WHITE PAPER

INDUSTRIAL DATA SPACE

DIGITAL SOVEREIGNITY OVER DATA

AUTHORS

Prof. Dr.-Ing. Boris Otto
Fraunhofer Institute for
Material Flow and Logistics IML
Joseph-von-Fraunhofer-Str. 2-4
44227 Dortmund, Germany

Prof. Dr. Sören Auer
Fraunhofer Institute for Intelligent
Analysis and Information Systems IAIS
Schloss Birlinghoven
53757 Sankt Augustin, Germany

Jan Cirullies
Fraunhofer Institute for
Material Flow and Logistics IML
Joseph-von-Fraunhofer-Str. 2-4
44227 Dortmund, Germany

Prof. Dr. Jan Jürjens
Fraunhofer Institute for
Software and Systems Engineering ISST
Emil-Figge-Str. 91
44227 Dortmund, Germany

Nadja Menz
Fraunhofer Institute for Open
Communication Systems FOKUS
Kaiserin-Augusta-Allee 31
10589 Berlin, Germany

Jochen Schon
Fraunhofer Institute for Intelligent
Analysis and Information Systems IAIS
Schloss Birlinghoven
53757 Sankt Augustin, Germany

Dr. Sven Wenzel
Fraunhofer Institute for
Software and Systems Engineering ISST
Emil-Figge-Str. 91
44227 Dortmund, Germany

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: Jan Cirullies
Design: Fraunhofer-Gesellschaft
Typesetting and page layout:
www.Ansichtssache.de

The original version of this paper is
available at www.industrialdataspace.org

SPONSORED BY THE
Federal Ministry
of Education
and Research

DLR Project Management Agency

INDUSTRIAL
DATA SPACE

Grant ID 01IS15054

THIS WHITE PAPER GIVES AN OVERVIEW ON AIMS AND ARCHITECTURE OF THE »INDUSTRIAL DATA SPACE«. ADDITIONALLY, SOME USE CASE AND THE INDUSTRIAL DATA SPACE USER ASSOCIATION ARE INTRODUCED.

TABLE OF CONTENTS

SUMMARY

The »Industrial Data Space« is a virtual data space using standards and common governance models to facilitate the secure exchange and easy linkage of data in business ecosystems. It thereby provides a basis for creating and using smart services and innovative business processes, while at the same time ensuring digital sovereignty of data owners.

The Industrial Data Space initiative was launched in Germany at the end of 2014 by representatives from business, politics, and research. Meanwhile, it is an explicit goal of the initiative to take both the development and use of the platform to a European/global level.

The Industrial Data Space comes as an initiative that is organized in two branches: a research project and a non-profit user association. The research project is funded by the German Federal Ministry of Education and Research (BMBF). It is of pre-competitive nature and aims at the development and pilot testing of a reference architecture model of the Industrial Data Space. The work of the research project is tightly connected with the activities of the user association named »Industrial Data Space e.V.«. The main goal of the user association is to identify, analyze and evaluate the requirements of user companies to be met by the Industrial Data Space. Furthermore, the user association contributes to the development of the reference architecture model and promotes its standardization.

THE MOST IMPORTANT USER REQUIREMENTS TO BE MET BY THE REFERENCE ARCHITECTURE MODEL ARE:

- **Data sovereignty:** It is always the data owner that specifies the terms and conditions of use of the data provided (terms and conditions can simply be »attached« to the respective data).
- **Decentral data management:** Data management remains with the respective data owner, if desired.
- **Data economy:** Data is viewed as an economic asset. It can be distinguished into three categories: private data, socalled »club data« (i.e. data belonging to a specific value creation chain, which is available to selected companies only), and public data (weather information, traffic information, geo data etc.).
- **Value creation:** The Industrial Data Space facilitates the creation and use of smart services and digital business models.
- **Easy linkage of data:** Linked-data concepts and common vocabularies facilitate the integration of data between participants.
- **Trust:** All participants, data sources, and data services of the Industrial Data Space are certified against commonly defined rules.
- **Secure data supply chain:** Data exchange is secure across the entire data supply chain, i.e. from data creation to data capture to data usage.
- **Data governance:** Participants jointly decide on data management processes as well as on applicable rights and duties.

THE REFERENCE ARCHITECTURE MODEL CONSISTS OF FOUR ARCHITECTURES:

- The **business architecture** addresses questions regarding the economic value of data, the quality of data, applicable rights and duties (data governance), and data management processes.
- The **security architecture** addresses questions concerning secure execution of application software, secure transfer of data, and prevention of data misuse.
- The **data and service architecture** specifies (in an application and technology independent form) the functionality of the Industrial Data Space, especially the functionality of the data services, on the basis of existing standards (vocabularies, semantic standards etc.).
- The **software architecture** specifies the software components required for pilot testing of the Industrial Data Space. Existing technologies are being used as far as possible.

The reference architecture model thereby serves as a blueprint for different implementations of the Industrial Data Space. Both the research project and the user association are eager to get in touch with similar projects and initiatives. A cooperation has already been established with working groups of the »Plattform Industrie 4.0« project. The activities of the research project build upon the results of previous research projects (OMM, for example) and existing standards (Resource Description Framework, RDF, or reference architecture model »Industrie 4.0«, (RAMI4.0, for example).

1

DIGITIZATION AND THE ROLE OF DATA

1.1 Digitization as a basic trend

The process of digitization currently underway has become the central development in society, businesses, and technology. Smart services provided via mobile applications not just represent technological innovation, but have changed the way humans work and live. While digitization can be seen as a result of other developments, it is at the same time an enabler of these developments. Here are some examples:

- **Globalization:** Globalization is not a new phenomenon. Many companies have long since operated in global markets, allowing them to leverage economies of scale and develop new potentials of growth. During the 1980s and 1990s, globalization basically referred to standardized products, which were traded on the basis of clearly defined supplier-customer relationships. Today, however, globalization is characterized by the existence of complex production and service networks (in industries such as mechanical engineering or automotive, for example) and high information transparency.
- **Mobility:** Customers expect to be provided with (smart) services anytime anywhere. Restrictions or limitations of any kind whatsoever are hardly accepted.
- **Sharing Economy:** Sharing resources has become a growing trend. The basis of this trend has been a shift in the value paradigms of people (particularly those living in the urban areas of highly industrialized countries in the Western hemisphere), in the course of which people value material goods (cars, houses etc.) lower than they used to in the past (mainly because they have managed to satisfy a large proportion of their needs already). Other developments (like crowd sourcing, for example) also give evidence of the change that has taken place with regard to the values relevant in society and for individuals.
- **Privacy:** The current situation regarding privacy issues and data protection is characterized by a certain contradiction: on the one hand, more and more people are using social media, mobile applications, or search engines, knowing they lose sovereignty regarding their private data; on the other hand, consumer protection authorities demand from social media operators to change their privacy policies and strengthen the rights of users.

Together these developments have the potential to fundamentally change entire industries. For example, it needs to be seen to what extent carmakers will preserve their dominant role as original equipment manufacturers (OEMs), or whether these companies will turn into mere suppliers of hardware for mobile service providers.

1.2 Data as the link between the »Smart Service World« and »Industrie 4.0«

The above example also shows that digitization is reshaping the interface between the provider/supplier and the customer. Providers like mytaxi or AirBnB are successful because they are able to support the entire customer process without media disruption. As these companies have no or just little resources and assets, they excel just by coordinating the processes of »suppliers«. As a consequence, the service range in the era of digitization, the so-called »Smart Service World«, is characterized by certain features:

- **Individualization:** Products and services increasingly take into account the personal needs and requirements of individuals and the activities important and relevant for them (work, health, traveling, shopping etc.).
- **End-to-end support:** In the past, products and services served to meet customer demands from the perspective of the supplier/provider. Today, and even more so in the future, products and services must address the entire customer process, and not just random points of interaction between the supplier and the customer. At the same time, end-to-end support must be ensured also between a company and its suppliers and their processes.
- **Hybrid products:** It is not just the clear discrimination between products and services which is increasingly disappearing, but also the discrimination between traditional offers and digital services. Examples such as mytaxi.de or AirBnB.com show that more and more traditional offers (getting from A to B, staying at a hotel etc.) are digitally enriched.
- **Business ecosystems:** To meet customer demands as comprehensively as possible, the collaboration between multiple players is getting increasingly important. Customers have trust in the suppliers/providers and brands the value systems of which show the highest degree of congruency with their own value systems.

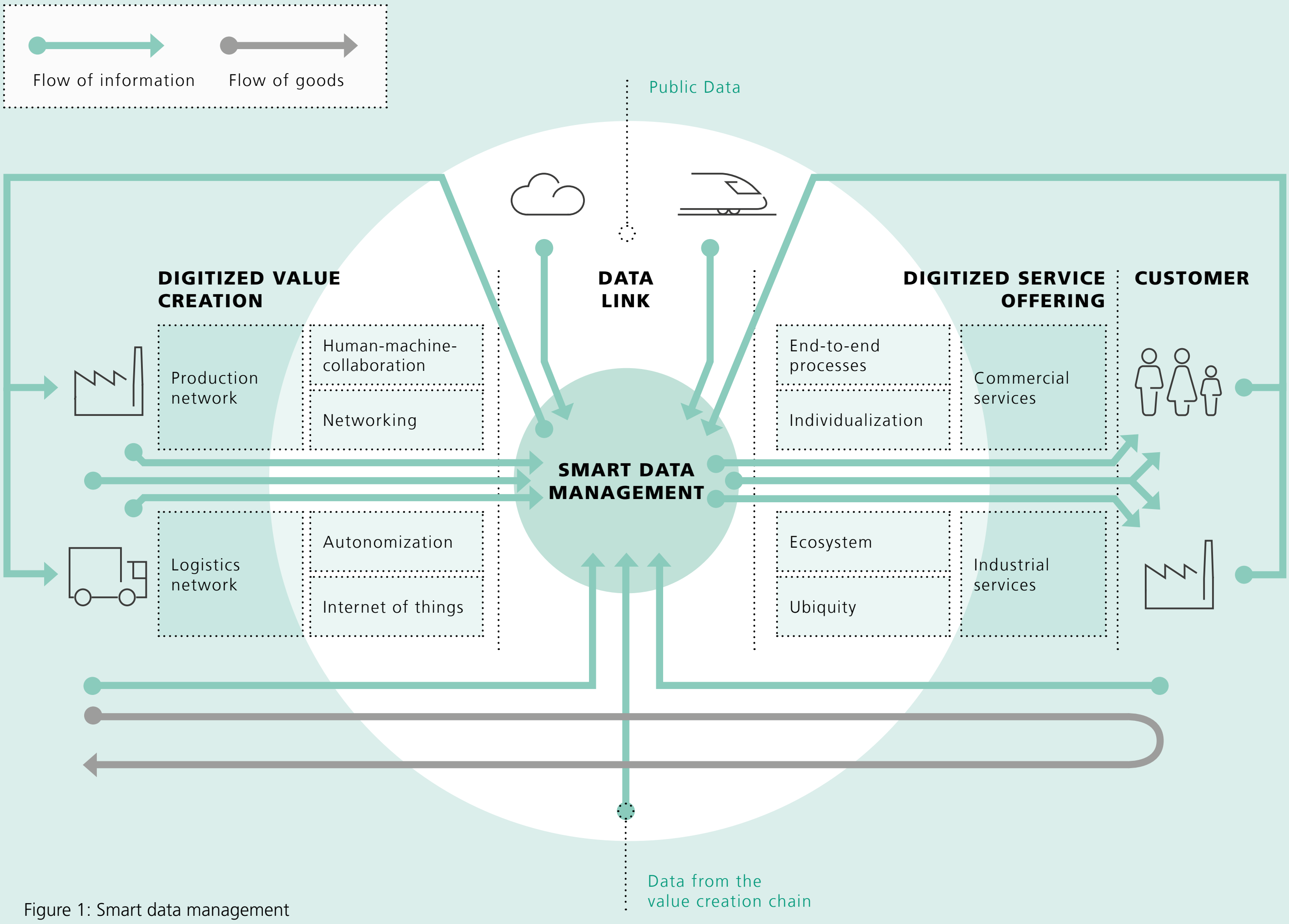


Figure 1: Smart data management

There is one success factor that is critical for products and services to meet customer requirements in the best possible way: data (i.e. customer data, product data etc.). Being able to manage data like any other company asset, in order to create the basis to offer smart services, is becoming more and more important for companies that want to excel in the market. The above mentioned features of these new services in the Smart Service World pose new challenges with regard to the processes required for rendering the services. Especially the increasing individualization of services leads to a growing complexity of production and logistics processes. Carmakers, for example, need to manage 1030 theoretical product variants (for example, the number of variants of single components such as headlights or outside mirrors is 40 and more). Taking into account the ever growing number of product features, ever shorter product lifecycles, shorter delivery times, legal guidelines, and value creation processes getting increasingly globalized, this complexity cannot be managed by traditional organizational principles and management approaches anymore.

In this context, »Industrie 4.0« represents an organizational principle for ambitious manufacturers that is based on four core features:

- **networking** of humans and machines
- **autonomization** of processes and systems
- end-to-end **information transparency**
- decision-making support offered by **assistance systems**

Consequently, data represents the link between industrial manufacturing and smart services. What is needed is a »Smart Data Management«, as shown in Figure 1.
The Industrial Data Space offers an architecture draft to support this new form of data management.

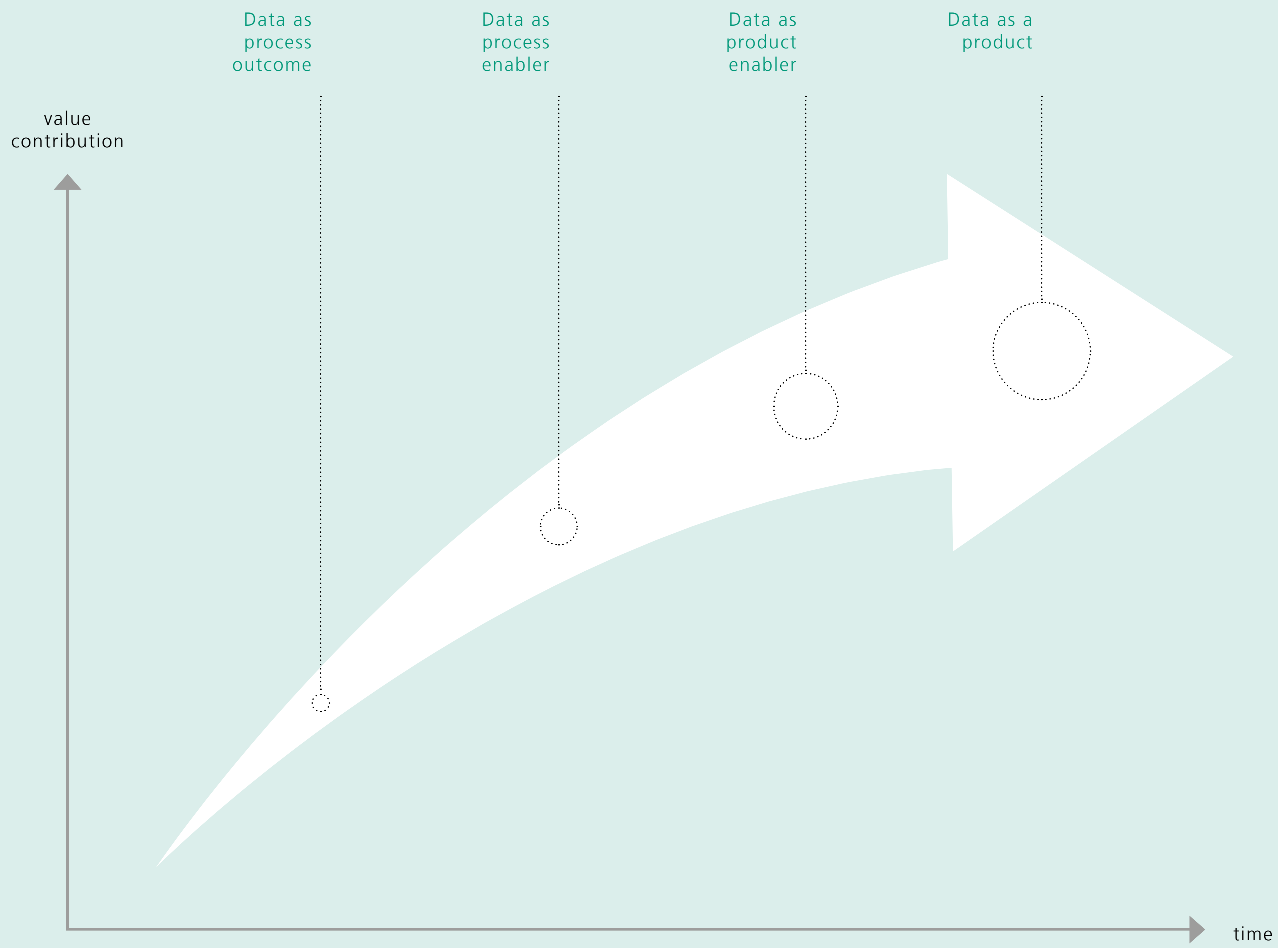


Figure 2: Development of the role of data for the performance of businesses

1.3 Data as an economic asset

The importance of data for businesses to be successful has continuously grown since the upcoming of electronic data processing and the automation of production processes (see Figure 2). Over time, data has played different roles in a company's business processes and overall performance:

- **Data as the result of a process:** In the early times of electronic data processing (the 1960s and 1970s), information systems and data were basically used to support business functions. For example, inventory management systems just served to support warehousing processes at a certain location; to check whether a certain item was in stock, a warehouse manager could make an inquiry in the system instead of walking over to the shelf to find out whether the respective item was still there. In those days, value was created for the enterprise only by the physical product, not by data.
- **Data as an enabler of processes:** With the proliferation of Manufacturing Resource Planning (MRP) and Enterprise Resource Planning (ERP) systems in the 1980s and 1990s, data turned into an enabler of company-wide business process management. Without the existence of consistent data, made available in almost real time, the implementation of standardized processes such as order-to-cash or procure-to-pay on a global (or at least regional) level would not have been possible. During this phase, data became a strategic resource for operational excellence in production, logistics, and customer service.
- **Data as an enabler of products and services:** Since the beginning of the new millennium, companies increasingly offer products and services which require high-quality data. Examples are miCoach by adidas, Hilti's leasing and fleet management models, or all kinds of smart services offered to consumers.
- **Data as a product:** In recent years, data marketplaces have emerged, on which requests for data APIs are billed by volume or time. This way, data is not just an enabler of products anymore, but has become a product itself.

As the role and function of data has changed with regard to a company's business processes and overall performance, so has the value of data. Enterprises increasingly demand methods allowing them to calculate the value of data. Existing approaches for doing so have been adopted from the field of material goods. They can be subdivided into three basic models:

- **Cost of production/purchase:** The value of data is determined by the cost for producing or purchasing it.
- **Use value:** The value of data is determined by its contribution to a company's business processes and overall performance (increase in customer satisfaction, reduced stock-keeping, or more efficient deployment of sales staff in business models including direct sales, for example).
- **Market value:** The value of data is determined by its price when sold in the market.

While all three models are used in practice, comprehensive and broadly accepted instruments are still missing. Furthermore, these models have still not sufficiently been rooted in accounting and auditing practices.

The value of data also depends on its nature. Three categories of data can be distinguished here:

- **Private data** is the property of one enterprise. This enterprise may offer its data to other enterprises (to terms and conditions which the data-owning enterprise may determine).
- **Club data** is made available and can be disposed of by a group of enterprises. These enterprises jointly decide on the management of the data.
- **Public data** is available to any enterprise. It is usually offered by a public authority.

Questions related to the economic valuation of data are addressed by the business sub-architecture of the reference architecture model of the Industrial Data Space.

2

INDUSTRIAL DATA SPACE

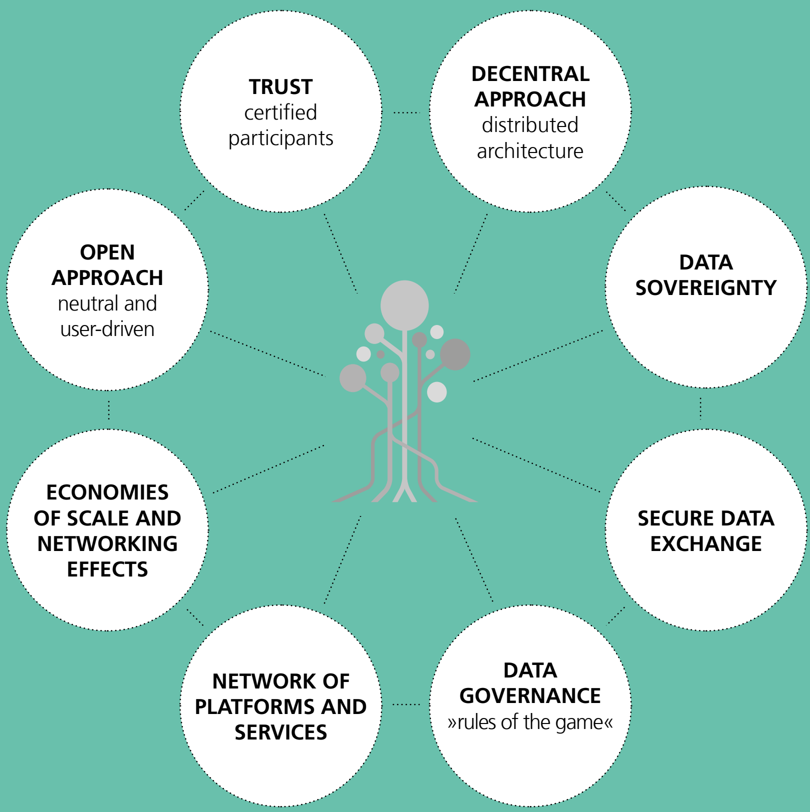


Figure 3: Key elements of the Industrial Data Space

2.1 Key elements

Guided by the demand for digital sovereignty, the Industrial Data Space aims at establishing a »network of trusted data«.

Figure 3 shows the key elements of the Industrial Data Space:

- **Data sovereignty:** It is always the data owner that determines the terms and conditions of use of the data provided (terms and conditions can simply be »attached« to the respective data).
- **Secure data exchange:** A special security concept featuring various levels of protection ensures that data is exchanged securely across the entire data supply chain (and not just in bilateral data exchange).
- **Decentral approach (distributed architecture):** The Industrial Data Space is constituted by the total of all end points connected to the Space via the Industrial Data Space Connector. This means that there is no central authority in charge of data management or supervision of adherence to data governance principles. In this respect, the Industrial Data Space represents an alternative architecture that is different from both centralized data management concepts (like so-called »data lakes«, for example) and decentralized data networks (which usually have no generally applicable »rules of the game«). What architecture will be used in the end depends on how beneficial each architecture turns out to be in economic terms for each individual application scenario. This is why the Industrial Data Space initiative presumes various coexisting architectures from the outset.
- **Data governance (»rules of the game«):** As the Industrial Data Space comes with a distributed architecture, and therefore has no central supervisory authority, data governance principles are commonly developed as »rules of the game«. These rules are derived from the requirements of the users and determine the rights and duties required for data management.
- **Network of platforms and services:** Providers of data can be individual enterprises, but also »things« (i.e. single entities within the »internet of things«, such as cars, machines, or operating resources) or individuals. Other Data Providers may be data platforms or data marketplaces currently being established in various industries. Furthermore, data services of various providers are made available via an »AppStore«.
- **Economies of scale and networking effects:** The Industrial Data Space provides data services for secure exchange and easy linkage of data. It thereby represents an infrastructure, as using the Industrial Data Space will facilitate the development and use of services (smart services, for example). While these services must rely on data services as offered by the Industrial Data Space, they are not an element of the range of services of the Industrial Data Space themselves. This is why economies of scale and networking effects will be critical for the success of the Industrial Data Space: The more participants the Industrial Data Space will have, the more it will become »the place to be« for Data Providers, Data Users, and data service providers alike.

– **Open approach (neutral and user-driven):** The Industrial Data Space is a user-driven initiative. Regarding the reference architecture model, it is based on a participatory development process, with design decisions being made jointly by the research project and the user association.

– **Trust (certified participants):** It is important for all participants in the Industrial Data Space to trust the identity of each Data Provider and Data User. This is why all »end points« may connect to the Industrial Data Space via a certified software (the »Industrial Data Space Connector«) only. The Connector also incorporates authentication and authorization functionality.

In sum, these key elements allow the Industrial Data Space to live up to its role as a link between the Internet of Things and the Smart Service World, while at the same time being capable to leverage economies of scale and follow a distributed, decentralized approach (see Figure 4).

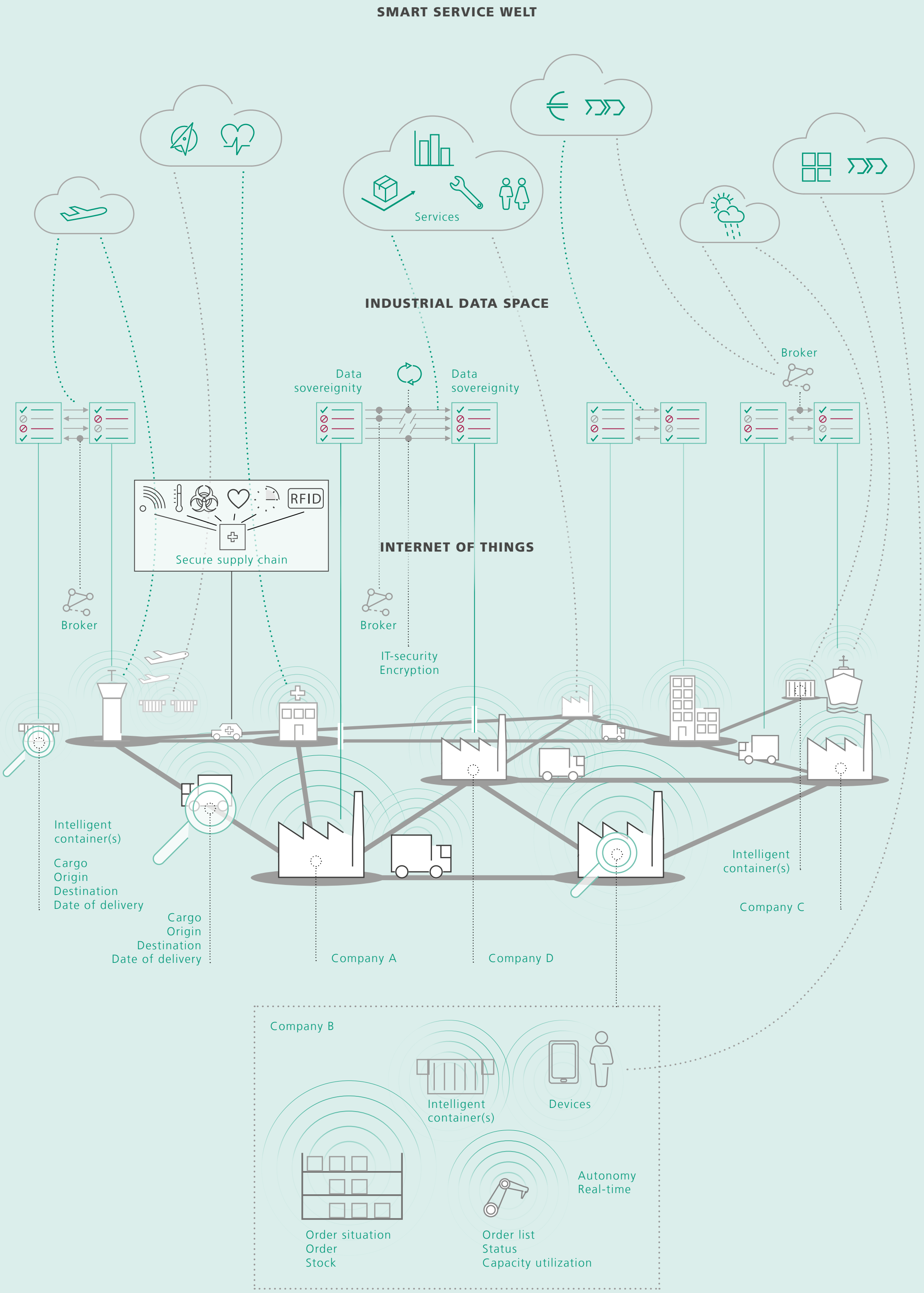


Figure 4: Industrial Data Space overview

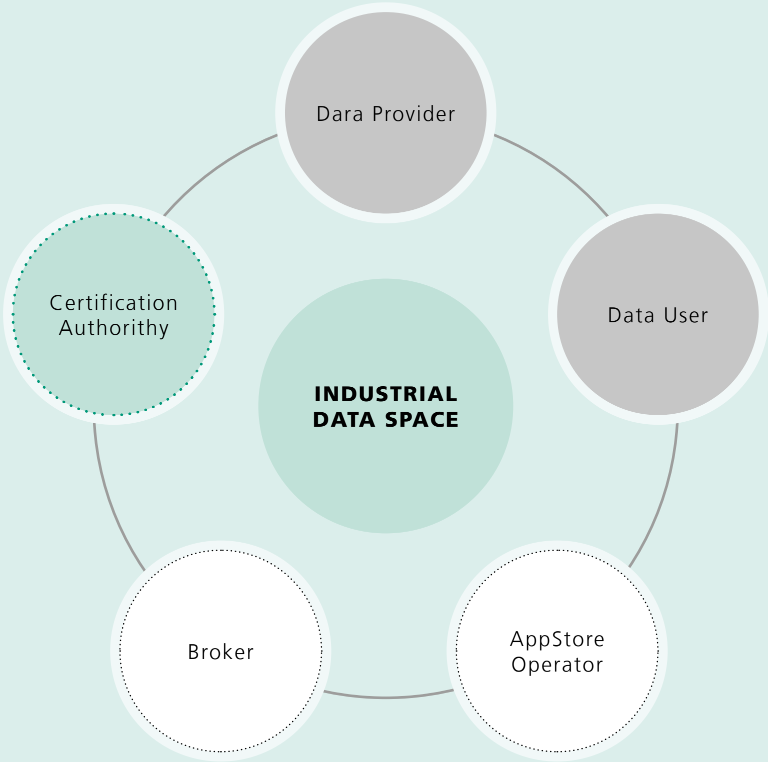


Figure 5: Role Concept

2.2 Role concept

The main goal of the Industrial Data Space is to facilitate the exchange of data between Data Providers and Data Users, which represent two major roles within the Industrial Data Space. However, for this data exchange to be secure, and the linking of data to be based on a simple concept, more roles are required. These roles are the Broker, the AppStore Operator, and the Certification Authority.

Figure 5 shows the five roles and how they are connected to each other within the Industrial Data Space. Each participant of the Industrial Data Space may take one or several roles. Furthermore, participants may appoint third parties for execution of certain activities.

In the following paragraphs, the roles are explained in detail:

2.2.1 Data Provider

A Data Provider possesses data sources and offers data from these sources to be used by other participants in the Industrial Data Space. Data sovereignty always remains with the respective Data Provider. In more detail, a Data Provider performs the following activities:

- provides descriptions of its data sources to be registered by the Broker for other participants in the Industrial Data Space to retrieve the data;
- preselects data from internal systems to be made available in the Industrial Data Space, processes and integrates data, and transforms it into a target data model; attaches terms and conditions of use to its data;
- makes data available to be requested by certain contractors;
- receives data service apps, vocabularies, schemes, and the Industrial Data Space Connectors over the Industrial Data Space AppStore.

2.2.2 Data User

A Data User receives data from other participants (the Data Providers) in the Industrial Data Space. In more detail, a Data User performs the following activities:

- retrieves data from certain contractors,
- receives data service apps, vocabularies, schemes, and the Industrial Data Space Connectors over the Industrial Data Space AppStore,
- preselects data from various sources (i.e. from different Data Providers), processes and integrates data, and transforms it into a target data model.

2.2.3 Broker

A Broker acts as a mediator between Data Providers offering data and Data Users requesting data. It also acts as a data source registry. In more detail, a Broker performs the following activities:

- provides Data Providers with functions to publish their data sources,
- provides Data Users with functions to search through the data sources of Data Providers,
- provides Data Providers and Data Users with functions to make agreements on the provision and use of certain data.

Furthermore, a Broker acts as a clearing house and supervises the exchange of data (without infringing upon the data sovereignty of the data owners). In more detail, a Broker performs the following activities in its function as a clearing house:

- supervises and records data exchange transactions,
- furnishes reports on the search for data sources and on data exchange transactions,
- supports the rollback of transactions in case of faulty or incomplete data exchange.

If requested by participants, a Broker may offer additional services, such as data quality related services or data analysis services (particularly in the case of large data volumes).

2.2.4 AppStore Operator

The Industrial Data Space promotes the development of a business ecosystem in which participants may develop software (especially data services) and make this software available via the AppStore.
The AppStore Operator performs the following activities:

- provides functions by which software developers may describe data services and make these services available to other participants,
- provides functions by which participants may retrieve and download data services,
- provides functions for payment and rating of data services.

2.2.5 Certification Authority

The Certification Authority makes sure that the software components of the Industrial Data Space meet the requirements jointly defined by the participants and rules and standards are observed. In more detail, the Certification Authority performs the following activities:

- supervises each certification procedure from the beginning (request for certification) until the end (approval/refusal of certification),
- approves reports made by test bodies,
- issues notices of approval/refusal of certification,
- issues certificates,
- ensures comparability of evaluations,
- maintains a catalog of criteria and (if need be) protection classes.

The Certification Authority collaborates closely with test bodies and accreditation bodies.

3

REFERENCE ARCHITECTURE MODEL OF THE INDUSTRIAL DATA SPACE

The reference architecture model of the Industrial Data Space consists of four architectures:

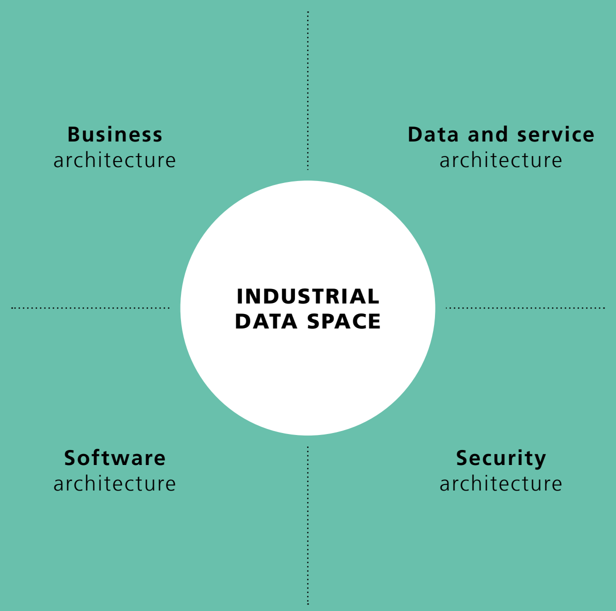


Figure 6: Architectures of the reference architecture model of the Industrial Data Space

Data Governance	Data source	Privat	Public	Common	
	Data stewardship	By Data Provider	By Data User	By broker	None
	Data use	Unlimited	Limited		
	Data good	Private data	Public data	Club data	
Collaborative data management	Data request	By subscription	On Demand		
	Identity of Data Provider	Visible	Invisible	Via broker	
	Data quality	Guaranteed by Data Provider	Rated by crowd	Rated by broker	Unrated
Business model	Access	Certified			
	Use model	Prosumer	Data Provider	Data User	
	Preismodell	Flat-Rate	Pay-per-Use		
	Pricing model	Data User	Intermediary	Sponsoring	

Figure 7: Design options within the business architecture

3.1 Business architecture

The business architecture comprises all concepts critical for the Industrial Data Space to be successful in economic terms. These concepts can be subdivided into three categories:

– **data governance:** rights and duties of the different roles within the Industrial Data Space;
– **collaborative data management:** inter-organizational processes for data management (publication, data use etc.);
– **business model:** evaluation of data, compensation for data use, terms and conditions of data use in additional services (smart services).

Each category offers a number of design options allowing flexible configuration of the business architecture for different usage scenarios. Figure 7 shows possible design options in the form of a morphological field. Here are three examples:

– Regarding data governance, the organizational establishment of data quality management (so-called »data stewardship«) can be implemented differently: responsibility for data stewardship can remain with the Data Provider (which seems to be a reasonable option in most cases, as the Data Provider usually knows best about the correctness of its data), the Data User, or the Broker.
– Regarding collaborative data management, how data may be requested can basically be organized by two different options: either by subscription or on demand.
– Regarding possible business models, two options for pricing are flat rate and pay-per-use.

The Industrial Data Space research project is implementing certain options as shown in Figure 7 in selected application scenarios.

Industrial Data Space AppStore	Basic Data Services Provisioning	Data Service Management and Use	Vocabulary Management	Software Curation
	Data Provenance Reporting Data Transformation Data Curation Data Anonymization	Data Service Publication Data Service Search Data Service Request Data Service Subscription	Vocabulary Creation Collaborative Vocabulary Maintenance Vocabulary/Schema Matching Knowledge Database Management	Software Quality and Security Testing

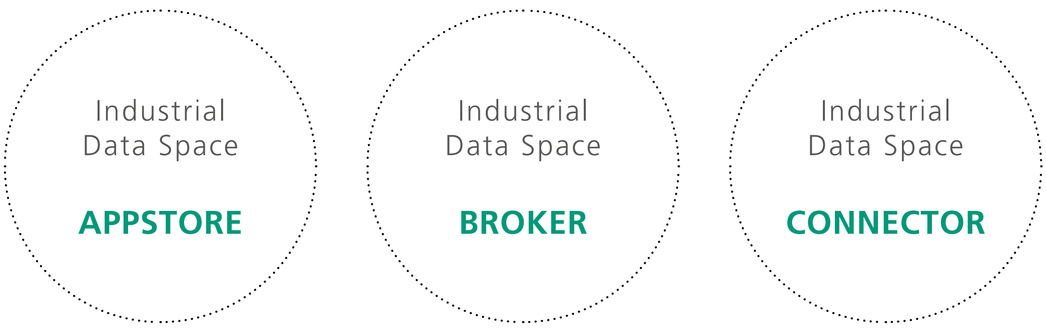
Industrial Data Space Broker	Data Source Management	Data Source Search	Data Exchange Agreement	Data Exchange Monitoring
	Data Source Publication Data Source Maintenance Version Controlling	Key Word Search Taxonomy Search Multi-criteria Search	»One Click« Agreement Data Source Subscription	Transaction Accounting Data Exchange Cleaning Data Usage Reporting

Industrial Data Space Connector	Data Exchange Execution	Data Preprocessing Software Injection	Remote Software Execution
	Data Request from Certified Endpoint Usage Information Maintenance (Expiration etc.) Data Mapping (from Source to Target Schema) Secure Data Transmission between Trusted Endpoints	Preprocessing Software Deployment and Execution at Trusted Endpoint	Data Compliance Monitoring (Usage Restriction etc.) Remote Attestation Endpoint Authentication

Figure 8: Data and service architecture

3.2 Data and service architecture

The data and service architecture constitutes the functional core of the Industrial Data Space. It specifies the functions to be implemented in the pilot applications. The data and service architecture does not however make decisions on the use of certain technologies or applications. The functions are arranged in eleven blocks, which in turn are assigned to one of the following functional components:



The Industrial Data Space AppStore comprises the following functional blocks:

- **Basic Data Service Provision:** provides basic services for Data Users and Data Providers; among them are services for transformation of data from a source scheme into a target scheme, traceability of data, or data anonymization.
- **Data Service Management and Use:** supports publication, search, and use of services; these functions can be compared to the AppStore functionality that can be found in the consumer market (Apple's AppStore, for example).
- **Vocabulary Management:** supports the joint management and maintenance of vocabularies.
- **Software Curation:** provides functions for data quality management and data service improvement; can be requested via the AppStore.

The Industrial Data Space Broker comprises the following functional blocks:

- **Data Source Management:** supports publication, maintenance, and version control of data sources.
- **Data Source Search:** supports the search for data sources (with the help of taxonomies, by free-text search, or by multi-criteria search).
- **Data Exchange Agreement:** supports the contractual agreement between Data Providers and Data Users regarding the exchange and use of data.
- **Data Exchange Monitoring:** supports the clearing process (transaction rollback, for example); reports on the usage of data sources.

The Industrial Data Space Connector comprises the following functional blocks:

- **Data Exchange Execution:** supports the entire data exchange process (from searching for certain data sources to maintenance of the terms and conditions of use on the part of the Data Provider to the provision of data).
- **Data Preprocessing Software Injection:** supports the provision and use of data preprocessing routines in a safe execution environment.
- **Remote Software Execution:** supports remote monitoring of the execution of software functionality and, in doing so, of adherence with data security provisions (to a predefined extent).

The functional blocks are shown as a part of the functional map of the data and service architecture in Figure 8.
The functional map offers support to both users and software suppliers of the Industrial Data Space with regard to a number of activities:

- **development plan:** functions can be aggregated in different versions of the implementation of the Industrial Data Space.
- **implementation plan:** functions can be implemented by means of different technologies (under consideration of existing applications), which may then be depicted in the functional map in different colors.
- **comparison of software suppliers:** participants in the Industrial Data Space may map the service offers of different software suppliers on the functional map in order to compare these offers.

The Industrial Data Space research project is developing initial versions of the data and service architecture. The architecture will then be maintained and developed further by the Industrial Data Space user association.

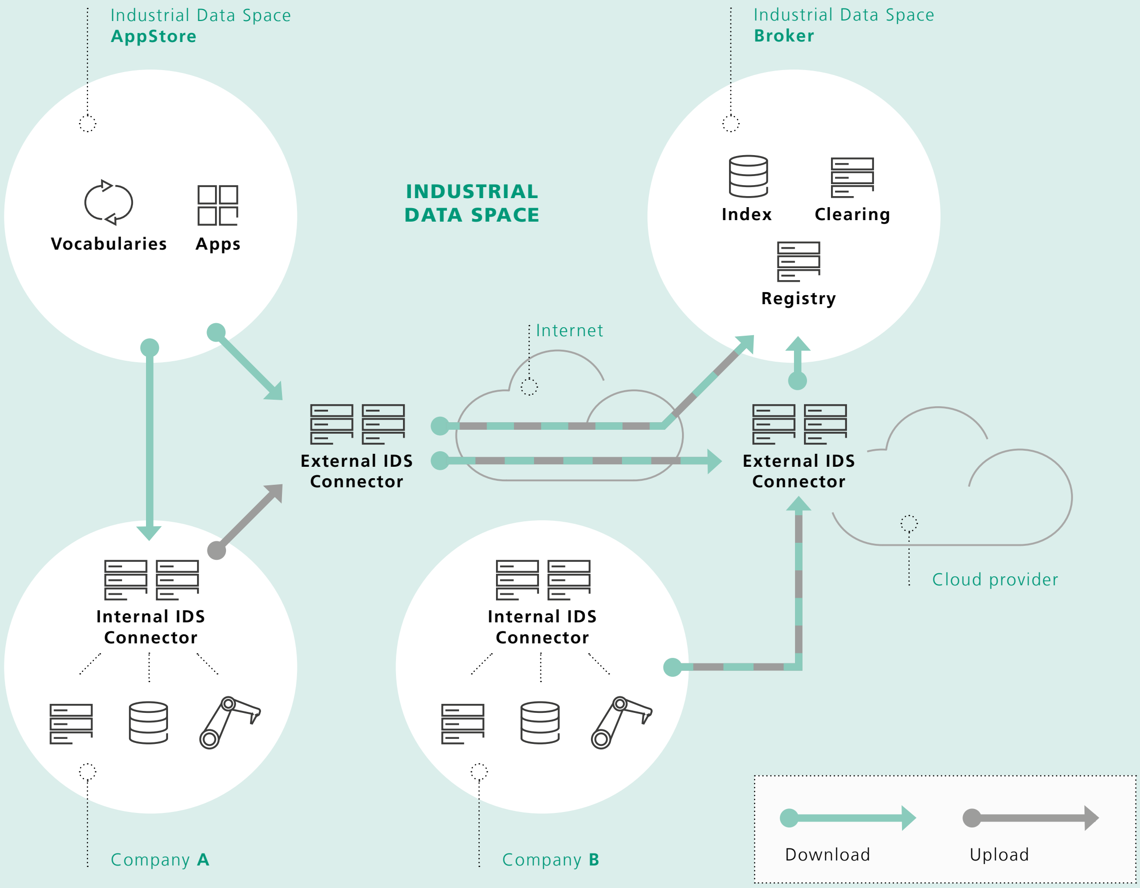


Figure 9: Software components

3.3 Software architecture

The software architecture specifies the implementation of the data and service architecture in the pilot applications of the Industrial Data Space research project. Figure 9 shows the software components to be implemented. A central software component is the Industrial Data Space Connector, which is actually implemented as two components: the »External Industrial Data Space Connector« and the »Internal Industrial Data Space Connector«.

3.3.1 External Industrial Data Space Connector

The External Industrial Data Space Connector (EXIC) facilitates the exchange of data between the participants in the Industrial Data Space. A single EXIC can be understood as an end point of the Industrial Data Space (i.e. the Industrial Data Space is constituted by the total of all EXICs). This means that a central authority for data management is not required. Typically, an EXIC can be operated in a secure environment (beyond a firewall, for example); this means that internal systems cannot be directly accessed. However, EXICs can also be connected to a machine, a car, or a transportation vehicle, for example. Basically it is possible for each company participating in the Industrial Data Space to use several EXICs. Another possibility is that intermediaries (data trustee services, for example) operate EXICs on behalf of one or several companies.

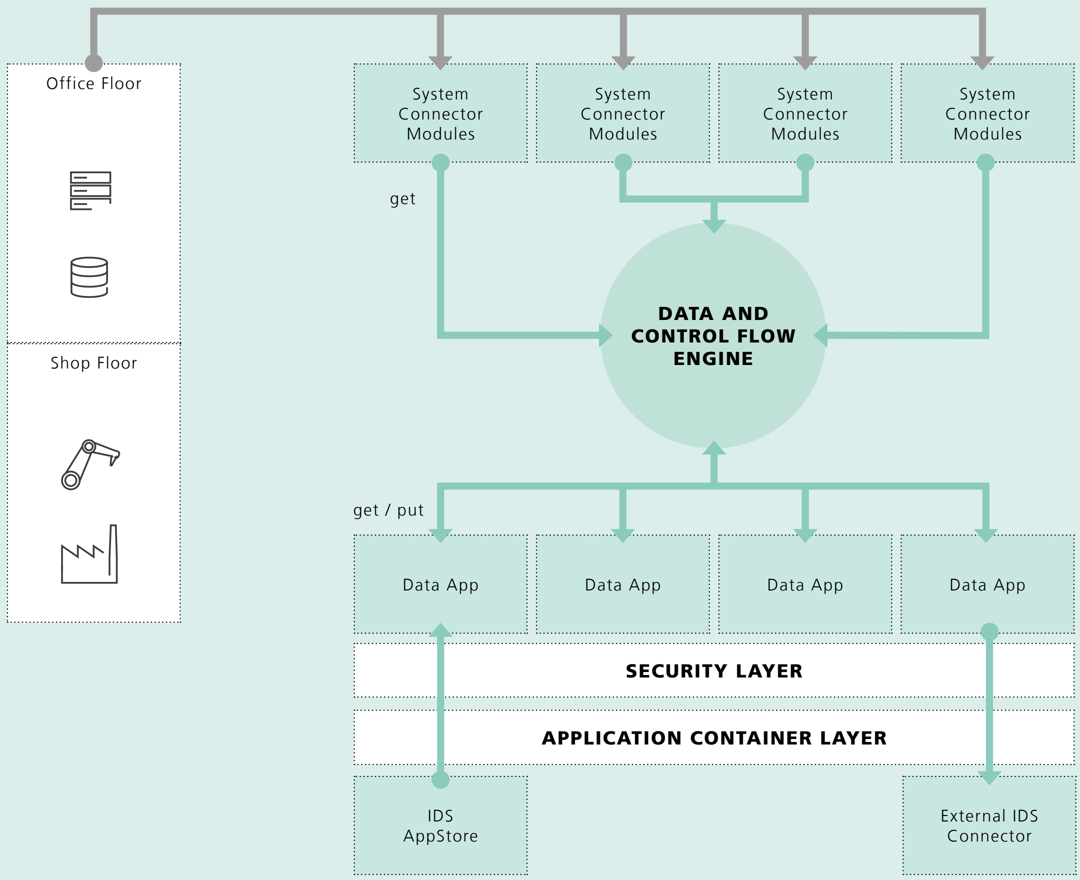


Figure 10: Architecture of the Internal Industrial Data Space Connector

3.3.2 Internal Industrial Data Space Connector

In terms of structure and functionality, the Internal Industrial Data Space Connector (INIC) is very similar to the EXIC. However, an INIC is typically operated within a protected enterprise network. INICs have access to internal data sources and make data from there available to EXICs (see Figure 10).
The connector architecture basically uses technologies for application container management, in order to ensure a safe execution environment for the connector functionality.
For reasons of performance and to simplify communication, data intensive evaluation and analysis operations should take place as closely to the respective data source as possible. Due to safety requirements or limitation of resources, it may be necessary to execute certain data operations on other EXIC instances (a cloud provider, for example). Therefore the Industrial Data Space must allow for flexible distribution of data operations on various INIC and EXIC instances

3.3.3 Industrial Data Space Broker and Industrial Data Space AppStore

The software components of the Industrial Data Space Broker bring together data offers and data requests, execute clearing functions, and create reports on the use of data sources.
The Industrial Data Space AppStore provides data services and supports the joint creation and maintenance of vocabularies.

3.4 Security architecture

The security architecture of the Industrial Data Space comprises a number of aspects. The combination of several security aspects which can individually be designed allows to implement different levels of security.

3.4.1 Network security

Communication between participants in the Industrial Data Space is protected against manipulation and tapping. All connections are encrypted, and end points must provide authentication, making »spoofing« (i.e. misuse of another identity) practically impossible.

3.4.2 Proof of identity

For reasons of accounting, network security, and data access control, participants in the Industrial Data Space must always be unambiguously identifiable. Each participant is described by means of attributes (i.e. identity information). Furthermore, participants may deposit a certain, verifiable »state of security« or a certain »reputational value«; this way approval for accessing certain data may be given not only on the basis of a user name and role, but by considering additional security aspects as well.

3.4.3 Data use restrictions

In order to get access to the data of a Data Provider, a Data User must take into account certain requirements. For example, it may be necessary that the Data User pays a certain amount of money before being allowed to use the requested data, or the Data User must confirm to adhere to certain minimum standards in terms of data protection. Furthermore, a Data Provider may specify a maximum period of time during which its data may be used, prohibit that its data be passed on to other users, or restrict data access to certain requests or levels of aggregation only. The modules for controlling data use are an elementary part of the Industrial Data Space Connector, allowing Data Providers to specify data use rights and levels of security as they deem appropriate.

3.4.4 Secure execution environment

The Industrial Data Space provides different levels of security. While it is basically possible to implement Industrial Data Space Connectors on unsafe platforms, it must be clear that in such cases certain basic characteristics of the Industrial Data Space – such as correct accounting, confidentiality of data, or correct data processing – cannot be guaranteed. By providing a secure execution environment for the Industrial Data Space Connectors, a much higher level of security can be provided. So the Industrial Data Space offers execution environments on different levels of security, which on the one hand presuppose higher security requirements, but on the other hand allow to benefit from extended functionality and get access to sensitive data.

The basic functionality of the security architecture is implemented on each level of security (i.e. these functions cannot be deactivated). Further reaching functionality depends on the hardware and configuration used. For example, certain features require a hardware trust anchor (Trusted Platform Module (TPM), for example). Figure 11 shows the highest possible level of security to be implemented in the software architecture, which allows, for example, trusted data processing on external Connectors.

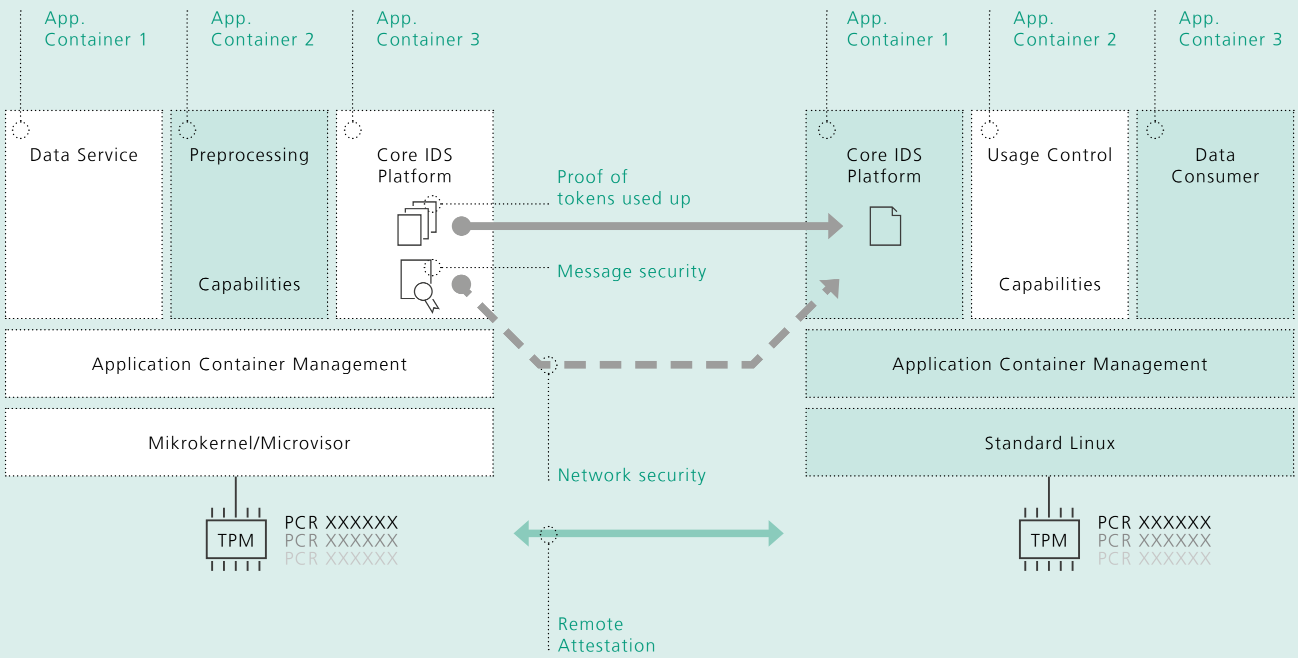


Figure 11: Security architecture of the Industrial Data Space

3.4.5 Remote attestation

The execution environments of a Connector are able to attest that two communication partners act within a known, trustworthy state (by TPM, for example). This way a Data Provider can be sure that a certain Data User has been certified by an Industrial Data Space Connector. If this is the case, the Data Provider may define individual terms and conditions for data use (for example, a maximum period of time during which certain data may be used in connection with deadlines for deleting personal data).

3.4.6 Application layer virtualization

A central element of the secure execution environment of the Industrial Data Space is virtualization on the application layer, allowing to implement individual functions in separate application containers. Depending on the security level of the underlying execution environment, an application container can be protected against unwanted access on the part of the platform operator, allowing a participant in the Industrial Data Space to extend its own trust domain to platforms other participants are on. An example could be to outsource data processing activities to a cloud instance of the Industrial Data Space Connector; depending on the security level of the Connector, participants could load their evaluation algorithms and data onto such thirdparty platforms while still being protected against unwanted access on the part of the respective platform operator (see participant on the left in Figure 11).

4

SELECTED APPLICATION CASES OF THE INDUSTRIAL DATA SPACE

The »Industrial Data Space« allows secure exchange and easy linkage of data in business ecosystems. Typical application scenarios of the Industrial Data Space are characterized by the following features:

- linking of data from several data sources,
- integration of data of different classes (master data and production status data, for example),
- combination of different categories of data (private data, club data, public data),
- participation of at least two companies,
- integration of more than two enterprise architecture levels (shop floor and office floor, for example),
- provision of »smart services«.

The activities of the Industrial Data Space research project are being conducted in close collaboration with user companies (already over 70 applications for taking part in the project have been submitted so far).

4.1 Truck and cargo management in inbound logistics

In many supply chains, data is stored redundantly by several companies. At the same time, data from individual stages of the supply chain is not available on other stages, leading to increased delivery times, safety stocks, and process costs. What is needed is increased supply chain transparency, allowing tracking of products, improved transportation services, and improved forecasting regarding order quantities and production quantities.

A frequent problem in inbound logistics is truck and cargo management. Here it is critical that truck data and cargo data is available at the time of arrival, as a number of parallel and subsequent activities (check-in of trucks, assignment of dock doors and personnel for cargo discharge, job order planning in production etc.) are dependent on this data. However, data on the arrival time of trucks often lacks completeness and correctness, as shipping companies use different types of freight carriers using different routes (hub-and-spoke concept).

The Industrial Data Space allows standardization and simplification of the exchange of data by making data of different classes and from different sources (i.e. order data, transport data, customer master data, supplier master data, product master data, plus additional data such as traffic information or truck GPS data) transparent and available to all companies across the supply chain.

Table 1 gives an overview of the basic elements of this application case.

Participants	Customers Suppliers Logistics service providers, carriers
Data affected	Supplier master data Customer master data Order and transport data Material master data Truck GPS data Traffic information
Business processes affected	Dynamic time window management Staff deployment planning Supply chain risk management Customer relationship management
Data sources involved	ERP systems Transport management systems GPS Web services (providing traffic information)

Table 1: Application case »Truck management in inbound logistics«

4.2 Development of medical and pharmaceutical products

As medical and clinical data is both highly sensitive and heterogeneous, such data usually is centrally gathered in just a few places (i.e. in »maximum-care« hospitals). This lack of data integration is one of the main reasons impeding the development, efficacy, and tolerability of new therapies. To conduct medical studies and assess new therapies, not just clinical data (genetics, therapy, diagnosis) and patient master data needs to be taken into account, but also context data present in highly diverse IT systems and in highly different structures and quality.
The Industrial Data Space allows aggregation of data from different sources, as well as transformation of this data for the purpose of further analysis. While this new way of combining heterogeneous data sources will accelerate clinical studies and promote the exchange of study results, it will also facilitate the review and evaluation of hypotheses and study results published in medical journals. The open interface of the Industrial Data Space allows seamless integration of existing systems for offering services for systematic data processing, as well as visualization of raw data and analysis results. For anonymization of personal medical data, and to ensure that access to such sensitive data is in compliance with data protection and privacy laws, special functions and services of the Industrial Data Space are being applied.
Table 2 gives an overview of the basic elements of this application case.

Participants	Health service providers Pharmaceutical companies and institutes Research centers Insurance companies Medical device manufacturers
Data affected	Personal medical data Clinical study data Epidemiological data Market data Environmental data
Business processes affected	Research and development Production and service Customer relationship management
Data sources involved	Medical data exchange platforms (»Elektronische FallAkte (EFA)«, for example) Management systems in hospitals and doctor's offices Medical engineering systems Data analysis systems Data warehouse systems

Table 2: Application case »Development of medical and pharmaceutical products«

4.3 Collaborative production facility management

Running and maintaining state-of-the-art production facilities requires up-to-date and complete data on the properties of individual machines and components, as well as status data on the utilization of facilities (i.e. from manufacturing processes). Many facility operators are facing high expenses for purchasing, using and analyzing such data, which is mainly due to limited availability of data and information concerning relevant machine status and manufacturing process parameters.
While standards such as OPC-UA are capable of efficiently integrating a number of facilities, there are still information barriers between diverse IT systems and platforms. With the ongoing advancement of the »industrial internet of things« (IIoT) into production processes, the situation will become more aggravated, as day-to-day IT processes additionally are characterized by substantial time differences between planning and operation, and by the inter-dependence of multiple, dynamically changing contingency factors, such as availability and wear-and-tear of production means. Problems typically occur when data (machine status data required by maintenance staff, or up-to-the-minute information about ongoing job orders, for example) needs to be securely exchanged across company boundaries.
The Industrial Data Space facilitates and simplifies cross-company exchange of facility data and product data, both between the manufacturers and the operators of facilities and across entire supply chains. The initiative comes as an alternative to existing approaches of cross-company data exchange, which usually lack interoperability and transferability. Furthermore, companies currently have no standardized tools to control the information flow. The Industrial Data Space has the potential to function as such a tool, allowing, for example, service providers to improve and extend their range of services by getting access to facility data previously not accessible (due to technical or confidentiality reasons, for example). Furthermore, manufacturing companies themselves may grant their customers access to certain information, thereby extending their range of products and services as well.
Table 3 gives an overview of the basic elements of this application case.

Participants	Production facility operators Manufacturers of production facilities and related components Maintenance service providers Software manufacturers
Data affected	Production facility master data Production data Contextual information (ambient temperatures etc.)
Business processes affected	Maintenance Production control Production facility management
Data sources involved	Machine control systems Manufacturing execution systems ERP systems

Table 3: Application case »Collaborative production facility management«

4.4 End-to-end monitoring of goods during transportation

Many companies must rely on critical and expensive goods, which may be transported under guarantee of special precautions only, as otherwise they would be damaged or destroyed. Among these goods are, for example, components for the auto industry (windshield wiper systems featuring rain sensors, for example), the pharmaceutical industry, or the chemical industry. Unfavorable ambient conditions such as ambient temperature being too high or too low, humidity, shock, vibration, light, air pressure, acoustic waves, or magnetic fields pose a multitude of risks to sensitive goods. These ambient conditions can be monitored during transportation by means of sensors, and the respective data can be transmitted via mobile radio communication. Thereby potential risks can be detected early enough, and appropriate measures for risk reduction can be taken more quickly.

The Industrial Data Space serves as a platform for customers and suppliers allowing end-to-end monitoring of ambient conditions goods are exposed to during transportation. Customers and suppliers are provided with data necessary to be informed at any time as to where certain goods are at a certain moment and in what condition these goods are. In doing so, the Industrial Data Space ensures that companies receive all data required, while at the same time ensuring data sovereignty on the part of the company sending the data.

Table 4 gives an overview of the basic elements of this application case.

Participants	Suppliers and customers Manufacturers of transportation vehicles Logistics service providers
Data affected	Sensor data Transportation order data Product and material data Customer and supplier master data
Business processes affected	Production control Warehouse management Quality management Customer complaint management
Data sources involved	Sensorics Transport management systems ERP systems Dangerous goods management systems

Table 4: Application case »End-to-end monitoring of goods during transportation«

Image source @ Robert Bosch GmbH

5

ORGANIZATION AND STRUCTURE OF THE INDUSTRIAL DATA SPACE INITIATIVE

The Industrial Data Space comes as an initiative that is organized in two branches: a research project and a non-profit association of users. Both the research project and the user association are closely collaborating with similar projects and initiatives, as well as with relevant standardization bodies.

5.1 Industrial Data Space research project

The research project is funded by the German Federal Ministry of Education and Research (BMBF). It basically aims at the pre-competitive establishment of the Industrial Data Space, with the following scientific and technological goals to be accomplished:

- design, specification, and development of a reference architecture model of the Industrial Data Space; the reference architecture model is a conceptual model specifying not just the (software) technical basis of the Industrial Data Space, but also the mechanisms required for data privacy, data governance, collaboration, and control in the process of exchanging data securely;
- prototype implementation of the reference architecture model in selected application cases;
- design and continuous development of a standardization map;
- design of the business model of the Industrial Data Space operator;
- design of the certification concept and the business model of the Industrial Data Space Certification Authority;
- development of a methodology allowing users of the Industrial Data Space to adapt their business strategies in compliance with the new requirements posed by digitization;
- development of recommendations for action for operating the Industrial Data Space;
- identification of new areas of research for the sustainable development and establishment of the Industrial Data Space.

In total, twelve Fraunhofer institutes participate in the project:

- Fraunhofer Institute for Applied and Integrated Security (AISEC), Garching by Munich
- Fraunhofer Institute for Applied Information Technology (FIT), Sankt Augustin
- Fraunhofer Institute for Communication, Information Processing and Ergonomics (FKIE), Wachtberg-Werthhoven
- Fraunhofer Institute for Open Communication Systems (FOKUS), Berlin
- Fraunhofer Institute for Intelligent Analysis and Information Systems (IAIS), Sankt Augustin
- Fraunhofer Institute for Industrial Engineering (IAO), Stuttgart
- Fraunhofer Institute for Experimental Software Engineering (IESE), Kaiserslautern
- Fraunhofer Institute for Material Flow and Logistics (IML), Dortmund
- Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB), Karlsruhe
- Fraunhofer Institute for Manufacturing Engineering and Automation (IPA), Stuttgart
- Fraunhofer Institute for Software and Systems Engineering (ISST), Dortmund
- Fraunhofer Institute for Secure Information Technology (SIT), Darmstadt

The research project was started on October 1, 2015 and has a duration of 36 months.

5.2 Industrial Data Space user association

Industrial Data Space e.V., located in Berlin, is a non-profit user association. The main goal of the association is to identify, analyze and evaluate user requirements to be met by the Industrial Data Space. Furthermore, Industrial Data Space e.V. contributes to the standardization of the project results and conducts public relation and communication activities.
Industrial Data Space e.V. was founded on January 26, 2016 in Berlin. Founding members are:

- Allianz SE
- Atos IT Solutions and Services GmbH
- Bayer HealthCare AG
- Boehringer Ingelheim Pharma GmbH & Co.KG
- Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.
- KOMSA Kommunikation Sachsen AG
- LANCOM Systems GmbH
- PricewaterhouseCoopers AG
- REWE Systems GmbH
- Robert Bosch GmbH
- Salzgitter AG
- Schaeffler AG
- Setlog GmbH
- SICK AG
- thyssenkrupp AG
- TÜV Nord AG
- Volkswagen AG
- ZVEI - Zentralverband Elektrotechnik- und Elektronikindustrie e.V.

Industrial Data Space e.V. is open to participation of researchers and user companies from outside Germany, in order to take both the development and use of the Industrial Data Space to a European/global level.

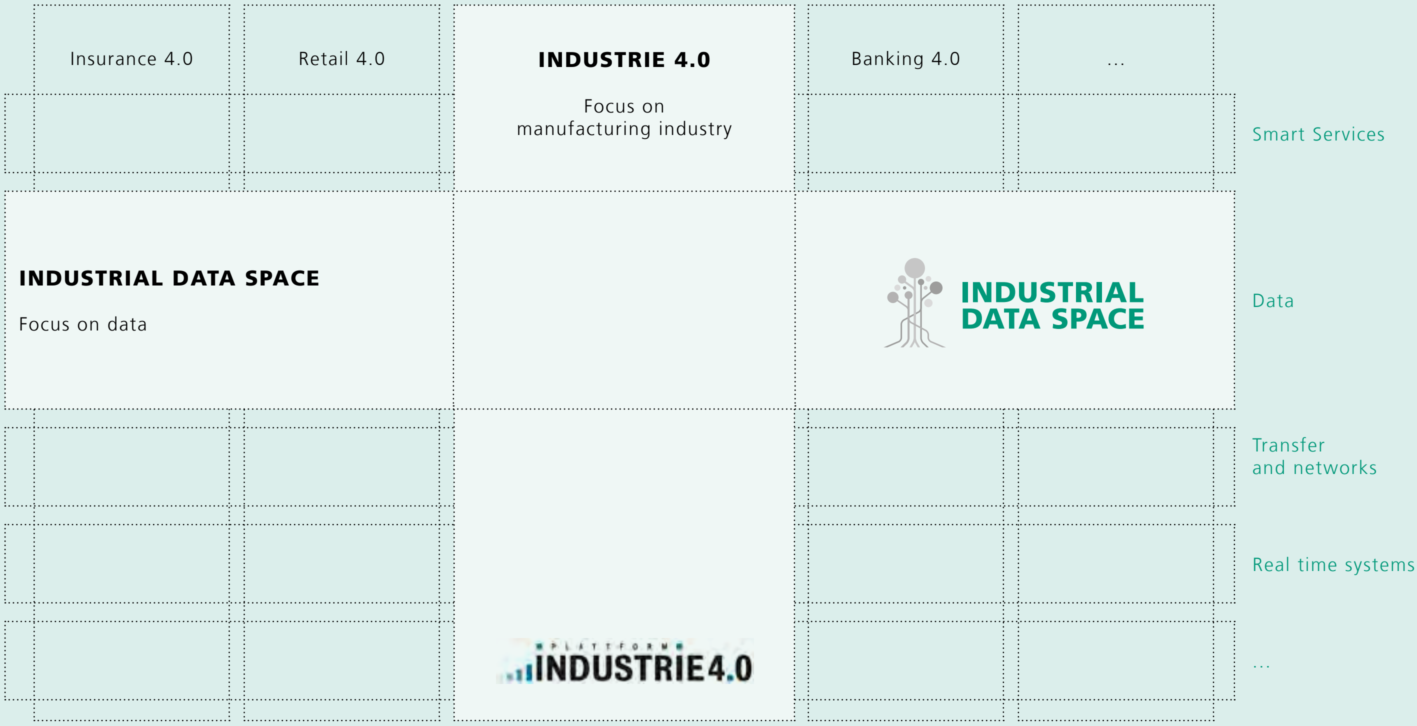


Figure 16: Collaboration between Industrial Data Space and Plattform Industrie 4.0

5.3 Cooperations

Activities for the development and promotion of the Industrial Data Space are being conducted in close collaboration with »Plattform Industrie 4.0« initiative. Whereas the latter is dealing with all aspects of digitization and has its focus on the manufacturing industry, the Industrial Data Space initiative focuses on the data (architecture) level and pursues a cross-industry approach.

Collaboration between the Industrial Data Space initiative and the Plattform Industrie 4.0 project is basically taking place in two working groups of Plattform Industrie 4.0:

- AG 1: Referenzarchitekturen, Standards und Normung, and
- AG 3: Sicherheit vernetzter Systeme.

Both the Industrial Data Space research project and the Industrial Data Space user association are eager to get in touch with similar projects and initiatives, as well as with relevant standardization bodies.

In designing and developing the reference architecture model, the Industrial Data Space research project makes use of existing technologies (dockers for system virtualization, for example) and results of previous research projects (»Theseus«, for example) to the extent possible.

6

OUTLOOK

The activities of the research project and the user association constitute the basis for the design, pilot application, and subsequent promotion and dissemination of the Industrial Data Space. Strategic areas to be addressed by future activities are:

- **Internationalization:** Both the development and use of the Industrial Data Space will be taken to a European/global level. Previous results accomplished in other countries will be integrated into the reference architecture model, as far as deemed appropriate.
- **Standardization:** While the Industrial Data Space takes advantage of existing standards to the extent deemed appropriate, it is also the goal of the initiative to function as a standard for the data economy on its own. Therefore both national and – particularly – international standardization bodies will be addressed.
- **Application scenarios:** The Industrial Data Space comes as an infrastructure providing basic data services. To leverage economies of scale and networking effects, it is critical that these data services be used in as many different application scenarios as possible.
- **Communication, information and training:** To ensure broad dissemination of the reference architecture model of the Industrial Data Space, multiple measures for communication, information, and training will be offered, taking into account the requirements of different industries and companies (in terms of size and level of maturity regarding digitization mainly).

GLOSSARY

AppStore	Part of the Industrial Data Space architecture. Provides apps (e.g. for data aggregation, data processing) that enhance connector functionality and can be operated by 3rd party.
Broker	Role in the Industrial Data Space; acts as a mediator between Data Providers offering data and Data Users requesting data, as a data source registry, and as a clearing house and supervisor of data exchange transactions.
Club Good	Data good (cf. Data) that is available - in contrast to public data - for Industrial Data Space participants only.
Connector	Interface for the decentral data exchange via the Industrial Data Space architecture. Internal connectors support data exchange within organizational units. External ones connect participants to the Industrial Data Space, and thus, must be certified.
Data governance	Organizational capability aiming at managing data as an economic asset; defines applicable rights and duties, and provides corresponding methods and tools.
Data space	Architecture model for data integration; characterized by distributed management of data from multiple data sources and by not using a common semantic model.
Data steward	Rolle in data governance; responsible for data quality management.
Data (IT context)	Formalized representation of information; reusable for the purpose of communication and processing.
Data (economic context)	Immaterial asset.
Data service	Software application supporting functions of data management.

Data owner	Legal entity or natural person holding property rights of data.
Data Provider	Role in the Industrial Data Space; offers data to be used by Data Users.
Data User	Role in the Industrial Data Space; uses data provided by Data Providers.
Data quality	Fitness of data for being used to serve a certain purpose.
End point	Participant in the Industrial Data Space; connected by installation of an Industrial Data Space Connector.
Linked data	Totality of data available in the World Wide Web; can be identified via a Uniform Resource Identifier (URI) and retrieved over HTTP; links to other data also via URIs.
Ecosystem	Multilateral form of collaboration and coordination of organizations and individuals having a common goal (oftentimes comprehensive services offers for certain customer groups), thereby leveraging complementary skills and competencies.
OPC Unified Architecture (OPC UA)	Industrial communication protocol for exchanging data between machines; specifies the definitions for data exchange and the semantic description of the data to be exchanged.
RAMI4.0	Reference Architecture Model »Industrie 4.0«; developed by VDE, VDI, and ZVEI for the digitization of industrial value creation chains.
Reference architecture	Template for a class of architectures to be modeled.
Reference architecture model	Conceptual model of a reference architecture.
Trusted Platform Module (TPM)	Chip designed after the TCG specification; adds basic security functionality to computers and similar devices.