Fraunhofer

REFERENCE ARCHITECTURE MODEL FOR THE INDUSTRIAL DATA SPACE

IN COOPERATION WITH

INDUSTRIAL DATA SPACE ASSOCIATION

Authors:
Prof. Dr.-Ing. Boris Otto, ISST
Dr.-Ing. Steffen Lohmann, IAIS

Prof. Dr. Sören Auer, IAIS
Gerd Brost, AISEC
Dr.-Ing. Jan Cirullies, IML/ISST
Andreas Eitel, IESE Thilo
Ernst, FOKUS

Dr.-Ing. Christian Haas, IOSB
Manuel Huber, AISEC
Christian Jung, IESE
Prof. Dr. Jan Jürjens, ISST
Dr. Christoph Lange, IAIS
Dr. Christian Mader, IAIS
Nadja Menz, FOKUS
Dr. Ralf Nagel, ISST

Heinrich Pettenpohl, ISST
Jaroslav Pullmann, FIT
Dr. Christoph Quix, FIT
Jochen Schon, IAIS
Daniel Schulz, IAIS
Dr. Julian Schütte, AISEC
Markus Spiekermann, ISST
Dr. Sven Wenzel, ISST

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

Industrial Data Space e.V.
Anna-Louisa-Karsch-Str. 2
10178 Berlin, Germany

Internet: www.fraunhofer.de
E-Mail: info@zv.fraunhofer.de

Coordination
Editorial: Dr.-Ing. Steffen Lohmann
Copyediting: Tom Fleckstein, Text-World
Design: Fraunhofer-Gesellschaft
Typesetting and page layout:
www.ansichtssache.de

SPONSORED BY THE
Federal Ministry of Education and Research

DLR Project Management Agency

INDUSTRIAL DATA SPACE

Grant ID 01IS15054

TABLE OF CONTENTS

1 INTRODUCTION

The Industrial Data Space is a virtual data space leveraging existing standards and technologies, as well as accepted governance models, to facilitate the secure exchange and easy linkage of data in a trusted business ecosystem. It thereby provides a basis for smart service scenarios and innovative business processes, while at the same time ensuring data sovereignty 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 proposes a Reference Architecture Model for this particular capability and related aspects, including requirements for secure data exchange in business ecosystems.

The Industrial Data Space is an initiative that is institutionalized by two main activities: a Fraunhofer research project entitled "Industrial Data Space" and the "Industrial Data Space Association". While the research project is concerned with the design and prototype implementation of the Reference Architecture Model, the association unites the requirements from various industries and provides use cases to test the results gained from its implementation.

The Industrial Data Space aims at meeting the following strategic requirements:

- **Data usage control:** In line with the central aspect of ensuring data sovereignty, a data owner in the Industrial Data Space may attach usage restriction information to its data before it is transmitted to a data consumer. The data consumer may use this data only if it fully agrees to that usage policy.
- **Decentralized approach:** The architecture of the Industrial Data Space does not require central data storage capabilities. Instead, it follows a decentralized approach, which means that data physically remains with the respective data owner until it is transmitted to a trusted party. Thus, the Industrial Data Space is not a cloud platform, but an architectural approach to connect various, different platforms (both operational and emerging ones). Nevertheless, participants in the Industrial Data Space may agree on trusted entities offering central data storage, if deemed necessary.
- **Multiple implementations:** The Industrial Data Space Connector, being a central component of the architecture, is implemented in different versions: a standard version, which runs in corporate or cloud environments, a mobile version running on devices with limited capacities, and an IoT version tailored to the requirements of Internet-of-Things based scenarios.
- **Standardized interfaces:** Both the architecture of the Industrial Data Space Connector and its communication API are subject to standardization.
- **Certification:** The Industrial Data Space materializes as a distributed network of Connectors, representing data endpoints. Each implementation of the Connector, as well as each organization seeking admission to the Industrial Data Space, has to undergo a certification process, ensuring trust and reliability across the entire business ecosystem.
- **Data economy:** The Industrial Data Space Connector allows the creation of novel, data-driven services making use of data apps (i.e., software components providing dedicated data-related service functionality). The participants in the Industrial Data Space can request these data apps from an app store.
- **Secure data supply chains:** The Industrial Data Space aims at enabling secure data supply chains (i.e., networks consisting of data providers and data consumers), ranging from the source the data originates from (e.g., a sensor on an IoT device) to the actual point of use (e.g., an industrial smart service for predictive maintenance).

1.1 GOALS OF THE INDUSTRIAL DATA SPACE

As the central deliverable of the research project, the Reference Architecture Model constitutes the foundation for 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 Industrial Data Space Association is a non-profit organization institutionalized under the German law of associations. Every organization is invited to participate, as long as it follows 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.

Funded by the German Federal Ministry of Education and Research (BMBF), the Industrial Data Space research project is pre-competitive. In cooperation with the non-profit Industrial Data Space Association, the initiative pursues the following mission:

> The Industrial Data Space stands for secure data exchange between its participants, while at the same time ensuring data sovereignty for the participating data owners. The Industrial Data Space Association defines a framework and governance principles for the Reference Architecture Model, as well as interfaces aiming at establishing an international standard. This standard is being developed using agile approaches and use-case scenarios. It forms the foundation for a variety of certifiable software solutions and business models, the emergence of which is in the stated interest of the Industrial Data Space Association.

1.2 PURPOSE AND STRUCTURE OF THE DOCUMENT

The purpose of this document is to introduce the Reference Architecture Model for 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.

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

The general structure of the Reference Architecture Model is illustrated in Figure 1.1. The Business Layer specifies and categorizes the different stakeholders (namely the roles) of the Industrial Data Space, including their activities and the interactions between them. The Functional Layer comprises the functional requirements of the Industrial Data Space and the concrete features derived from them (in terms of abstract, technology-agnostic functionality of logical software components). The Process Layer provides a dynamic view of the architecture; using the BPMN notation, it describes the interactions between the different components of the Industrial Data Space. 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 (e.g., participants active, Data Endpoints deployed, Data Apps advertised, or datasets exchanged). 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.

In addition, the Reference Architecture Model contains three cross-sectional perspectives (Security, Certification, and Governance) in order to consider the implementation of core concepts of the Industrial Data Space with regard to each of the layers (Figure 1.1).

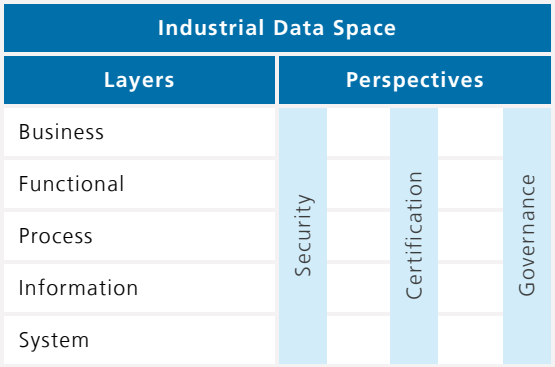


Figure 1.1: General structure of Reference Architecture Model

2

CONTEXT OF THE INDUSTRIAL DATA SPACE

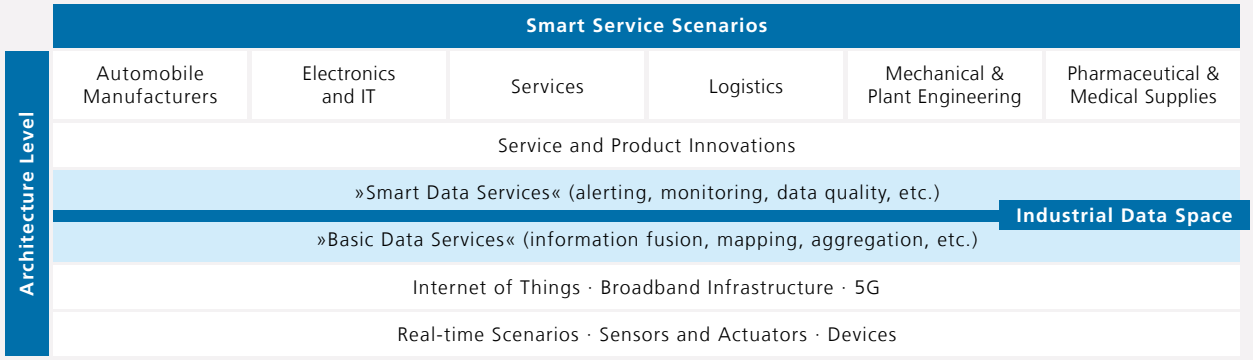


Figure 2.1: Typical enterprise architecture stack

The Industrial Data Space initiative contributes to the design of enterprise architectures in commercial and industrial digitization scenarios. Figure 2.1 shows a typical architecture stack of the digital industrial enterprise.

By providing an architecture for secure exchange of data, the Industrial Data Space bridges the gap between lower-level architectures for communication and basic data services and more abstract architectures for smart data services. It therefore supports the establishment of secure data supply chains from the lowest layer (i.e., the data source) to the highest layer (i.e., data use), while at the same time ensuring data sovereignty for data owners.

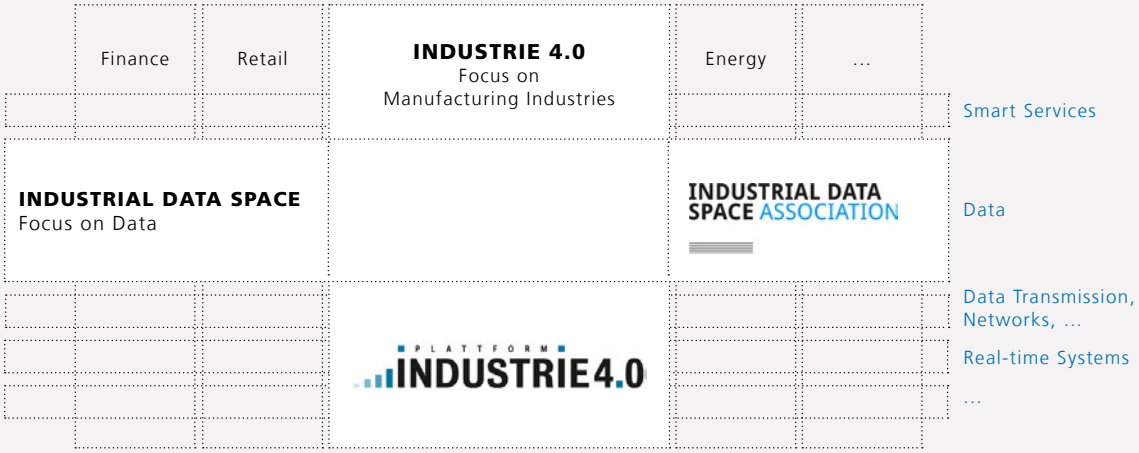


Figure 2.2: Relations with Plattform Industrie 4.0

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, in particular the Reference Architectures working group. 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 (see Figure 2.2). 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 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

- Big Data Value Association,[1]
- FIWARE Foundation,[2]
- Industrial Internet Consortium,[3]
- OPC Foundation,[4] and
- Plattform Industrie 4.0.[5]

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

1 http://www.bdva.eu
2 https://www.fiware.org/foundation
3 http://www.iiconsortium.org
4 https://opcfoundation.org
5 http://www.plattform-i40.de

3

LAYERS OF THE REFERENCE ARCHITECTURE MODEL

The Reference Architecture Model comprises five layers, each one addressing specific stakeholder concerns and viewpoints.

3.1 BUSINESS LAYER

The Business Layer specifies the roles of the participants in the Industrial Data Space. This is mainly done from a business perspective; i.e., the roles are defined on the basis of their respective activities to create added value for other participants. Furthermore, the Business Layer 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 of the different roles, which is important in the subsequent sections to identify the components of the architecture.
Participation in the Industrial Data Space requires the use of software that is compliant with the Reference Architecture Model. However, the Industrial Data Space is not limited to the software of a specific provider, as it uses an open architecture. This implies that a service can be offered by multiple organizations, including general services in the Industrial Data Space infrastructure, such as a metadata broker or a digital distribution platform (below defined as "App Store"). At the same time, an organization may offer services that relate to several roles.
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. Certain roles may require certification of the organization that wants to assume that role, including certification of the software the organization uses. In general, certification of organizations participating and software used in the Industrial Data Space is considered a measure to increase mutual trust among the participants (especially with regard to central roles, such as the App Store Provider, the Identity Provider, or the Clearing House, which should act as trusted intermediaries). The Certification Scheme applied is described in detail in Section 4.2.

Data Owner

The Data Owner (Section 4.3.3) has the legal rights and complete control over its data. Usually, a participant acting as a Data Owner assumes the role of a Data Provider at the same time. However, there may be cases in which the Data Owner is not the Data Provider (e.g., if the data is technically managed by a different entity than the Data Owner; an example would be a company using an external IT service provider for data management).
The only activity of the Data Owner is to authorize a Data Provider to publish its data. Any authorization should be documented in a contract, including a policy describing the permissions granted to that specific data. This contract must not necessarily be a paper document, but may also be an electronic file.

Data Provider

The Data Provider exposes data to be exchanged in the Industrial Data Space. In most cases, the Data Provider is identical with the Data Owner, but not necessarily (see above). 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.
Exchanging data with a Data Consumer is the main activity of a Data Provider. To facilitate a data request from a Data Consumer, the Data Provider must provide metadata about its data to a broker first. However, a Broker is not necessarily required for a Data Consumer and a Data Provider to establish a connection.
In addition to exchanging data, other activities may be performed by the Data Provider and the Data Consumer. For example, at the end of a data exchange completely or partially performed, the successful (or unsuccessful) completion of the

transaction (from the perspective of the Data Provider) can be logged at a Clearing House (to facilitate billing, conflict resolution, etc.). Furthermore, a Data Provider can use Data Apps to enrich, transform, or improve their data in some way (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 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 using a Broker. The Broker then provides the required metadata for the Data Consumer to connect to a Data Provider. Alternatively, the connection between the Data Provider and the Data Consumer can be established directly (i.e., without involving a Broker). In cases in which the connection information of the Data Provider is already known to the Data Consumer, the Data Consumer may retrieve the data (and the corresponding metadata) directly from the Data Provider.
Similar to the Data Provider, the main activity of the Data Consumer is to exchange data. And like the Data Provider, the Data Consumer may log the transaction details of a successful (or unsuccessful) data exchange at a Clearing House, use Data Apps to enrich, transform, or improve the data received from the Data Provider (or its own data) in some way, and use a Service Provider to connect to the Industrial Data Space (if it does not deploy the technical infrastructure for participating in the Industrial Data Space itself).

Broker Service Provider
The main duty of the Broker Service Provider is to manage a metadata repository that provides 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). 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 in connection with that role; 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 to receive metadata from the Data Providers. This metadata should be stored in some internal repository for being queried in a structured manner. While the core of the metadata model must be specified by the Industrial Data Space (by the Information Model, see Section 3.4), a Broker Service Provider may extend the metadata model to manage additional metadata elements.
For metadata retrieval, it should be possible to query the stored metadata in a structured manner. Although the query interface is standardized in the Industrial Data Space, a Broker Service Provider may provide specific extensions. 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 act as a trusted, intermediate entity between the Data Provider and the Data Consumer.
The Clearing House should log all activities performed in the course of a data exchange. After (part of) a data exchange has been completed, both the Data Provider and the Data Consumer should confirm transmission and reception of the data, respectively, by logging the transaction at the Clearing House. Based on the logged data, the transaction can be billed then. 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 should also provide reports on the performed (logged) transactions for billing, conflict resolution, etc.

Identity Provider
For secure operation, and to avoid unauthorized access to data in the Industrial Data Space, there has to be a service to verify identities. An identity needs to be described by a set of properties (e.g., characterizing the role of the identity within an organization). The Identity Provider should offer a service to create, maintain, manage and validate identity information of and for participants in the Industrial Data Space. More details about identity management can be found in Section 4.1.

App Store Provider
The App Store provides applications that can be deployed in the Industrial Data Space to enrich the data processing workflows. An option would be to have the artifacts of an App Store certified by a Certification Body, following the certification procedures defined in Section 4.2.
The App Store Provider is responsible for managing information about Data Apps offered by App Providers. App Providers should describe their Data Apps using metadata, in compliance with a metadata model describing the semantics of the services. 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 (Section 3.5). In addition, Data Apps can be certified by a Certification Body, which would increase the trust in such apps (especially with regard to Data Apps dealing with sensitive information). All Data Apps need to be published in an App Store for being accessed and used by Data Consumers and Data Providers. Each Data App should include metadata describing it (e.g., its functionality and interfaces).

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 Industrial Data Space Vocabulary (Section 3.4). However, other (domain specific) vocabularies can be provided as well.

Software Provider
A Software Provider provides software that implements 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 and used on the basis of individual agreements between a Software Provider and a software user (e.g., a Data Consumer, a Data Provider, or a Broker Service Provider). The difference between an App Provider and a Software Provider is that App Providers distribute their apps exclusively via the App Store, whereas Software Providers use their usual channels for distribution of their products (which means 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 to participate in the Industrial Data Space itself, it can transfer the data to be made available in the Industrial Data Space to a Service Provider hosting the required infrastructure for other organizations. If this is the case, this Service Provider assumes the role of a Data Provider, Data Consumer, Broker Service Provider, etc. and performs the corresponding activities.
The role of the Service Provider covers also providers offering additional services (e.g., data analysis, data integration, data cleaning, or semantic enrichment) to improve the data exchanged in the Industrial Data Space. From a technical point of view, such service providers can be considered Data Providers and Data Consumers at the same time (e.g., as a Data Consumer they receive data from a Data Provider, provide their specific service, and then turn into a Data Provider and offer the data in the Industrial Data Space). Unlike these services, 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.

Industrial Data Space Association
The Industrial Data Space Association is a non-profit organization promoting the continuous development of the Industrial Data Space. It supports and governs the development of the Reference Architecture Model. The Industrial Data Space Association is currently organized across several working groups, each one addressing a specific topic (e.g., architecture, use cases and requirements, or certification). Members of the Association are primarily large industrial enterprises, IT companies, SMEs, research institutions, and industry associations.
As the Industrial Data Space Association is not directly involved in the data exchange activities of the Industrial Data Space, its role will not be further addressed by the other layers.

Certification Body and Evaluation Facility
The Certification Body and the Evaluation Facility are in charge of the certification of the participants in the Industrial Data Space, the core software components, and the providers of compliant software. 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 between them. As some of the roles (Industrial Data Space Association, Certification Body, and Evaluation Facility) are not actively involved in the exchange of data, they are omitted from the illustration. 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 Provider, Data Consumer, and Data Owner. The role of a Core Participant can be assumed by any organization that owns, wants to provide, and/or wants to consume data.
The product relevant for these roles is data. Added value is created by providing or consuming 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 Provider, Vocabulary Provider, and Identity Provider. Only trusted organizations should assume these roles. Added value is created by these roles as they promote trust throughout the Industrial Data Space and provide metadata to the other participants.

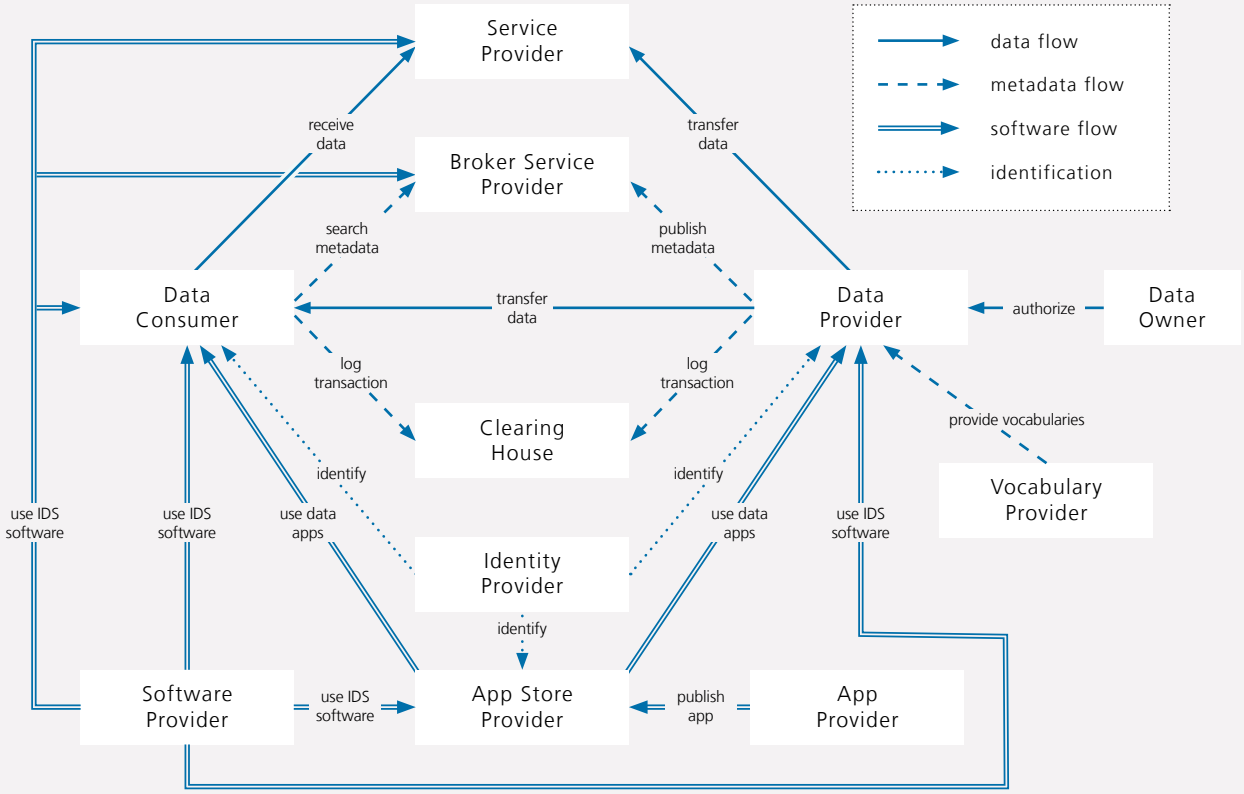


Figure 3.1: Roles and interactions in the Industrial Data Space

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 other participants of the Industrial Data Space. Roles subsumed under this category are Service Provider, Software Provider, and App Provider.

Added value is created by providing 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 a Service Provider. The process of providing software used for establishing the endpoints of an exchange of data 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 and the Industrial Data Space Association. These two bodies make sure that only compliant organizations enter this well-defined business ecosystem.

Value is created by the roles of this category through performing the certification process and issuing certificates (both for organizations that want to assume a role and for software components used). As the governance of the Industrial Data Space is a permanent, yet rather hidden activity, it does not create a direct value. However, it sustainably increases and protects the overall value of the Industrial Data Space.

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 list entitled Functional Overview, containing all functional requirements identified. Figure 3.2 shows the Functional Architecture of the Industrial Data Space, grouping the single items of the list into functional entities to be provided by the Industrial Data Space. As can be seen in the figure, Trust & Security encapsulates the other functional entities; this is because trust and security are key concepts of the Industrial Data Space, having an impact on all the other functional entities.

The central functional entity of the Industrial Data Space is the Connector. It facilitates the exchange of data between participants. The Connector is basically a dedicated communication server for sending and receiving data in compliance with the Connector specification (Section 3.5.1). A single Connector can be understood as a node in the peer-to-peer architecture of the Industrial Data Space. This means that a central authority for data management is not required.

Connectors can be installed, managed and maintained both by Data Providers and Data Consumers. Typically, a Connector is operated in a secure environment (e.g., beyond a firewall). This means that internal systems of an enterprise cannot be directly accessed. However, the Connector can, for example, also be connected to a machine or a transportation vehicle. Each company participating in the Industrial Data Space may operate several Connectors. As an option, intermediaries (i.e., the Service Provider) may operate Connectors on behalf of one or several participating organizations (Section 3.1.2). The data exchange with the enterprise systems must be established by the Data Provider or the Data Consumer.

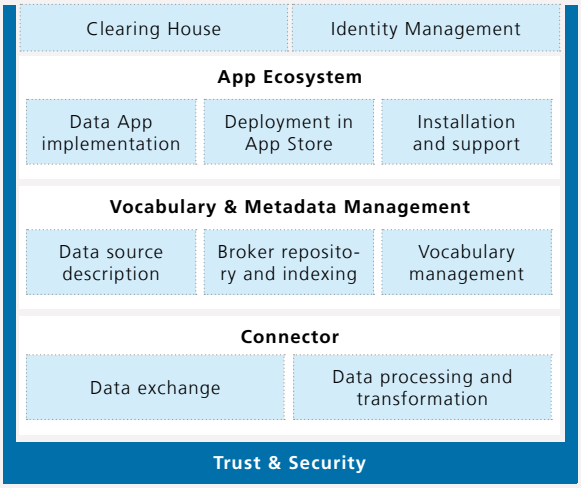


Figure 3.2: Functional Architecture of the Industrial Data Space

Data Providers can offer data to other participants of the Industrial Data Space. The data therefore has to be described by metadata. The metadata contains information about the Data Provider, syntax and semantics of the data itself, and additional information (e.g., pricing information or usage policies).
To support the creation of metadata and the enrichment of data with semantics, vocabularies can be created and stored for other participants in the Vocabulary and Metadata Management component. If the Data Provider wants to offer data, the metadata will automatically be sent to one or more central metadata repositories hosted by the Broker. Other participants can browse and search data in this repository.
Connectors can be extended with software components that help transform and/or process data. These Data Apps constitute the App Ecosystem. Data Apps can either be purchased via the App Store or developed by the participants themselves. App Providers may implement and provide Data Apps using the App Store.
Every participant possesses identities required for authentication when communicating with other participants. These identities are managed by the Identity Management component.
The Clearing House logs each data exchange between two Connectors.
Each functional entity described above is characterized by different requirements. In the following, the numbers of the items are added as they appear in the Functional Overview in Appendix B (e.g., [IDSFO-1]).

3.2.1 Trust and Security

Although requirements related to trust and security are usually non-functional, they are addressed by the Functional Layer, since they represent fundamental features of the Industrial Data Space. The Trust & Security entity can be split into three main aspects: Security, Certification, and Governance, representing the three cross-sectional perspectives of the Reference Architecture Model.

Security

Connectors, App Stores, and Brokers can check if the Connector of the connecting party is running a trusted (certified) software stack [Appendix B, IDSFO-71]. Any communication between (external) Connectors can be encrypted and integrity protected [IDSFO-42]. 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 [IDSFO-53]. To reduce the impact of compromised applications, appropriate technical measures must be applied to isolate Data Apps from each other and from the Connector [IDSFO-86]. Data Providers and Data Consumers can decide about the level of security of their respective Connectors by deploying Connectors supporting the selected security profile [IDSFO-76]. More information about security is given in Section 4.1.

Certification

The core components of the Industrial Data Space, and especially the Connector, require certification from the Certification Body [IDSFO-102], along with the organizations participating in the Industrial Data Space [IDSFO-60], in order to establish trust among all participants. More information about the certification process is given in Section 4.2.

Governance

So far, no general requirements related to governance could be identified. However, since data governance is an important topic for the Industrial Data Space, such requirements are likely to occur in the future. More information about data governance is given in Section 4.3.

3.2.2 Connector

Participants should be able to run the Connector software in their own IT environment [IDSFO-63]. Alternatively, they may run a Connector on mobile or embedded devices [IDSFO-65]. The operator of the Connector must be able to define the data workflow inside the Connector [IDSFO-32]. Users of the Connector must be identifiable and manageable [IDSFO-79]. Passwords and key storage must be protected [IDSFO-88]. Every action, data access, data transmission, incident, etc. should be logged [IDSFO-4, IDSFO-56, and IDSFO-95]. Using this logging data, it should be possible to draw up statistical evaluations on data usage etc. [IDSFO-5]. Notifications about incidents should be sent automatically [IDSFO-77].

Data Exchange

The Connector must receive data from an enterprise backend system, either through a push mechanism or a pull mechanism [IDSFO-33]. The data can be provided via an interface or pushed directly to other participants [IDSFO-43]. To do so, each Connector must be uniquely identifiable [IDSFO-44]. Other Connectors may subscribe to data sources, or pull data from these sources [IDSFO-31]. Data can be written into the backend system of other participants [IDSFO-52].

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 [IDSFO-34]. 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) [IDSFO-35].

3.2.3 Vocabulary and Metadata Management

Participants must have the opportunity to describe [IDSFO-2, IDSFO-98], publish [IDSFO-8], maintain [IDSFO-9] and manage different versions of metadata [IDSFO-10]. Metadata should describe the syntax and serialization [IDSFO-97] as well as the semantics [IDSFO-96] of data sources. Furthermore, metadata should describe the application domain of the data source [IDSFO-94]. The operator of the Connector must be able to define the price, the price model [IDSFO-3], and the usage policies [IDSFO-7] regarding certain data.

Broker and Indexing

The operator of a Connector must be able to provide an interface for data and metadata access [IDSFO-37]. Each Connector must be able to transmit metadata of its data sources to one or more brokers [IDSFO-66]. Every participant must be able to browse [IDSFO-26] and search [IDSFO-25] 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 [IDSFO-59].

Vocabulary Management

To create metadata, the operator may use vocabularies which help structure metadata. The operator can use standard vocabularies, create own vocabularies [IDSFO-11], 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 [IDSFO-17], selection [IDSFO-1], matching [IDSFO-15], updating [IDSFO-12], suggestion of vocabulary changes by users [IDSFO-13], version management [IDSFO-14], deletion [IDSFO-92], duplicate identification [IDSFO-91], and unused vocabularies [IDSFO-90]. Vocabulary hubs need to be managed [IDSFO-16].

3.2.4 App Ecosystem

The App Ecosystem describes the lifecycle of each Data App (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 [IDSFO-83].

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.) [IDSFO-22]. Data Apps must explicitly define their interfaces, dependencies, and access requirements [IDSFO-82].

Providing Data Apps

Any authorized Data App developer may initiate a software provision process (App Store publication) [IDSFO-23]. Prior to publication in the App Store, Data Apps must pass an optional evaluation and certification process controlled by the Certification Body [IDSFO-20]. The App Store should support authorized users in their search for a matching application in an adequate fashion [IDSFO-67]. Access of privileged users (e.g., administrators or operators) should require strong authentication (e.g., 2-factor authentication) [IDSFO-81].

Installing and Supporting Data Apps

A dedicated Connector service should support authorized users in (un-)installing Apps not originating from an official App Store [IDSFO-18]. 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 [IDSFO-80].

3.2.5 Identity Management

Every Connector participating in the Industrial Data Space must have a unique identifier [IDSFO-39]. Each Industrial Data Space Connector must have a valid certificate [IDSFO-61]. Each Connector must be able to verify the identity of other Connectors (with special conditions being applied here; e.g., security profiles) [IDSFO-75].

3.2.6 Clearing House

Any transaction of participants can be logged [IDSFO-55]. Privileged access to the Clearing House should require strong authentication (e.g., 2-factor authentication) [IDSFO-85].

3.3 PROCESS LAYER

The Process Layer describes the interactions between the different components of the Industrial Data Space. It therefore provides a dynamic view of the Reference Architecture Model. In the following, three processes are described that involve all of the roles introduced in the Business Layer section, and which cover the main activities of the Industrial Data Space:

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

To provide data in the Industrial Data Space, a Data Provider first must describe a data source (e.g., a backend system of the enterprise) in accordance with the Information Model (Section 3.4), using (generic and/or domain-specific) vocabularies offered by a Vocabulary Provider. The metadata includes a data usage policy that states the constraints for using the data. The set of metadata resulting from this is part of the configuration of the Connector, and a prerequisite for the Connector to be deployable. The Connector must be configured to provide data for the data source specified. This might include activities such as defining a connection to a data source, deploying a System Adapter in the Connector, or configuring and using data processing and transformation apps. The result of this process step is a configuration model, which constitutes the basis for the deployment of the Connector. After deployment, the Connector sends metadata regarding the Data Source to the Broker Service. The Broker Service indexes the metadata and returns an acknowledgement of receipt to the Connector. This acknowledgement may include an identifier generated by the Broker Service for identification of this particular data source or Connector.

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 BPMN diagram of this process is illustrated in Figure 3.3.

3.3.2 Exchanging Data

The process of exchanging data starts with a Data Consumer sending a metadata request to a Broker Service. The Broker Service then compiles a set of metadata describing a data source in the Industrial Data Space, and sends this information back to the Data Consumer. If the data source is already known to the Data Consumer, the request to the Broker Service can be omitted, and the Connector can be configured to directly connect to the corresponding Connector of the data source. The process of exchanging data may comprise complex sub-processes. These sub-processes are not displayed here in detail, mainly because of two reasons: first, the establishment

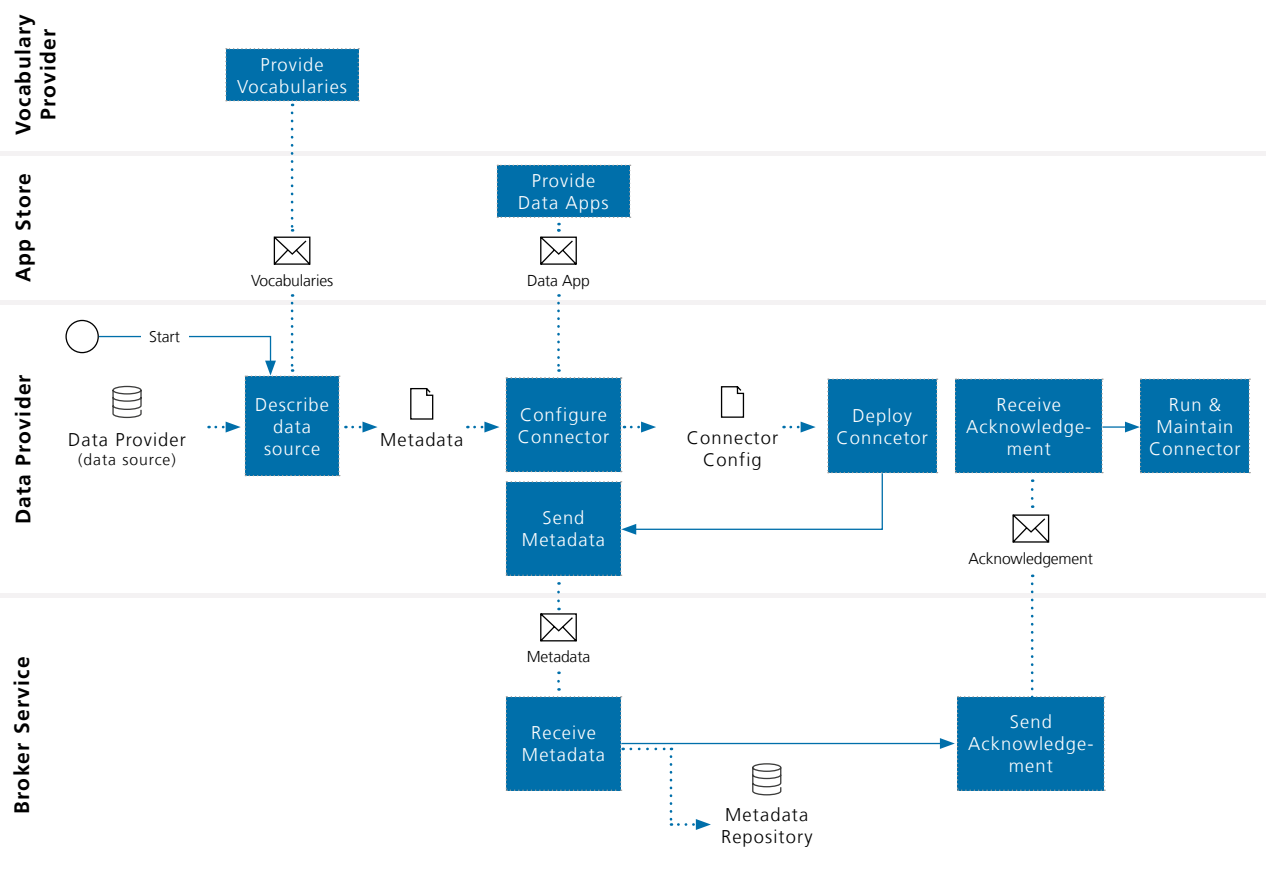


Figure 3.3: Process of providing data

of a legal agreement between a Data Provider and a Data Consumer is 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); second, the orchestration of the data flow inside the Connector can be very complex, as 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).

Data usage policies are an important element of legal agreements and are therefore modeled as first-class objects in the Information Layer (Section 3.4). In the process diagram, data usage policies are part of the metadata provided by the Broker Service.

After all prerequisites are fulfilled, the actual data exchange process can be initiated by the Data Consumer querying data from the remote Connector of the Data Provider. The query is then processed by the Connector of the Data Provider, and the result is sent back to the Data Consumer. Communication between the Connectors can be asynchronous; i.e., the Data Consumer does not need to wait in idle mode for the result to arrive, but will be notified by the Data Provider as soon as the

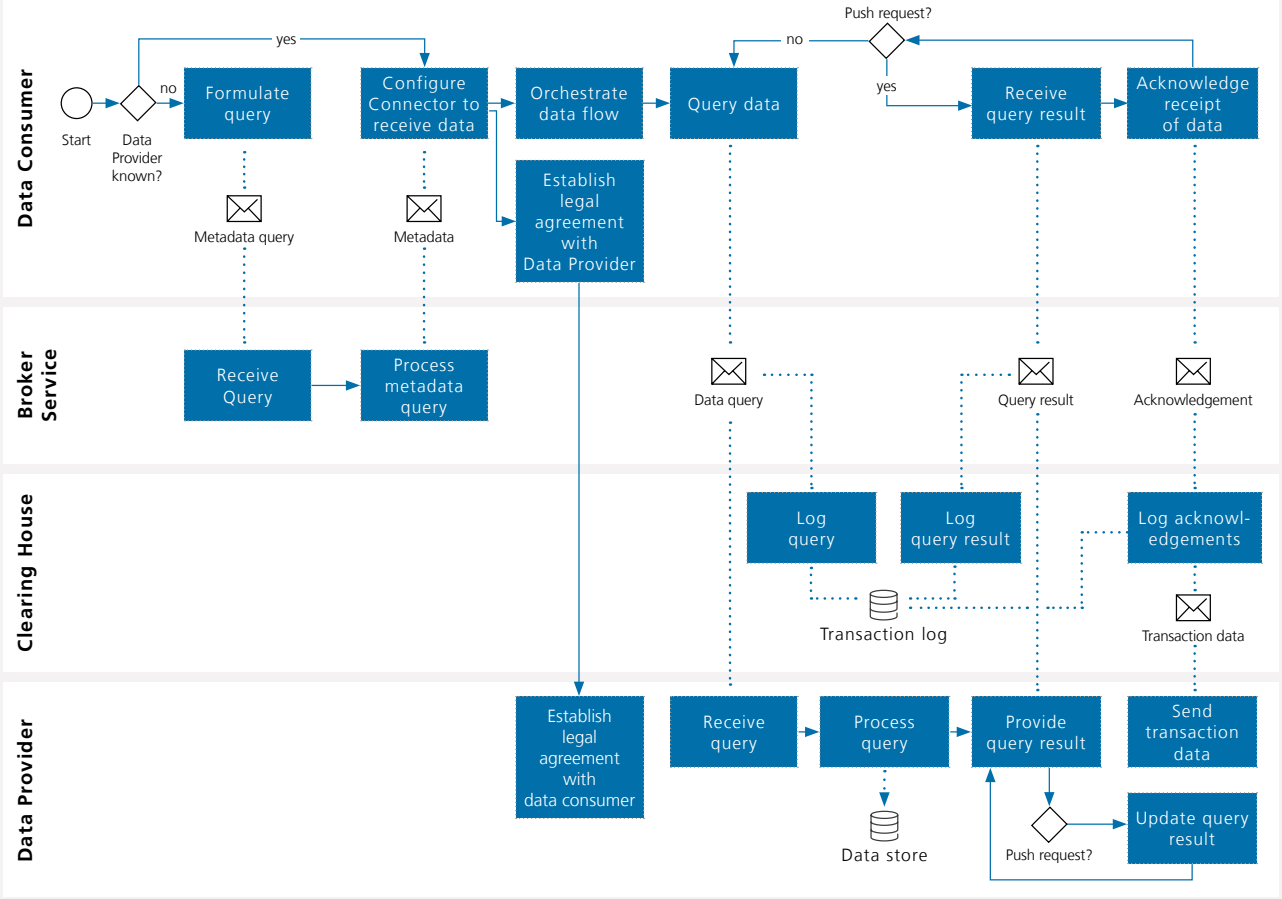


Figure 3.4: Process of exchanging data

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 has 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 step of the process is the Clearing House logging the successful completion of the transaction. For that, 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 information has been requested and which result has been sent, the query and the result (or its metadata) are also logged by the Clearing House. Figure 3.4 illustrates the BPMN model of this 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 a data transformation may be a Data App parsing a single string field with address information and producing a data structure consisting of street

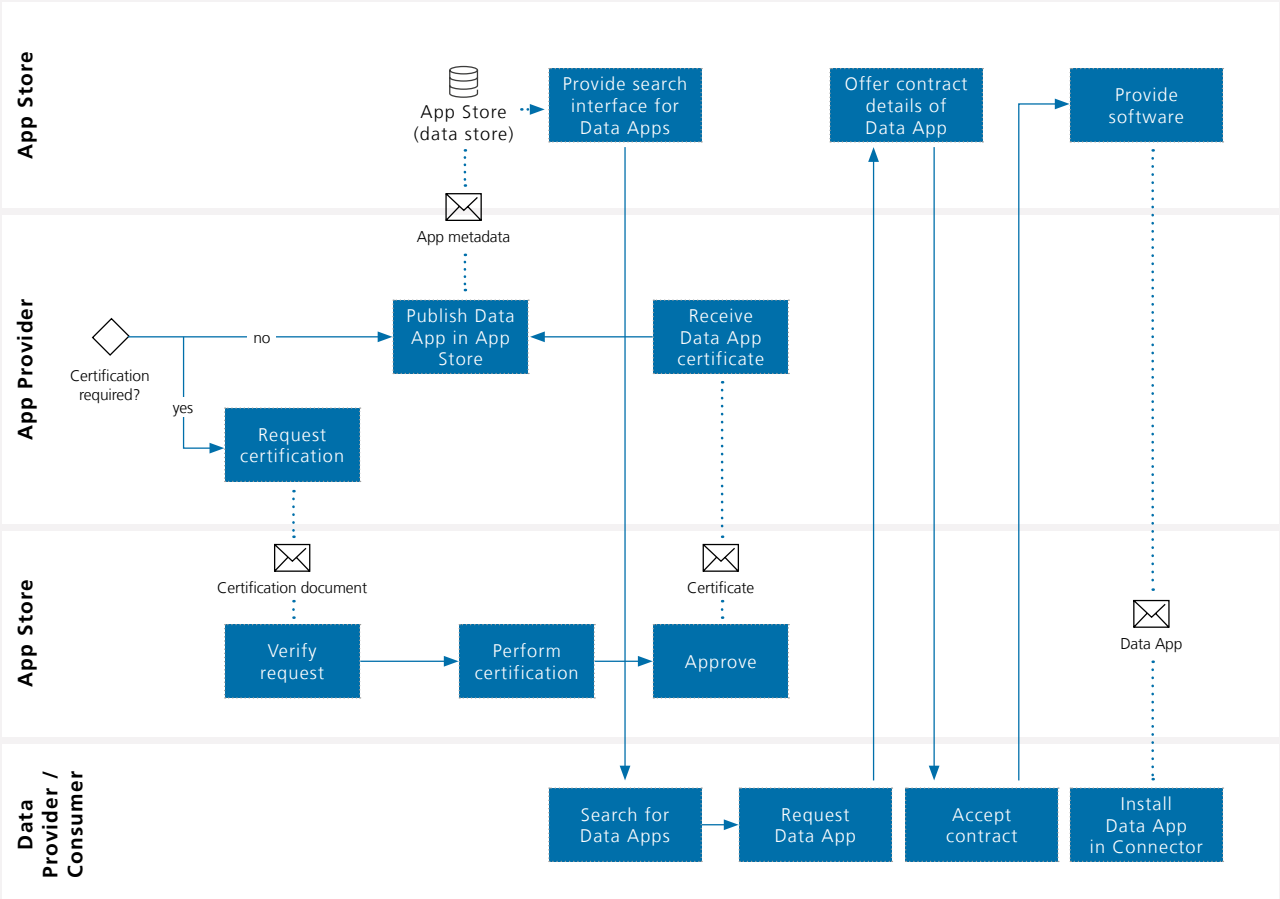


Figure 3.5: Process of publishing and using Data Apps

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 to an address or a street section).

Data Apps may require certification from a Certification Body (see first step of the process shown in Figure 3.5). Upon successful certification, the Certification Body sends the App Provider a certificate, which is required for the Data App to be published in the App Store.

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. If a user finds a suitable Data App, the App can be requested from the App Store. 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, usage restrictions, 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).

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 defines a domain-agnostic information model of the Industrial Data Space. It constitutes a central agreement shared by both participants and components (regardless of how they are implemented), facilitating compatibility and interoperability.
The Entity Relationship (ER) modeling paradigm applied for the design of the Information Layer is a standard method for business domain analysis and documentation. It distinguishes three levels of abstraction:

1. conceptual model: high-level overview capturing the main domain entities, their attributes, and their relationships, without committing to a particular technology;
2. logical model: specification of the conceptual model with regard to concrete data structures and their respective constraints;
3. physical model: specification of the logical model in terms of a technology-dependent solution.

ER diagrams are used throughout this section to capture the conceptual Information Model by a technology-independent notation. Concepts (structures) or terminals of multivalued relations are modeled via ER entities, whereas plain, single-valued nodes are modeled via ER attributes, as depicted in Figure 3.6.
The Reference Architecture Model aims at identifying a generalized, invariant conceptual model of the entities of the Industrial Data Space, while its detailed definition is delegated to the specification of the Industrial Data Space Vocabulary. The Industrial Data Space Vocabulary encodes the structures and constraints of the logical data model on top of linked-data principles. It is subject to perpetual updates and will be disclosed in a separate document. The Industrial Data Space Vocabulary is not in the focus of the document at hand, which is why only a brief overview is given in the following section. Some vocabulary terms are used as a reference to concepts of the Information Model.

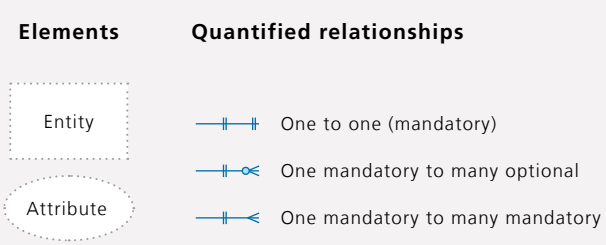


Figure 3.6.: Legend of the used ER notation

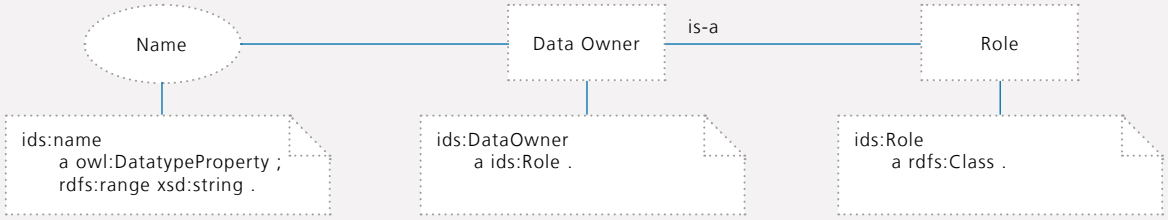


Figure 3.7: Sample mapping of Information Model onto Industrial Data Space Vocabulary

3.4.1 Industrial Data Space Vocabulary

The Industrial Data Space Vocabulary provides a formal, machine-readable representation and specification of concepts envisaged by the Information Model. It leverages a stack of W3C standards based on the Resource Description Framework (RDF). This simple, yet powerful knowledge representation formalism expresses information in a generic way as a set of triples. Triples are ternary statements consisting of a subject, a predicate, and an object resource. Each one of them can be identified by a URI (Universal Resource Identifier), while objects may hold typed literal values as well. The adjacent standards – RDF Schema, OWL, and XML Schema – allow the definition of concepts (classes), predicates (relationships and attributes), and data types.
Because of their different purpose and granularity, a one-to-one mapping between the concepts of the Information Model and the concepts of the Industrial Data Space Vocabulary is not possible. Where applicable, conceptual entity types of the Information Model are mapped to classes, attributes to datatype properties, and entity relations to object properties of the Industrial Data Space Vocabulary (Figure 3.7).
The design of the Industrial Data Space Vocabulary is based on the following principles (which are adapted from Linked Data principles):

1. **Universal identification using resolvable URIs:**
   A resource can be identified by a unique (resolvable) Universal Resource Identifier (URI). In the context of the Industrial Data Space, resources are instances of Industrial Data Space Vocabulary classes representing concepts of the Information Model. The resolution of the URI should point to a description of the resource identified.
2. **Resource description using RDF:**
   A resource can be described using the Resource Description Framework (RDF). The semantics of resource descriptions are captured via RDF Schema and OWL vocabularies. Existing standard vocabularies can be reused.
3. **Support for domain-specific vocabularies:**
   Apart from vocabularies modeling the Industrial Data Space, domain-specific third-party vocabularies can be shared among participants through publication on vocabulary hubs. Reuse of these vocabularies promotes a common understanding as well as interoperability and data processing across company and industry boundaries.
4. **Integration and lifting of legacy data models:**
   It may be necessary to convert (lift) non-RDF legacy data models (RDBMS, XML, or JSON Schema) into an RDF format (e.g., by means of model conversion implemented via Data Apps). If that is the case, the data itself would naturally integrate with the metadata and become accessible for advanced querying and processing.

Following these principles, participants in the Industrial Data Space can develop a common understanding and seamlessly share and integrate data along their value chains.

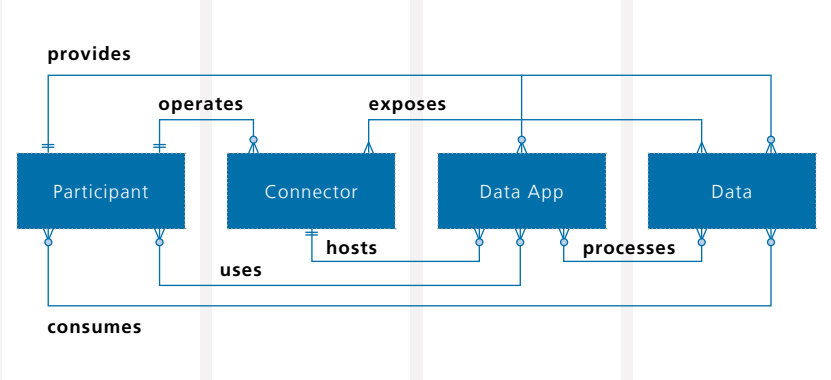


Figure 3.8: Main entity types of Information Model

3.4.2 Information Model

The Information Model can be divided into four sub-models (according to the main entity types of the Information Model as illustrated in Figure 3.8):

1. Participant Model, focusing on the participants (community of stakeholders) in the Industrial Data Space;
2. Connector Model, focusing on the infrastructure components of the Industrial Data Space;
3. Data App Model, focusing on the Data Apps providing reusable software packages for data integration and processing (extension mechanism); and
4. Data Asset Model, focusing on the central commodity of the Industrial Data Space.

1) Participant Model

Section 3.1.1 introduced the roles of the Industrial Data Space from a business perspective. Formal modeling of participants is crucial for runtime operation of all services (e.g., to specify details of the membership account, disclose service operators, indicate the provenance of data, or sign agreements). Figure 3.9 outlines the logical ER model of an individual Industrial Data Space Participant. The physical model is implemented by the Industrial Data Space Vocabulary class ids:Participant. Its definition was influenced by the UDDI Business Entity model[6].

6 http://www.uddi.org/pubs/uddi_v3.htm

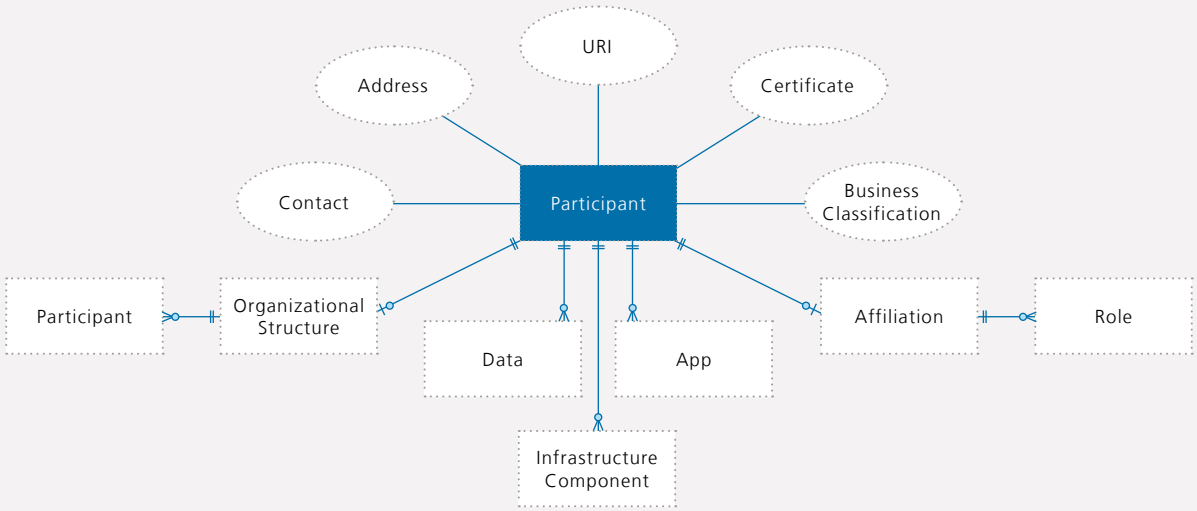


Figure 3.9: Participant model

Identification and Naming

An Industrial Data Space Participant is identified by a unique ID (URI), an address, and, for organizations, a contact. Resources provided by an Industrial Data Space Participant are expected to share the participant's unique address space (internet domains) and identification scheme (XML namespaces) in order to prevent naming conflicts across the Industrial Data Space.

Organizational Structure

Corporations may indicate an organizational structure and a link to subsidiary companies or units acting as related, but more or less independent Industrial Data Space Participants. This approach allows e.g. sharing authorization certificates along a trust chain and enforcing company-wide policies. The Organization Ontology[7] provides an appropriate vocabulary for modeling organizational structures.

Business Classification

Industrial Data Space Participants may indicate the type of business and the domain in which they operate by making references to established business catalogs and registries. Prospective customers can search for products of participants (i.e., data and apps) based on the business classification. Schemes for classification of services and businesses to use are D&B D-U-N-S® Number[8], SIC[9], NAICS[10], or UNSPSC[11], among others.

7 https://www.w3.org/TR/vocab-org
8 http://www.dnb.com/duns-number/what-is-duns.html
9 https://www.osha.gov/pls/imis/sic_manual.html
10 http://www.census.gov/eos/www/naics/
11 https://www.unspsc.org

Affiliation and Involvement

The type of involvement in the Industrial Data Space is described by the participant's affiliation with (i.e., membership in) the Industrial Data Space and by the role(s) the participant may assume. A taxonomy of roles is depicted in Figure 3.10.

While some roles in the Industrial Data Space are permanent (e.g., Data Provider), others are volatile and may change from case to case (e.g., Data Consumer). Role associations therefore are implicit (e.g., via a statement of ownership given in a provenance record of a Data Asset) and serve merely conceptual modeling purposes.

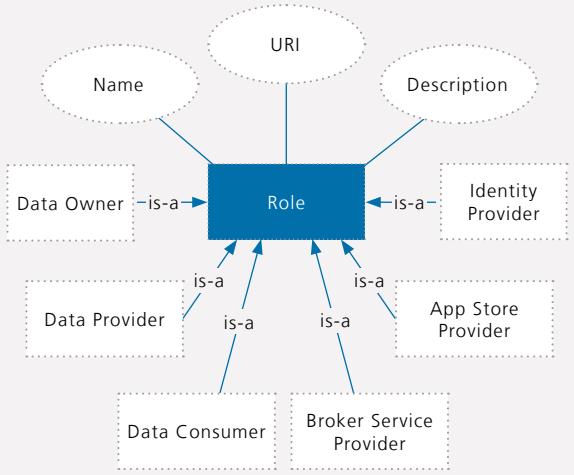


Figure 3.10: Taxonomy of roles

2) Connector Model

Figure 3.11 shows a taxonomy of the main infrastructure components of the Industrial Data Space.

The Connector is the central component of the Industrial Data Space infrastructure, as it is the defined point of data exchange and policy enforcement. It is the basis for the implementation of more specialized components, such as the Broker or the App Store. The conceptual information model of the Connector is depicted in Figure 3.12.

Identification

By default, and in accordance with linked-data principles, a Connector is uniquely identified by a dereferencable HTTPS URL, which resolves to a live metadata document describing the Connector's setup and capabilities. This identity is confirmed by the (X509) certificate attached.

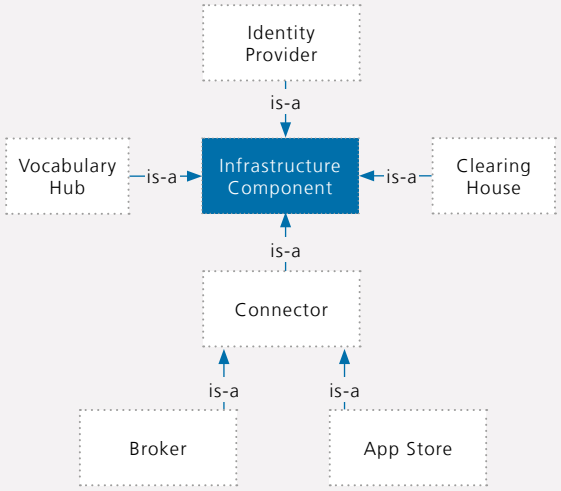


Figure 3.11: Taxonomy of main infrastructure components

Deployment Context

Apart from the description of the Connector itself, the deployment context of the Connector is specified. This comprises e.g. geo-location information (e.g., country of deployment or applicability of national law) and deployment information (on-premise vs. cloud). Furthermore, the responsible Participant operating the Connector (Service Provider) is referenced.

Data Endpoint

The Connector exposes Data Assets via a number of Data Endpoints, which are modeled and referenced accordingly (see below for details).

Security Profile

The Security Profile is a mandatory entity that models the security-relevant characteristics of a Connector. It supports, among other things, attribute-based access control and verifies Connectors in terms of being eligible to take part in data transaction processes. The security profile (Section 4.1.8) manifests some high-level attributes, as depicted in Figure 3.13.
The Authentication Level indicates the level of trust of a component's certificate (ranging from self-signed certificates to certificates issued by the Certification Body). The App Isolation Level refers to the isolation capabilities of the container technology deployed. The Execution Control Level states the level of control over deployed resources when executing containerized resources. Hardware Security references the hardware security technology in use, if any (TPM, HSM etc.). The Remote Attestation Level refers to the set of components that are subject to remote attestation (only system components or Data Apps as well). With the Software Assurance Certificate, the Certification Body certifies the approval of a Connector's software stack.

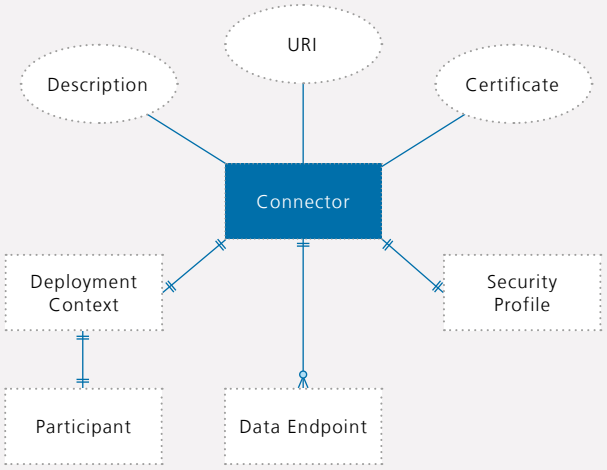


Figure 3.12: Information Model of Connector

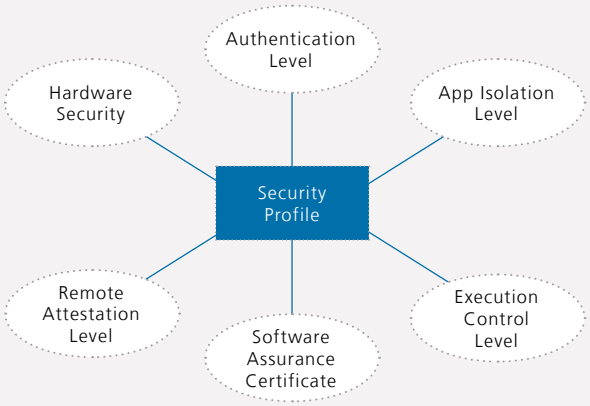


Figure 3.13: Security profile of Connector

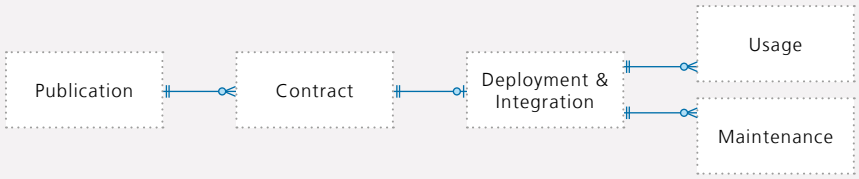


Figure 3.14: Model layers of Data App supply process

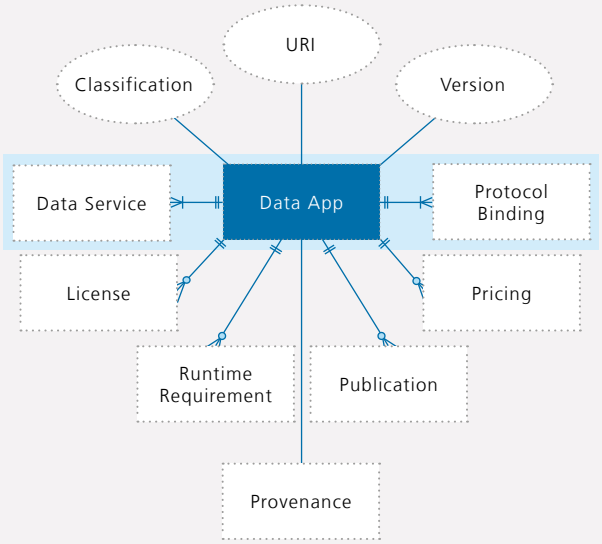


Figure 3.15: Data App model

3) Data App Model

Data Apps are self-contained and self-descriptive software packages extending the functionality of the standard Connector with customized data integration, processing, and publishing capabilities. Following a successful certification process, Data Apps are distributed via the App Store for being deployed inside Connectors. At each stage of the Data App supply process, different types of information models are required, as shown in Figure 3.14 and elaborated in the following paragraphs.

Publication

Data Apps are published in and distributed via the App Store. The Publication entity represents a versioned snapshot of the entire Data App's metadata description. Figure 3.15 shows the model of the Data App linked to a multivalued Publication entity.

Each Data App available for being deployed can uniquely be identified by a URI. Data App revisions are distinguished by Version numbers. The Provenance entity maintains the history of the respective Data App, including the name of its developer (of the Software Provider, respectively), the results of its certification, and information on modifications. Each Data App can be associated with multiple License models and corresponding Pricing models. Requirements on the target runtime (Connector features, computational resources, etc.) are described by the Runtime Requirement entity.

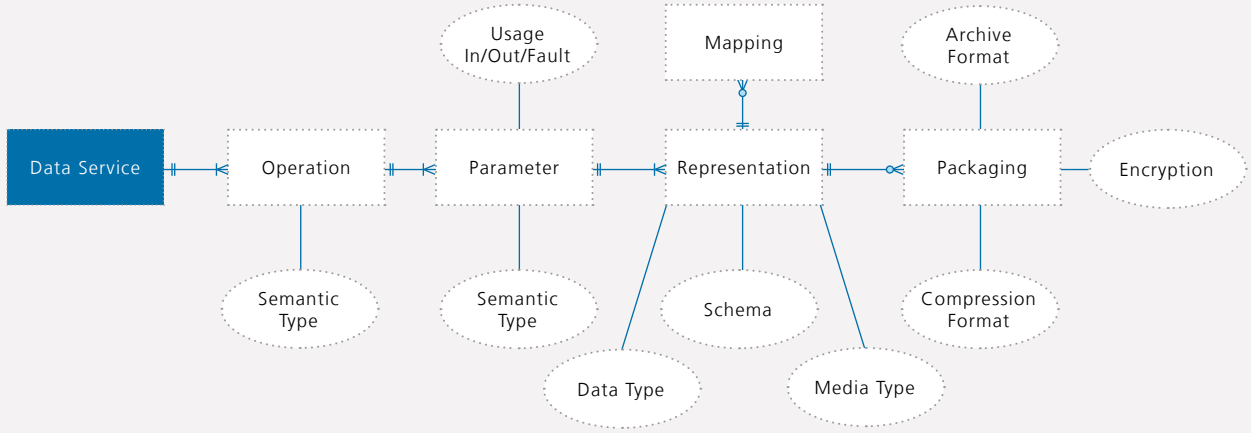


Figure 3.16: Data Service model

Data Service

The Data Service entity models the effective functionality of a Data App. It encapsulates a range of operations (ids:Operation) upon a homogeneous Data Asset exchanged via the operation's parameters (ids:Parameter). As outlined in Figure 3.16, the semantic type of an operation indicates the processing of and effect on input data in an interoperable way ("transform", "anonymize", "read", etc.). Likewise, the semantic type indicates the intention and interpretation of a parameter (geo:CityName) in terms of abstract operations (ids:Operation)as well as their input, regular output, and fault parameters (ids:Parameter), as outlined in Figure 3.16.

The annotation of a semantic type supports the retrieval of relevant Data Services at an abstract, functional level. The Representation entities express the concrete serialization of parameter data in terms of a syntactic data type (xsd:String), media type (text/plain) and schema reference (XML- and JSON schema). The Mapping entity optionally expresses possible mappings of the underlying representation format to RDF (RML[12] statements). It allows the Data Consumer to perform a "lifting mapping" to generate an RDF representation of the data. The parameter data may further be archived, compressed, and encoded for optimized transmission, as indicated by the Packaging entity

12 http://rml.io

Deployment option	Description
On-premise deployment	The Service Provider deploys the Data App inside an on-premise Connector on behalf of the Data Provider. This is assumed to be the default case.
On-premise injection	The Service Provider deploys the Data App inside an on-premise Connector on behalf of the Data Consumer (asking for customized data preprocessing, according to contract specifications; e.g., edge computing).
Remote integration	The 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 3.1: Main deployment options

Data App API
The abstract, technology-independent service representation is suitable for a fine-grained semantic search, e.g., according to the classified operation semantics (e.g. "geo name resolution"), the syntactic or semantic type of parameters. Concrete programming interfaces of the Data Apps are defined via service bindings to communication protocols (ids:ProtocolBinding) like HTTP, CoAP, MQTT, etc.

Deployment and Integration
A Data App that was downloaded is usually deployed inside a Connector. Depending on the deployment option, different configuration, monitoring, and runtime information models might be required. Table 3.1 lists the main deployment options possible.

Usage and Maintenance
The usage model comprises Data App runtime aspects (e.g., resource consumption, quality feedback, anonymous usage, and error statistics), while the maintenance model supports and tracks update events and troubleshooting. It should be noticed that both models are beyond the scope of the current Reference Architecture Model.

4) Data Asset Model

Data is the central asset of the Industrial Data Space. The Data Asset (content) as defined here represents an intentional, selective view upon arbitrary data that is "focused" and consistent in terms of several dimensions listed in the following. As explained later, the particular exposure (access and processing) of the abstract Data Asset is represented by a Data Service in terms of operations and IO parameters. This abstract interface is turned into an addressable, concrete Data Endpoint serving the data interchange as depicted in Figure 3.17.

Following a resource-oriented paradigm inspired by the REST architecture style, a Data Endpoint delivers representations of the underlying Data Asset resource as defined by the parametrized Data Service interface. A Data Provider has the option to advertise own Data Endpoints by registering them with the Broker, thereby increasing market visibility and business potentials.

This section first looks at the generic, domain-agnostic aspects of data, which need to be taken into account when identifying Data Assets, creating metadata descriptions, and architecting interfaces of Data Endpoints as depicted in Figure 3.18. It then presents the Information Model accompanying the Data Assets at all stages of their lifecycle.

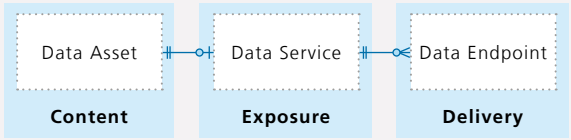


Figure 3.17. Data Asset provisioning layers

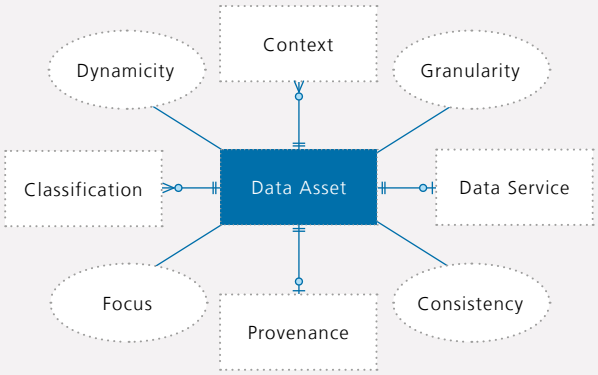


Figure 3.18. Conceptual Data Asset model

Dynamicity

Data can differ significantly in terms of dynamicity. Finite datasets (modeled via the ids:DataSet class) can be contrasted with continuously growing, infinite sources of dynamic data (sensor measurements, log entries, etc.). The time variance of data needs to be explicitly modeled (sample rate, collection updates, etc.) and considered when selecting an appropriate delivery pattern and communication protocol (PULL vs. PUB-SUB).

Level of granularity	Description and properties
Data stream	Continuous, opaque byte stream - Source: webcam, media stream - Infinite, continuously growing - Access by time (range, instant) or volume - Filtering, no grouping, no sorting
Value set	Set of discrete values - Source: sensor readings - Items (values) are anonymous (no identity) - Finite (historical value sets) or infinite (live values) - Access by volume and count - Access by index, if ordered - Access by time, if time-ordered (time series) - Filtering, grouping and sorting, if structured
Resource(s)	Single resource, collection of resources - Source: document archives, collections of resources - Items (resources) have an identity (name) - Identity uniqueness level: global or local (per collection) - Finite (closed collections) or infinite (live collections) - Access by ID - Access by volume and count - Access by index, if ordered - Access by time, if time-ordered - Filtering, grouping, sorting, if structured or on file level

Table 3.2: Typical levels of granularity of data and its properties

Granularity

The type of data exchanged via a Data Endpoint may vary with regard to granularity, as depicted in Table 3.2.

The respective "state of aggregation" of data, as depicted above, has an impact on the mode of accessing and selecting a particular item or range of items, in particular:

– **Access by ID:** a continuum of named data items is segmented by the item's names
– **Access by volume:** a continuum of data is segmented by volume (e.g., every 5 MB)
– **Access by time:** a continuum of time-ordered data is segmented by a time instant (e.g., at 1 h, 30 s, 215 ms) or range
– **Access by count:** a continuum of ordered data items is segmented by counting (e.g., every 10,000 items), provided there is an initial item (identified by index or time)
– **Access by index:** a continuum of ordered data items is segmented by position (e.g., every 10th item), provided there is an initial item (identified by index or time)

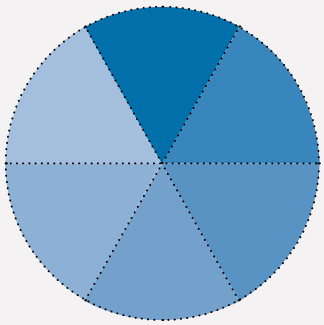
Context

The above analysis does not consider the context of data so far. One of the modeling aspects remaining after a cross-sectional generalization of data is its context (i.e., the reference to spatial, temporal, and socio-economical coordinates of the data's origin). Accurate and expressive context modeling gives answers to questions like "where", "when" and "what" regarding a specific Data Asset, and is seen as a prerequisite for the assessment of its relevance and business value with respect to the needs of Data Consumers.

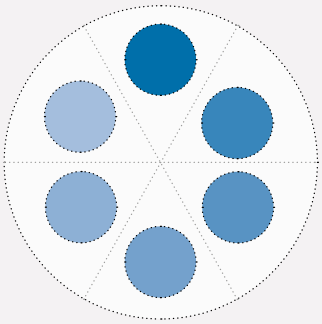
Among the standards available, the Industrial Data Space Vocabulary may leverage the W3C Time Ontology[13] for temporal context modeling. Spatial context descriptions may reference a named location (symbolic coordinates) as well as a geographic area or arbitrary shape using the NeoGeo vocabulary for RDF representation for GeoData[14]. As the original context is part of the data's provenance, the Provenance Ontology[15] may apply here as well.

Focus

Each Data Asset (and thus its Data Endpoints) should have a clear focus. The Data Provider should avoid overstretching the Data Endpoint interface by adding a multitude of query parameters for exploiting several data dimensions via a single access point, as this would obscure the clarity of a Data Endpoint in terms of its semantics. It should instead represent a data exchange point with unambiguous, predictable and focused semantics, as contrasted in Figure 3.19.



Unfocused Data Endpoint



Separate focused Data Endpoints

Figure 3.19: Data Endpoint focus

13 https://www.w3.org/TR/owl-time/
14 http://geovocab.org/doc/neogeo.html
15 https://www.w3.org/TR/prov-dm/

Consistency
Data Assets published via an individual Data Endpoint must be homogeneous and must remain consistent over time with regard to granularity, coverage, context, data structure (scheme compliance), and conceptual classification, in order to allow constant processing and interpretation. Any changes made to one of these dimensions (if backward compatible) should lead to the creation of a new revision of the Data Endpoint or a new, independent Data Endpoint.

Provenance
The Provenance entity keeps track of modifications applied to the Data Asset's state, i.e. its initial creation, filtering, aggregation etc.

Data Supply Stages
Similarly to Data Apps, the information model of Data Assets can be decomposed into several parts related to provisioning phases, as depicted in Figure 3.20.

Publication
The Publication entity represents the publication of a Data Endpoint by a Connector, which is a versioned snapshot of the Data Endpoint description at a given point in time, as depicted in Figure 3.21. The purpose of the attributes URI, Version, and Classification is in line with the Data App model.

Figure 3.20: Data supply stages

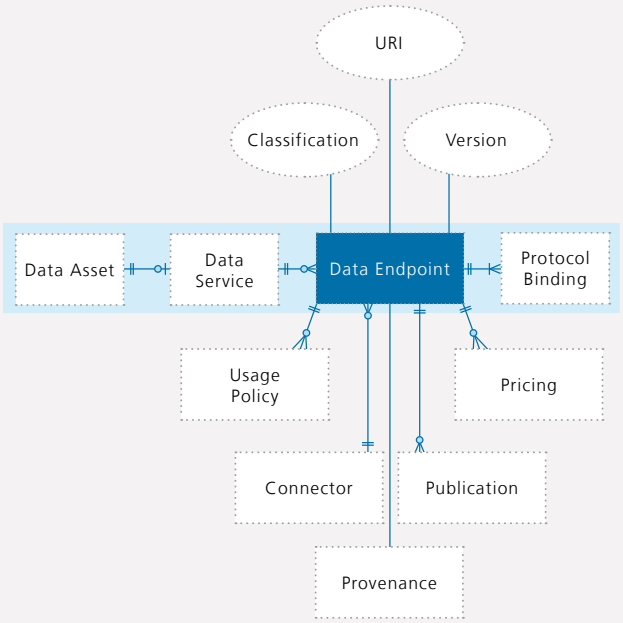


Figure 3.21: Conceptual model of Data Endpoint

Data Service

While any meaningful Data Service can be published as a Data Endpoint, the Industrial Data Space Vocabulary identifies classes of data publishing apps and operations, as depicted in Figure 3.22. Data Endpoints either expose (ids:DataSource) or consume (ids:DataSink) data.

A Data Source may support different interaction patterns in parallel (e.g., passively expose data for being retrieved by a Data Consumer (ids:PassiveDataSource) and actively deliver the data based on a prior subscription (idsv:ActiveDataSource). The latter method is particularly suited for random data events.

Figure 3.23 outlines a simple subscription model. Based on a contract, the subscription expresses an obligation to deliver data at a particular rate from an active Data Source to a number of subscribed Data Sink targets.

Inspired by the REST architecture style, the operations of a passive Data Source cover the functional range of HTTP method equivalents (e.g., ids:RetrieveDataOperation and ids:ListDataOperations vs. HTTP GET) in a protocol-agnostic fashion. The Data Endpoint may support the following advanced data access methods:

- **filtering**, i.e., extracting a subset of the data that matches a filter expression (the filter is applied to the structured content of an item or, if deemed appropriate, to file properties of the materialized item, such as file extension, full name, file type, etc.);
- **grouping** structured items by a common property value;
- **sorting** structured items by a property value;
- **pagination**, i.e., the dataset is split into segments (pages) to be sequentially navigated by the client.

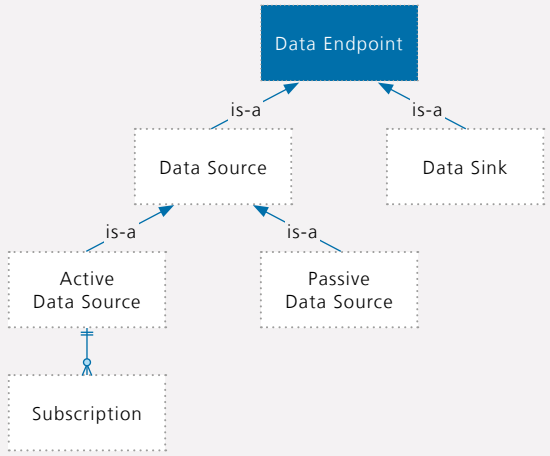


Figure 3.22: Data Endpoint taxonomy

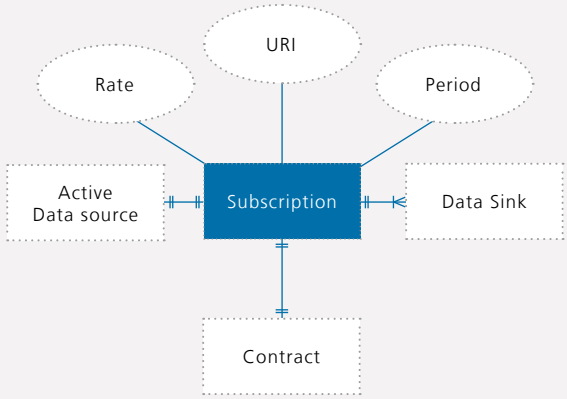


Figure 3.23: Subscription entity

Provenance

The provenance model records a Data Asset's history and context of creation and the processing it undergoes, allowing an assessment of its quality, reliability, and trustworthiness. The Industrial Data Space Vocabulary delegates to the W3C Provenance Data Model[16] for purposes of provenance encoding.

Usage Control

The usage control model declaratively states the restrictions on processing and exploitation of transferred data enforced on the side of the Data Consumer. The conceptual usage control model envisaged in the Industrial Data Space is outlined in Figure 3.24.

So far, no decision has been made on a particular usage policy language and control framework. Adopting Open Digital Rights Language (ODRL)[17], which presumably will be standardized by the W3C Permissions & Obligations working group[18], is considered a promising option, among others. The ODRL vocabulary of actions (subject to permission and prohibition rules), constraints, and duties could be augmented by an extension profile tailored to the purposes of the Industrial Data Space.

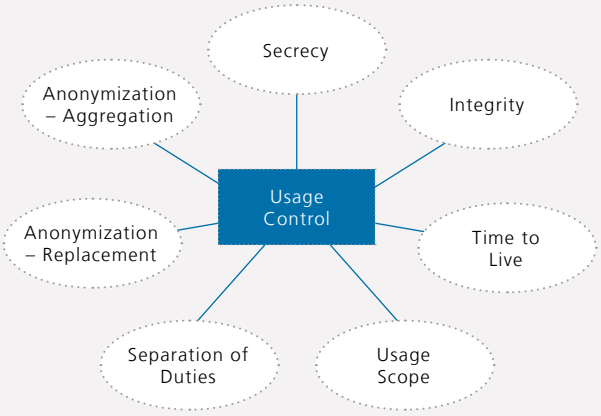


Figure 3.24: Conceptual usage control model

Pricing

Pricing models may comprise both quantitative (by data volume or number of API calls) and qualitative (by content layering) billing options. The results of the ongoing, evidence based research, analysis and modeling of pricing models will be documented in the next iteration of this document. It will take into account partial models, as provided e.g. by the eClassOWL ontology[19] or Open Digital Rights Language (ODRL)[20].

16 https://www.w3.org/TR/prov-overview/
17 https://www.w3.org/ns/odrl/2/
18 https://www.w3.org/2016/poe/wiki/Main_Page
19 http://www.heppnetz.de/projects/eclassowl/
20 https://www.w3.org/ns/odrl/2/

Contract

The information model of a data exchange contract (Figure 3.25) establishes a temporary link between the Data Provider and the Data Consumer, with a reference to a specific data offer (Publication). It requires from the Data Provider to maintain the purchased Data Endpoint version and conditions of service during the period of the contract's validity, and to optionally install a preprocessing Data App according to the customer's specifications (preprocessing entity).

Transfer

The transfer stage of the data provisioning chain involves the actual exchange of data. It is based on a previously (ad-hoc) signed data contract. The information model of metadata being transferred along with the data (ids:TransferredDataset) comprises the timestamp, media type, and size attributes (Figure 3.26). It further references the sender (Data Provider) and the receiver (Data Consumer). Depending on the contract's pricing model (pay per use), a data transfer may entail billing processes.

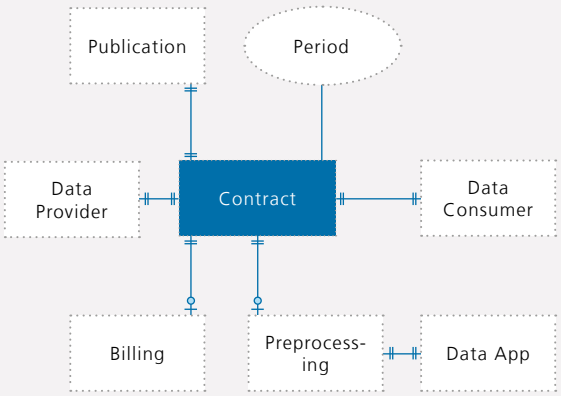


Figure 3.25: Conceptual data contract model

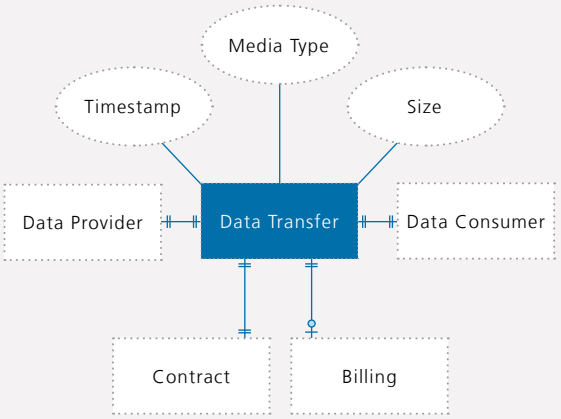


Figure 3.26: Conceptual data transfer model

3.5 SYSTEM LAYER

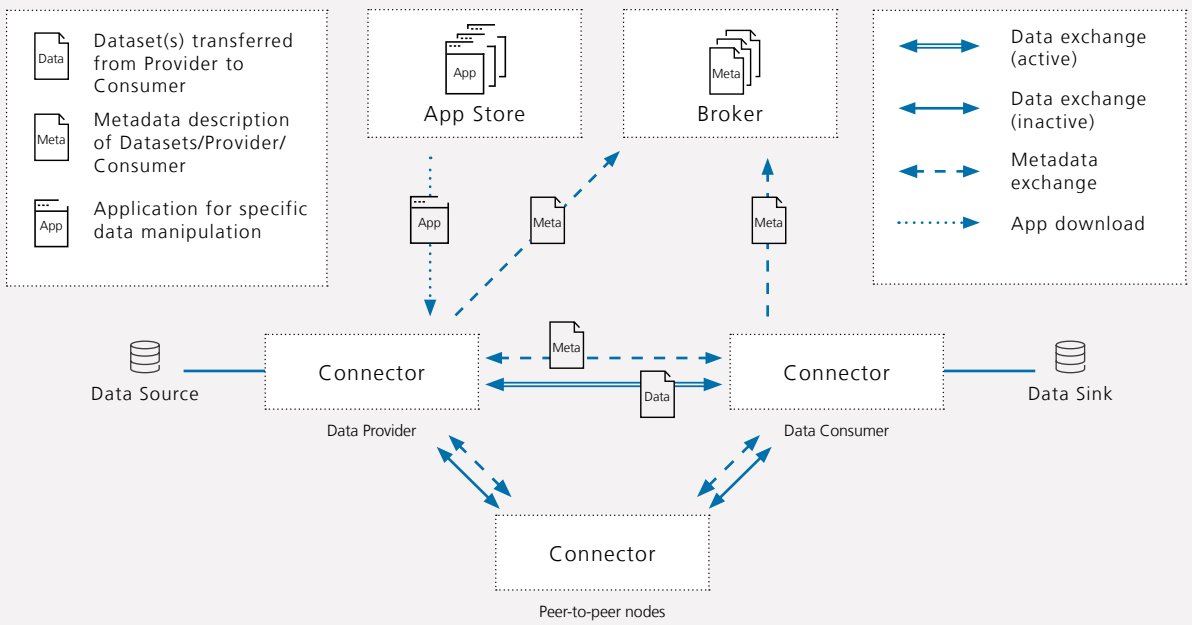


Figure 3.27: Interactions of components on System Layer

With regard to the Business Layer and the Functional Layer of the Reference Architecture Model, a number of roles and functional requirements have been introduced and defined. The roles are now mapped onto a concrete data and service architecture in order to meet the requirements, resulting in what is the technical core of the Industrial Data Space. From the requirements identified, three major technical components can be derived:

- Connector,
- Broker, and
- App Store.

The interaction of these components on the System Layer is depicted in Figure 3.27. A technical system to support the Certification Body has not been defined yet. Connector, Broker, and 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: the Data Endpoints). The Connector is responsible for the exchange of data, as it executes the complete data exchange process (Section 3.3.2). The Connector thus works as an 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 only metadata is submitted to the Broker, whereas the actual data is transferred between the Connectors of the Data Provider and the Data Consumer (peer-to-peer network concept).There may be different types of implementations of the Connector, based on different technologies and featuring different functions. Two basic examples are the Base Connector and the Trusted Connector (Section 4.1).
A Connector can be classified as external or internal. 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 (the so-called "Demilitarized Zone" (DMZ). From a DMZ, direct access to internal systems is not possible. An External Connector should be reachable using the standard Internet Protocol (IP) and operated 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-premise 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 secured environment for individual Data services. Data Services are Data Apps that are deployed inside Connectors. To ensure privacy of sensitive data, data processing should be done 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 transmitted to other participants should be transferred to External Connectors, where it is available for authorized recipients.
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 (to be certified if required), 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) (to be certified if required).

In addition, Data Apps can be divided into two further categories:

- System Adapters establish 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) is to add metadata to data.
- Smart Data Apps execute any kind of data processing. Normally, the data provided from or to a Smart Data App is already annotated with metadata (as described in the Information Layer section).

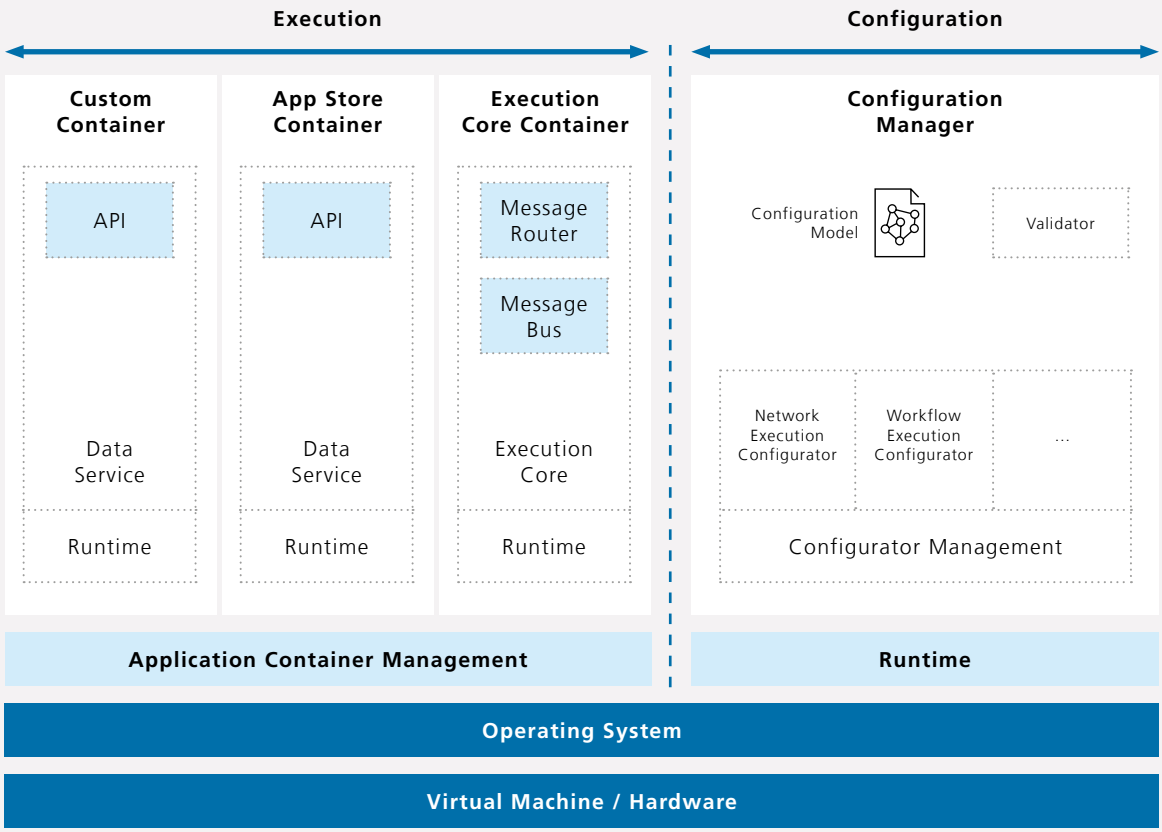


Figure 3.28: Reference Architecture of Connector

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 3.28 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 3.28 can be grouped into two phases: Execution and Configuration.

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 dependencies between them. Using alternative Application Container Management, extended

control of Data Services and containers can be enforced. During development and in the case of systems with limited resources, Application Container Management can be omitted. Differences in deployment can be handled by specialized Execution Configurators (see below).

- An Execution Core Container provides components for app orchestration and communication (e.g., Message Router or Message Bus to a Connector).
- A Message Router executes multiple workflows and invokes Data Services according to defined workflow steps. Additionally, it is responsible for sending data to and receiving data from the Message Bus. Participants have the option to replace the Message Router component by alternative implementations of various vendors. Differences in configuration can be handled by specialized execution configurator plugins. If a Connector in a limited or embedded platform consists of a single Data Service or a fixed workflow only (e.g., on a sensor device), the Message Router can be replaced by a hard-coded workflow, or the Data Service is exposed directly.
- The Message Bus stores data between services or Connectors. Usually, the Message Bus provides the simplest method to exchange data between Connectors. Like the Message Router, the Message Bus can be replaced by alternative implementations in order to meet the requirements of the operator. The selection of an appropriate Message 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 provides the executable activities for a workflow to be executed by a Message Router. A Data Service defines a public API which is invoked from a Message Router. This API is formally specified in a meta-description that is imported to 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 the 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 the execution of a deployment. The deployment task 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 deployment is executed, 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 deployed. 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 Workflow defines the control and data flow between the Message Router, the Data Services, and the Message Bus (for multiple data pipelines).
– Metadata describes the data types for input and output used by different Connector components. This may include Data Services, Workflows, and different types of Message Queues or topics of a Message Bus. Data Services can provide metadata descriptions, which can be imported into the Configuration Model.
– Networking means the definition of 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 Workflows or Data Services are provided to external participants. This information is submitted to Brokers.
– The Lifecycle summarizes information on single Workflows 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 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 kind/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 Workflows 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 workflows 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 container or workflow 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 step of Connector miniaturization is 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 development 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 miniaturization may include the use of a common runtime for all components or simplified versions of the Message Router and the Message Bus. If messages are to be sent to a fixed recipient only, a simple Message Bus client library may be sufficient. Similarly, it may be sufficient to hard-code a single fixed workflow instead of using a configurable component. To save communication overhead, remote-procedure calls might be replaced by simple API calls inside the common runtime.

4

PERSPECTIVES OF THE REFERENCE ARCHITECTURE MODEL

In addition to the five layers, the Reference Architecture Model consists of three cross-sectional perspectives (Security, Certification, and Governance), which are described in detail in the following sub-sections.

4.1 SECURITY PERSPECTIVE

As stated in Section 1.1, a strategic requirement of the Industrial Data Space is to provide secure data supply chains (i.e., to ensure a high level of protection and confidence when exchanging data between participants). The Security Architecture provides means to identify participants, protect data communication, and control the usage of data, even after the data has been transmitted to a Data Consumer.
For these purposes, the Industrial Data Space offers a Trusted Connector on top of the Base Connector (Section 3.5). The Trusted Connector ensures the validity of the Security Architecture and its related concepts. The security features described in the following provide the basis of the Trusted Connector.

4.1.1 Security Aspects on the Different Architectural Layers

Business Layer
Security has an impact on the definition of roles and on potential business processes. To enforce individual business models in the Industrial Data Space, the Business Layer relies on the System Layer to enable secure business transactions.

Functional Layer
In some cases, security requirements may have an impact on certain functionality, or even prevent it from being used. However, security is also an enabling factor. Without security, many use cases would not be possible (e.g., offering sensitive data for trusted business partners). Usage control enables Data Providers to attach usage policies to data sources or items in order to define how a Data Consumer may use the data.

Process Layer
To enable security features, new processes need to be defined and processes in place need to be adjusted, respectively. For example, to enable trustworthy identification and authentication of participants using a central Public Key Infrastructure (PKI), a participant must apply for a public key certificate that is being registered in a central PKI and deployed on 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 enables participants to use a common vocabulary and semantics to express concepts and relationships between them. In doing so, it is possible to express 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 System Layer of the Industrial Data Space is predominantly formed by the Connectors, it is the Connectors where the security features are realized. The Trusted Connector is an exemplary implementation based on the security aspects mentioned in this section. It is built to demonstrate the security concepts developed in the project and serves as a technological basis for use case implementations.

4.1.2 Security Principles

In line with the general goals of the Industrial Data Space, the development of the Security Architecture follows three main principles:

Reliable Technologies

To the extent possible and reasonable, existing standards and best practices are to be taken advantage of. The aim of the Security Architecture is not to offer a new solution for problems already solved, but to combine existing, reliable approaches in a useful and meaningful way and bridge gaps where necessary.

Scalable Approaches

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 such participants must be reliable and verifiable for others. Certain minimum security requirements (e.g., encrypted communication) therefore need to be met by all participants.

Security Pays Off

Provided a participant is in line with the preceding principle, it may decide about the level of security to be applied for it. It should be noticed, however, that Data Sources can mandate a certain set of security features that have to be fulfilled. This means that a higher security level enables access to Data Sources of higher quality and to services of higher value.

4.1.3 Key Security Aspects

The Security Architecture addresses five key aspects: secure communication, identity management, trust, trusted platform, and access and usage control. Each of these aspects relates to several of the Architectural Layers.

Secure Communication

Secure communication protects transactions against eavesdropping, manipulation, and impersonation while data is being transmitted between two participants. To facilitate confidential and integrity protected communication, a number of technical standards and best practices (e.g., WebSockets over TLS) is available. To provide specific functionality (e.g., remote attestation), the Industrial Data Space Communication Protocol (IDSCP) is designed and implemented.

Identity Management

Each participant possesses identities that are used for authentication when communicating with another participant. The Industrial Data Space uses standard technologies like OAuth 2.0, JSON Web Tokens, and X.509 based certificates for identity management. All these technologies are well-established and wide-spread in the context of Web Services and the Internet of Things, and there are numerous standard products available in the market supporting them.

Trust Management

The Security Architecture has a strong focus on concepts for establishing trust between participants in the Industrial Data Space.

Trusted Platform

Connectors have a security profile that is a composite of various characteristics. Every participant may use a Connector with a certain security profile, which can be verified by Connector instances of other participants. Central aspects here are isolation of Data Apps deployed and remote integrity verification.

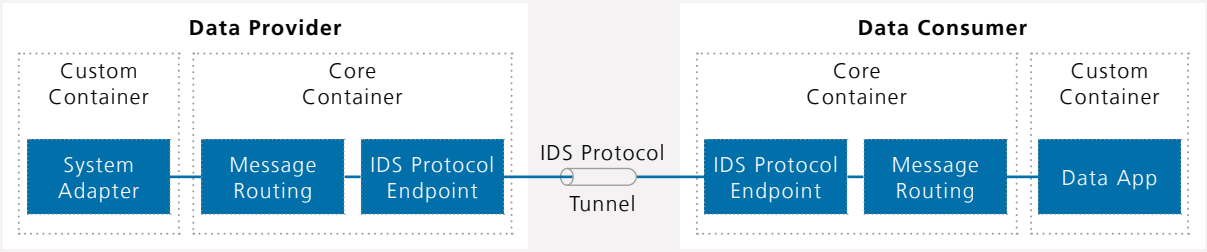


Figure 4.1: Industrial Data Space Communication Protocol

Access and Usage Control

Access control is mandatory to regulate access to data while it still remains at its source. Unlike concepts of access control typically known, the Industrial Data Space also provides means to attach usage restriction information to datasets. These policies, specifying obligations that need to be fulfilled by the Connector the data is sent to, are enforced during the data lifecycle. This way, data usage can be controlled even after the data has been sent by the Data Provider.

4.1.4 Secure Communication

To ensure confidentiality and authenticity of the data transmitted, 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 general Security Architecture. The Security Architecture defines the Industrial Data Space Communication Protocol (IDSCP), which must be supported by Trusted Connectors, and can be supported by any other Connector too. The purpose of the IDSCP is to establish confidential, authenticated communication, exchange data and metadata 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 policies attached).

The last aspect, exchange of data, provides the basic function of data usage control: the data can be attached with a set of usage policies specifying how the data may be used after delivery.

4.1.5 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).

An identity may have several attributes, which are linked to that identity.
For proper operation and access control decisions, information about the identity of a participant alone may not be enough, as every entity may also possess attributes as part of its identity. Examples of such attributes are
- certification level,
- certification timeframe,
- certified security features (e.g., secure server room),
- membership status, or
- domain and business area.

To manage these attributes, an attribute provider is needed for assigning attributes to entities. As these attributes may change over time, it may be useful to provide attributes dynamically (instead of e.g. embedding them in static X.509 certificates). Taking these aspects into consideration, a relatively simple identity management architecture is proposed, supporting both certificate-based identification and flexible, dynamic attribute management, (as depicted in Figure 4.2).
The Certificate Authority (CA) issues certificates for all entities. These certificates are used to establish communication between participants (e.g., to verify the identity of an Attribute Server). The Attribute Server is an identity management server that connects identities with dynamic attributes and issues identity tokens to requesting parties.

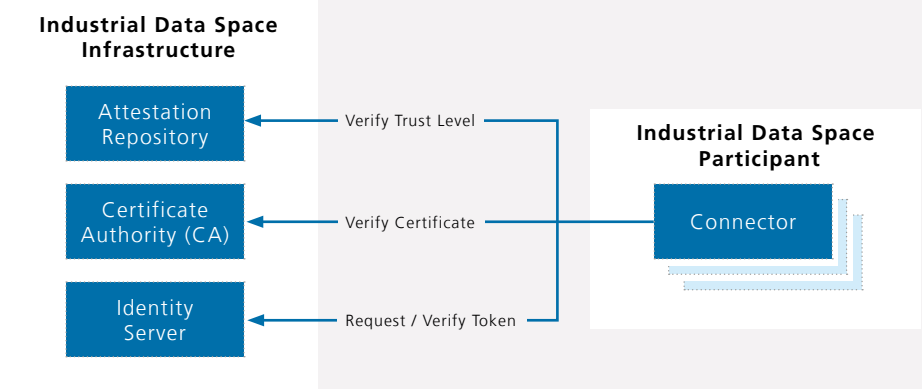


Figure 4.2: Identity management architecture

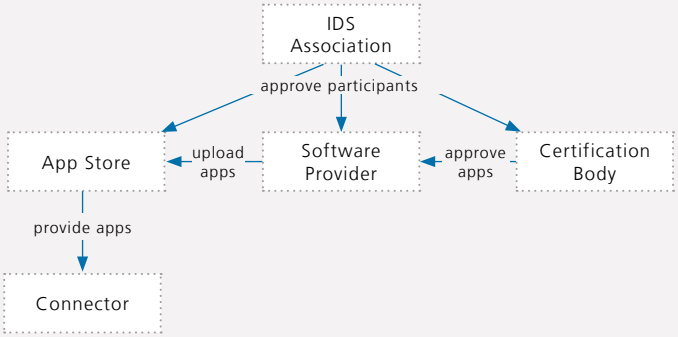


Figure 4.3: Technical roles in the Industrial Data Space ecosystem

4.1.6 Trust Management

To establish trust across the entire business ecosystem (i.e., to protect Industrial Data Space 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.

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 business roles defined on the Business Layer. In particular, six entities with different security levels are of interest to the Security Architecture (Figure 4.3). In the following, these entities and the related roles are described. They map directly to the roles described on the Business Layer.

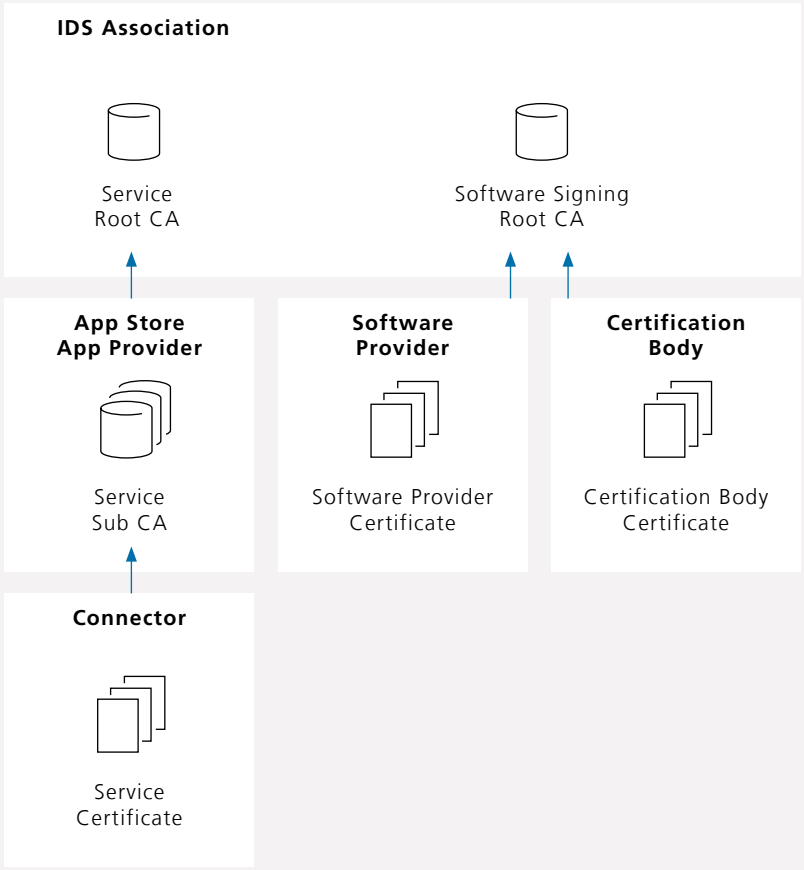


Figure 4.4: Mapping of technical roles and PKI layout

Identity Provider

The Identity Provider acts as an agent for the Industrial Data Space 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 Provider or App Provider). Only if equipped with such an identity, an entity is allowed to participate in the Industrial Data Space (e.g., by offering Data Apps). The Identity Provider may exclude participants from the Industrial Data Space, if instructed to do so. Furthermore, the Identity Provider can authorize certain entities to act as Certification Bodies.

As a trusted entity, the Identity Provider manages the PKI roll-out. 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). Every entity is assigned either end-certificates or sub/root CA certificates. The two hierarchies protect the interests of the six entities, which use and manage the PKI as described in the following (Figure 4.4).

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 enabled 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 the 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 its software from an App Store. Connectors can only connect with approved App Stores for requesting downloads and updates. The App Store is a Connector itself, which 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 on the App Store (i.e., on 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 (hence 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 upload it to the App Store. Each App Provider can be uniquely 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 software but does not create a certificate.

Connector Manifestations
An Industrial Data Space 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 others (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.

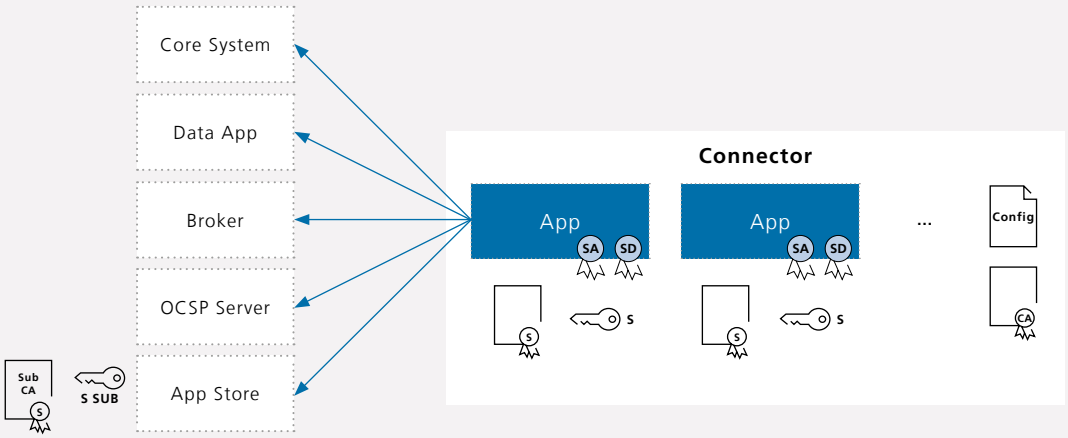


Figure 4.5: Connector roles and manifestations

CA Certificates

In order to verify PKI signatures (e.g., for authentication or for downloaded Data Apps), the Connector stores the trusted root certificates (Service Root CA and Software Signing Root CA) in a way their integrity is preserved (Figure 4.5).

Apps

Apps offered in the Industrial Data Space are usually running in isolated containers. The Connector creates a key pair for every app it downloads. The private key protects the app's persistent data. When downloading a software 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 Association 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.

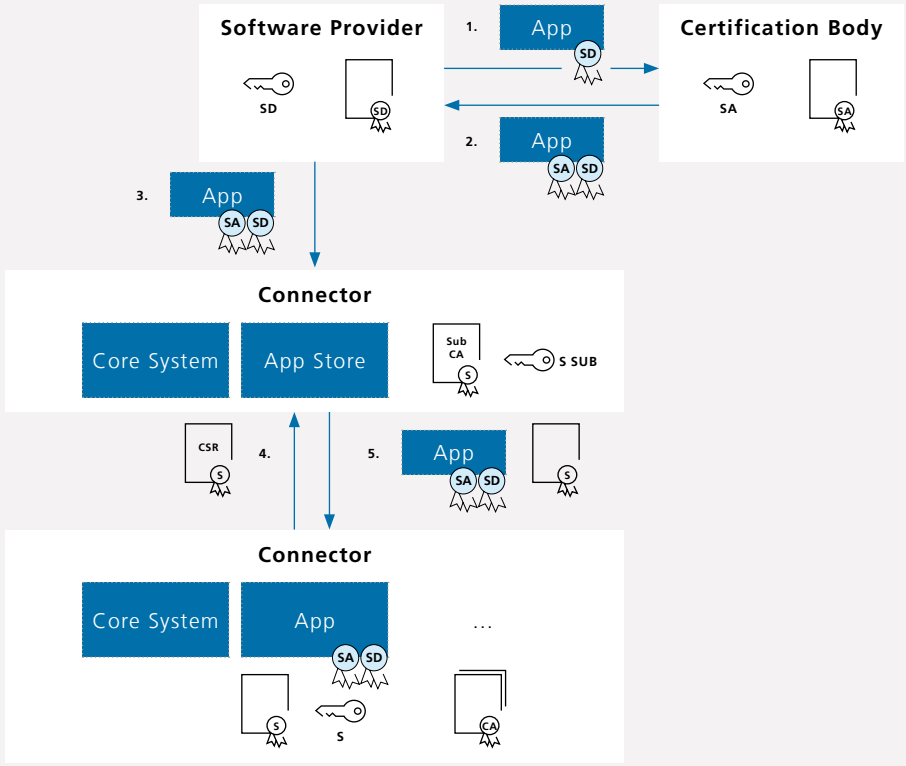


Figure 4.6: Software development, approval, and download process

App Development and Deployment

The following steps describe the app lifecycle, from app development to app deployment onto a Connector (Figure 4.6):

The Identity Provider signs a key pair and a certificate for each Software Provider on behalf of the Industrial Data Space 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.

If the Certification Body approves the app, a second signature is added to it.

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).

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.

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.

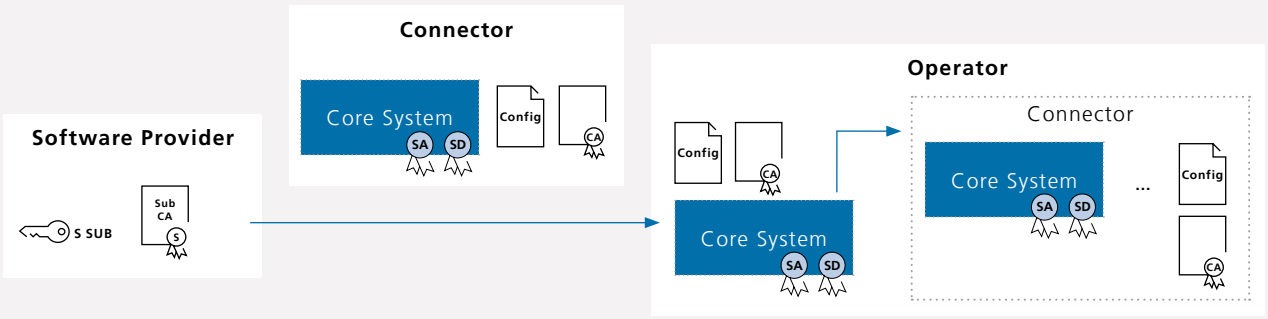


Figure 4.7: Delivery of Trusted Connector

Delivery of Trusted Connectors

After initial deployment the Connector is delivered to the Operator in a completely preconfigured state (Figure 4.7). 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.

4.1.7 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 parties exchanging data, a common understanding of each other's security profiles needs to be established. The Connector supports mutual verification of security 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 the Data Apps have access only to data that is meant for them and cannot exceed given execution boundaries.
- 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 (a hardware-based trust anchor is mandatory for proving the existence and integrity of a given software stack).

Isolation and Remote Execution Guarantee

Isolation is a form of integrity enforcement for the runtime environment of an app. Apps can be isolated against each other by deploying each app into a separate container (or all apps of a specific Software Provider into one container), as illustrated in Figure 4.8. This allows implementation of additional security features, such as time-to-live policy enforcement for complete container instantiations.

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

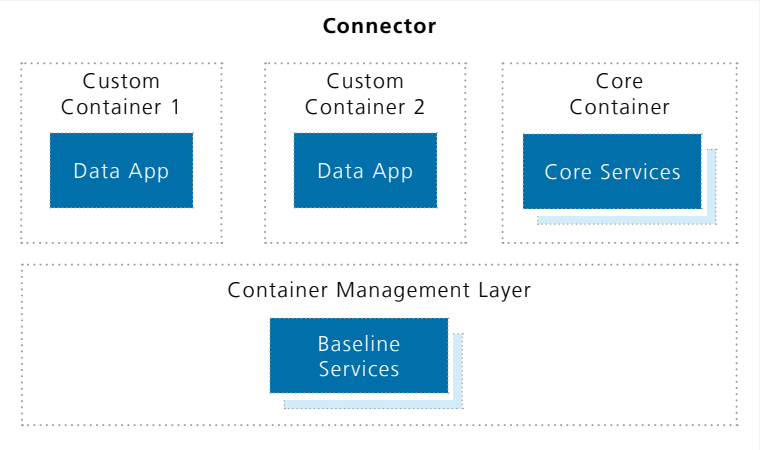


Figure 4.8: Container isolation

may take 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, 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 might occur:

- A Connector might 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.

Dimension	Implementation				
Trusted Platform Module (TPM)	Without TPM		TPM 1.2		TPM 2.0
Authentication	Without certificate	"self-signed" - certificate	Ca-based certificate of internal CA	Ca-based certificate of external CA (cross-certified)	Ca-based certificate of IDS CA
Container Management Layer (CML)	–		Baseline CML (e.g., Docker)		Hardened TrustX CML
Remote Attestation	No RAT		CML & Core Container Attestation		CML & Core Container & Container Attestation
Isolation / Execution Control	–		Basic Runtime Monitoring		Controlled Remote Execution
Software Assurance Level	Unknown software stack			IDS-certified software stack	

Table 4.1: Security Profile options

4.1.8 Connector Security Profiles

Security Profiles are attributes of Connectors and can be used as such for attribute-based access control. Each Connector must provide its Security Profile upon request (however, the profile may be empty, as already mentioned above). The Security Profile

- describes the Connector security configuration in place,
- allows the Data Consumer to decide whether or not it is willing to trust the data provided by a certain Data Provider, and
- allows the Data Provider to decide whether or not it is willing to provide sensitive data to a certain Data Consumer.

A Security Profile may consist of a number of options, as listed in Table 4.1.

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

4.1.9 Access and Usage Control

Industrial Data Space Connectors provide mechanisms to regulate access to data. To define access conditions for data and services, the following criteria can be specified:

- specific identity of Connector(s): only access requests from one specific Connector (or from a number of specific Connectors, respectively) are granted;
- Connector attributes: only access requests from a Connector that possesses specific attributes are granted;
- Security Profile requirements: only access requests from a Connector meeting specific security requirements are granted (e.g., having a TPM >= 1.2 and doing application isolation with trusted container management).

Using static security levels would make it necessary to anticipate all future needs of any participant whatsoever. Since the Industrial Data Space is designed to grow over time and map flexibly to the individual security needs of every participant, it offers the possibility to base access control decisions on fully customized criteria. Access 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 and the security properties of the technical setup of the Connector. This is described in the section on the security profile earlier in this document.

Beside access control, the Reference Architecture Model also supports data usage control. The purpose of usage control is to bind policies to individual messages or data sources, and restrict the way data contained in these messages may be processed, aggregated and forwarded. At configuration time, these policies support developers and administrators in setting up correct data pipes, which comply with these policies and do not leak data via side channels. At runtime, usage control enforcement prevents Connectors from treating data in an undesired way (e.g., by forwarding personal data to public endpoints).

The following are examples of requirements where data-centric usage control is necessary:

- secrecy: classified data must not be forwarded to Connectors which do not have the respective clearance;
- integrity: critical data must not be modified by untrusted Connectors, as otherwise the integrity of this data cannot be guaranteed anymore;
- time to live: data may be persisted only if it is clear that it will be deleted after a certain period of time;
- anonymization by aggregation: personal data may only be used as aggregates by untrusted parties; a sufficient number of distinct records must be aggregated in order to prevent deanonymization of individual records;
- anonymization by replacement: data allowing personal identification (e.g.; faces shown on photos) must be replaced by an adequate substitute (e.g., pixelized) in order to guarantee that individuals cannot be deanonymized;
- separation of duty / conflict of interest: two datasets from conflicting entities (e.g., two automotive OEMs) must not be processed by the same node;
- scope of usage: data may only serve as input for data pipes inside the Connector, but must not leave the Connector to be sent to an external endpoint.

It is important to note that the sole purpose of usage control is to allow specification of such constraints and enforcing them in the running system. It is a prerequisite of usage control that the enforcement mechanism itself is trusted (i.e., usage control itself does not establish trust in an endpoint, but rather builds upon an existing trust relationship and facilitates the enforcement of legal or technical requirements, such as enforcement of service level agreements (SLAs) or data privacy regulations).

The Reference Architecture Model of the Industrial Data Space supports integration of usage control frameworks, but it does not dictate a specific product or policy language. The following exemplary policies illustrate rules which can be expressed in the respective policy language:

1. Any personal data to be published to an external source must be anonymized to an aggregation level of 10 distinct records before.
2. A data source may only be read by a Connector with a certain ID given in the certificate.
3. If data is persisted, it must be deleted after 30 days.

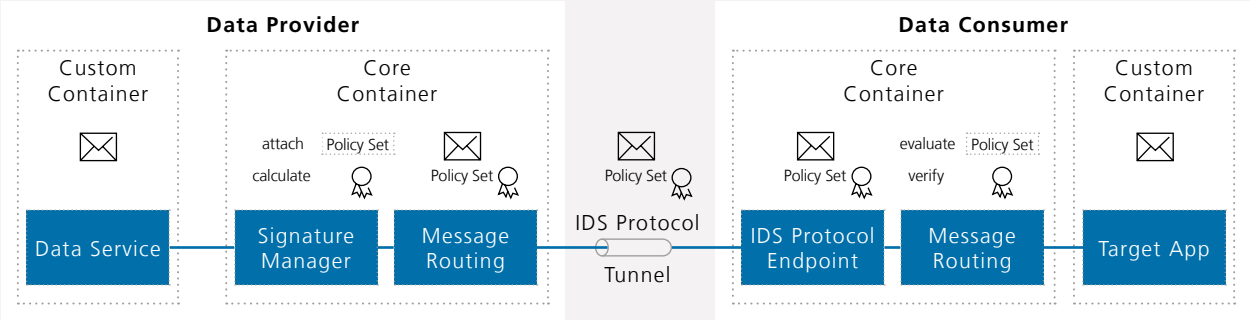


Figure 4.9: "Sticky policies"

Policy 2) is an access control policy that needs to be enforced before data access (once data is offered to a consumer, access control policies cannot be enforced any more). Policy 1) controls data flow, while policy 3) specifies data lifetime; both need to be enforced during the data lifecycle, as they specify how data may be handled.

Policies are attached to data items using a technique called "sticky policies" (Figure 4.9). When policies are attached to a dataset, the whole set is signed by the Data Owner (i.e., by the Connector hosting the data and administering the policies). As far as data streams are concerned, a hybrid approach should be applied, with the two Connectors negotiating usage policies for a data source. Whenever data is transmitted from one Connector to another, a negotiation set is attached to the data, or the data source itself is tagged.

In general, three building blocks are required to implement data usage control:

- Enforcement: To restrict data usage, usage control enforcement must be implemented where usage actually takes place (typically on the client side). To do so, enforcement components need to be integrated into the respective systems. These components are typically technology-dependent and domain-dependent. They can be split into two basic categories: components intercepting data flow or usage (e.g., printing a file) and components executing specific actions (e.g., deleting all copies on the client).
- Decision: Common access control policies are typically binary (i.e., access is either allowed or not). To regulate the usage of data, however, this approach is not sufficient. Usage control policies may additionally need to consider contextual information (e.g., on the location or the task executed) and make fine-grained decisions (e.g., only an average value may be exposed), including obligations (e.g., data has to be deleted after 30 days). All of this makes decision-making complex. However, the policy language and evaluation is typically independent of domains and technologies. It can be implemented on the server side or the client side, depending on the technology used (e.g., policies can be stated in a simple, domain-specific, human-readable language).
- Management: The management of the policy lifecycle is an important, yet challenging task. Before a policy can be enforced, it needs to be specified, negotiated and deployed. As multiple policies can be deployed simultaneously, conflict detection and resolution may be necessary. Furthermore, policies can be changed and revoked. All these processes need to be clearly defined and managed in order to avoid undesired behavior during runtime.

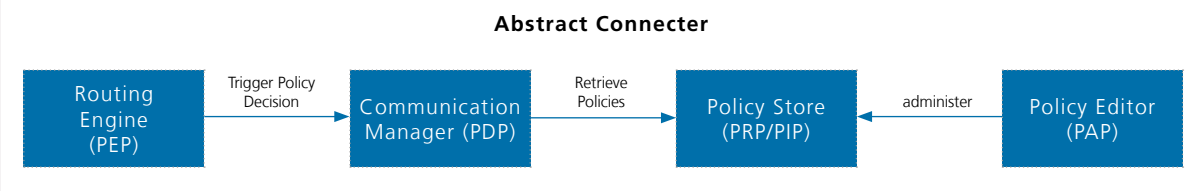


Figure 4.10: Abstract Connector

The Reference Architecture Model reflects these necessary components in an abstract manner (Figure 4.10). The Trusted Connector incorporates a Routing Engine, in which usage control policies are enforced (Policy Enforcement Point, PEP). Furthermore, a Communication Manager for evaluation of usage control decisions (Policy Decision Point, PDP) and a Policy Store keeping policies for being retrieved by other components (Policy Retrieval Point, PRP) are in place, as well as a Policy Editor for policy administration (Policy Administration Point, PAP).

Integrating Usage Control into a Connector implementation requires some prerequisites: Considering the container architecture, the integration points depend on what is to be enforced. Possible integration points can be distributed only at some specific points in the architecture or simply where all data flows through. Both approaches have their advantages and disadvantages, define what can be enforced by a policy at the end and how the performance of the system is influenced. At such integration points, the data must not be encrypted or processed by anyone else before.

As described in 3.5.1, a Connector implementation can use a routing engine and a message bus. In case a routing engine is used, this would be a perfect interception point, if all applications submit their data via routes and if the policies can be enforced, based on the messages submitted between the routes. An example for such an implementation is the Trusted Connector (Figure 4.10). Typically, routing engines offer a possibility to intercept all communications by providing some kind of interception strategy. When an interception takes place, the PEP has access to the message that has been sent as well as to the information about the sender and receiver. This allows the handling of data (e.g., modification) according to active policies between every route point. The drawbacks of the approach are that other libraries are also able to add interceptors and that the PEP interceptor is not at a guaranteed fixed position before all others. In such a case, other interceptors are able to modify the data in a way that a PEP is unable to process it (e.g., by encryption) or can use the data before the PEP was active. Preventing this is only possible by ensuring a certain interceptor sequence or that only certain PEP interceptors are instantiated.

Basic Connector implementations can also contain some kind of message bus. Like a routing engine, they typically offer interception strategies with comparable drawbacks. Depending on the policies to be enforced, adding interception points with PEPs to a message bus, may be necessary in addition.

If neither a routing engine nor a message bus is used, the general advices stated above should be considered when implementing usage control in a custom Connector.

4.2 CERTIFICATION PERSPECTIVE

The Certification Scheme of the Industrial Data Space defines the processes, roles, objects, and criteria involved in the certification of hardware and software artifacts as well as organizations in the Industrial Data Space. While certification of organizations focuses on trust and security, certification of components also evaluates compliance with other requirements defined in the Reference Architecture Model.
This section provides an overview of how the central entities and roles defined for the Reference Architecture Model (Section 3) are linked with the Certification Scheme. After a general description of how certification affects the different layers of the Reference Architecture Model, this section discusses which roles are in charge of carrying out the certification process, which entities and components are targets of the certification process, 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. They are defined in subsection 4.2.2, along with their interactions and responsibilities within the Certification Scheme.
Organizations assuming roles under the three categories Core Participant, Intermediary, and Software / Service Provider (Section 3.1.2) are potential targets of certification. Subsection 4.2.3 describes for each role whether it requires certification and what the focus of certification is.

Functional Layer
The functional requirements of the Industrial Data Space are the core requirements expected to be implemented by the 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 (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 an Industrial Data Space core component comprises evaluation of its security as well as evaluation of its compliance with the Reference Architecture Model (regarding functionality, protocols, etc.). Whenever relevant, evaluation of a core component's compliance will also ensure compatibility with the Information Model, as defined in the Information Layer.

System Layer
The System Layer defines the interactions between the components, detailed requirements for the Connector, and specific types of Connector implementations. The System Layer is the predominant layer in focus of 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.

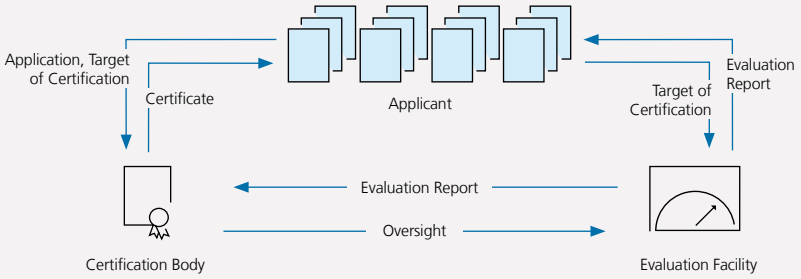


Figure 4.11: Certification Scheme

4.2.2 Roles in the Certification Process

The Certification Scheme of the Industrial Data Space comprises the roles shown in Figure 4.11. These roles were introduced under the "Governance Body" category in the Business Layer. The tasks of these roles with regard to the certification process are described in the following sections.

Certification Body

The Certification Body manages the entire certification process and supervises the actions of the Evaluation Facility. Organizations are granted a certificate only if both the Evaluation Facility and the Certification Body have come to the conclusion that all preconditions for certification are met.

Responsibilities of the Certification Body include

- ensuring correct implementation and execution of the Certification Scheme;
- analyzing already existing certificates (e.g., of organizations or of hardware security components) and deciding about their validity for and acceptance by the Certification Scheme;
- checking and commenting on evaluation reports received from Evaluation Facilities;
- making the final decision about granting or denial of a certificate;
- making the decision about approval or exclusion of Evaluation Facilities from executing Industrial Data Space evaluations (based on ongoing monitoring);
- monitoring all certification-relevant external developments (e.g., new attack methods which might break certified security measures);
- taking care of the future evolution of the Certification Scheme.

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 any 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 sufficient.

The Certification Body itself may be accredited by the national accreditation body (e.g., DAkkS in Germany[21]), which supervises a set 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.

21 http://www.dakks.de

Evaluation Facility
To carry out the technical evaluation during a certification process, an Evaluation Facility is contracted by the respective Applicant.
Responsibilities of the Evaluation Facility include
- obtaining approval from the Certification Body to perform evaluations;
- applying the criteria specified in the Certification Scheme according to generally accepted standards and best practices (including the execution of any necessary tests and on-site checks);
- documenting the results in an evaluation report;
- providing the evaluation report to the Certification Body.

Applicant
The Applicant plays an active part in the certification process. As such, the respective organization has to
- provide the necessary resources for the certification process (in terms of financing and personnel);
- formally apply for certification (with the Certification Body) in order to trigger the certification process;
- contract an Evaluation Facility approved by the Certification Body to carry out the evaluation according to the Certification Scheme;
- 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.

This applies to both organizations that develop software components intended to be deployed within the Industrial Data Space (i.e., prospective Software Providers) and organizations that intend to become Participants in the Industrial Data Space. All applicants need to actively submit an application to start the certification process and have the duties as listed above. During the certification process, the primary focus of the evaluation will be either on the product or on the organization itself.

4.2.3 Targets of Certification – Entities

Core Participants
The Data Provider is responsible for the integrity, confidentiality, and availability of the data it publishes and provides. Evaluation and certification of the security mechanisms employed by the Data Provider should provide a sufficient degree of security against the risk of data integrity, confidentiality, or availability being undermined by attacks.
Data Owners are assumed to often act as a 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.
For a Data Owner not acting as a Data Provider at the same time, evaluation and certification of the technical, physical, and organizational security mechanisms employed provide a sufficient degree of security against the risk of data integrity or confidentiality being undermined by attacks.
As an organization that has access to data provided by a Data Owner, the Data Consumer also assumes responsibility for the confidentiality and integrity of that data (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, the Data Consumer has 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 the Data Consumer 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 initiative, it is highly critical to eliminate all risks involving manipulation of identities. The integrity and availability of identity-related information processed by the Identity Provider is therefore of utmost importance. Only 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 able to provide a sufficient degree of security against these risks.

Broker Service Providers, providers of Clearing House services, the App Store Provider, and the Vocabulary Provider deal only with metadata, transactions, or apps (i.e., 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 metadata is lower (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 extended periods of time. Therefore, evaluation and certification tailored to the specific risk profiles of and security mechanisms employed by Broker Service Providers, providers of Clearing House services, App Store providers, and Vocabulary Providers is proposed in order to ensure a sufficient degree of security against the risks mentioned. As far as the App Store Provider is concerned, there is an additional risk in terms of an attacker successfully substituting legitimate apps with modified versions, thereby threatening the payload data indirectly. However, technical measures in the App Store implementation (e.g., only apps cryptographically signed by the app developer are accepted and distributed) seem more effective for reducing this risk than organizational measures on the part of the App Store Provider.

Software and Service Providers

Providers of compliant software usually have no contact with sensitive data, but execute tests with appropriate, non-sensitive test data. Therefore, in most cases no certification of the organizational security is required. If access to actual data of the Industrial Data Space is necessary, the Software Provider assumes the role of Data Consumer or Data Provider for as long as such access is needed. In that case, the certification requirements of the corresponding roles apply.

Service Providers are employed by other participants of the Industrial Data Space in order to outsource certain tasks (e.g., publishing data). As they adopt the other role's duties and responsibilities, they should be subject to certification.

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 transfer of data from the Data Provider to the Data Consumer. As such, the necessary trust in the correct and complete implementation of the functionality required by the Reference Architecture Model and the Connector specification can only be ensured by independent evaluation and certification from an approved Evaluation Facility and the Certification Body of the Industrial Data Space.

Broker Service Providers do not have access to primary data, but only to metadata provided by Data Providers, which is generally considered less sensitive. Likewise, Broker Service Providers do not assign or enforce access rights, but merely support data exchange. Nevertheless, 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 with the required functionality as defined by the Certification Body is therefore evaluated and certified.

The activities of the Clearing House encompass the provision of reports on the performed transactions for billing, conflict resolution, etc. As such, all implementations of the clearing component need to be evaluated and certified according to the functional and security 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 needs to be evaluated and certified according to the functional and security requirements as defined by the Certification Scheme.

Data Apps have direct contact with primary data, which means that a compromised Data App 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. Therefore, not every Data App to be made available in the Industrial Data Space requires certification. Nevertheless, certification should be required for apps of high importance to the Industrial Data Space community, and for apps 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 (Chapter 4.1.5), 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. 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 Industrial Data Space defines the roles, functions, and processes from a governance and compliance point of view. It defines the requirements to be met by an innovative data ecosystem to achieve corporate interoperability. This chapter provides an overview of how central questions of governance are defined on the different layers of the Reference Architecture Model (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, interoperability, and the use of new 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 the requirements of the participants;
- 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 (Chapter 3.1) contributes to the development of business models which can be applied by the participants in the Industrial Data Space. In particular, it describes the different roles participants may assume. The Business Layer directly refers to the Governance Perspective by considering the business point of view regarding data ownership, provision, brokerage, and consumption, and by describing core service concepts.

Functional Layer

The Functional Layer (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 (Chapter 3.2.6), Identity Management (Chapter 3.2.5), and Trust & Security (Chapter 3.2.1), which are entities for which the relation to the topic of governance is obvious, the described functionality of entities such as the App Ecosystem, Vocabulary and Metadata Management, and the Connector has an impact on the Governance perspective, and vice versa. Vocabulary and Metadata Management (Chapter 3.2.2) plays a primary role in defining common rules for exchanging business-related metadata in a standardized way.

Process Layer

The Process Layer (Chapter 3.3) describes the interactions between the different components of the Industrial Data Space, offering a dynamic view of the architecture and the different steps in terms of providing and exchanging data as well as using Data Apps. The Governance perspective is influenced by each of the three processes – 1) providing data, 2) exchanging data, and 3) publishing and using Data Apps – described in the Process Layer section, as they define the scope of the Governance Perspective regarding the technical architecture.

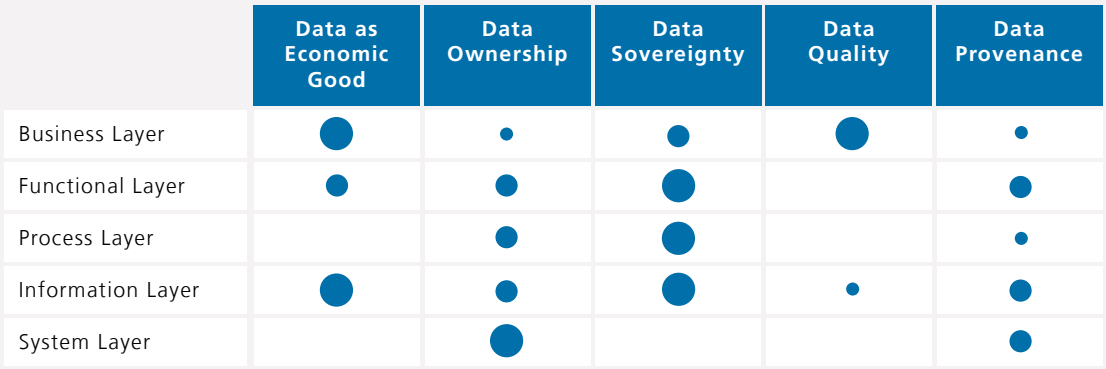


Figure 4.12: Impact of governance related topics on different architectural layers

Information Layer
The Information Layer (Chapter 3.4) provides a common model and vocabulary for the participants to express their concepts. It defines a framework for standardized collaboration and 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 central relevance for arranging and describing data in the Industrial Data Space.

System Layer
The System Layer covers Governance aspects 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. An overview of the impact of each topic on the different architectural layers is given in Figure 4.12 (large circle = strong impact; medium circle = medium impact; small circle = weak impact).

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. In particular, companies need a means to ensure data sovereignty.
The Industrial Data Space aims at an architectural approach that facilitates the exchange of data within business ecosystems while ensuring data sovereignty. In doing so, it offers a basic architecture for organizations that want to optimize their data value chains. The main goal is to enable companies to leverage the potential of their data within a secure and trustful ecosystem. The Industrial Data Space does neither make any statements on legal perspectives, nor does it restrict companies to any predefined patterns. Instead, it offers the possibility to design 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. Furthermore, it is necessary to take into account that the Data Owner may not be the Data Provider (Chapter 3.1.1), also (or especially) in digital ecosystems.
Data ownership is an important aspect when it comes to offering data and negotiating contracts in a digital ecosystem, especially because data can easily be duplicated. The Industrial

Data Space covers the topic of data ownership by providing a secure and trusted approach for authorization and authentication within a decentralized architecture, where Data Providers as well as Service Providers can be identified and controlled by an Identity Provider (Chapter 3.1.1). Decentralized data exchange through Connectors, in contrast to other architectures of data networks (e.g., data lakes), ensures full sovereignty over the configuration of data offerings on the part of Industrial Data Space participants. In addition to these self-control mechanisms, the architecture considers clearing and logging of data transfers through a Clearing House (Chapter 3.2.5). Data ownership thus is indeed relevant on every layer of the architecture.
The Industrial Data Space intends to build upon and apply existing law. It will not cover any purely technology-oriented solutions to prevent data duplication or misuse of data goods, but supports these important aspects over the full data exchange 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. While data ownership mainly refers to data provision, data sovereignty rather considers data access, including permissions, restrictions, and control.
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, and that misuse on the customer side has to be restricted.
Data exchange takes place through secured and encrypted transfer including secured (Chapter 4.1) and certified (Chapter 4.2) authorization and authentication. The Data Provider may add a metadata description using the Industrial Data Space Vocabulary. In doing so, the conditions for ensuring 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 existing legal regulations, 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

The Industrial Data Space covers data quality aspects because of the correlation between stable data quality and maximizing the value of data as an economic good.
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 due to its brokerage functionality. Especially in competitive environments, this transparency may force Data Providers to take the maintenance of their data goods more seriously. By extending the functionality of the Connectors with self-implemented Data Apps (Chapter 3.2.4), the Industrial Data Space lays the foundation for automated data (quality) management.

4.3.6 Data Provenance

As it creates transparency and offers 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, supporting these two aspects by encouraging traceability.
For example, the Clearing House (Chapter 3.1.1) logs all activities performed in the course of data exchange and requests confirmations from the Data Provider and the Data Consumer. By this, data provenance information is always recursively traceable.
The Industrial Data Space hereby provides the possibilities to implement and use appropriate concepts and standards. However, it does not force its participants to use these concepts and standards. Therefore, it is up to the individual Industrial Data Space Participant to provide correct information (i.e., metadata) on the provenance of data.

APPENDIX A: 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 or for having their software product certified by the Certification Body
Broker	Intermediary managing a metadata repository that provides information about the Data Sources available in the Industrial Data Space; multiple Brokers may be around at the same time, maintaining references to different, domain-specific subsets of Data Endpoints
Certificate 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)
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 Industrial Data Space Vocabulary
Data Asset	Content exposed for interchange 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

APPENDIX A: GLOSSARY

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 via 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
Data Sovereignty	A natural person's or corporate entity's capability of being entirely self-determined with regard to its data
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
Industrial Data Space 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
Industrial Data Space Vocabulary, 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	The Industrial Data Space materializes as a distributed network of Data Endpoints (i.e., instantiations of the Industrial Data Space Connector), allowing secure exchange of data and guaranteeing Data Sovereignty
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

APPENDIX B: FUNCTIONAL OVERVIEW

The following list contains the functional requirements to be met by the Reference Architecture Model of the Industrial Data Space. If a number is missing, this indicates that the respective requirement has turned out to be irrelevant during the validation process and was therefore removed from the list.

Select vocabulary [IDSFO-1]
Each Connector operator should be able to select a vocabulary from the Vocabulary Hub in order to describe the data offered.

Describe Data Source [IDSFO-2]
Each Connector operator should be able to describe the properties of data, including data format (IDSFO-96, IDSFO-97), date and time of creation and owner of the data (IDSFO-98), price information and access permissions, domain (IDSFO-94), etc.

Define pricing model [IDSFO-3]
Each Data Provider should be able to define the pricing model and the price, such as pay per transfer, pay for access per day/month/year, etc.

Account data usage [IDSFO-4]
Each Connector operator should be able to account the usage of the data transferred and received.

Statistics of data usage [IDSFO-5]
Each Connector operator should be able to request statistics regarding the usage of the data transferred and received.

Define usage policy [IDSFO-7]
Each Connector operator should be able to define how data must be used. For example, the usage policy may prohibit forwarding of data to other participants or merging of data with other data.

Offer data [IDSFO-8]
Each Data Provider should be able to offer data to the general public, IDS Participants, or groups of IDS Participants (IDSFO-93).

Maintain source description [IDSFO-9]
Each Connector operator should be able to maintain the Data Source description. Any modification results in a new version of the description in order to stay consistent with data already transferred.

Manage versions of source descriptions [IDSFO-10]
Each Connector operator should be able to publish different versions of a Data Source and mark versions as "deprecated".

Create vocabulary [IDSFO-11]
Each Industrial Data Space Participant should be able to create vocabularies. Access to a vocabulary can be restricted to selected Participants.

Update vocabulary [IDSFO-12]
Each vocabulary can be edited, updated, and extended by its creator and, if allowed, by other users. Any modification results in a new version of the vocabulary in order to stay consistent with its users.

Manage versions of vocabularies [IDSFO-14]
Each creator of a vocabulary should be able to manage and publish different versions and mark versions as "deprecated" (IDSFO-10).

Match vocabularies [IDSFO-15]
Each Industrial Data Space Participant should be able to define mappings between related vocabulary terms.

Manage knowledge database [IDSFO-16]
The Vocabulary Hub operator should be able to manage the knowledge database. More specifically, the operator should be able to 1) update and maintain local copies of standard vocabularies (IDSFO-89) and 2) identify and delete unused or duplicate vocabularies (IDSFO-90, IDSFO-91, IDSFO-92).

Search for given vocabularies [IDSFO-17]
Each Industrial Data Space Participant should be able to search for vocabularies in the Vocabulary Hub.

Installation support for custom Data Apps [IDSFO-18]
A dedicated connector service should support authorized users in (un-)installing custom Data Apps not originating from a certified App Store.

Certification process for Data Apps [IDSFO-20]
Prior to publication in an App Store, each Data App should undergo an optional evaluation and certification process executed by the Certification Body.

App annotation support [IDSFO-22]
Prior to publication in an App Store, each developer of a Data App should be able to annotate the Data App with metadata (on the functionality and interfaces provided, the pricing model, license information, etc.). This annotation is either a manual activity or a semi-automatic procedure assisted by dedicated services of the Industrial Data Space (e.g., wizards).

Publish software in App Store [IDSFO-23]
Each authorized software developer should be able to initiate a software supply process (App Store publication).

Search for Data Sources [IDSFO-25]
Each Industrial Data Space Participant granted access rights should be able to search for Data Sources.

Browse Data Sources [IDSFO-26]
Each Industrial Data Space Participant granted access rights should be able to browse Data Sources.

Buy data [IDSFO-30]
Each Industrial Data Space Participant granted access rights should be able to buy data. To do so, the Participant must be identifiable in order to balance the corresponding account. To initiate a transaction, a broker is not necessarily required, as the data can be bought directly from a Connector also.

Gather data from Participants [IDSFO-31]
A Data Sink is a Data Endpoint intended to retrieve (= active mode) or receive (= passive mode) data from other IDS Participants. It should adhere to a retrieval configuration (frequency, amount, etc.) or enforce acceptance tests when receiving data from subscribed data sources. Furthermore, a Data Sink should subscribe to a data source and request updates. Participants should be able to choose between different update models.

Define workflow [IDSFO-32]
Each Connector operator should be able to define the data workflow inside the Connector (message router). It starts with a Data App (System Adapter) or with an input by the execution core container. The data will then be transferred to Data Apps following a defined workflow.

Read data from backend systems [IDSFO-33]
Each Connector must be able to receive data from an enterprise backend system, either through a push-mechanism or a pull-mechanism.

Data processing [IDSFO-34]
Each data processing app (subtype of a Data App) should be able to provide a single, clearly defined processing functionality to be applied on input data for producing an expected output. It should operate in a stateless and transparent way. The processing of varying input (streams) should have no side-effects. It should provide interfaces allowing integration in data processing workflows. Among other things, associated metadata should provide information on the programming interface and the semantics of the data and the processing logic applied (e.g. anonymization).

Transform data [IDSFO-35]
Each 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). Mapping of input dataset and output dataset should be 1:1; i.e., no aggregation or state persistence should be involved. Annotated metadata should explicitly indicate the type of transformation and the IN/OUT parameter types supported.

Expose data [IDSFO-37]
Each Connector operator should be able to expose data (e.g., to generate a Data Source). As part of this process, the operator should be able to control the access policy (e.g., access granted to all Participants or restricted to a certain group of Participants).

Identify Connectors [IDSFO-39]
Each Connector must have a unique identifier (URI). Each Participant should be able to identify other Industrial Data Space Connectors and check their configuration (e.g., security profile or published Data Endpoints). The Connectors should support identity management (e.g., enabled by exchanging Web Tokens) to make it possible to authenticate incoming and outgoing connections.

Establish secure connection [IDSFO-42]
Any communication between (external) Connectors should be encrypted and integrity protected. Established secure protocols (e.g., HTTPS with TLS) should be used for transferring data over public networks. Deprecated versions of the protocols should not be supported. This principle matches current best practices for data exchange. To achieve a higher degree of security, instances of the "Trusted Connector" should use the IDS Communication Protocol supporting remote attestation.

Share data with Participants [IDSFO-43]
A Data Source is a Data Endpoint intended to share data with other Participants. The data can be requested by the receiver and the Connector requests the data from an App or backend system. The data can either be pushed to the receiver, or the receiver can pull the data, or the receiver can subscribe to the Data Endpoint. Both data and metadata is known and will be transferred simultaneously. Each data transfer and access should be logged.

Identify Data Sources [IDSFO-44]
Each Participant should be able to retrieve information about a data source by dereferencing its URI (e.g., http://connector-name/datasource-id).

Write data to backend systems [IDSFO-52]
Exchange of data between Connectors and (proprietary) backend systems located at a Participant's premises should be possible.

Enforce usage policies [IDSFO-53]
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. Each Participant should be able to define usage control policies and attach them to the respective data. Policies may include restrictions (e.g., prohibiting persistence of data or transfer of data to other parties).

Data Exchange Clearing [IDSFO-55]
Each Participant's data transfers can be subject to (cost) accounting. This ensures that contracts between Participants are fulfilled.

Data Usage Reporting [IDSFO-56]
Each Connector should be able to deliver a data usage report, covering inbound, outbound, and internal data flows.

Browse Participant list [IDSFO-59]
Each Participant should be able to browse the list of Participants that provide data with the help of a Broker.

Certification of Participants [IDSFO-60]
Each Participant must undergo a certification process executed by the Certification Body.

Become authenticated [IDSFO-61]
Each Connector must undergo a certification process executed by the Certification Body.

Run connector in own data center [IDSFO-63]
Each Participant should be able to run the Connector software in its own IT environment (e.g., Linux/x86 platforms).

Run connector on mobile/embedded device [IDSFO-65]
There should be versions of Connectors that run on mobile and embedded devices.

Register metadata at IDS Broker [IDSFO-66]
Each Connector should be able to transmit the metadata of its Data Sources to one or more Brokers. This transmission is configured by the Connector operator.

Search support within App Store [IDSFO-67]
The App Store should be able to support (authorized) Participants in searching for Apps using various levels of expressiveness and complexity, such as

- GUI-based browsing/navigation (taxonomy, categories),
- free-text search (words, phrases),
- constrained search language (simplified, natural language with a restricted/constrained vocabulary and syntax like a user-oriented Search-DSL),
- formal Search-DSL; i.e., a formal, less descriptive, normalized form of the above;
- structured search (standard query languages like SQL or SPARQL).

Remote Attestation [IDSFO-71]
Each Connector, App Store, and Broker should be able to check if the Connector of the connecting party is running a trusted (certified) software stack.

Certify connector environment [IDSFO-73]
A certification process should be in place that allows certification of each Connector environment.

Authenticate Connector [IDSFO-75]
Each Participant should be able to verify the identity of any other Participant.

Define level of security for Connector [IDSFO-76]
Each Data Provider and Data Consumer should be able to decide about the level of security of their respective Connectors themselves by deploying Connectors supporting the respective security profile.

Incident Monitoring [IDSFO-77]
Each Connector operator should be able to monitor the data flow between Apps deployed on the Connector and receive notifications about incidents (e.g., in case an App does not respond).

Manage and Identify users of Connector [IDSFO-79]
It should be possible to manage access rights in order to control permission to configure a Connector. For instance, creation and deletion of accounts of Connector administrators as well as any changes regarding permission should be logged.

Installation and management support for Apps [IDSFO-80]
A dedicated Connector service should support authorized users in searching, installing, and managing (e.g., removal or automatic updates) Apps offered by an App Store.

Apps must explicitly define their interfaces, dependencies, and access requirements [IDSFO-82]
Apps deployed within a Connector should be isolated from each other and from the underlying host system, reducing the impact of compromised applications. The only means of interaction and integration is through their documented interfaces. Integration must be approved by the Connector operator.

Isolation of Data Apps within a Connector [IDSFO-86]
To reduce the impact of compromised applications, appropriate technical measures must be applied to isolate Data Apps from each other and from the Connector.

Passwords and keys storage [IDSFO-88]
Authentication information stored on a Connector must be protected (e.g., hashed or encrypted passwords). For instance, passwords may not be stored as plain text, and keys used for signatures and decryption of data may not be stored on an External Connector.

Update and maintain local copies of standard vocabularies [IDSFO-89]
The Vocabulary Hub operator should be able to update and maintain local copies of standard vocabularies (such as DCMI Metadata Terms). Any modification results in a new version of the vocabulary, which can be uniquely identified.

Identify unused vocabularies [IDSFO-90]
The Vocabulary Hub operator should be able to identify unused vocabularies and mark them for deletion.

Identify duplicate vocabularies [IDSFO-91]
The Vocabulary Hub operator should be able to identify duplicate vocabularies and mark them for deletion.

Delete vocabularies marked for deletion [IDSFO-92]
The Vocabulary Hub operator should be able to delete vocabularies marked for deletion. This is required if vocabularies are identified as unused or duplicate (see IDSFO-90 and IDSFO-91).

Describe groups of IDS Participants [IDSFO-93]
In the context of publishing data (IDSFO-8), it should be possible to formally describe groups of Participants.

For example:
– "all automotive companies",
– "everyone interested in weather data", or
– "everyone who has used/purchased my data before".

Describe domain of Data Source [IDSFO-94]
Application domains should be described in a structured way (i.e., not just by textual keywords). The Information Layer should support this by linking to (and reuse of)
– domain-specific taxonomies (e.g., eCl@ss, which is specific to the domain of "products and services");
– digital library classification schemes, such as the Dewey Decimal Classification or the "Gemeinsame Normdatei";
– Wikipedia categories, which constitute a crowd-sourced approach of categorization that covers many domains of common interest (e.g., political administration regions);
– custom taxonomies developed by specific Participants using controlled vocabulary creation tools, such as VoCol or PoolParty.

Audit logging [IDSFO-95]
Any data transfer requires approval and should therefore be logged. Logs reveal, for example, who authorized the outbound transfer (e.g., cryptographic signature), who received the data, or what data was transferred and when. It should also be logged how and when access control decisions are made. This also applies to policies attached to data items. When data arrives at the target Connector, appropriate logging should be in place to document the data transfer and further data processing steps.

Describe semantics of Data Source [IDSFO-96]
Each Connector operator should be able to describe the semantics of a Data Source, such as the standard the data provided conforms to (e.g., GS1/XML).

Describe syntax/serialization of Data Source [IDSFO-97]
Each Connector operator should be able to describe the syntax and serialization format of a Data Source (e.g., XML or JSON).

Describe basic metadata of Data Source [IDSFO-98]
Each Connector operator should be able to provide basic metadata of a Data Source (including, e.g., name of data owner, data deployment or creation time, or classification information).

Certification of core components [IDSFO-102]
Each core component of the Industrial Data Space, especially each Connector, must undergo a certification process executed by the Certification Body.