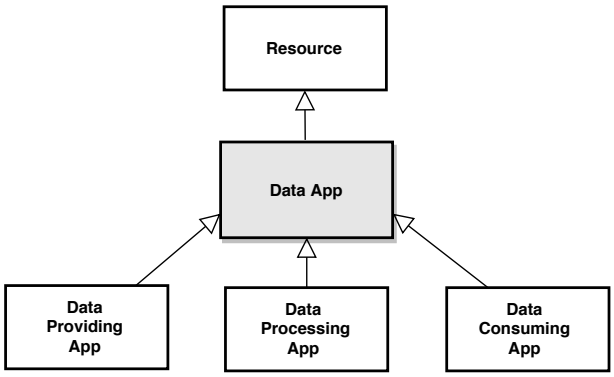


Figure 3.34: Data App taxonomy

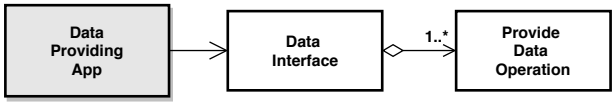


Figure 3.35: Outline of the Data Providing App concept

Process data Operation type	Description
Annonymize Data	Type of Operation used in reference example DAP1. The input and output are image files of traffic situations. Processing removes personally identifiable information (license plate).
Aggregate Data	Type of Operation used in reference example DAP3. The Input and output are event messages of a predefined type. The evaluation of sensor measurements by a complex event processing rule results in the generation of new, higher-order events.
Transform Data	Type of Operation used to transform a structured input into a semantically equivalent, but syntactically incompatible Representation.

Table 7: Examples of Process data Operation types

DEPLOYMENT

The Deployment dimension deals with concrete installations of Data Apps. A previously incomplete Data Service template becomes instantiated into a physically accessible Service model (endpoint) based on parameters of the host environment (IP address, port etc.). Data Providing and Data Consuming Apps may easily be turned into resource Endpoints by complementing their description in accordance with the 3C-Principle (e.g. by addition of the missing Product layer). The tasks to be supported by an Information Model of a Data App Deployment are, among others, the tracking of administration provenance (modifications applied to the Data App), logging of execution parameters (downtimes, usage of computational resources, service availability etc.) and the support of maintenance life-cycle (security updates etc.).

SUMMARY

This section elaborated upon the concept of a Data App, a re-usable software and auxiliary artifacts delivering a data-centric functionality. Data Apps were analysed along three dimensions. The resource dimension considers Data Apps as a tradable digital commodity according to the 3C-Principle. The Functionality dimension expresses its data handling capabilities, whereas the Deployment dimension deals with the runtime aspects of a concrete Data App deployment. Depending on the data flow and interactions supported Data App were categorized as a Data Providing App, Data Processing App, and Data Consuming App.

FACET 4: INFRASTRUCTURE

Figure 3.36 outlines a taxonomy of the main Infrastructure components of the Industrial Data Space. The Connector is its core building block, a communication server providing and consuming data by means of Data Apps via a number of resource endpoints. The Broker component is a meta-data registry of Data Asset offerings, whereas the App Store is a registry of Data App offerings and a secure registry for their distribution. The Vocabulary Hub serves the maintenance of shared vocabularies and related (schema) documents. The Identity Provider manages and validates the digital identity of Industrial Data Space Participants. The Clearing House provides clearing and settlement services B2B interactions within the Industrial Data Space.

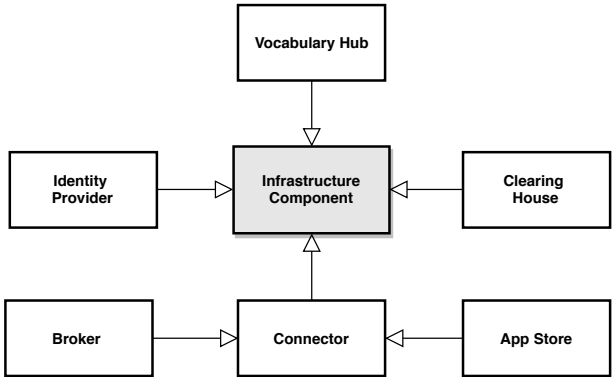


Figure 3.36: Taxonomy of infrastructure components

CONNECTOR
Being the dedicated point of data exchange and usage policy enforcement, the Connector is the central component of the infrastructure. It constitutes the basis for the implementation of other, more specialized components, such as the Broker. Each Connector may expose an arbitrary number of resource endpoints, offerings of digital commodities that are optionally advertised by publication at the meta-data registries, the Broker, or App Store respectively.

The Deployment Context of a Connector comprises the geo-location information (e.g., country of deployment or applicability of national law), deployment type (on-premises vs. cloud). Furthermore, the responsible Participant operating the Connector (Service Provider) is referenced.

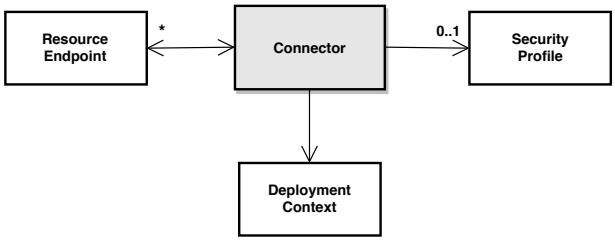


Figure 3.37: Outline of the Connector concept

A Connector may specify the supported Security Profile in order to indicate a level of technical trustworthiness. The Security Profile is composed of several security options, which are outlined in Figure 3.37, among others the capability of a remote integrity verification, applications isolation level, etc. Predefined configurations of Security Profiles should be supplied in order to identify common security levels of Connectors (e.g. Base Connector, Trusted Connector etc.).

The capabilities of enforcing data usage control by the Connector are modeled as part of the Security profile. These cover information whether and how certain usage control policies (e.g., mandatory deletion of data after a certain period of time) are automatically enforced by the Connector (i.e., on the technical level) or supported by governance processes during the data consumption process (i.e., on the organizational level).

FACET 5: PARTICIPANTS
A participant is a legal or natural person assuming a role (or more than one role) in the Industrial Data Space. For certain, critical roles to assume, participants must undergo a certification. Certification of participants is considered a measure to establish trust across the Industrial Data Space.

EXAMPLES
Instances of participants involved in the traffic scenario are outlined below.

EXAMPLE PAT1: MULTI-NATIONAL LOGISTICS COMPANY
MAIER Logistics is a multinational logistics company with hundreds of trucks driving throughout Europe. The company is interested in live traffic monitoring data, as it wants to provide its drivers with up-to-the-minute traffic information to allow for efficient routing and timely issuing of hazard warnings. In this scenario, MAIER Logistics is an organization that runs several sites, such as MAIER Deutschland, Musterstraße 5, Köln, Deutschland, or MAIER UK, Example Road 5, Liverpool, United Kingdom. The organization complies with the ISIC classification rev. 4 and has ISIC code 4923 (freight transport via road). For this scenario, the company's distribution departments are relevant, being the organizations which control and monitor outbound distribution via trucks. The distribution departments are part of the MAIER Logistics Organization. Each distribution department has a specific site. The German Distribution department is located at MAIER Logistics Distribution Cologne, Musterallee 323, Köln, Deutschland. MAIER Logistics Distribution Cologne assumes the role of a Data Consumer in data Example DAT3. It has a valid certificate and a unique identity. As a Data Consumer, it receives notifications with hazard warnings and congestion information. The information received is processed by a custom software of the department, which sends the information to the trucks using geo-location information.

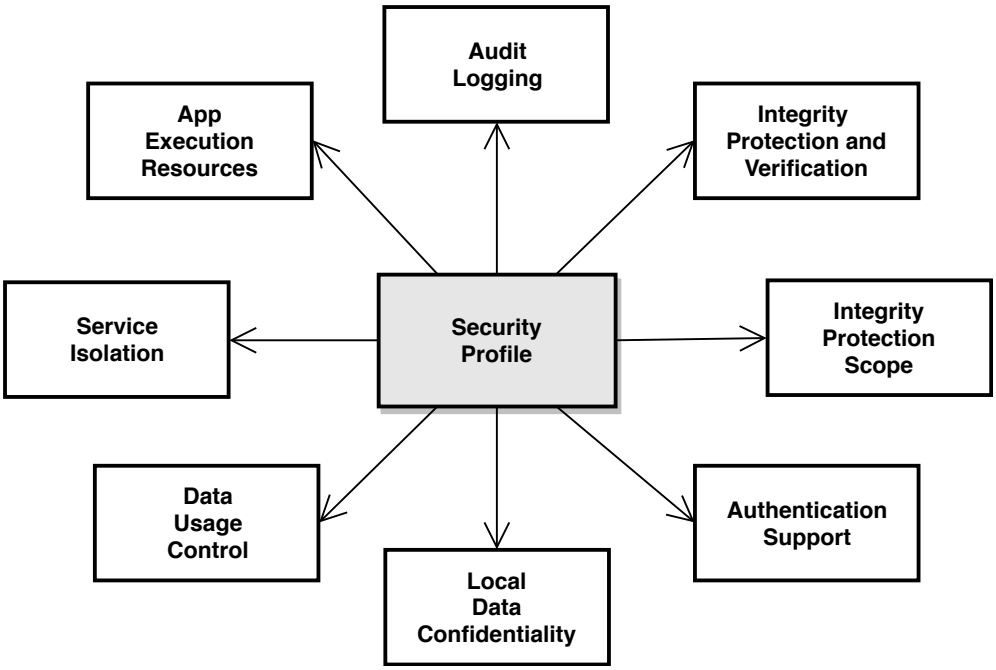


Figure 3.38: Outline of the Security Profile concept

ORGANIZATION AND PERSON
A Participant can be an organization (or organizational unit) or an individual. If the participant is an organization, it may consist of sub-organizations, departments, or other organizational structures. Corporations may indicate an organizational structure by linking to subsidiary companies or organizational units acting as related, but more or less independent participants. This approach allows sharing authorization certificates along a trust chain and enforcing company-wide policies. If the participant is an individual, he or she may assume a specific role in the corresponding organization.

BUSINESS CLASSIFICATION
Participants may indicate the type of business and the domain in which they operate by making references to established business classifications, i.e., business catalogs or registries. The classification can be used, for example, to search for data assets according to business category. For formal representation of business classifications, e.g. NAICS identifiers can be used. These are part of the extended core of the Industrial Data Space Information Model. It will therefore be possible to support additional classification schemes (such as D&B D-U-N-S® Number, ISIC, or UNSPSC) in future revisions of the extended core model.

SITE
Each Participant can be assigned to one or more unambiguously defined Sites. Site information comprises the name and address of the site as well as geo-location information. It is particularly important in cases in which specific rules (e.g., national law) apply, affecting, for example, the data usage control policy.

IDENTITY
By default, and in accordance with linked-data principles, a participant can unambiguously be identified by a dereferencable HTTPS URL, which references to a live meta-data document describing the participant. This identity is confirmed by a (X509) certificate.

FACET 6: REGULATIONS
This section refers to contracts and policies governing the interactions of participants and how they use data assets.

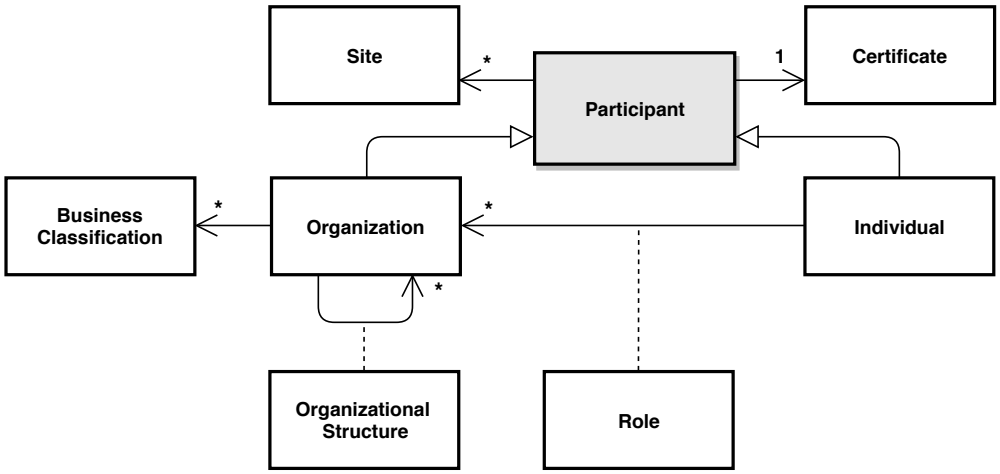


Figure 3.39: Outline of the Participant concept

USAGE CONTRACT

A pivotal part of the Product concept introduced by the Commodization view of resources in Section Commodization View is the formal expression of Usage Contracts pertaining to the Product. The Usage Contract defines a validity Period and formal Rules agreed upon by Participants involved in the provision, or subsequent usage of the Product. The Rules specify Actions that an involved Party (Participant) is obliged, permitted or prohibited to perform with respect to an Asset (resource or a collection of resources). Formal Constraints state the applicability of Rules and refine the interpretation of Actions. Given the reference data example DAT1, a Permission allowing for an unrestricted usage of the data holds when the Data Consumer met her Obligation to cite the data source. The Reference data example DAT2 prohibits the resale of commercial data segments via a Prohibition on Data Consumer. With respect do data example DAT3 a Duty may express the Obligation on Data Provider to maintain a particular Quality of Service (QoS) level, i.e. publish the live sensor data at a particular rate and warrant a reliable delivery (QoS level „at least once“).

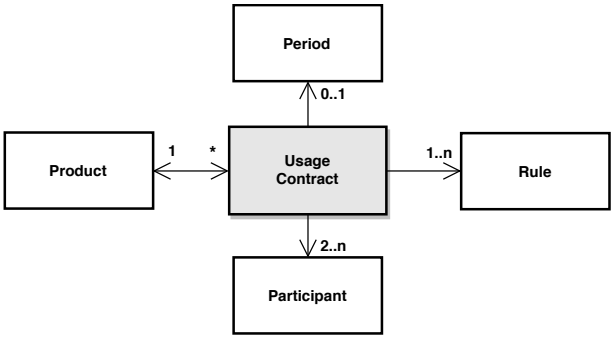


Figure 3.40: Outline of the Usage Contract concept

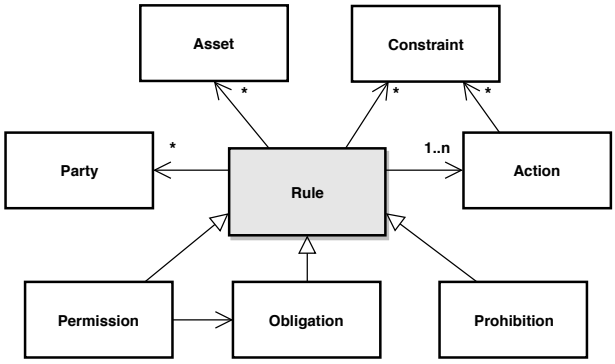


Figure 3.41: Outline of the Rule concept

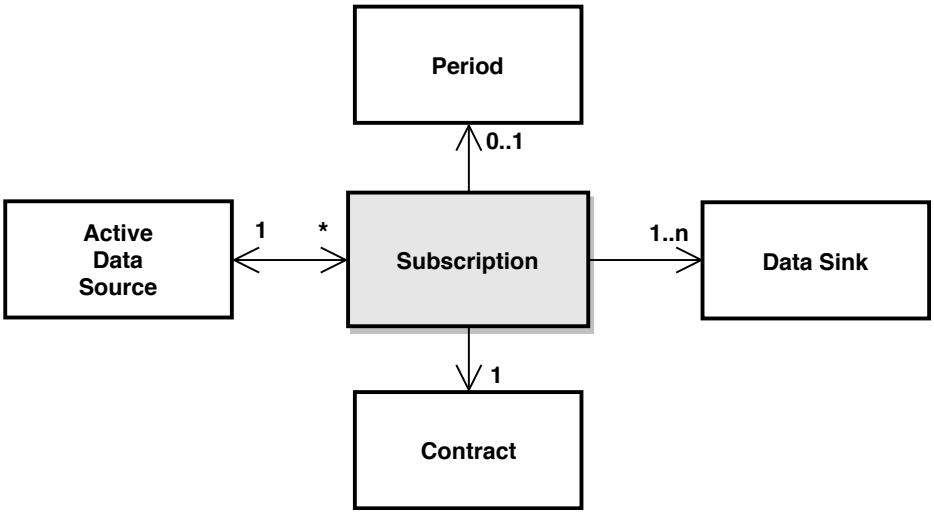


Figure 3.42: Outline of the Usage Contract concept

Usage Contracts formalize the expectations on behavior of involved Participants in a declarative, technology-agnostic way. The perpetual control and enforcement of such specification level policies may involve the inception of governance processes or, when appropriate the deployment of a technological solution. Data Usage Control Frameworks like IND²UCE define implementation-level policy languages in terms of technology-dependent events and actions, i.e. access to or modification of single files. Appropriate Policy Mappings should be specified to cope with the obvious conceptual gap between both policy levels and to enable a reliable and affordable technological enforcement of (parts of) Usage Contracts.

Usage Contracts therefore should indicate an enforcement strategy and, in case of a technological enforcement, the Policy Mappings to be supported by the target Connector. As mentioned in Section Connector, a Connector should disclose its Usage Control capabilities as part of its Security Profile.

SUBSCRIPTION
The Subscription concept expresses an Obligation to deliver data at a particular Quality of Service mandated by the Usage Contract from an active Data Source to a number of subscribed Data Sink targets within the given Period.

FACET 7: INTERACTIONS
The Interactions facet deals with concepts underlying business interactions among the IDS Participant, i.e. the interchange and consumption of Resources according to defined Regulations. The internal maintenance and operation processes of the IDS Infrastructure are considered afterwards. Both facet are subject to an ongoing change and are presented in a limited extent.

DATA TRANSFER
Each Resource interchange in the Industrial Data Space is modeled as an instance of the Data Transfer concept. It specifies a minimalistic meta-data model supporting security, traceability and usage control purposes. The Data Transfer refers to the originating and target Resource Endpoints, the time-stamp and the Payload (Resource) being distributed. The message is optionally signed and contains an authentication token (Trusted Connector). The Data Transfer carries a reference to the underlying Usage Contract, a source of formerly aggreed usage policies, optionally augmented by a dynamic instance of a Usage Policy.

FACET 8: MAINTENANCE
The Maintenance facet deals with the concepts describing the internal processes of maintenance and operation of the IDS Infrastructure, including the maintenance and dissemination of shared informational Resources, e.g. ontologies.

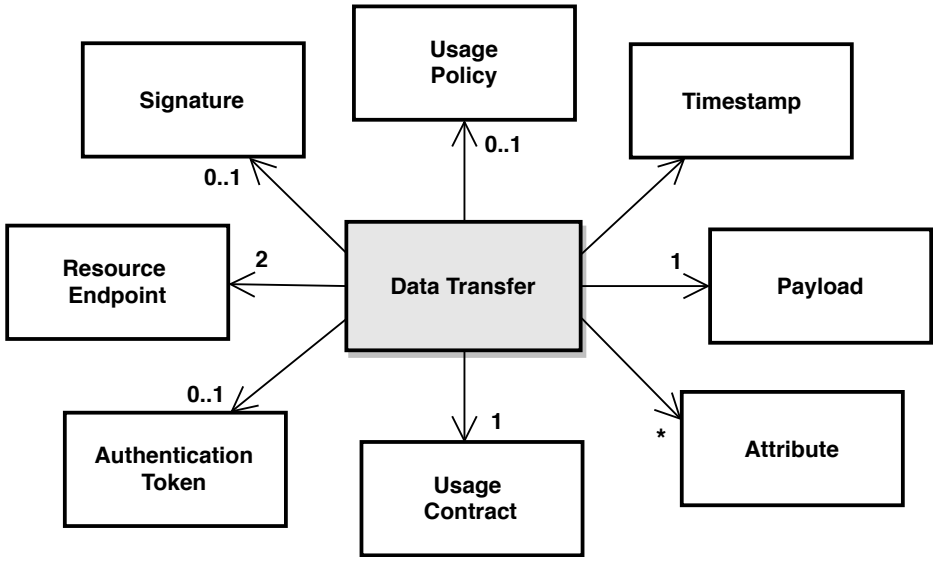


Figure 3.43: Outline of the Data Transfer concept

LIFE-CYCLE TRACKING

Infrastructure components of the Industrial Data Space are subjects to administrative operations, i.e. a life-cycle management defined by a set of States. The transition among these states are triggered by standardized Activities performed by administrative Agents. A record of life-cycle events should be maintained, e.g. to analyze and prevent failure conditions. Likewise the meta-data descriptions of Resources and Participants evolve over the time demanding for a life-cycle management and versioning. Such, e.g. historical versions of contracts have to be maintained alongside with recent revision or the representation of a Participant, created at some point may become temporarily suspended or permanently blocked. The concept of an Entity with Lifecycle was introduced to represent entities that are subject to evolution which needs to be tracked.

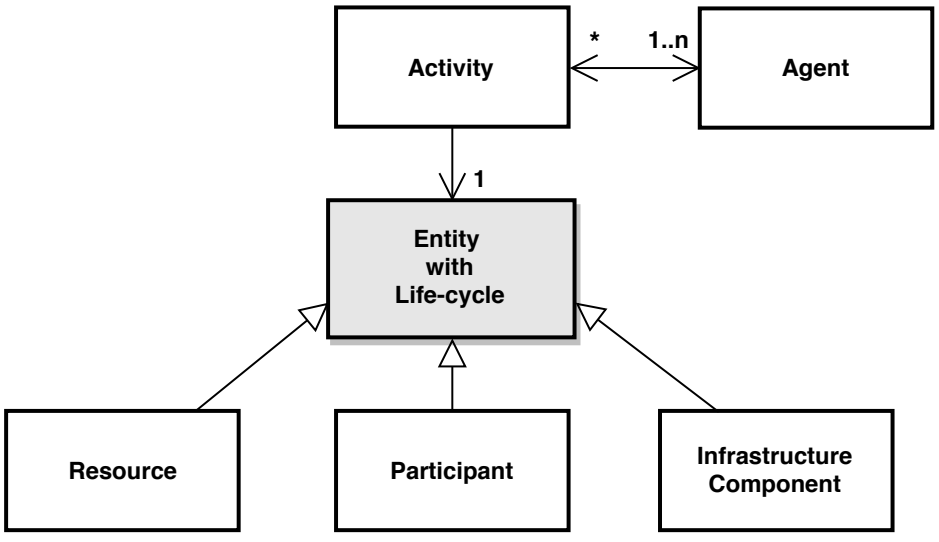


Figure 3.44: Outline of the Entity with Lifecycle concept

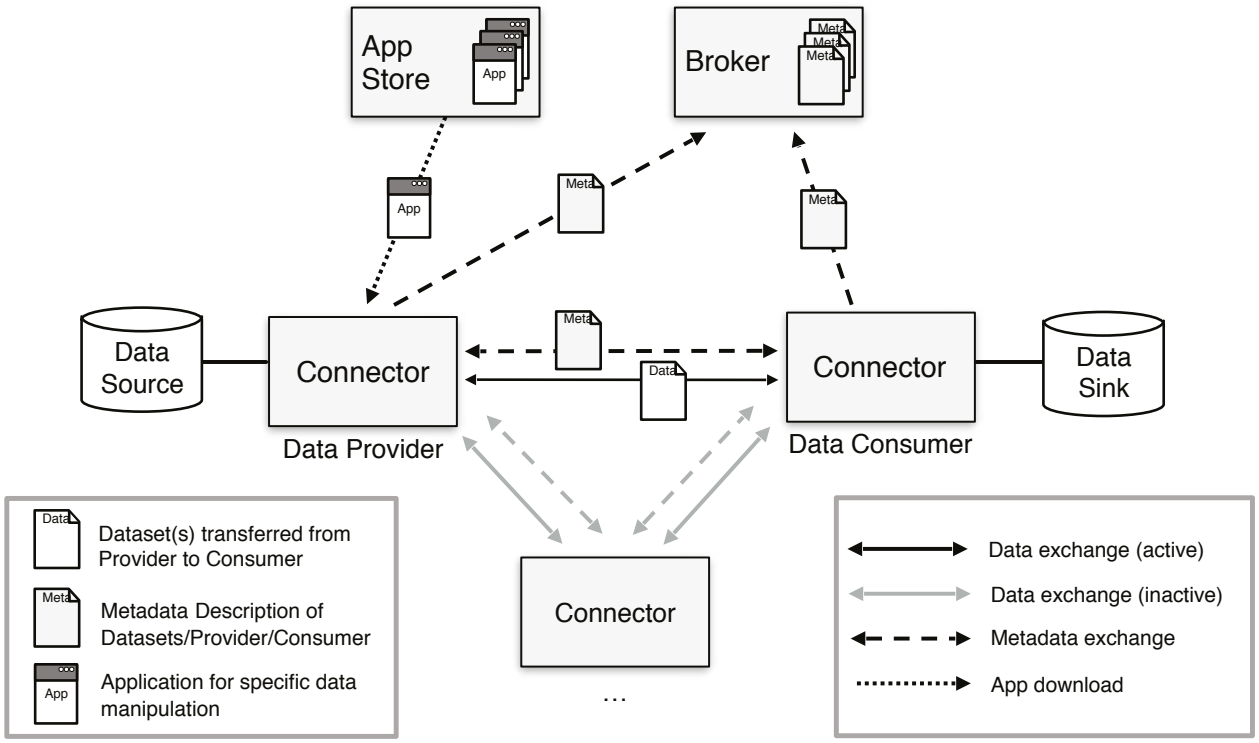


Figure 3.45: Interaction of technical components

3.5 SYSTEM LAYER

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 Industrial 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 3.45.

The Connector, the Broker, and the App Store are supported by four additional components (which are not specific to the Industrial 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).

A distributed network like the Industrial 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 (see Section 3.3.2). The Connector thus works at the interface between the internal data sources and enterprise systems of the participating organization and the Industrial 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-topeer 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 (see Section 4.1).

Connectors can be distinguished into External Connectors and Internal Connectors. An External Connector executes the exchange of data between participants of the Industrial Data Space. Each External Connector provides data via the Data Endpoints it exposes. The Industrial 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.

3.5.1 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 preprocessing (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 Industrial 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 3.46 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 3.46 can be assigned to two phases: Execution and Configuration.

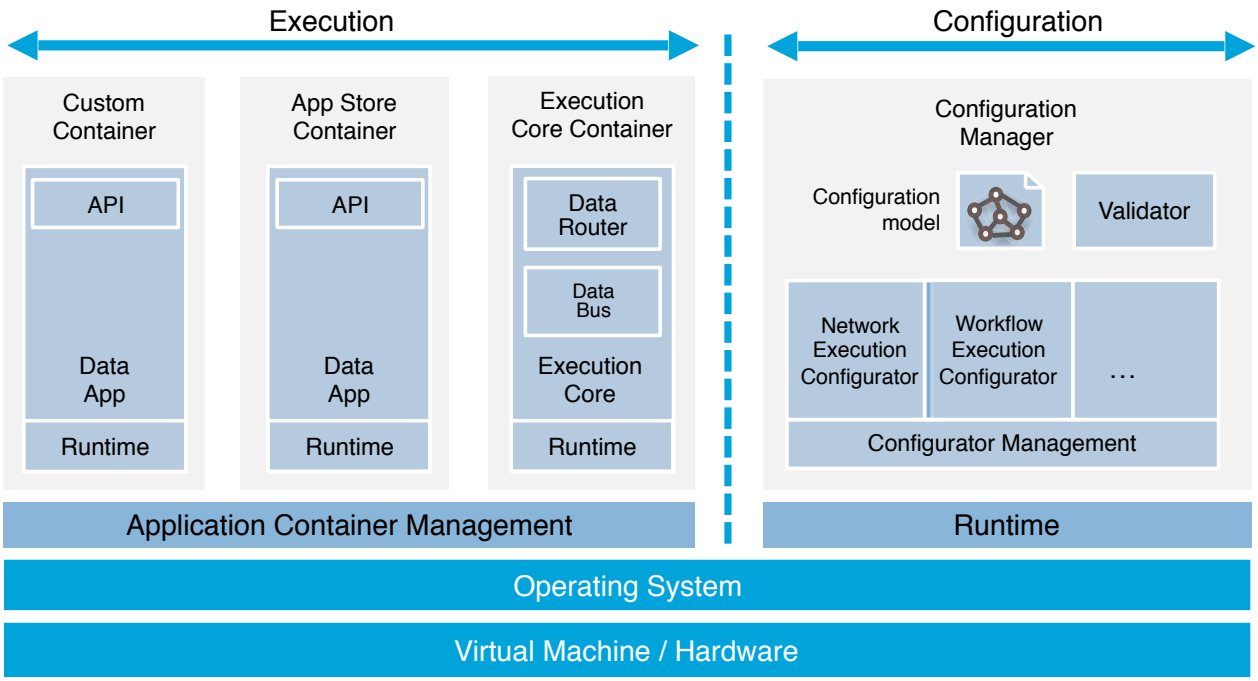


Figure 3.46: Reference Architecture of Connector

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 Industrial 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.

3.5.2 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 canceled.
- The Security settings contain information about e.g. which SSL certificates should be used for connections or which public key infrastructure should be used.

3.5.3 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.

04

PERSPECTIVES OF THE REFERENCE ARCHITECTURE MODEL

DIRECTLY RELATED TO THE FIVE LAYERS OF THE REFERENCE ARCHITECTURE MODEL ARE THREE CROSS-SECTIONAL PERSPECTIVES: SECURITY, CERTIFICATION, AND GOVERNANCE. THESE ARE DESCRIBED IN DETAIL IN THE FOLLOWING SUB-SECTIONS.

4.1 SECURITY PERSPECTIVE

As stated in Section 1.1, one strategic requirement of the Industrial Data Space is to provide secure data supply chains. This is critical for establishing trust among participants that want to exchange data and use Data Apps. The Security Architecture provides means to identify participants, protect communication and data exchange transactions between them, and control the use of data after it has been sent.For these purposes, the Industrial Data Space offers a Trusted Connector as an extension of the Base Connector (see Section 3.5). The Trusted Connector ensures that the specifications and requirements of the Security Architecture materialize in everyday interactions and operations of the Industrial Data Space. The security aspects described in the following constitute the basis of the Trusted Connector.

4.1.1 SECURITY ASPECTS ON THE DIFFERENT ARCHITECTURAL LAYERS

BUSINESS LAYER
Security aspects are crucial for the definition of roles and the possible interactions taking place between them in the Industrial Data Space. To enforce various business models in the Industrial Data Space, the Business Layer relies on the System Layer to enable secure business transactions.

FUNCTIONAL LAYER
Security requirements may restrict certain transactions or operations in the Industrial Data Space, or even prevent them. However, security is also an enabling factor. Without security, many use cases would not be possible (e.g., offering sensitive data to trusted business partners). The concept of data usage control allows Data Providers to attach data usage policy information to their data in order to define how a Data Consumer may use the data.

PROCESS LAYER
To take security aspects into account on the Process Layer, it is important that existing processes are permanently monitored, validated, and redesigned, if need be. For example, to allow trustworthy identification and authentication of participants using a central Public Key Infrastructure (PKI), a participant must apply for a public key certificate that is registered in a central PKI and deployed inside its Connector. For dynamic attribute support, an identity management server needs to verify attributes before issuing access tokens. The same is true for trustworthy operations of an App Store, for which data must be verified and signed by a trusted entity before it can be uploaded.

INFORMATION LAYER
The Information Layer provides the prerequisites for participants to use a common vocabulary and common semantics to express concepts and relationships between them. In doing so, it is possible to specify access and usage control policies in a way that they are understood by all participants. The same is true for access control requirements defining minimum security profiles, which must be met before access is granted.

SYSTEM LAYER
As the Connector is the central technical component on the System Layer, it is predominantly the Connector where the security features of the Industrial Data Space are implemented. Being an extension of the Base Connector, the Trusted Connector takes up all relevant security specifications and requirements, and serves as the technological basis for use case implementations.

4.1.2 GENERAL SECURITY PRINCIPLES

The development of the Security Architecture follows two general principles:

USE OF EXISTING STANDARDS AND CONSIDERATION OF BEST PRACTICES
To the extent possible and reasonable, existing standards and best practices are to be used and leveraged in the development of the Security Architecture. The aim of the Security Architecture is not to offer new solutions for problems already solved, but to combine existing, reliable approaches in a useful and meaningful way and bridge gaps where necessary.

ALLOW SCALABILITY OF SECURITY LEVELS
The Industrial Data Space does not enforce a single level of security to be applied for all participants. This way, also organizations with limited resources and technical means are able to participate (at least as Data Consumers). However, also the security level of these participants must be reliable and verifiable for others. Certain minimum security requirements (e.g., encrypted communication) therefore need to be met by all participants.

Provided a participant is in line with general security requirements, it may decide about the level of security to be applied for it itself. It should be noticed, however, that data sources may presuppose a certain minimum level of security to be met by potential Data Consumers. This means for Data Consumers: the higher the security level they choose for themselves to be applied, the better the access to high-quality data sources and high-value data services.

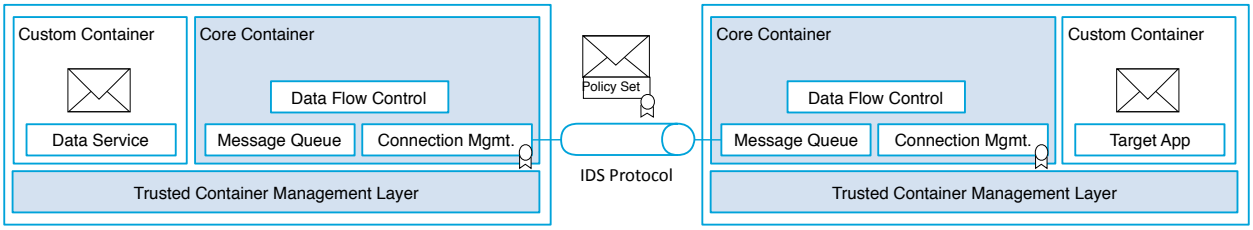


Figure 4.1: IDS Communication Protocol

4.1.3 KEY SECURITY CONCEPTS

The Security Architecture addresses six key aspects:

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

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

- point-to-point encryption (between Connectors), using an encrypted tunnel, and
- end-to-end authorization (authenticity and authorization based on actual communication endpoints. i.e., Data Apps).

Data from one External Connector to another is sent over the Internet or via a virtual private network (VPN), the specification of which is beyond the scope of the Security Architecture. The Security Architecture defines the IDS Communication Protocol (IDSCP), which must be supported by Trusted Connectors, and can be supported by any other Connector as well. The purpose of the IDSCP is to establish confidential, authenticated communication, exchange data between the Data Provider and the Data Consumer, and establish mutual remote attestation (if supported by the

Connectors involved). Trusted Connectors must communicate with each other over an encrypted tunnel (e.g., TLS), as depicted in Figure 4.1. The IDSCP is a high-level protocol established via WebSocket Secure (WSS). It contains several "conversations", which can be initiated by either side and must be confirmed by the other side to be entered. Currently, two conversations are provided: remote attestation and metadata exchange. The protocol itself is performed inside a tunneled connection. The protocol supports and enables several communication aspects:

- identification and authentication,
- remote attestation,
- exchange of metadata, and
- exchange of data (together with usage policy information attached).

The last aspect, exchange of data, provides the basic mechanism of data usage control, as it is possible to attach data usage policy information in order to specify how the data may be used by the Data Consumer.

IDENTITY MANAGEMENT

To be able to make access control related decisions that are based on reliable identities and properties of participants, a concept for Identity and Access Management (IAM) is mandatory. The following aspects are central for the concept:

- identification (i.e., claiming an identity),
- authentication (i.e., verifying an identity), and
- authorization (i.e., making access decisions based on an identity).

The Certificate Authority (CA) issues certificates for all entities. These certificates are used to establish communication between all participants.
An identity may have several attributes, which are linked to that identity. A "Dynamic Attribute Provisioning Service (DAPS) is used to provide dynamic, up-to-date attribute information about participants and Connectors.

MAPPING OF CERTIFICATION OF PARTICIPANTS AND CERTIFICATION OF CONNECTORS TO IDENTITY MANAGEMENT

As an expected outcome of the certification concept, there will be two entities of certification: Organizations (receiving an Organizational Certificate) and Connectors (receiving a Connector Certificate). If an organization (e.g., a company) is successfully certified, it is allowed to participate in the Industrial Data Space. The organization is rewarded a certification level by the Certification Body. Upon successful certification, a technical certificate is issued to the organization (a X.509 certificate) to confirm certain attributes like organizational name, certification status, etc. This technical certificate can be used to trigger processes such as applying for connector certificates, applying for certificate renewal, etc. An organization can apply for Connector certificates (also X.509 certificates) for every Connector deployed (depending on the successful approval by a Connector approval/certification partner).

An organization needs to be awarded certification before becoming an official participant. After successful certification, an Organizational Certificate is issued, containing the certification level. Connectors deployed need to be certified after Connector certification criteria, triggering the issuing of a Connector Certificate.

EXAMPLE X.509 CERTIFICATE FOR ORGANIZATIONS

- Version Number
- Serial Number
- Signature Algorithm ID
- Issuer Name (e.g., IDS Association CA)
- Validity period
- Subject Name (Organizational name)
- Subject Public Key Info
- Extensions
- Certificate Signature Algorithm
- Certificate Signature

All mandatory and optional attributes are to be defined before operationalizing a PKI concept. It is important to note that any modification of attributes leads to revocation and reissuing of the certificate. For this reason, the number of attributes that are contained in a certificate needs to be kept at a minimum. Dynamic attributes are kept by a separate Identity Provider instance handing out dynamic identity information.

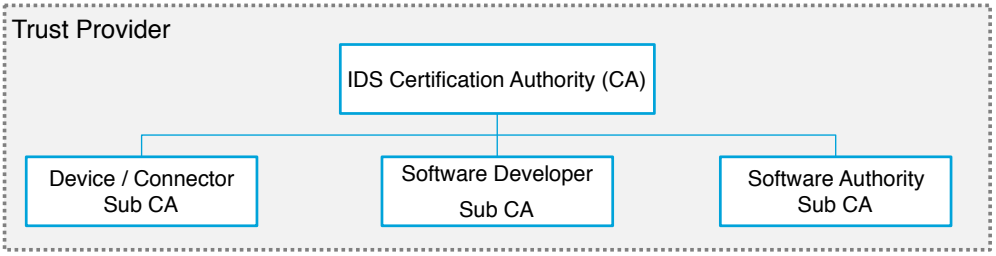


Figure 4.2: Exemplary PKI structure

PROPOSED PKI STRUCTURE

In general, a PKI can have several layers to achieve separation of duties (i.e., every Sub CA is responsible for a specific topic). Depending on the business and deployment model applied, several Sub-CAs may exist.

This allows for specific parties to issue certificates for specific purposes. It is also possible to support multiple instances (e.g., multiple Connector Sub CAs).

CONNECTOR CERTIFICATE DEPLOYMENT

After obtaining the (technical) Organizational Certificate, an organization may apply for one or more Connector Certificates (the issuing of which may be triggered by the International Data Spaces Association, for example). Exemplary attributes for Connectors to be embedded in the X.509 certificate are base attributes (as in the Organizational Certificate) and extensions.

Like in the case of Organizational Certificates, the number of attributes of Connector certificates should be kept at a minimum in order to reduce the risk of certificate revocation due to modifications made to attributes.

Once received, the certificate can be deployed onto the Connector.

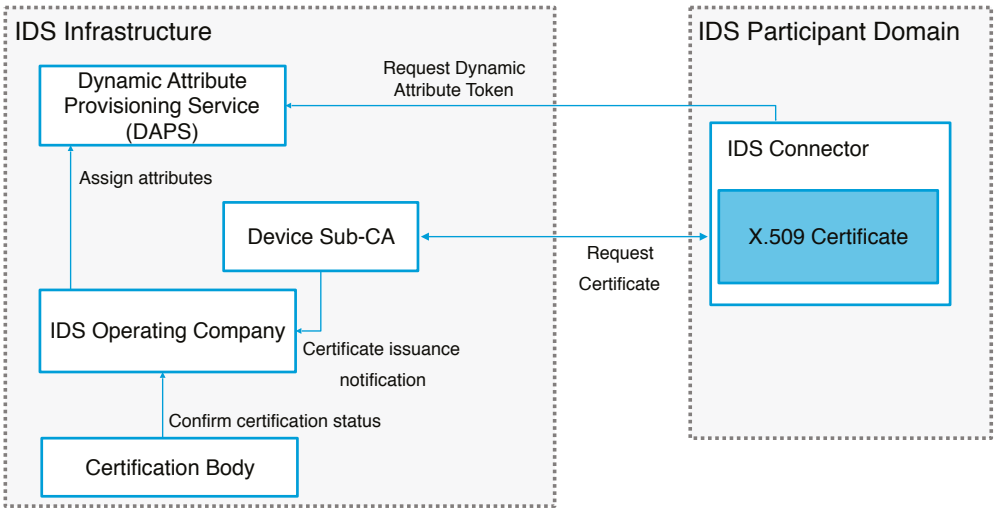


Figure 4.3: Embedding certificate into Trust infrastructure

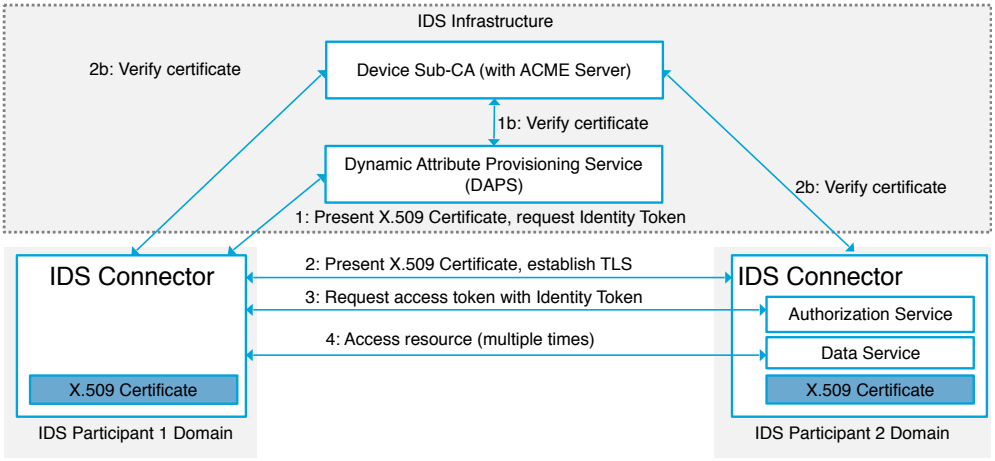


Figure 4.4 : Resource access workflow

Data exchange is always performed in an authenticated manner, using the Connector Certificate.

USING THE DYNAMIC ATTRIBUTE PROVISIONING SERVICE (DAPS) FOR IDENTITY MANAGEMENT

Using a service to hand out attributes in a dynamic fashion reduces the need for certificate revocation and enables more flexible attribute handling for participants in the Industrial Data Space. This allows dynamic assignment of attributes and status flags to Connector instances. Examples of status flags are:

- Withdraw a security status if known vulnerabilities have not been fixed.
- Upgrade the certification status without reissuing a X.509 certificate.
- Assign membership status to a workflow with contractors.

This concept avoids revocation of the certificate in most cases and increases flexibility to include new attributes if the need arises.

USING AN AUTHORIZATION SERVICE FOR RESOURCE ACCESS CONTROL

Using an Authorization Service (and, thus, access tokens) allows use case dependent modeling of access control decisions. Delegation of access decisions is possible. In complex workflows, multiple Connectors can use a dedicated Authorization Service to delegate resource access decisions.

An exemplary workflow for accessing a resource (e.g., a Data Service) using dynamic attributes and access tokens could look like this:

1. A Dynamic Attribute Token (DAT) is requested from the Dynamic Attribute Provisioning Service, presenting the Connector's X.509 certificate. Depending on the verification policy specified, the attribute can be verified at the CA.
2. Before accessing a resource, a TLS tunnel is established using the same X.509 certificate. Again, depending on the policy specified, the certificate can be verified at the CA.
3. Optional) If using several Access Tokens (ATs), a token request is performed at a separate Authorization Service in the domain of a use case operator or the domain of the Connector's (or, more specifically, resource's) owner.
4. The resource is requested by handing in either the Dynamic Attribute Token (DAT) or the Access Token (AT).

Due to the small size of access tokens, it is possible to incorporate these tokens into any resource request and support stateless access management.

TRUST MANAGEMENT

To establish trust across the entire business ecosystem (i.e., to protect participants from fraud and ensure they abide by the designated rules), the Industrial Data Space makes use of cryptographic methods. One such method is the Public Key Infrastructure (PKI). A central principle of a PKI is that every entity is allocated with secret keys, allowing each entity to authenticate against other participants. Thereby, a hierarchy is created, with the Identity Provider on top issuing certificates to the other entities, which in turn may issue certificates to other entities, and so on. In the following, the PKI rollout is described for mapping roles and entities required for the deployment of the Industrial Data Space.

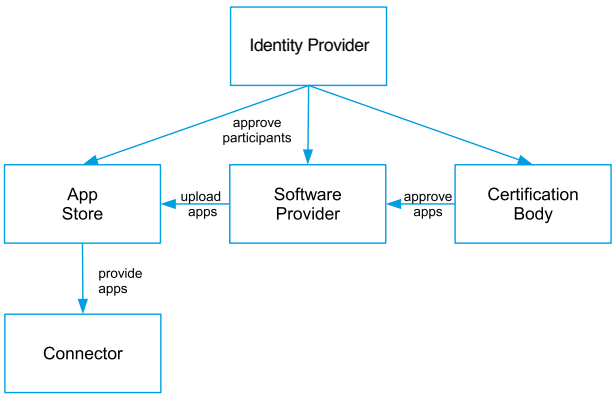


Figure 4.5: Technical roles in the Industrial Data Space

PKI ROLLOUT

To guarantee secure identity management, the Industrial Data Space defines technical roles for implementing a PKI system that is flexible enough to support all roles defined on the Business Layer. In particular, six entities with different security levels are relevant for the Security Architecture. In the following, these entities and the technical roles related to them are described.

IDENTITY PROVIDER

The Identity Provider acts as an agent for the International Data Spaces Association. It is responsible for issuing technical identities to parties that have been approved to become participants in the Industrial Data Space. The Identity Provider is instructed to issue identities based on approved roles (e.g., App Store or App Provider). Only if equipped with such an identity, an entity is allowed to participate in the Industrial Data Space (e.g., to provide data or publish Data Apps). The Identity Provider may exclude participants from the Industrial Data Space, if instructed to do so. Furthermore, the Identity Provider may authorize certain entities to act as Certification Bodies.

As a trusted entity, the Identity Provider manages the PKI rollout. It determines the properties of the Certificate Authority and takes care if certificates expire or must be revoked. There are two separate PKI hierarchies: one for software signatures (Software Signing Root CA) and one for the Connectors (Service Root CA). An entity is assigned either an end-certificate or a sub/root CA certificate. The two hierarchies protect the interests of the six entities, which use and manage the PKI as described in the following (Figure 4.6).

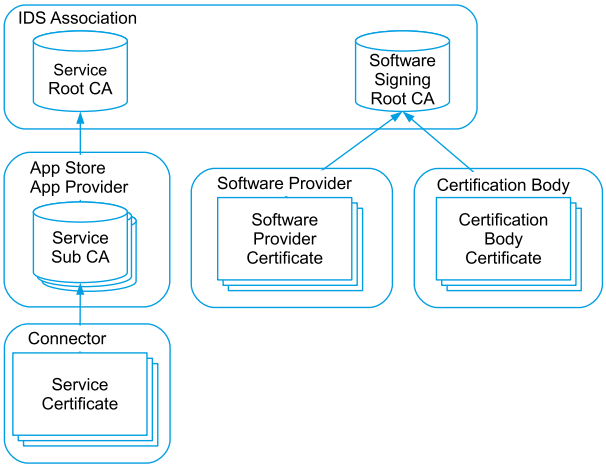


Figure 4.6: Mapping of technical roles and PKI layout

SOFTWARE PROVIDER

Software Providers produce and distribute software stacks for Connectors. They equip Connectors with an initial software system (for rollout and deployment). To every Software Provider seeking admission to the Industrial Data Space, the Identity Provider issues a service sub CA request. Approved Software Providers use the service sub CA during rollout and deployment of the Connector in order to provide it with an initial, valid and preconfigured system.

CONNECTOR
A Connector is allowed to communicate with other Connectors only if acquired from an approved Software Provider. Connectors download Data Apps from the App Store. For each Data App downloaded, the Connector creates a service key pair and a Certificate Signing Request (CSR). While the private key is used to identify the Data App and to protect its data, the CSR is sent to the App Store, which uses it to issue a certificate. This also allows entities to check whether the license of a certain Data App is still valid (see e.g. remote attestation). Furthermore, the private key and the certificate are used for establishing a secure channel with other Connectors. During rollout, the Software Provider deploys an initial system onto the Connector and signs the Connector's corresponding service CSRs for the initial system.

APP STORE
A Connector downloads the software it requires from an App Store in the form of Data Apps. Connectors can only connect with the App Store for requesting downloads and updates. As the App Store is a Connector itself, it additionally stores its own sub CA. When a new provider sets up an App Store, the Identity Provider signs a sub CA request issued by the provider. The provider deploys this sub CA inside the App Store (i.e., inside the respective Connector). This sub CA is used by the App Store to ensure the validity of services downloaded by other Connectors. This means that if an App Store signs a CSR (i.e., issues a certificate), a Connector receives a certificate for a downloaded Data App.

APP PROVIDER
App Providers must seek approval of Data Apps from the Certification Body. Upon successful certification of a Data App, the App Provider may publish the Data App by uploading it to the App Store. Each App Provider can be unambiguously identified by a certificate issued by the Identity Provider.

CERTIFICATION BODY
When an App Provider uploads a Data App, the App Store not only checks if the Data App comes from an approved App Provider, but also if the software meets certain quality and security standards. Therefore, App Providers must send the Data App to a Certification Body for inspection. The Certification Body checks the validity of the App Provider's signature. If the signature is valid, the source code of the respective Data App is inspected. If the Data App meets the quality and security standards, the Certification Body signs the Data App with the certificate's private key. To do so, it does not need a sub CA, as it only signs the software, but does not create a certificate.

CONNECTOR MANIFESTATIONS
A Connector can run different services and communicate with other Connectors. Using the PKI, a Connector protects the persistent storage of its services and the communication with other Connectors (in terms of authenticity, confidentiality, etc.). The following items characterize a Connector in the Industrial Data Space:

CONFIGURATION
Among other things, the configuration specifies from where the Connector downloads new services, or which Brokers or Online Certificate Status Protocol (OCSP) Servers it uses. Configuration is required in order to boot the system. It is deployed during deployment.

CA CERTIFICATES
In order to verify PKI signatures (e.g., for authentication or for Data Apps that were downloaded), the Connector stores the trusted root certificates (Service Root CA and Software Signing Root CA) in a way their integrity is preserved (Figure 4.7).

APPS
Apps offered in the Industrial Data Space are usually running inside isolated containers. The Connector creates a key pair for every app it downloads. The private key protects the app's persistent data. When downloading an app from the App Store, the Connector creates a CSR using the public key. The App Store signs the CSR and issues a certificate. The Connector uses this certificate to make sure that the app it is running is valid (i.e., licensed, not expired, etc.).

An app is a generalization of the following types of software:

- Core System: Every Connector runs exactly one Core System. The Core System, together with its certificate, is deployed during the Connector's deployment after being retrieved from the Software Provider providing the Connector. The Core System's certificate identifies the underlying hardware device. The Core System can connect to other Connectors (e.g., to communicate with the App Store for app downloads). When a Connector establishes a communication channel with another Connector,

it uses the Core System's private key and certificate for authentication.

- Data App: A Data App is any data processing or data collecting app, or a System Adapter.
- Broker: A Broker is a Connector providing a broker service.
- OCSP Server: A Connector is considered an OCSP Server if it runs the OCSP Server app.
- App Store: An App Store has a service sub CA. The Industrial Data Space signs this CSR in order to approve every new App Store. The CSR identifies the App Store and makes it possible to sign the service CSRs from the Connectors requesting apps.

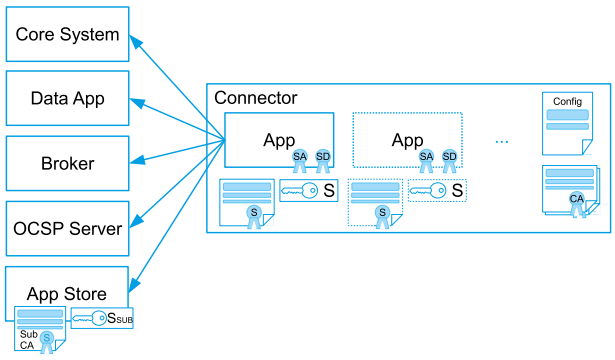


Figure 4.7: Connector roles and manifestations

APP DEVELOPMENT AND DEPLOYMENT

The following steps describe the app lifecycle, from app development to app deployment onto a Connector (Figure 4.8):

1. The Identity Provider signs a key pair and a certificate for each Software Provider on behalf of the International Data Spaces Association. When the app is fully developed and ready for being offered, the Software Provider signs the app using its private key, before the signed app is sent to a trusted Certification Body.
2. If the Certification Body approves the app, a second signature is added to it.
3. The Software Provider uploads the app to an App Store. The App Store only accepts valid (i.e., correctly signed) apps (since the App Store is a Connector with corresponding root CAs, it is able to verify all signatures).
4. A Connector downloading the app (e.g., a Data App) connects with the App Store. The Connector creates a service key pair and a CSR, requests a service download, and sends the CSR to the App Store. The App Store signs the CSR using the service sub CA and returns it to the Connector.
5. The Connector downloads the service and checks its signatures. If the signatures are found to be valid, the Connector installs the service. From now on the downloading Connector can check the validity of the downloaded service based on the certificate received.

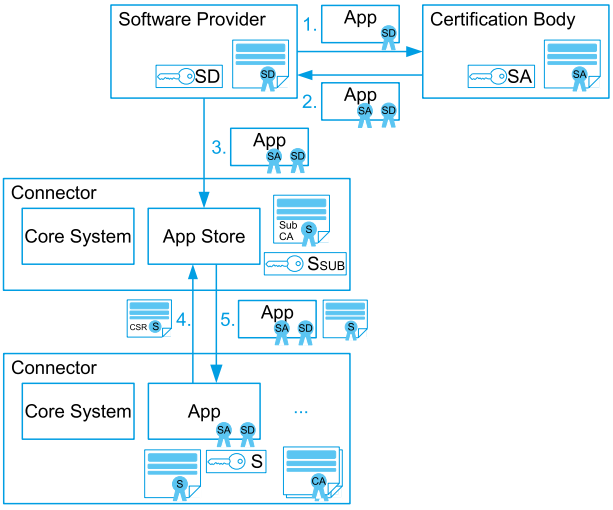


Figure 4.8 Software development, approval, and download process

DELIVERY OF TRUSTED CONNECTORS

After initial deployment, the Connector is delivered to the Operator in a fully preconfigured state (Figure 4.9). For deployment of the Connector, every approved Software Provider has a sub CA key pair and CSR (similar to an App Store Provider) to sign the initial system. When the Identity Provider signs the CSR of the sub CA, it confirms the requesting Software Provider as being compliant with Industrial Data Space regulations and policies. The Operator of a Connector (e.g., a Data Provider) may change the configuration, the root certificates, and even the Core System as deemed appropriate.

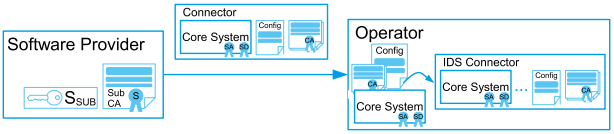


Figure 4.9: Delivery of Trusted Connector

TRUSTED PLATFORM

The Industrial Data Space consists of multiple manifestations of the Connector Architecture (as used by e.g. the Broker or the App Store). This is why a trusted platform is a central element of trustworthy data exchange. A trusted platform comprises certain key aspects:

- To be able to specify minimal requirements for participants that want to exchange data, a common understanding of each other's Security Profiles needs to be established. The Connector supports mutual verification of these profiles.
- To enable trustworthy execution of Data Apps and guarantee system integrity, strong isolation of components is necessary. The Connector's Application Container Management supports full isolation of Data Apps deployed, and limitation of illegitimate communication channels. This means that Data Apps have access only to data that is explicitly meant for them.
- To establish a trustworthy relationship with another participant, and to verify Connector properties, remote integrity verification is required. The Connector features a hardware-based trust anchor and a trustworthy software stack.

ISOLATION AND REMOTE EXECUTION GUARANTEE

Isolation is a form of integrity enforcement for a Data App's runtime environment. Data Apps can be isolated against each other by deploying each one inside a separate container (or all Data Apps of a specific Software Provider into one container), as illustrated in Figure 4.10. This allows implementation of additional security features, such as time-to-live policy enforcement for complete container instantiations.

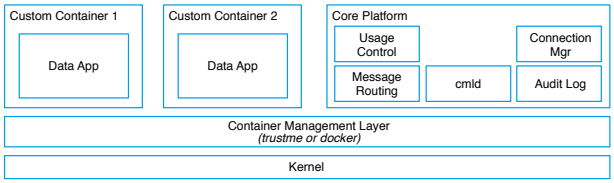


Figure 4.10 Container isolation

The Connector should provide some mechanism to isolate Data Apps, system apps, and the core platform from each other, in order to prevent applications from interfering with each other. Each Connector has a Security Profile attached to it, describing its isolation capabilities.

However, the Security Profile may be empty in cases in which the Connector does not provide isolation between Data Apps. Users of Data Apps may make data access control decisions based on the set of isolation capabilities stated in the Security Profile.

REMOTE INTEGRITY VERIFICATION

During system setup, trust remains strictly limited to each party's domain. Two levels of trust are supported in the Industrial Data Space:

- verification of each party's identity by exchanging credentials that originate from an entity both parties trust (e.g., credentials signed by a trusted PKI, or identity tokens issued by a trusted identity provider);
- verification of the integrity of each Connector's software stack by applying integrity measurement using trusted platform modules, and by remote attestation (for remote integrity verification, trust into the identity of a party is a mandatory requirement).

Verifying the integrity of a Connector software stack (and its configuration) is required for deploying trusted Data Apps. If platform integrity was not verified (either through certification or by technical measures), one or more of the following problems would occur:

- A Connector could pretend to run a certified and trusted software stack in order to feign an unjustifyingly high level of trust.
- A Connector might not run Data Apps as expected (i.e., the Data Apps do not receive the desired amount of resources in terms of CPU and memory, and neither execution nor communication is trustworthy); if that was the case, the data consumed and provided by Data Apps running on an untrusted and unattested Connector platform would not be reliable.
- Edge-computing use cases, where consumers push their Data Apps to the data source (i.e., onto remote Connector), would be difficult to realize, because correct execution of these Data Apps could not be guaranteed.

To enable a Connector to get technically reliable information about the integrity of the software stack and the runtime configuration of another Connector, Connectors may support remote attestation for more secure Connector instantiations. Trustworthy measurement is possible using TPM 1.2/2.0 in a Connector.

4.1.4 CONNECTOR SECURITY PROFILES

Using static security levels would make it necessary to anticipate all possible needs of every participant, now and in the future. Since the system is designed to grow over time and remain flexible with regard to the individual security needs of every participant, the Industrial Data Space offers the possibility to base access control decisions on fully customized criteria. Access control policies can be based on a set of attributes of the requesting Connector. Beside a unique identifier, these attributes include a set of properties describing the security level of Data Apps as well as the security properties of the technical setup of the Connector. A set of security properties is called a Security Profile.

A Security Profile comprises attributes of the Connector and may be used in an attribute-based access control policy. Each Connector must provide its Security Profile upon request. The set may also be empty though.

A Security Profile contains the following properties:

- It describes the Connector's current security configuration.
- It allows Data Consumers to decide whether they are willing to rely on data provided by a Data Provider's endpoint.
- It allows Data Providers to decide whether they are willing to make sensitive data available to a Data Consumer.

A Security Profile may consist of the following options:

Aspect	Level 0	Level 1	Level 2
Integrity protection and verification	None	Local Integrity Protection	Remote integrity verification
Connection authentication	None	Server side only authentication	Mutual authentication (both sides present ID token / certificate)
Service isolation	None	Process group Isolation	Least privilege based Isolation
Integrity protection / verification scope	None	Kernel & Core Container	Kernel & Core Container & Application Containers
App Execution Resources	None	Local enforcement	Remote verification
Data usage control support	None	Usage control policy enforcement	Remote policy compliance verification
Audit logging	None	Local logging capabilities & integrity protection	Remote audit log tracing
Local data confidentiality	None	Secure data erasure	Local full data encryption

A Connector with every security option set to Level 2 would be a Connector with the maximum possible set of security features and trust level.

EXPLANATION OF OPTIONS

Aspect	Explanation
Integrity protection and verification	Integrity protection and verification of the software stack installed; local integrity protection would be something like trusted boot; remote integrity verification means support for remote attestation.
Connection authentication	Authentication before opening a valid connection; this can be done by just verifying the server identity or doing mutual authentication.
Service isolation	Service isolation via a process group (as done with Docker) or by least privilege with clear separation and support for additional security modules (as done by trustme).
Integrity protection / verification scope	Defines the level up to which the software stack can be verified (Level 1 means up to the Core Container installed, Level 2includes all services installed).
App Execution Resources	Resource control for deployed services; local enforcement guarantees specific resources to services and makes sure the services do not exceed the resources.
Audit logging	Local logging, including integrity protection, is the baseline for auditing and clearing; remote audit log tracing provides means to perform external audit verification.
Local data confidentiality	Describes the means by which local data is protected.

Security Profiles are supported by the Information Model of the Industrial Data Space (see Section 3.4.2) and can be expressed in a standardized, machine-readable form, using the Industrial Data Space Vocabulary.

4.1.5 DATA ACCESS CONTROL

The Connector provides mechanisms to regulate access to data. To specify data access conditions, the following criteria can be applied:

- only access requests from one specific Connector (or from a number of specific Connectors, respectively) are granted;
- only access requests from a Connector that possesses specific attributes are granted;
- only access requests from a Connector meeting specific Security Profile requirements are granted.

Using static security levels would make it necessary to anticipate all possible needs of every participant, now and in the future. Since the Industrial Data Space is designed to grow over time and remain flexible with regard to the individual security needs of every participant, the Industrial Data Space offers the possibility to base access control decisions on fully customized criteria. Access control policies can be based on a set of attributes of the requesting Connector.

Beside a unique identifier, these attributes may include a set of properties describing the security level of Data Apps as well as the security properties of the technical setup of the Connector (this is described in the previous section on Connector Security Profiles).

4.1.6 DATA USAGE CONTROL

The Industrial Data Space is about creating a business ecosystem in which organizations can exchange and use data in a secure environment. In this respect, data sovereignty is a key success factor for the Industrial Data Space. Data sovereignty aims at granting Data Owners full control over their data, including control over the use of the data by Data Consumers. To ensure data sovereignty, the concept of data usage control provides the necessary technical mechanisms.

Data usage control can be seen as an extension of data access control (see Figure 4.11). It is about the specification and enforcement of restrictions regulating what is allowed to happen with data, and what is not. Data usage control thus is concerned with requirements that pertain to data processing (obligations) rather than data access (provisions). It is primarily relevant in the context of intellectual property protection, privacy protection, compliance with regulations, and digital rights management.

The following examples illustrate security requirements that cannot be achieved by data access control, but require data usage control in addition:

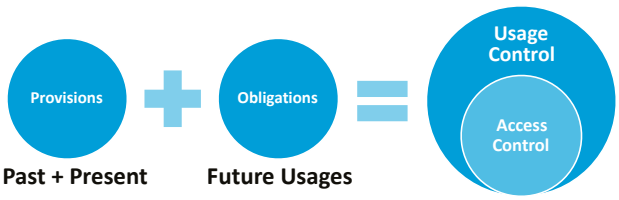


Figure 4.11: Data usage control – an extension of data access control

- Secrecy: Classified data may not be forwarded to nodes that do not have the respective clearance.
- Integrity: Critical data may not be modified by untrusted nodes, as otherwise their integrity can no longer be guaranteed.
- Time to live: A prerequisite for persisting data is that it must be deleted after a given period of time.
- Anonymization by aggregation: Personal data may only be used in an aggregated form by untrusted parties. A sufficient number of distinct records must be aggregated to prevent deanonymization of individual records.
- Anonymization by replacement: Data that allows personal identification (e.g., faces shown in camera images) must be replaced by an adequate substitute (e.g., a pixelized image) to ensure that the identity of individuals is not revealed.
- Separation of duties: Two data sets from competitive entities (e.g., two automotive OEMs) may never be aggregated or processed by the same service.
- Scope of usage: Data may only serve as input for data pipes within the Connector (i.e., it may never leave the Connector and be sent to an external endpoint).

Companies may have both an intrinsic and an extrinsic motivation to apply data usage control. On the one hand, they may want to prevent misuse of their own data, protect their intellectual property, or preserve the value of the data (intrinsic motivation). On the other hand, they may have to comply with legal or regulatory frameworks, such as the European Union General Data Protection Regulation EU-GDPR (extrinsic motivation). Hence, companies need to prevent misuse of company data or data entrusted to the company by third parties.

Data usage control is a way to track and trace data as it is used by different systems, and to collect evidence of violations of previously agreed data usage constraints. With that in mind, enforcement solutions may be either organizational/legal or technical. An organizational rule may state, for example, that employees may not use removable data storage media (such as USB sticks). Alternatively, a technical solution (e.g., a group policy of the operating system) could prevent employees from using removable data storage media. While there may be scenarios for which organizationally/legally and technically enforced rules could be used interchangeably, other scenarios might suggest to use both enforcement approaches together in order to complement each other. In the long run, it can be assumed that organizational/legal enforcement will increasingly be substituted by technical enforcement (as illustrated in Figure 4.12).

Figure 4.12: Technical enforcement vs. organizational/ legal enforcement

In the following, general concepts of data usage control are presented.

SPECIFICATION AND MANAGEMENT

One important aspect of data usage control is the specification and management of data usage restrictions. Data Owners have to express the restrictions to be imposed on their data in a more or less formal way. For technical enforcement, the specification must produce a machine-readable output. The Policy Administration Point (PAP) is the entry point for the specification of data usage policies which usually is accessible via a graphical user interface.

A Policy Management Point (PMP) manages data usage policies. This component is concerned with the policy lifecycle. Tasks include the instantiation, negotiation, deployment, and revocation of data usage policies at the Policy Decision Point (PDP), as well as conflict detection and resolution in the case of conflicting policies.

Data usage policies can be deployed in two different ways. One option is to attach machine-readable data usage policy information directly to the data (so-called “sticky policies”). Such sticky policies can be implemented in different ways. Usually, data is encrypted and can only be decrypted by the Data Consumer if it accepts the usage policy. The second option is to store data usage restriction information independently of the data exchanged (e.g., in a central component, such as a PDP). In this case, the management component has to distribute the data usage policy across all systems involved.

ENFORCEMENT

System actions need to be monitored, and potentially intercepted, by control points (i.e., Policy Enforcement Points, PEPs) enforcing data usage policies at runtime. These actions are presented to the decision engine (i.e., the Policy Decision Point, PDP), which either approves or denies their execution. In addition to just approving or denying a system action, the PDP's decision may also require modification of the action (e.g., modification of data in transit). In addition, there may be the need to perform additional system-specific actions as part of the policy enforcement, which is encapsulated in an execution handling component (i.e., the Policy Execution Point, PXP) triggered by the decision engine.

DECISION AND INFORMATION

Enforcement of data usage policies relies on a decision. A Policy Decision Point (PDP) responds to incoming requests (e.g., system actions) from a PEP by making a decision. Usage policies are used to specify the desired behavior of the decision engine. Hence, the decision-making process is called "policy evaluation".

Policy evaluation may depend on additional information that may not be present in the system action itself. This includes contextual information, such as information about data flow properties or the geographical location of an entity. A Policy Information Point (PIP) provides missing information for decision-making.

The concept of data usage control, including its components and the interactions between them, is illustrated in Figure 4.13.

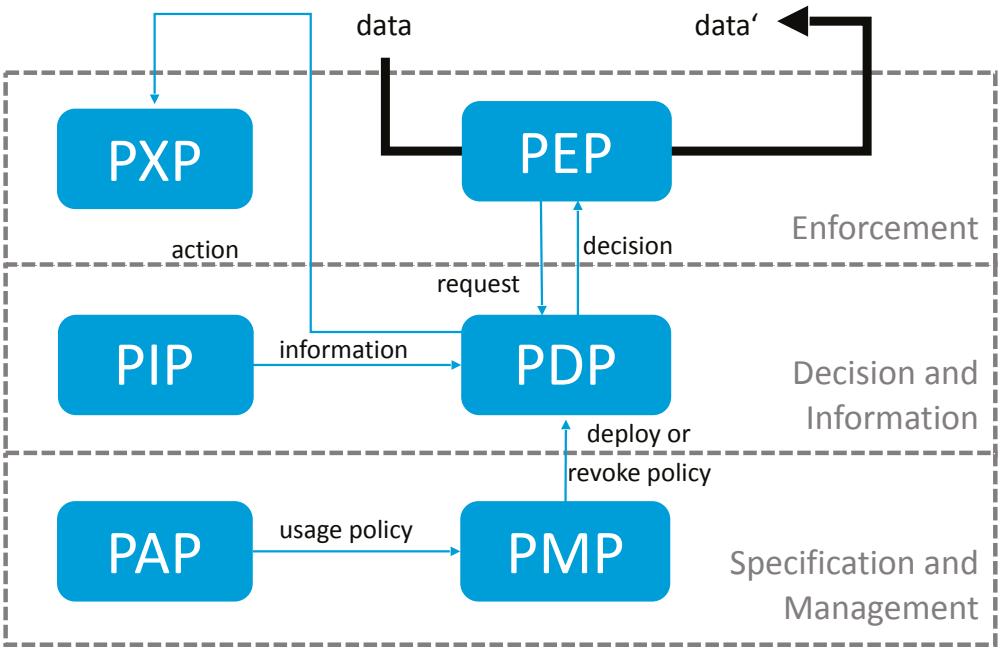


Figure 4.13: Components and their interaction in data usage control

4.1.7 USAGE CONTROL ASPECTS ON THE DIFFERENT ARCHITECTURAL LAYERS

Being a central concept of the Security Perspective of the Reference Architecture Model, data usage control is a cross-sectional concept affecting all five architectural Layers.

BUSINESS LAYER
Data usage control is essential for Data Owners to keep control over their data by specifying data usage restrictions in the form of a data usage policy. A data usage policy typically states usage restrictions for Data Providers and Data Consumers to follow. But also App Providers are affected by data usage control, as they have to integrate control points (i.e., PEPs) into their software.

FUNCTIONAL LAYER
Since data usage control has a substantial impact on general security requirements, it affects also certain aspects of the functionality of the Industrial Data Space. As Data Owners specify data usage restrictions in data usage policies, these usage policies must be addressed when describing the properties of data by metadata. In addition, enforcement of data usage policies must be ensured, which affects Connectors, but also Data Apps.

PROCESS LAYER
Data usage control must be reflected in several basic processes of the Industrial Data Space, but mainly in data provision and data exchange processes:

- Data provision: The Data Provider has to make sure that the usage restrictions specified by the Data Owner are followed when describing the metadata about the data source in the form of a usage policy.
- Data exchange: Both the Data Provider and the Data Consumer have to comply with the data usage restrictions specified by the Data Owner. Compliance can be enforced by organizational/legal and/or technical measures.

INFORMATION LAYER
The Information Model contains modular metadata, including usage policy information. For data usage control, two policy forms are basically available: declarative, specification-level usage policies and technology-dependent, implementation-level policies. The Industrial Data Space uses Open Digital Rights Language (ODRL), an extension of the W3C standard, as a standard language for specifying data usage policies. In order to instantiate and enforce declarative, specification-level policies within individual target environments, a mapping to organizational and technical enforcement measures is required. Figure 4.14 illustrates a mapping example for different target environments, such as Integrated Distributed Data Usage Control Enforcement (IND²UCE), Label-based Usage Control (LUCON), or some unspecified technical enforcement (N).

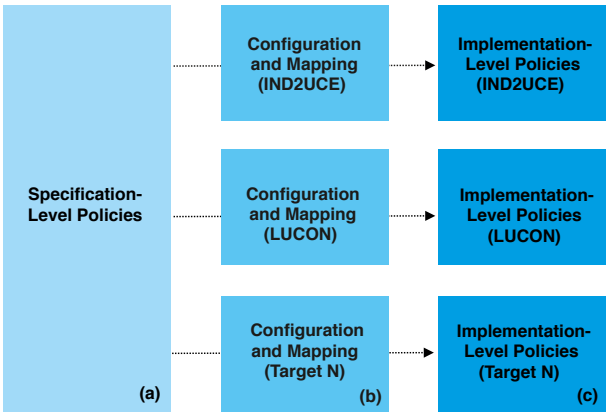


Figure 4.14: Example of mapping between specification-level policies and implementation-level policies

The usage control policies in the Information Model are only the declarative, specification-level policies that a Data Provider sends to the Broker Service Provider. The Software Provider implementing the Connector provides standard policy templates for the enforcement technologies used, which may be concretized afterwards by the Data Owner and the Data Provider. Technology-dependent, implementation-level policies can, but need not, be represented as references within the metadata. Independent of the references available, the two participants interacting with each other have to instantiate the specification-level policy and deploy it onto their own enforcement technologies. Depending on the technology used, the PMP handles policy deployment and revocation at the PDPs affected.

SYSTEM LAYER
The central technical component for data usage control in the Industrial Data Space is the Connector. The Connector connects the Data Provider and the Data Consumer to facilitate the exchange of data. Figure 4.15 illustrates how data usage control is integrated into the functionality of the Connector. Connectors implemented as prototypes so far use Apache

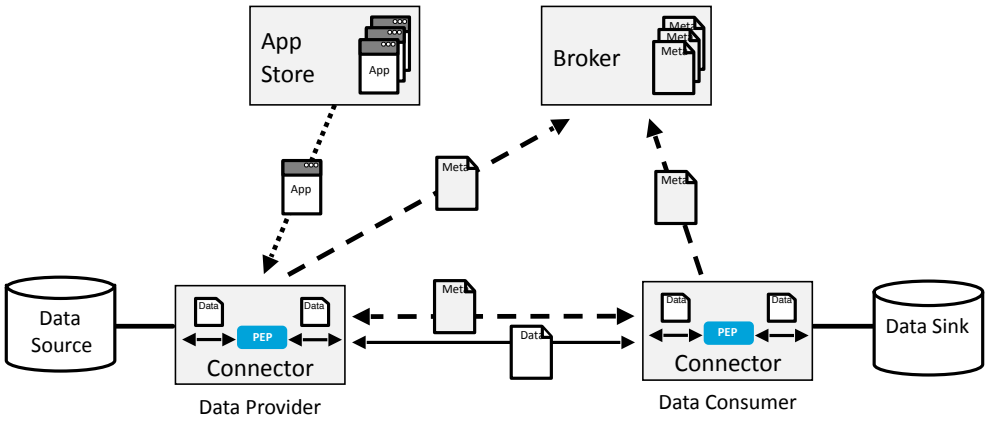


Figure 4.15 Data usage control on System Layer

Camel for data usage control across different systems and applications. To control the usage of data, the flow of data between the services and applications must be intercepted. Apache Camel allows integration of interceptors being executed each time before and after a processor is executed. By implementing PEPs as interceptors, the data flow can be controlled in the respective Message Router of the Connector (see Figure 4.16). Regarding the flow of data between a Data Provider and a Data Consumer, for example, one PEP controls the data exiting the Connector on the Data Provider side, while another PEP controls the data entering the Connector on the Data Provider side.

Depending on the policy desired, this kind of enforcement may not be sufficient to establish comprehensive data usage control and cover all possible use cases. For example, an App Provider may develop Data Apps that communicate with external systems. In this case, the Data Apps would require integration of additional control points (i.e., PEPs).

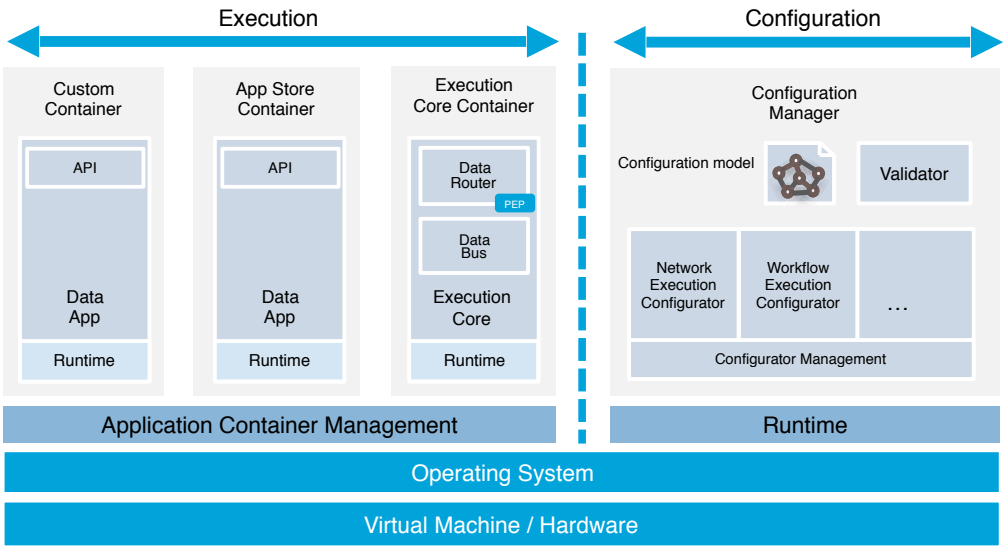


Figure 4.16: Example of integrating PEPs as interceptors into the Apache Camel Message Router

4.2 CERTIFICATION PERSPECTIVE

For participants and technical core components to provide a sufficiently high degree of security regarding the integrity, confidentiality, and availability of data exchanged in the Industrial Data Space, evaluating and certifying both is vital for the functioning of the Industrial Data Space. This necessitates the definition of a framework ensuring a consistent evaluation and certification process for all participants and components.

The Certification Scheme, following best practices from other, internationally accredited certification concepts, encompasses all processes, roles, rules, and standards governing the certification of participants and core components. While the certification of organizations and individuals aiming to participate in the Industrial Data Space focuses on security and trust, the certification of components also refers to compliance with specific requirements of the Industrial Data Space to ensure interoperability. This section provides an overview of how the central entities defined in the Reference Architecture Model (see Section 3) are linked with the Certification Scheme. After a general description of how certification is relevant at each of the five Layers of the Reference Architecture Model, it is outlined which roles are involved in the certification process, which entities and components are targets of certification, and how both sides interact with each other.

4.2.1 CERTIFICATION ASPECTS ON THE DIFFERENT ARCHITECTURAL LAYERS

BUSINESS LAYER
The Certification Body and the Evaluation Facility are the two roles in charge of the certification process. Their interactions and responsibilities with regard to certification are described in subsection 4.2.2.

Organizations assuming a role under one of the three categories Core Participant, Intermediary, and Software/Service Provider (see Section 3.1.2) are potential targets of certification. Subsection 4.2.3 describes for each role what level of certification is required and what the focus of the certification is.

FUNCTIONAL LAYER
The functional requirements of the Industrial Data Space are the core requirements expected to be implemented by the technical core components (e.g., the Connector or the Clearing House). Therefore, compatibility of each such implementation with these functional requirements forms the basis of the compliance part of a core component's certification. The security part of the certification focuses on security-specific requirements. As for the Security Perspective (see Section 4.1), these security-specific requirements are mainly related to the System Layer.

PROCESS LAYER
Whenever relevant for the compliance part of a component's certification, a component is also evaluated in terms of whether it fully supports all processes it is involved in, as defined by the Reference Architecture Model.

INFORMATION LAYER
Certification of a core component comprises also its compliance with the Reference Architecture Model regarding functionality, protocols, etc.. Whenever relevant, evaluation of a core component's compliance also refers to its compatibility with the Information Model defined at the Information Layer.

SYSTEM LAYER
The System Layer defines the possible interactions between the components, detailed requirements for the Connector, and specific types of Connector implementations. The System Layer is the predominant layer regarding the security part of a component's certification.

An overview of the core components that are targets of certification is presented in subsection 4.2.4.

4.2.2 ROLES IN THE CERTIFICATION PROCESS

The Certification Scheme of the Industrial Data Space comprises the roles shown in Figure 4.17. These roles were introduced under the "Governance Body" category specified at the Business Layer. The tasks of these roles with regard to the certification process are described in the following paragraphs.

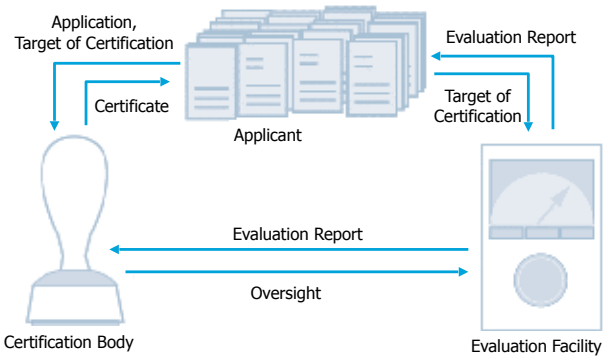


Figure 4.17 Certification process

It should be noted that all roles described in this section are specific to the Industrial Data Space (i.e. terms such as "Certification Body" should not be misunderstood to refer to an existing organization already granting certificates).

CERTIFICATION BODY

The Certification Body manages the entire certification process, defines standard evaluation procedures, and supervises the actions of the Evaluation Facility. It grants a certificate only if and when both the Evaluation Facility and the experts of the Certification Body have come to the conclusion that all preconditions for certification are fulfilled.

Responsibilities of the Certification Body include the following:

- defining the Certification Scheme in cooperation with the International Data Spaces Association by specifying evaluation procedures and certification criteria to be applied to participants and technical components;
- ensuring correct implementation and execution of the Certification Scheme;
- ensuring permanent compliance of participants and components with the Certification Scheme;
- analyzing existing „base" certificates (e.g., for organizations or for software and hardware security components) and deciding whether they can be accepted for evaluation within the Certification Scheme;
- reviewing and commenting on evaluation reports received from Evaluation Facilities;
- making the final decision about issuing or denying a certificate;
- authorizing and initiating the process of the generation and revocation of a X.509 certificate (these digital certificates represent the evaluation certificate and allow automated trust checks between partners prior to data exchange within the Industrial Data Space);
- deciding about approval or exclusion of Evaluation Facilities (to be) involved in evaluation procedures in the Industrial Data Space (based on ongoing monitoring activities);
- monitoring of external certification-relevant developments (e.g., new attack patterns which might circumvent certified security measures);
- driving the continuing development of the Certification Scheme based on practical quality assurance experiences.

Certificates issued in the Industrial Data Space have a limited validity period. In order to renew a certificate before it expires, re-certification is required, taking into account relevant developments that have happened in the meantime. Similarly, re-certification is required if changes are made to the target of certification (in case of minor changes, "lightweight", low-cost re-certification may be deemed sufficient). How major and minor changes are defined follows the definition used by widely accepted certification standards, such as ISO 27001 or Common Criteria.

The Certification Body itself may be accredited by the national accreditation body (e.g., DAkkS in Germany1), which supervises a number of certificate-granting institutions. Whether this will be arranged in the case of the Certification Body of the Industrial Data Space is still to be determined.

EVALUATION FACILITY
The Evaluation Facility is contracted by an Applicant. It is responsible for carrying out all technical and/or organizational evaluation activities during a certification process. The Evaluation Facility issues an evaluation report for the participant or technical component to be certified, listing details regarding the evaluation activities performed and regarding the security level specified. The latter determines the depth and scope of the evaluation activities performed.

Responsibilities of the Evaluation Facility include the following:

- taking care of the evaluation of a participant or component after obtaining approval from the Certification Body;
- applying the evaluation criteria specified by the Certification Scheme according to generally accepted standards and best practices (including execution of tests and on-site checks deemed necessary);
- documenting the results of the evaluation in an evaluation report;
- providing the evaluation report to the Certification Body.

The term "Evaluation Facility" refers both to authorized auditors for management system evaluation (for certification of participants) and to approved evaluators for product evaluation (for certification of components). Hence, multiple Evaluation Facilities may be present in the Industrial Data Space, but in each evaluation process only one Evaluation Facility is involved.

APPLICANT
The Applicant is not just the subject of the evaluation and certification process, but plays an active part in it. As such, the respective organization has to

- contract an Evaluation Facility approved by the Certification Body to carry out the evaluation process according to the Certification Scheme;
- formally apply for certification (at the Certification Body) in order to initiate the certification process;
- provide the necessary resources for the certification process in terms of financing and personnel;
- communicate openly and efficiently with, and provide all necessary information and evidence to, the Evaluation Facility and the Certification Body;
- respond adequately to any issues occurring in the course of the evaluation process.

Each Applicant has to submit an application for certification to start the process outlined above. This applies to organizations that develop software components and applications intended to be deployed within the Industrial Data Space (i.e., prospective Software Providers and App Providers) and to organizations that intend to become core participants or intermediaries in the Industrial Data Space. Depending on the specific case, the primary focus of the evaluation is either on the product or on the organization.

4.2.3 TARGETS OF CERTIFICATION
ENTITIES

CORE PARTICIPANTS
Data Providers are responsible for the integrity, confidentiality, and availability of the data they make available. Evaluation and certification of the security mechanisms employed by Data Providers is intended to provide a sufficient degree of security against the risk of data integrity, confidentiality, or availability being undermined by attacks.

Data Owners often act as Data Provider at the same time. In the case of the Data Owner and the Data Provider being different entities (i.e., the Data Owner does not publish the data itself, but hands over this task to a Data Provider), both the Data Owner and the Data Provider are responsible for integrity and confidentiality of the data. Responsibility for the availability of the data, however, rests solely with the Data Provider in this case, provided the Data Owner has handed over the data to the Data Provider. Regarding entities acting as a Data Owner only, evaluation and certification of the technical, physical, and organizational security mechanisms employed by them is intended to provide a sufficient degree of security against the risk of data integrity or confidentiality being undermined by attacks.

Data Consumers also have to assume responsibility for the confidentiality and integrity of data they receive from a Data Provider (i.e., in terms of making sure the data cannot leave the Industrial Data Space in an uncontrolled manner and cannot be corrupted before being used). Furthermore, Data Consumers have to make sure the data cannot be used for purposes other than permitted. Against all these risks, evaluation and certification of the technical, physical, and organizational security mechanisms employed by Data Consumers is intended to provide a sufficient degree of security.

INTERMEDIARIES
Since preventing sensitive data from ending up in the wrong hands is a central goal of the Industrial Data Space, it is critical to eliminate all risks involving manipulation of identities. The integrity and availability of identity-related information processed by the Identity Provider therefore is of utmost importance. Again, evaluation and certification of the security mechanisms employed by the respective organization (in combination with technical security measures in relation with the software components used for processing identity-related information) is intended to provide a sufficient degree of security against these risks.

The Broker Service Provider, the Clearing House, the App Store, and the Vocabulary Provider have one thing in common: they do not get in touch with sensitive payload data which the Industrial Data Space is designed to protect. The risk associated with possible breaches of confidentiality, integrity, and availability of e.g. metadata is rather low (with the exception of Clearing House transaction data, which, however, lies beyond the scope of the Industrial Data Space). Nevertheless, an attacker succeeding in exfiltrating or corrupting metadata, or impeding the availability of metadata, would be able to cause considerable damage to the Industrial Data Space or targeted participants, especially if such successful attacks would remain undetected over an extended period of time. Therefore, evaluation and certification tailored to the specific risk profiles of and security mechanisms employed by the Broker Service Provider, the Clearing House, the App Store, and the Vocabulary Providers is mandatory to ensure a sufficient degree of security against the risks mentioned. As far as the App Store is concerned, there is an additional risk in terms of an attacker successfully replacing legitimate, certified Data Apps with malware, threatening the payload data directly. To reduce this risk, technical measures on the level of the App Store implementation (e.g., only Data Apps cryptographically signed by the App Provider are accepted and distributed) seem to be more effective than organizational measures on the part of the App Store.

SOFTWARE PROVIDERS AND SERVICE PROVIDERS
As providers of software compliant with the requirements of the Industrial Data Space do not get in touch with sensitive data, usually no certification of the organizational security of Software Providers is required. If access to data of the Industrial Data Space is necessary, the Software Provider may assume the role of a Data Consumer or Data Provider for as long as such access is needed. In that case, the certification requirements of the respective role apply.

If a participant does not deploy the technical infrastructure required to participate in the Industrial Data Space itself, it can outsource certain tasks (e.g., publishing data) to a Service Provider hosting the required infrastructure. If this is the case, the Service Provider assumes the role of a Data Provider, Data Consumer, Broker Service Provider, etc., and performs the corresponding activities. Since the Service Provider then inherits the responsibilities and risks related to these roles, the certification requirements of the respective role apply here as well.

4.2.4 TARGETS OF CERTIFICATION
CORE COMPONENTS

Being the point of access to the Industrial 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 Industrial Data Space.

As the Broker does not have access to primary data (but only to metadata provided by Data Providers, which is generally considered less sensitive), and as it does not assign or enforce access rights (but merely supports data exchange), integrity and availability of metadata (i.e., correct and secure storing and handling of metadata) is of high importance for the Industrial Data Space. Compatibility of the Broker with the required functionality as defined by the Certification Body must therefore be evaluated and certified.

The Clearing House's activities comprise the provision of reports on the transactions performed for billing, conflict resolution, etc. As such, all implementations of the Clearing House have to be evaluated and certified according to the requirements as defined by the Certification Scheme.

The Identity Provider is required for secure operation of the Industrial Data Space. Since data sovereignty is a core value proposition of the Industrial Data Space, identity management is an essential system function. Therefore, the Identity Provider also has to be evaluated and certified according to the requirements as defined by the Certification Scheme.

Data Apps and Services have direct contact with primary data, which means that a compromised Data App or Service may compromise the integrity of data. However, confidentiality and availability of data is ensured by the measures defined in the Security Architecture of the Industrial Data Space, which strongly limit the potential damage caused by Data Apps and Services. Therefore, not every Data App or Service to be made available in an App Store of the Industrial Data Space requires certification. Nevertheless, certification should be required for apps and services of high importance to the Industrial Data Space community, and for apps and services having a high risk potential (e.g., anonymization apps for privacy protection). Requiring certification only for a small subset of apps ensures smooth and rapid evolution of the range of apps offered (especially since apps may have a significantly faster paced release cycle than other software components, and thus require frequent re-evaluation).

For certain Security Profiles (see Chapter 4.1.4), additional hardware security components are required to achieve an appropriate level of protection for access to sensitive data. In addition to the core software components of the Industrial Data Space, these hardware components must therefore be considered in the context of certification as well. In the interest of trustworthiness, and to avoid double certification, the use of third-party certified hardware components will be required (e.g., trusted platform modules certified in accordance with the Protection Profiles BSI-CC-PP-0030-2008 or ANSSI-CC-PP-2015/07). Certification activities of the Industrial Data Space regarding these components will be limited to checking the validity of existing base certificates.

4.3 GOVERNANCE PERSPECTIVE

The Governance Perspective of the Reference Architecture Model defines the roles, functions, and processes of the Industrial Data Space from a governance and compliance point of view. It thereby defines the requirements to be met by the business ecosystem to achieve secure and reliable corporate interoperability. This chapter provides an overview of how central questions of governance are defined on each Layer of the Reference Architecture Model (see Chapter 3). In particular, it describes how the Industrial Data Space enables companies to define rules and agreements for compliant collaboration.

While the Industrial Data Space enables all participants to act in compliance with negotiated rules and processes, it does not make any restrictions or enforce predefined regulations. The architecture of the Industrial Data Space should be seen as a functional framework providing mechanisms that can be customized by the participating organizations according to their individual requirements.

In more detail, the Industrial Data Space supports governance issues by

- providing an infrastructure for data exchange, corporate interoperability, and the use of new, digital business models;
- establishing trustworthy relationships between Data Owners, Data Providers, and Data Consumers;
- acting as a trustee for mediation between participants;
- facilitating negotiation of agreements and contracts;
- aiming at transparency and traceability of data exchange and data use;
- allowing private and public data exchange;
- taking into account individual requirements of the participants; and
- offering a decentralized architecture that does not require a central authority.

4.3.1 GOVERNANCE ASPECTS ON THE DIFFERENT ARCHITECTURAL LAYERS

BUSINESS LAYER
The Business Layer (see Chapter 3.1) facilitates the development and use of new, digital business models to be applied by the participants in the Industrial Data Space. It is thereby directly related to the Governance Perspective by considering the business point of view regarding data ownership, data provision, and data consumption, and by describing core service concepts such as data brokerage.

FUNCTIONAL LAYER
The Functional Layer (see Chapter 3.2) defines the functional requirements of the Industrial Data Space, and the concrete features resulting from them, in a technology-independent way. Beside the Clearing House and the Identity Provider, which are entities for which the relation to governance is obvious, the functionality of certain technical core components (e.g., the App Store or the Connector) relates to the Governance Perspective.

PROCESS LAYER
Providing a dynamic view of the architecture, the Process Layer (see Chapter 3.3) describes the interactions taking place between the different components of the Industrial Data Space, . The three major processes described in the Process Layer section (providing data, exchanging data, and publishing and using Data Apps) are directly related to the Governance Perspective as they define its scope regarding the technical architecture.

INFORMATION LAYER
The Information Layer (see Chapter 3.4) specifies the Information Model, which provides a common vocabulary for the participants to express their concepts. It thereby defines a framework for standardized collaboration and for using the infrastructure of the Industrial Data Space for establishing individual agreements and contracts. The vocabulary plays a key role in the Governance Perspective because of its relevance for describing data by metadata in the Industrial Data Space.

SYSTEM LAYER
The System Layer (see Chapter 3.5) relates to the Governance Perspective due to its technical implementation of different security levels for data exchange between the Data Endpoints in the Industrial Data Space.

The following subsections describe five topics that are addressed by the Governance Perspective. These topics play an important role when it comes to the management of data goods.

4.3.2 DATA AS AN ECONOMIC GOOD

As data can be decoupled from specific hardware and software implementations, it turns into an independent economic good. While this opens up new opportunities, it creates challenges as well. To ensure competitiveness of organizations, a solution is required that facilitates new, digital business models.

The Industrial Data Space offers a platform for organizations to offer and exchange data and digital services. In doing so, it offers a basic architecture for organizations that want to optimize their data value chains. The main goal is to enable participants to leverage the potential of their data within a secure and trusted business ecosystem. The Industrial Data Space thereby covers the information system perspective and provides the components that enable participants to define individual business cases.

The Industrial Data Space neither makes any statements on legal perspectives, nor does it restrict participants to any predefined patterns. Instead, it offers the possibility to design digital business models individually and as deemed appropriate.

4.3.3 DATA OWNERSHIP

In the material world, the difference between the terms "possession" and "property" is an abstract, yet necessary construct. It is accepted that moving a good from one place to another and changing possession of the good does not necessarily have an impact on the property rights. Regarding the specific concept of the Industrial Data Space, it is necessary to take into account that the Data Owner and Data Provider may not be identical (see Chapter 3.1.1).

Data ownership is an important aspect when it comes to offering data and negotiating contracts in a digital business ecosystem, especially because data can easily be duplicated. The Industrial Data Space makes sure the topic of data ownership is comprehensively addressed by providing a secure and trusted platform for authorization and authentication within a decentralized architecture. This allows Data Providers as well as Service Providers to be identified and controlled by an Identity Provider (see Chapter 3.1.1).

Decentralized data exchange through Connectors, in contrast to other architectures of data networks (e.g., data lakes or cloud services), ensures full data sovereignty for Data Owners. In addition to these self-control mechanisms, the architecture allows logging of data transfer information at a Clearing House (see Chapter 3.2.5). Data ownership thus is indeed relevant on every layer of the architecture.

As the Industrial Data Space intends to build upon and apply existing law, it will not include any purely technology-oriented solutions to prevent data duplication or misuse of data goods. However, it supports these important aspects over the entire data lifecycle. Furthermore, it supports the arrangement of collaborative solutions by providing an appropriate technical infrastructure.

4.3.4 DATA SOVEREIGNTY

Data sovereignty is a natural person's or corporate entity's capability of being entirely self-determined with regard to its data. The Reference Architecture Model presented in this document particularly addresses this capability, as it specifies requirements for secure data exchange and restricted data use in a trusted business ecosystem.

The Industrial Data Space promotes interoperability between all participants based on the premise that full self-determination with regard to one's data goods is crucial in such a business ecosystem. Data exchange takes place by means of secured and encrypted data transfer including authorization and authentication. The Data Provider may attach metadata to the data transferred using the IDS Vocabulary. In doing so, the terms and conditions to ensure data sovereignty can be defined unambiguously (e.g., data usage, pricing information, payment entitlement, or time of validity). The Industrial Data Space thereby supports the concrete implementation of applicable law, without predefining conditions from a business point of view, by providing a technical framework that can be customized to the needs of individual participants.

4.3.5 DATA QUALITY

Because of the correlation between good data quality and maximizing the value of data as an economic good, the Industrial Data Space explicitly addresses the aspect of data quality. Due to this premise, the Industrial Data Space enables its participants to assess the quality of data sources by means of publicly available information and the transparency it provides with regard to the brokerage functionality it offers. Especially in competitive environments, this transparency may force Data Providers to take data maintenance more seriously. By extending the functionality of the Connector with self-implemented Data Apps (see Chapter 3.2.4), the Industrial Data Space lays the foundation for automated data (quality) management.

4.3.6 DATA PROVENANCE

By creating transparency and offering clearing functionality, the Industrial Data Space provides a way to track the provenance and lineage of data. This is strongly linked to the topics of data ownership and data sovereignty. The implementation of data provenance aspects is part of the IDS Vocabulary (see Chapter 3.2.3), which is maintained by the participants during the process of data exchange. Additionally, the Clearing House (see Chapter 3.1.1) logs all activities performed in the course of a data exchange transaction, and requests confirmations of successful data exchange from the Data Provider and the Data Consumer. In doing so, data provenance is always recursively traceable.

The Industrial Data Space thereby provides the possibility to implement and use appropriate concepts and standards. However, it does not force participants to use these concepts and standards. It is therefore up to the individual participant to provide correct information (i.e., metadata) on the provenance of data.

APPENDIX: GLOSSARY

Term	Definition
App Store	Secure platform for distributing Data Apps; features different search options (e.g. by functional or non-functional properties, pricing model, certification status, community ratings, etc.)
Applicant	Organization formally applying for being certified by the Certification Body
Broker Service Provider	Intermediary managing a metadata repository that provides information about the Data Sources available in the Industrial Data Space; multiple Broker Service Providers may be around at the same time, maintaining references to different, domain-specific subsets of Data Endpoints
Certification Authority	Trusted third-party entity issuing digital certificates (e.g., x509 certificates); may host services to validate certificates issued
Certification Body	Governance body certifying components and entities seeking admission to the Industrial Data Space; aside from having the final word on granting or denying a certificate, it is responsible for maintaining the Certification Scheme (including its catalog of requirements), for overseeing and approval of Evaluation Facilities, and for ensuring compatibility of evaluation procedures carried out by different Evaluation Facilities
Certification Scheme	Scheme defining the processes, roles, targets, and criteria involved in the certification of components and entities; maintained by the Certification Body
Clearing House	Intermediary providing clearing and settlement services for all financial and data exchange transactions within the Industrial Data Space
Connector	Dedicated communication server for sending and receiving data in compliance with the general Connector specification; different types of Connectors can be distinguished (Base Connector vs. Trusted Connector, or Internal Connector vs. External Connector)

Term	Definition
Data App	Self-contained, self-descriptive software package that is distributed via the App Store and deployed inside a Connector; provides access to data and data processing capabilities; the interface of a Data App is semantically described by the IDS Vocabulary
Data Asset	Content exposed for exchange via Data Endpoints according to a parametrized Data Service interface; Data Assets are expected to be focused, homogeneous, and consistent over time with regard to granularity, coverage, context, data structure, and conceptual classification
Data Consumer	Core participant in the Industrial Data Space requesting and using data provided by a Data Provider
Data Endpoint	Data interface for data publication (Data Source) and data consumption (Data Sink), respectively
Data Exchange Agreement	Contractual agreement between a Data Provider and a Data Consumer regarding the exchange of data in the Industrial Data Space
Data Owner	Core participant owning the legal rights for, and having complete control over, the data it makes available in the Industrial Data Space; defines the terms and conditions of use of its data
Data Provider	Core participant exposing Data Sources via a Connector; a Data Provider may be an enterprise or other organization, a data marketplace, an individual, or a “smart thing”
Data Sink	Data Endpoint consuming data uploaded and offered by a Data Provider
Data Source	Data Endpoint exposing data for being retrieved or subscribed to by a Data Consumer

APPENDIX: GLOSSARY

Term	Definition
Data Sovereignty	The capability of an entity (natural person or corporate) of being entirely self-determined with regards to its data
Dynamic Attribute Provisioning Service (DAPS)	Issues Dynamic Attribute Tokens (DATs) to verify dynamic attributes of participants or Connectors
Dynamic Attribute Token (DAT)	Contains signed dynamic attributes for participants and Connectors
Evaluation Facility	Governance body providing services related to the certification of components and entities (certification targets) seeking admission to the Industrial Data Space; responsible for detailed technical evaluation of targets in consistence with the Certification Scheme and its catalog of requirements; reports evaluation results to the Certification Body
Governance	Concept defining the rights and duties (“rules of the game”) for formal data management, ensuring quality and trust throughout the Industrial Data Space; mission critical to the Industrial Data Space, as a central supervisory authority is missing
Identity Provider	Intermediary offering services to create, maintain, manage and validate identity information of and for participants in the Industrial Data Space
Information Model	Set of vocabularies and related schema information for the semantic description of Industrial Data Space entities (e.g., Data Endpoints or Data Apps), data provenance, or licensing information; the core IDS Vocabulary is domain-independent; it can be extended and/or reference third-party vocabularies to express domain-specific aspects
Industrial Data Space	Distributed network of Data Endpoints (i.e., instantiations of the Industrial Data Space Connector), allowing secure exchange of data and guaranteeing Data Sovereignty

Term	Definition
Participant	Stakeholder in the Industrial Data Space, assuming one or more of the predefined roles; every participant is given a unique identity by the Identity Provider
Security Profile	Defined set of a Connector's security properties; specifies several security aspects (e.g., isolation level, attestation, or authentication), expressing the minimum requirements a Data Consumer must meet to be granted access to the Data Endpoints exposed
System Adapter	Data App used for integration of custom Data Sources and legacy systems with a Connector
Usage Policy	Set of rules specified by the Data Owner restricting usage of its data; covers aspects like time-to-live or forwarding conditions (e.g., anonymization or scope of usage); transmitted along with the respective data, and enforced while residing on the Connector of the Data Consumer
Vocabulary Hub	Server providing maintenance facilities for editing, browsing and downloading vocabularies and related documents; mirrors a set of external third-party vocabularies ensuring seamless availability and resolution

LEGAL OFFICE:

International Data Spaces Association
Anna-Louisa-Karsch-Str. 2
10178 Berlin
Germany

HEAD OFFICE:

International Data Spaces Association
Joseph-von-Fraunhofer-Str. 2 – 4
44227 Dortmund

Phone: +49 (0) 231 9743 - 619
info@industrialdataspace.org