FRAUNHOFER INSTITUTE FOR SOFTWARE AND SYSTEMS ENGINEERING ISST

ISST-REPORT

DATA ECOSYSTEMS

CONCEPTUAL FOUNDATIONS, CONSTITUENTS AND RECOMMENDATIONS FOR ACTION

■ ISST-Report

DATA ECOSYSTEMS

CONCEPTUAL FOUNDATIONS, CONSTITUENTS AND RECOMMENDATIONS FOR ACTION

Water, light, plants and animals: the interplay of these factors in a natural ecosystem is an excellent role model for state-of-the-art value chains within economy, as ecosystems are characterized by the fact that not one of the systems members is able to optimize their well-being on their own. Any ecosystem has to come together and act as a whole, in order to gain profits. Thus, ecosystems create a balance between the reciprocal benefits of its members. But what do ecosystems have in common with the digital economy?

The proliferation of digital technologies and artificial intelligence solutions is accelerating business models that are characterized by steadily increasing data traffic. In this sense, a "data ecosystem" is an ecosystem in which data is the strategic resource for the success of the entire system. Making the most of this valuable resource must be the goal of business ecosystems. The Fraunhofer Institute for Software and Systems Engineering ISST is setting standards for a (controllable) data sharing. Its scientists are researching the value of data as well as handling data sovereignly. With many other Fraunhofer institutes, Fraunhofer ISST closely collaborates with partners from industry, research and politics within the initiative "International Data Spaces" to design and implement the framework for sovereign data sharing of the future. Within this endeavor, Fraunhofer ISST focuses on four application sectors: Healthcare, Logistics, Data Business and Automotive. This paper will elaborate on the rising topic of data ecosystems and how organizations can generate more value of existing data assets by using inter-organizational cooperations.

ISST – Series of Reports:

Within the Series of "ISST Reports", the Fraunhofer Institute for Software and Systems Engineering ISST, publishes its white paper. Thematically, it examines trends and technologies in computer sciences and takes up innovative subjects from some of the Institutes research projects. They provide insights into the current state of research concerning "Data Ecosystems", Fraunhofer ISST's main research topic.

AUTHORS

Prof. Dr. Boris Otto
Dominik Lis
Prof. Dr. Jan Jürjens
Dr.-Ing. Jan Cirullies
Sebastian Opriel
Prof. Dr. Falk Howar
Dr. Sven Meister
Markus Spiekermann
Heinrich Pettenpohl
Frederik Möller

ISST-REPORT

ISSN 0943-1624

October 2019

EDITOR

Prof. Dr.-Ing. Boris Otto
Prof. Dr. Jakob Rehof

IMAGE SOURCE

Cover: ©damedias stock.adobe.com

CONTACT

Fraunhofer Institute for Software and Systems Engineering ISST
Emil-Figge-Straße 91
Germany - 44227 Dortmund
info@isst.fraunhofer.de
+49 231 97677-0

Table of Content

Table of Figures

1

DATA ECOSYSTEMS AS DRIVER FOR INNOVATION AND GROWTH

1.1
Lessons Learned from Nature and its Applicability to the Digital World

Natural phenomena are often a source for clever ideas and innovative solutions. Complex interrelationships or techniques that occur in nature give us the right direction for practical applications and new technologies. Observations of the flight behavior and wing span for example led to the generation of important insights in aerodynamics for the aviation industry. We strive to learn from biological ecosystems in order to gain knowledge that can be applied into the digital world. Animals, plants and other influencing factors such as light or water grow and live to participate in a way that leads to the formation of a healthy ecosystem.

But how can the concept of natural ecosystems with a diversity of organisms be applied to business practices?
The exemplary ecosystem ”forest” is an excellent model for complex value creation structures in inter-organizational business ecosystems in the digital world. There is a balance in nature that allows all participants to benefit of interdependencies that occur in the forest. With regard to the digital world, it is required to identify the correlating mechanisms to incentive relationships that benefit all participating organizations while preserving everyone’s interests. The raw material of innovative services is ”data”. It forms the raw material of digitalization and artificial intelligence (AI). In order to fully unfold the potential of AI, data has to be made available. Inter-organizational data sharing is therefore essential for the creation of business ecosystems, in which all participants form common value propositions based on data. This paper describes the novel concept of data ecosystems, its constituents, and how organizations can identify potentials to generate greater value with existing data assets.

1.2
New Forms of Value Creation with Data Sharing

The proliferation of digital technologies and AI accelerates a shift in business models which is characterized by an increasing importance of data as a strategic resource. While traditional business models rest on tangible assets, data is the raw material not only for information and knowledge, but for innovative services and customer experiences. Besides the shift from tangible to ”smart” products and from controlling the physical to orchestrating the data value chain, there is one additional fundamental change in the digitalized economy. Innovation increasingly takes place in ecosystems in which various members such as businesses, research organizations, intermediaries such as electronic marketplaces, governmental agencies, customers, and competitors band together to jointly achieve innovative value offerings.

Ecosystems are characterized by the fact that no one member is capable of creating innovation on its own, but the ecosystem as a whole needs to team up. In other words: Every individual member has to contribute in order to benefit. Ecosystems function in an equilibrium state of mutual benefits for all members.

A Data Ecosystem is characterized by diverse relationships in a network of multiple actors such as organizations, companies, individuals or technical components e.g. machines or software. The actors of the ecosystem engage in data sharing to contribute in pursuing common goals and value propositions. In this context, data represents the strategic resource for the success of the ecosystem.

In a data ecosystem, data is the strategic resource for the success of the whole system as it is understood as a stand-alone asset that will be exchanged and monetarized within the ecosystem. That offers the participating actors new growth opportunities through networking with other participants and acts as a driver for innovative services and customer experience.

Data sharing opens up new opportunities for progress and the formation of cooperations with other companies or actors from which every participant in the data ecosystem benefits. The various activities of the different members in a data ecosystem lead to a complete coverage of the data value chain. This includes the stages of data generation, curation, exchange, storage and analysis as well as the use of the resulting knowledge for comprehensive business decisions. Through the sustainable data exchange the participating actors are able to develop further and to operate value co-creation which leads to new digital value proposition.

1.3
Challenges in Exploiting the Potential of Data Ecosystems

While the emergence of data ecosystems offers new opportunities for the different ecosystem participants, many social, environmental and business challenges have to be addressed in order to pave the way for these opportunities to materialize.
Among the most significant challenges are:

- Trust: New methods are needed to increase trust in data sharing so that more data would be available for new applications. What is needed is a framework that includes building blocks for data sharing, data management, data protection techniques, privacy-preserving data processing and distributed accountability and traceability. In addition to providing technology for platform developers, the framework should provide incentive and threat modelling tools for data sharing business developers and strategists, who consider opening data for new cooperation and business.

- Data Sovereignty: The framework should also support compatibility with the latest and emerging legislation, like the EU's General Data Protection Regulation (GDPR) and free flow of non-personal data, as well as ethical principles, like IEEE Ethically Aligned Design. This will increase trust in industrial and personal data platforms, which will enable larger data markets combining currently isolated data silos and increase the number of data providers and users in the markets. The result should aim to be platform-agnostic to be applied in multiple domains with platforms based on different technologies.

- Interoperability: The main objective should be to support a trusted data ecosystem providing easy-to-use privacy mechanisms and solutions that guarantee citizens and business entities can fully manage data sharing and privacy. The challenge is thus to provide a corresponding overall technical architecture which needs to take into account the key reference platforms and technologies to support data sharing, to improve existing solutions and architectures, to define the overall reference architecture, and to design platform-agnostic trusted data sharing building blocks and interoperability.

- Data Governance: Data Ecosystems highly depend on access to data and interactions of actors providing or using data or similar resources such as application programming interfaces (APIs). The role of data governance in these complex networks between organizations is an under-researched field. There is a lack of concepts and mechanisms to mandate responsibilities among participants of a data ecosystem.

It is essential to study inter-organizational mechanisms that allow participative interactions, incentives to influence the dynamics and evolution of the ecosystem.

- **Compliance with Antitrust Legislation:** To avoid the risk of data monopolies, the following needs to be ensured:
  - Improving the mobility of non-personal data across borders in the single market, which is limited today in many member states by localization restrictions or legal uncertainty in the market;
  - Ensuring that the powers of competent authorities to request and receive access to data for regulatory control purposes, such as for inspection and audit, remain unaffected; and
  - Making it easier for professional users of data storage or other processing services to switch service providers and to port data, while not creating an excessive burden on service providers or distorting the market.

- **Data Economics:** Data business, i.e. viewing data as an economic asset, will bring additional motivation for data providers and owners to open up their data for various applications. Personal data is becoming a new economic asset class, a valuable resource for the 21st century that will touch all aspects of society. The rapid development of the Personal Data Service (PDS) market will provide big changes in the way individuals, business and organizations deal with each other, as individuals assert more control over their data or service providers process personal data.

Overcoming these challenges requires a concerted effort between all stakeholder groups.

1.4 The Necessity of an Offense Conduct in Data Management

A defense approach is characterized by minimizing data-related risks with the key objectives of assuring security, privacy, integrity, data quality, regulatory compliance and governance. These efforts are mainly focused on the single organizational entity to provide control mechanisms and optimize internal data activities e.g. extraction, standardization, storage and access. In the era of digitalization and data ecosystems, companies must develop their data management from "Defense" to "Offense".

The offense data approach aims at utilizing data to improve the competitive position and profitability with data. Typical activities in this approach are the optimization of data analytics, modelling and visualization. The offense approach allows for more flexibility in the conduct of data management activities in data intensive environments e.g. customer-focused business functions or real-time analytics (DalleMule and Davenport 2017). This development transcends organizational boundaries, as internal data is increasingly used externally, and vice versa. In this context, organizations face new challenges regarding the management of data because innovative, data-driven business models often require data sharing with other participants.

DEMAND

The research project DEMAND is part of Fraunhofer ISST's project landscape to understand organizations engaging data-driven business models and develop a future-oriented data management approach for effectively managing the resource "data".

2

DATA – FROM BY-PRODUCT TO VALUABLE RESOURCE

2.1 Historical and Conceptual Background

Over time, the role, perception and characteristic of data has changed. Data emerges as a core foundational element in the digital transformation of organizations. In the early times of electronic data processing around 1960 and 1970, information systems and data supported business functions. Data was characterized by being byproduct outcome of the essential processes and functionalities. With the beginning and proliferation of Enterprise Resource Planning (ERP) or Manufacturing Resource Planning (MRP) systems, data functioned as an enabler of standardized business processes. Since the beginning of 2000, companies increasingly offer products and services that require qualitative data resources to increase and improve product maturity. In the course of digital transformation, the value proposition of data has changed drastically. Powerful intermediaries of data e.g. digital platforms have emerged that transformed data into a product itself. As a result, data developed from a by-product of business processes and transactions to a potentially valuable asset, which can function independently as a product.

With the global exponential data growth and the technological capabilities to store and process the immense amounts of data, companies seek to derive value of their data resources. This development with the perception of data being a resource or an independent economic good can be seen as a paradigm shift in terms of managing data. Figure 1 illustrates the development of data over time. From a research, as well as from an economic and social perspective, data is increasingly playing an important role. The platform economy intensifies data sharing across organizations, external suppliers and customers. Formerly isolated infrastructure landscapes are developing towards a network of business partners enabling inter-organizational data interactions. New questions regarding the governance, management, security and ownership emerge in this context (Bharadwaj et al., 2013; Otto and Österle 2016).

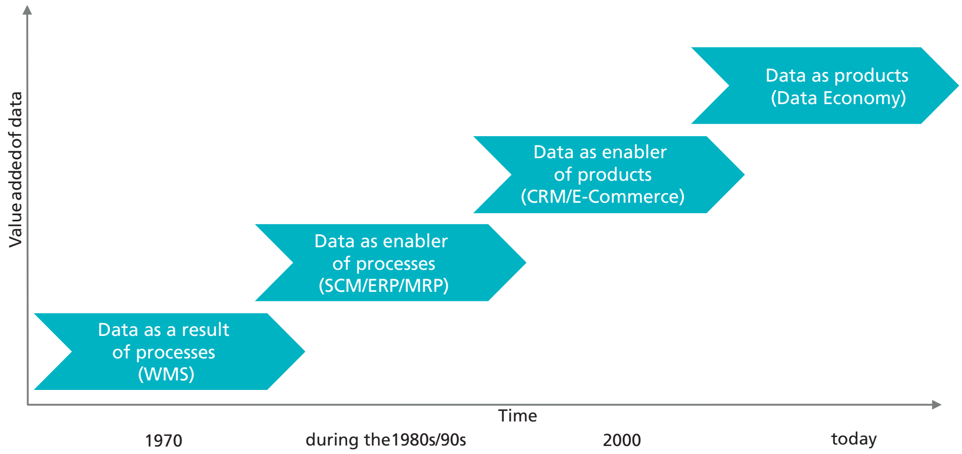


Figure 1: The Changing Role of Data over Time

Depending on the research discipline one stems from, there are multiple conceptualizations and definitions on the term "data". Data describes the properties of business objects, meaning material or immaterial objects found in the real world. The ISO/IEC 2382-1 standard provides a vocabulary of fundamental information technology terms and defines data as a "formalized representation of the properties of business objects for further processing, interpreting, and communication" (ISO/IEC 1993). It has to be noted that the user of data in this regard can be a human being or an automated process.

> **Data** is a "formalized representation of the properties of business objects for further processing, interpreting, and communication" (ISO/IEC 1993).

The **d**ata, **i**nformation, **k**nowledge and **w**isdom model, also known as "DIKW hierarchy" (Rowley 2007) corresponds with the commonly used notion of data as a raw material (Oppenheim et al. 2003). The model represents the conceptual relationships between these four terms with the fundamental assumption that processed data leads to information, which can be used to create knowledge, and gain wisdom. In the context of the notion of data as a resource or AI, the model can be used to describe the building blocks for information services and/or knowledge generation processes. Figure 2 shows the relationships between the different concepts (Bärenfänger 2017).

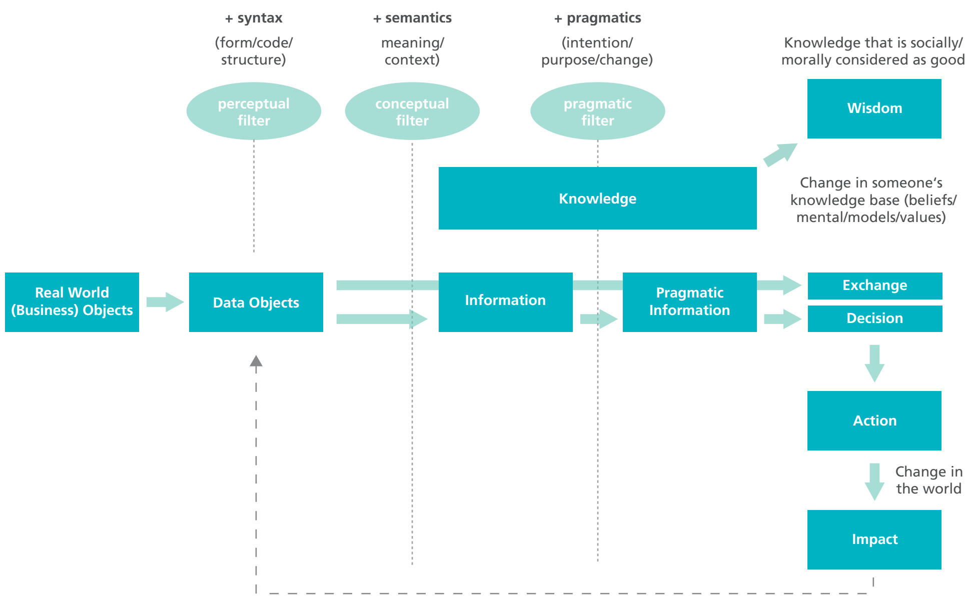


Figure 2: DIKW Model

A further common distinction between data information relates to syntax and semantics. Data is considered to be context-free, i.e. it is represented by its form, coding and structure. In contrast, information is meaningful because it emerges when data is put into context, e.g. when data is used. Context-information can be referred to as semantics. Furthermore, pragmatics then describe how information is to be used for a specific purpose or desired change.

2.1.1
Types and Forms of Data

Over the last years, the variety of forms and types of data has increased significantly. The notion of "Big Data" was coined to refer to this development. Criteria such as volume, variety and velocity – sometimes referred to as the three "V" – were introduced to categorize the various forms of appearances of data. Over time, the 3V-model was extended with two additional properties (value and veracity) to indicate the importance of quality data and incorporate value-adding aspects i.e. data analytics or data-driven services.

Data includes unstructured data from heterogeneous sources such as sensor or social media data. Structured data stored in relational databases only add up to a little amount of enterprise data assets. In fact, most data lies in unstructured form and it is estimated that 80% of all data assets are unstructured (Schneider 2016). The rapid development of Internet of Things (IoT) devices plus the pace and variety at which data is generetated increase the complexity of interoperability and challenge traditional static data models.The strategic relevance of data has been recognized in the emerging data driven economy and its adequate governance will differentiate who will benefit of theirdata assets accordingly (Berghaus and Back 2016).

Examples of these types of data:

- Archives comprise systems of records of scanned documents, forms, correspondences etc.
- Documents comprise office documents and presentations, spreadsheets, XML files etc.
- Media are images, videos, pictures etc.
- Data storage includes SQL databases, NoSQL databases, repositories, file systems etc.
- Business applications comprise ERP, customer relationship management (CRM), supply chain management (SCM) systems etc.
- Public web refers to open data which is publicly available
- Social media refers to logs and streams from Facebook, Twitter, LinkedIn etc.
- Machine logs includes event logs from servers, machines, mobile devices etc.
- Sensor data refers to data streams from industrial machinery, meters, medical devices etc.

"FRAUNHOFER ISST – TAKE AWAY"
THE DIVERSITY OF DATA IS CHALLENGE AND GIFT AT THE SAME TIME AS THE VOLUME AND VARIETY COMPLICATES ADEQUATE FORMS OF MANAGING DATA. THE COMBINATION OF VARIOUS DATA SOURCES CAN LEAD TO THE CREATION OF INNOVATIVE DATA-DRIVEN SERVICES THAT WERE NOT POSSIBLE BEFORE

2.2 How to Tame the Data-Dragon?

With the growing complexity of existing data resources, organizations are increasingly faced with new challenges in terms of adequately organizing and managing data. Established data management practices follow a "defense" approach by building capabilities to effectively manage and control internal data. Data ecosystems require organizations to establish an equilibrium between the opposing approaches "defense" and "offense". A balance between having control over data resources and willing to share data to design and deliver common value propositions is key to successfully engage in data ecosystems. In this context, the role of data governance is increasingly gaining awareness as the novel settings derived from digitalization (e.g. data becoming more significant as a valuable resource, interactions in ecosystems) raise new questions regarding the adequate management of data (Al-Ruithe et al. 2018; Alhassan et al. 2016).

From a conceptual representation of the logical organization of data, different levels of aggregation are distinguished (Chen 1976; Levitin et al. 1998; Yoon et al. 2000; Yoo et al. 2010). Data elements form the lowest level of aggregation. Data elements are instantiations of the attributes of data objects (such as a customer's family name). Records form the second level of aggregation. A record is the instantiation of a data object. For example, a customer master data record is constructed using all of the characteristics of the customer business object such that business processes (marketing, service and accounts receivable) will run smoothly for customers. At the third level of aggregation, tables (such as a customer master table) aggregate multiple records. Databases in turn aggregate several tables. A customer management database could contain all customer master data as well the associated marketing data. The totality of all company databases ultimately forms the data resource of an organization. Figure 3 summarizes these relationships (Otto and Österle 2016).

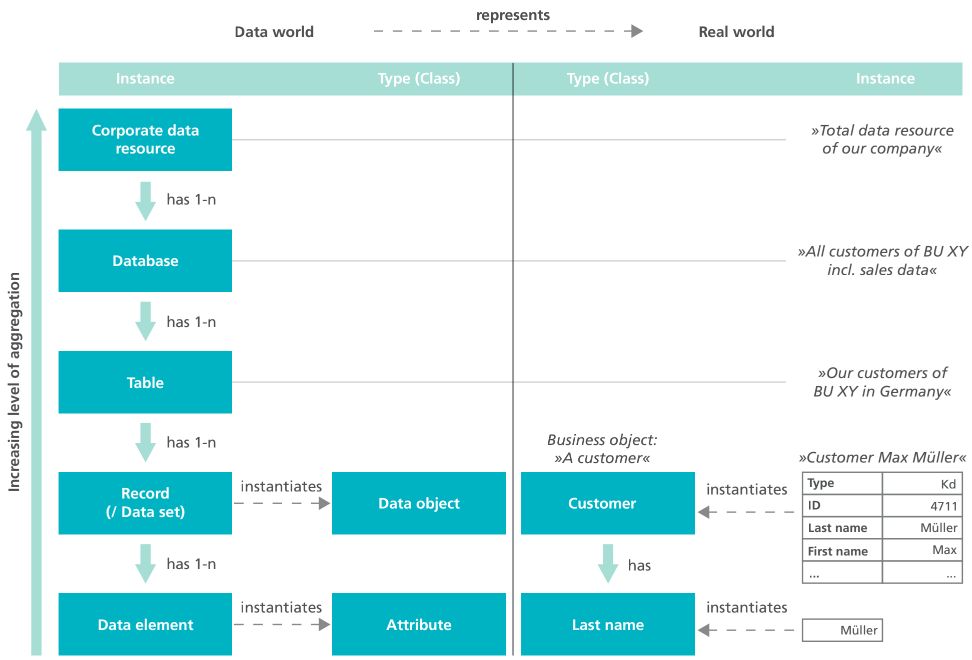


Figure 3: Organization of Data

2.2.1
Data Quality

The importance of managing data and treating it as an asset has been early recognized (Goodhue et al. 1988). Early research tracks on data placed strong emphasis on data quality adapting the knowledge of the total quality domain (Wang and Strong 1996). Data quality is defined as a context dependent, multidimensional concept. Context dependency means that data quality is in the "eye of the beholder" (i.e. the user). Multidimensionality refers to the fact that there is no single criterion by which data quality can be fully ascertained. Therefore, data quality is often described as "fitness for use". Wang and Strong (1996) proposed a hierarchical framework for organizing data quality dimensions. The framework consists of four categories:

- **Intrinsic data quality** refers to the extent to which a data value is in conformance with the actual or true value of an object. Example: The value for the weight of a certain product as it is stored in an ERP system equals the actual weight of the product when weighed after goods entry.

- **Contextual data quality** refers to the extent to which data is applicable to a certain context or purpose as defined by the user. Example: In the pharmaceutical industry different functional views of the business object "product" exist. While the research and development department focuses on active ingredients and the recipe, the logistics department focuses on the product's dimensions, weight, and packaging.

- **Representational data quality** refers to the extent to which data is presented in an intelligible and clear manner. Example: The character string "315016295" is likely to be meaningless to a user if not accompanied by the information that it is the DUNS (Data Universal Numbering System) number of Volkswagen AG.

- **Accessibility related data quality** refers to the extent to which data is available or obtainable. Example: For global spend analysis central purchasing departments need purchase order information with regard to certain suppliers and all enterprises affiliated to them from all organizational units (Hinderer et al. 2009; Madnick et al. 2002).

Data quality is closely related to the value of data. As data quality refers to fitness for use, data quality stands in a direct positive relation to the value of data, i.e. the higher the quality of data, the higher its value.

2.2.2
Data Management, Data Quality Management and Data Governance

The Data Management Association (DAMA) defines data management as "a function that develops and executes plans, policies, practices and projects that acquire, control, protect, deliver and enhance the value of data" (DAMA 2017). However, this definition is not unambiguous, as DAMA points out that data management can also be seen as a program for implementing data management functions, as a discipline, or a profession.

Data management can further be divided into a set of sub-functions of which one is data quality management (DQM). DQM aims at maximizing data quality by using methods and approaches for defining, measuring, analyzing, improving, and controlling data quality.

Typically, though, the creation and use of data are spread across an organization, i.e. the unit responsible for data capturing and creation is in often cases not the same unit that uses the data. In other words, data supply and demand are organizationally disconnected.

This is the motivation to establish data governance in organizations. Data governance is a framework of decision-making rights and responsibilities regarding the management and use of data, aiming at maximizing the value of data. The majority of data governance contributions in research and practice, though, focuses on company-internal data.

While the general purpose of data governance to define and control data management holds true for data ecosystems as well, the design of data governance differs. From an internal perspective of one single organization, the execution and allocation of decision rights for the management and use of data is manifested within organizational structures. They ensure that relevant guidelines and principles regarding data assets are in place and monitored. However, traditional instruments for assigning decision rights and accountabilities in terms of data can not be applied in ecosystem scenarios. Figure 4 shows the relationship between these concepts (Otto and Österle 2016).

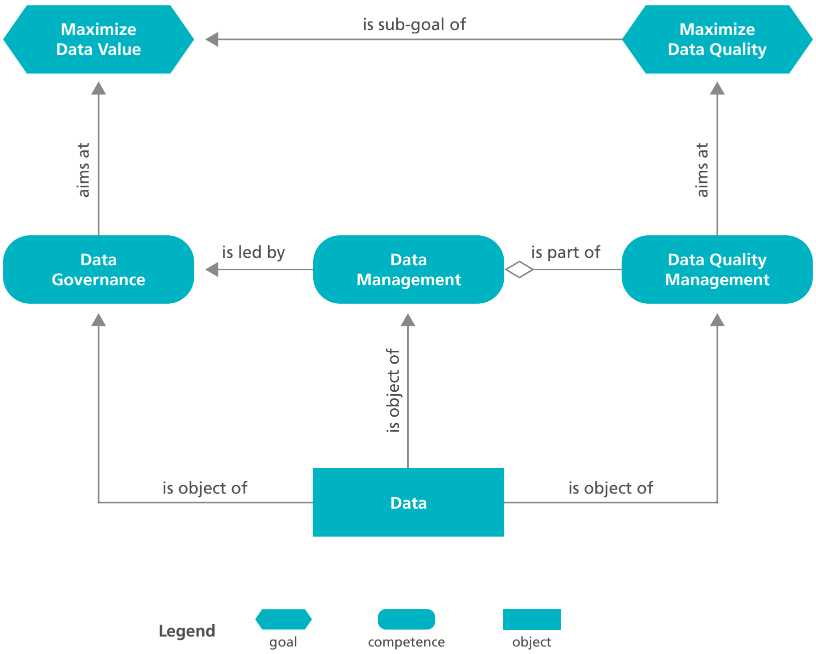


Figure 4: Position of Data Management and Related Concepts

"FRAUNHOFER ISST – TAKE AWAY"
THE EMERGENCE OF DATA ECOSYSTEMS AND DATA-DRIVEN VALUE CREATION DEMONSTRATE THE URGENCY FOR ORGANIZATIONS TO HAVE A MORE SOPHISTICATED APPROACH OVER THE MANAGEMENT OF THEIR DATA RESOURCES. ENABLE DATA GOVERNANCE STRUCTURES TO BUILD YOUR ORGANIZATION DATA-CENTRIC.

2.3
The Emergence of Ecosystems

In the course of digitalization, various concepts emerged around the central term of "ecosystem", such as business ecosystem (Moore 1993), platform ecosystem (Tiwana 2013; Schreieck et al. 2016) or digital ecosystem (Nachira 2002; Chang and West 2006). The continuous growth of different approaches and perspectives on ecosystems lead to a certain degree of ambiguity in the terminology of these concepts (de Reuver et al. 2018). The different analogies are based on similar principles, since they originate from the discipline of biological ecosystems. Biological ecosystems are natural and continuously evolving systems characterized by dynamics, the capability of reshaping themselves in order to react to natural disturbances and by relationships between involved species.

2.3.1
Business Ecosystem

In general "an ecosystem is defined as a loosely coupled, domain clustered environment where each species conserves the environment, is proactive and responsive for its benefit" (Chang 2006). Moore (1993) popularized the term "business ecosystem" as a novel organizational form reshaping traditional market hierarchies and business networks. His observation uncovers the then already existing notion of complementary innovation to provide value to the customer jointly (Moore 2006). This distinct form of organization demands the inclusion of "capital, partners, suppliers, and customers to create cooperative networks" (Moore 1993). The business environment ideally involves actors e.g. suppliers, customers or other business partners who are closely related in terms of shared co-value creation.

Later on, Teece (2007) defines the business ecosystem as "the community of organizations, institutions, and individuals that impact the enterprise and the enterprise's customers and supplies".

2.3.2
Digital Ecosystem

The increasing investments in information technologies lead to a penetration of digital technologies in companies and all areas of societal life. This development reflects the holistic concept of "digital ecosystems", where the technical infrastructure is combined with "digital species" such as software, smart services or digital things form a collaborative network.

The emergence of the platform ecosystems enables relationships of various actors and stakeholders within the technical infrastructure of a platform that enables interconnections on a technological level through (API) (de Reuver et al. 2018). It is useful to distinguish between conceptual viewpoints on the ecosystem. The technical view mostly resembles the common understanding of what makes a "digital ecosystem" or "platform ecosystem". The organizational view is in line with Moore's definition of business ecosystems.

2.3.3 Platform Ecosystem

The concept of platforms as an agent for conducting transactions of various kinds between two or more parties is well established and has been around before the digitization age (Staykova and Damsgaard 2015). In general, transaction between two parties, which is controlled by a third party, takes place in a two-sided platform (King 2013).

There is a threefold distinction of platform structures in one-sided, two-sided, or multi-sided. The early years of Facebook resemble a one-sided platform structure, as there only was one user group active on Facebook, without additional advertising or app-developers (Yablonsky 2018). Two-sided platforms mediate between two customer groups, e.g. users on a social media platforms and app developers. If the number of participating sides exceeds two, one speaks of multi-sided Platforms (Staykova and Damsgaard 2015). Multi-sided platforms enable direct (i.e., same-side network effects) and indirect (i.e., cross-side effects) network effects (Kazan and Damsgaard 2013; Gawer et al. 2014), also called externalities (Evans et al. 2013).

Direct network effects increase (or decrease) the value of a platform if the growth of users in one side results in higher value for the same side. For example, the value of Facebook for its users directly correlates with the numbers of users itself, as more users, in general, are equivalent to a higher potential for each user to conduct transactions (Ruutu et al. 2017). On the other hand, indirect network effects result from one side profiting from the growth of the other side (Kazan and Damsgaard 2013; Staykova and Damsgaard 2015; Ruutu et al. 2017). For example, the Apple app store becomes more attractive to consumers if there is a growing number of app developers providing additional applications, while at the same time users benefit more from a wider range of applications (Evans 2013).

2.3.4 Data Ecosystem

The concept of »data ecosystems« has been mainly discussed so far from a company-internal big data perspective, referring to the ecosystem as the entirety of internal actors involved in making use of big data (Demchenko et al. 2014). Another discussion threat coins the notion of personal data ecosystems, but none of the above embraces the definition of business ecosystems as a new organizational form to innovate around customer needs with data as the key resource.

In general, business ecosystems are beneficial for the systems as well as for the individual ecosystem members. Every member has to contribute resources in order to benefit and, thus, the behavior of one member has impact to all other members.

Existing approaches towards a better understanding of data ecosystems embody a multitude of concepts. Additionally, varying attempts emerged to what specifically constitutes to the domain of data ecosystems. The continuous growth of different perspectives is characterized by a level of ambiguity and fuzziness around that central term.

Author(s)	Data Ecosystem Definitions
Oliveira and Lóscio (2018)	"a set of networks composed by autonomous actors that directly or indirectly consume, produce or provide data and other related resources (e.g., software, services and infrastructure). Each actor performs one or more roles and is connected to other actors through relationships, in such a way that actors collaboration and competition promotes Data Ecosystem self-regulation."
Ding et al. (2011)	"data-based system where stakeholders of different sizes and roles find, manage, archive, publish, reuse, integrate, mash-up, and consume data in connection with online tools, services, and societies."
Pollock (2011)	"An ecosystem has data cycles, in which intermediate consumers of data such as builders of apps and data wranglers may share back their cleaned, integrated, and pack-aged data into the ecosystem in a reusable way. This cleaned and integrated data is often more valuable than the original source."
Poikola (2011)	"a multi-level and multidimensional entity where raw material, as far as distribution and developing are concerned, is the target of cooperation"
Zubcoff et al.(2016)	"is made up of many actors and small organizational structures that should recognize data like the raw material that is in a cycle and is capa-ble of feeding the ecosystem, providing bene-fits to all parties."
Zuiderwijk et al. (2016)	"Data Ecosystems consist of "all activities for releasing and publishing data on the Internet, where data users can conduct activities such as se-arching, finding, evaluating and viewing data and their related licenses, cleansing, analyzing, enriching, combining, linking and visualizing data and interpreting and discussing data and providing feedback to the data provider and other stakeholders".

Table 1: Definition Approaches of the Term Data Ecosystem

The Fraunhofer ISST approach to data ecosystems is distinguished by a strong end-customer focus and emphasis on the importance of data as a resource. We define principles that guide our perspective on data ecosystems:

- Data ecosystems are strong drivers for innovation
- Data ecosystems create benefits for each individual actor and for the ecosystem as a whole system
- Each actor is required to integrate resources in the form of data in order to profit from the ecosystem engagement
- The data sovereignty of each individual actor needs to be preserved
- A technical data infrastructure is required to enable fair data sharing among actors

If a business ecosystems requires contributions of data as a central resource, this ecosystem is a »data ecosystem«. Key to all these scenarios is the sharing of data within ecosystems. End-to-end customer process support can only be achieved if ecosystem partners team up and jointly utilize their data resource.

Data Ecosystems are not limited to a specific domain. Data sharing is independent of domain and context so that many industries can benefit from it.

3

CHANCES FOR INNOVATION THROUGH DATA ECOSYSTEMS

Fraunhofer ISST conducts multiple research projects in various domains i.e. healthcare, logistics or industrial with the goal of helping partners in becoming more offensive in their data management approach to utilize the potential of their partners in forming an ecosystem. We believe that data-driven innovation is powered through data sharing in inter-organizational relationships. It stimulates co-value generated benefits to all involved parties and results in improved service propositions to the end-customer. Ongoing projects demonstrate the potential from the perspective of various actors in different industries.

3.1 Data Ecosystem for Healthcare Innovation

Especially within the healthcare sector sensors, electronic patient records or intelligent sensor networks play a vital role. One example are Cyber Physical Systems (CPS) which is an integration of sensor systems, data capturing techniques, data processing methods and real-life processes. Furthermore mobile health, meaning the usage of apps to document health related data, also lead to large amount of citizen or rather patient related data. The treatment of epilepsy, which is one of the most frequent neurological illness, is one of many examples of significant research conducted by Fraunhofer ISST and partners. Early detection and complete documentation of epileptic seizures is essential to prevent patients from harming themselves and others. Enabling data sharing between various partners from the healthcare ecosystem is a vital aspect to improve therapeutical processes such as:

- Vital parameters (such as body temperature) of individual patients,
- Health records,
- Data from hospitals, doctors and other institutions,
- Medication data.

Vital parameters of patients as well as its medical history and their analysis can help to forecast seizures, adjust medical treatment (especially medication) to individual needs as well as increase the safety of patients.

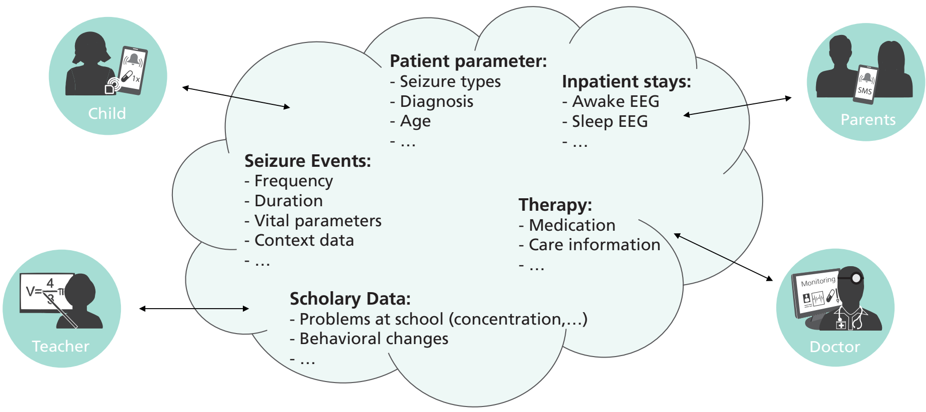


Figure 5: Data Collection of Vital Parameters for Epilepsy Fit Detection

What are the additional benefits with Data Ecosystems?
It is clear that integration and analysis of this data yields enormous potential for better medication, healthcare services and personal health. However, it must be ensured that while aiming at individual and community benefits, the privacy interests of the individual are ensured. The resulting data ecosystem – consisting of patients, medical device vendors, healthcare providers, insurance companies and pharmaceutical companies – has to ensure the data sovereignty of the patient as well as individual interest of the

industry. As healthcare is driven by several regulatory frameworks, fitting medical and business interests together is not a trivial task. Only with mutually agreed rules of the game will be ecosystem as whole and, thus, individual members be able to benefit.

EPItect

EPItect is a research project funded by the German Federal Ministry of Education and Research. The project aims at the development of an innovative ear sensor system to be carried by epileptics in order to timely recognize the occurrence of seizures with the help of AI-algorithms.

3.2 Supply Chain Visibility with Data Ecosystems in Logistics

Today's products are often of enormous complexity. A middle class car, for example, comes at 10^{30} theoretical variants. In fact, many mass products are unique – driven by the customer demand for individualized products. As a consequence of that, supply chain complexity is increasing, too. Just-in-sequence assembly scenarios require components and parts delivered to the assembly line in the correct sequence of individual production orders. Thus, in order to meet the demand of individualized products and at the same time contain production costs, visibility along the supply chain is key to successfully adapt to irregular issues within the supply network.

Root causes of capacity shortages that might lead to supply bottlenecks often are not located at the 1st tier supplier, but further upstream within the far-reaching network. To gain visibility, data is required from various partner.

The success of the supply chain network heavily depends on the availability of data and its effective processing. The generated data within a supply chain network is very diverse. Its provision is crucial to successfully enable a data ecosystem in logistics. Exemplary data include:

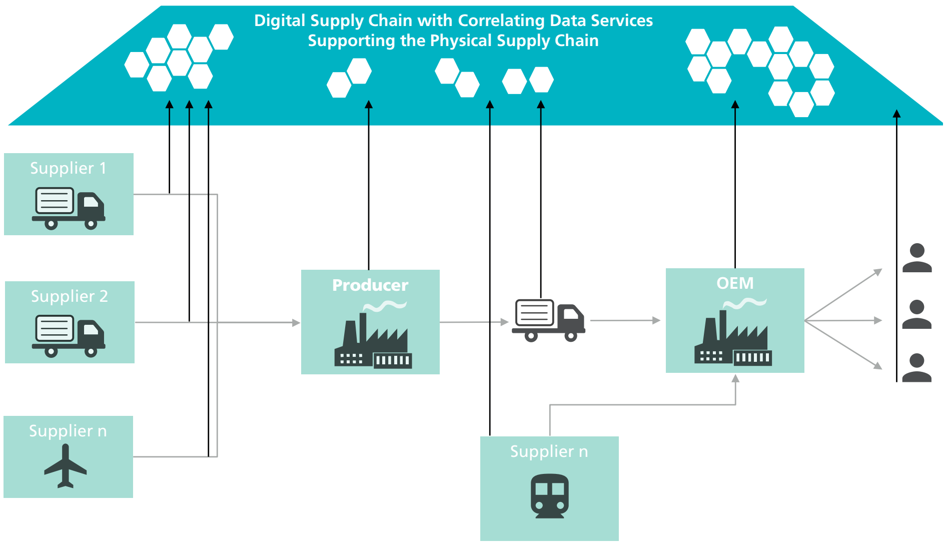
- Data on production programs and manufacturing plans,
- Data on upcoming demands and available capacities in the supply network,
- Data on risks that may disrupt or already have disrupted the supply chain on an upstream supply tier,
- Data on supply chain events such as delays in transportation, missing parts and components as well as defects,
- Short-term planning and scheduling data such as estimated time of arrival (ETA) or order sequences.

As this data is created and owned by different organizations, the overall goal of supply chain visibility can only be achieved if mutual agreement between the different partners exists to engage in data sharing. Fraunhofer ISST conducts research projects with the focus of enabling safe data sharing among partners in the supply chain. A data ecosystem for collaborative supply chain visibility can consist of the following members.

The goal of data sharing in the context of logistics is to generate benefits to all members through improved transparency by sharing data amongst all members such as:

- OEMs
- 1st tier suppliers
- 2nd tier suppliers and further upstream suppliers
- Logistics service providers
- Warehouse operators
- Carriers and trucking companies
- Providers of loading equipment (pallets, bins etc.)
- Public agencies (e.g. traffic management service providers, open data sources)

What are the additional benefits with Data Ecosystems?
Today's approaches for data sharing are usually OEM-centric, meaning that OEMs usually dominate requirements regarding the use of systems and interoperability standards. In order to move towards a data ecosystem mode of operation, pure competition and "power play" must be replaced by collaboration. The data ecosystem in logistics is characterized by strong supplier and service networks that enable data sharing for the improvement of flexibility, planning and control of the supply network. The resulting data ecosystem is then capable of creating a continuous, near real-time digital representation of the entire supply network as shown in Figure 6. This is a prerequisite to sustain the competitive advantage not only of one individual firm in the supply chain, but of the supply network as a whole.



own illustration adapted from © Fraunhofer IML

**Figure 6:
Digital Supply Chain**

RIOTANA®
Real-time IoT Analytics

RIOTANA is an IoT architecture developed by Fraunhofer ISST and stands for Real Time Internet of Things Analytics. With RIOTANA, large amounts of data can be collected and analyzed in real time. The designed architecture forms the basis for the creation of digital twins, which contribute significantly to supply chain transparency.

3.3
Innovative Predictive Maintenance Services with Data Ecosystems

Seen from a historical perspective, manufacturing companies have been among the largest producers of data since the use of machinery and electronics accounts for large data volumes. The trend continues with the fourth industrial revolution and is intensified by the profound networking of production facilities.

The sensor-equipped manufacturing process is often a data-intensive scenario. To gain new insights into the production process and the finished goods, data is continuously collected on the condition of machines, products or components. Using Big Data analyses, manufacturers can access historical and real-time sensor data to accurately evaluate the measurement parameters. It enables service technicians to optimize maintenance cycles, predict incidents and optimal runtime for customers. In addition, external data can be obtained over horizontal business networks to perform data analytics with greater sets of data.

In the context of Industrie 4.0, this scenario opens up new possibilities by linking the core business of industrial companies with data-enabled services. The strength lies in the combination of technical domain expertise in the form of production and product know-how and the ability to filter information from the generated data volumes and exploit new business opportunities. The predictive maintenance example is also a great example of how data ecosystems arise and how different actors are willing to share data because of advantageous win-win-situations.

Figure 7 gives an illustration on the predictive maintenance example. This ecosystem consists of the plant manufacturer, the maintenance service provider and the plant operators. For instance, plant operator 1 (PO1) is interested in ensuring a failure-free operation. PO1 is willing to share operational machine data to a maintenance service provider. Exemplary data can be performance data of the robot/machine, produced quantities or technical values. The maintenance service provider is then able to perform data analytics on the manufacturer's operational data to detect anomalies or possible disturbances unknown to the manufacturer. The results can then be used to maintain the machines at an early stage before a potential failure can occur. All of the plant operators are aware of the fact that the maintenance service provider is capable of performing more mature and accurate recommendations through AI algorithms, if he possess a larger database on operational machine data.

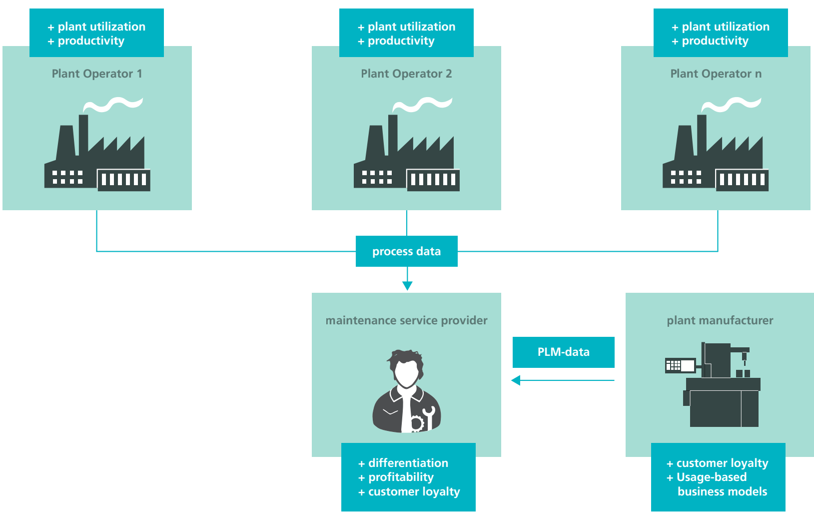


Figure 7: Business View of a Data Ecosystem [1]

[1]Legend: PLM – Product Life cycle Management.

3.4
Data Marketplaces

Data marketplaces are gaining increasing popularity in science and practice. From a practical perspective, the number of data platforms increased in the last years (Muschalle et al. 2013). The primary goal of these platforms is the trade of raw- and processed data, as well as offering data-related services. Platforms serve as an intermediary instance that provide the technical infrastructure for exchanging data between data providers and data buyers. This intermediary instance is often called "digital broker" (Koutroumpis et al. 2017).

Basically, a data marketplace can be understood as a digital platform on which data products are traded (Fricker and Maksimov 2017; Lange et al. 2016). These platforms must allow anyone (or at least a large number of potentially registered customers) to upload and manage their data products. Traded data products can occur as both static data and data streams and can be offered via various access types, such as individual file downloads, subscriptions, or APIs (Fricker and Maksimov 2017).

Additional Benefit with Data Ecosystems
Within the data ecosystem, data marketplaces provide a digital platform that facilitates the exchange of data across organizational boundaries and connect the different participants within the data ecosystem as depicted in Figure 8. With the technological further development of web services and service-oriented architectures, data marketplaces are gaining in importance alongside traditional data providers. The awareness of the need for data marketplaces is growing and drives organizations from both business and politics to deal with the design and development of data marketplaces (Accenture 2018). With the increasing importance of data as an asset, data producers are also attempting to derive greater monetary benefit from their data resources. In these data marketplaces, each participant can buy, sell and maintain its data products through different licensing models. The licensing models or contracts are either standardized or negotiated and set the rules for data authorization, access and pricing (Koutroumpis et al. 2017).

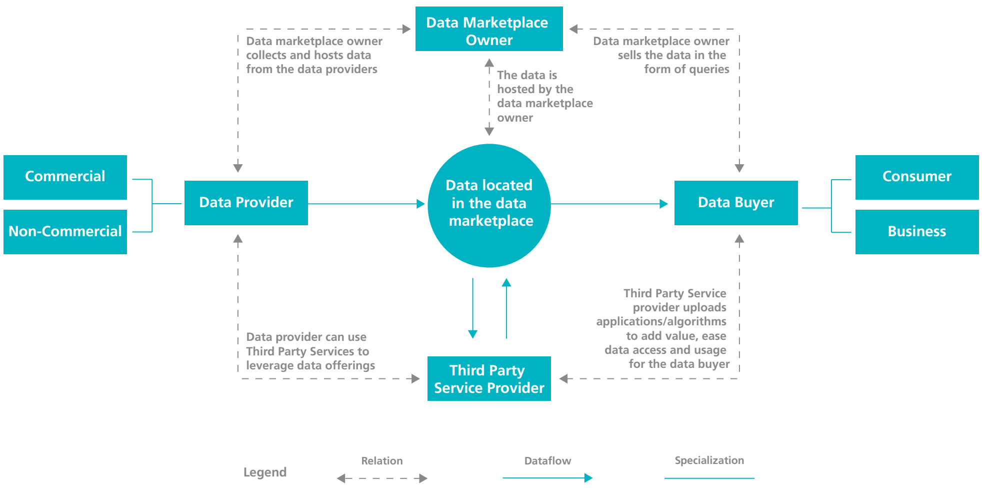


Figure 8: Data Marketplace Connecting Participants of Data Ecosystems

One example for a data marketplace is the Data Intelligence Hub (DIH) by Deutsche Telekom. The DIH functions as a neutral intermediary between data provider and data consumer while considering aspects of data sovereignty through decentralized data storage.

3.5 Training Data for Machine Learning

AI is expected to disrupt many industries. It is envisioned to become the basis for applications and systems with unprecedented capabilities (e.g., fully autonomous vehicles). It will also change the way we build these applications and systems fundamentally: Today, the software that defines the behavior of such systems is programmed. Any knowledge or intelligence of this software has to be encoded (e.g., in form of rules) by programmers. Systems that are built on machine learning technology, on the other hand, are not programmed but trained. The necessary knowledge about the domain of application does not have to be encoded explicitly by a developer but can exist implicitly in the form of labeled training data.

An example is a function that recognizes traffic signs in pictures taken by an onboard camera of an autonomous vehicle. The algorithm for detecting and classifying a traffic sign could be written manually – requiring to formalize rules that characterize color, shape, size, possible positions of these signs. When training this function based on machine learning, all that is needed is a set of images with different road signs labeled with the position of type of the contained road signs.

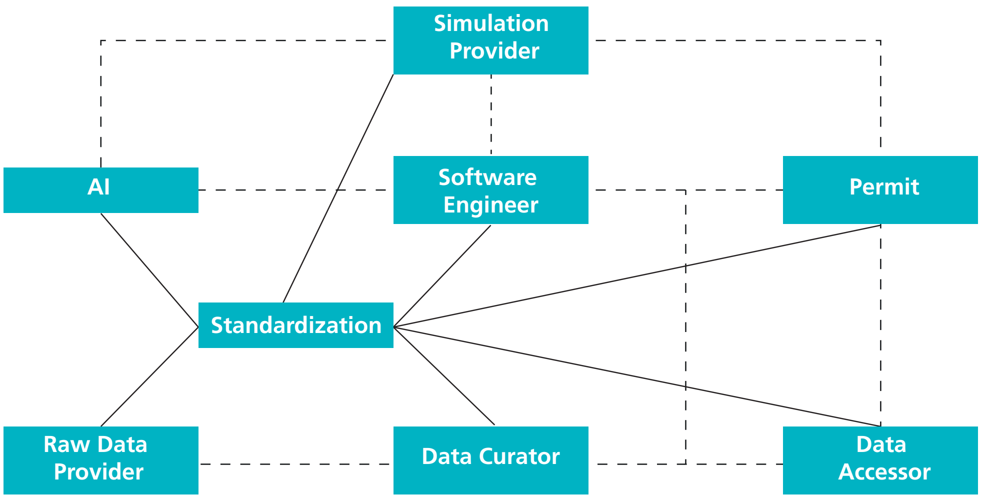


Figure 9: ML Data Ecosystem

Additional Benefit with Data Ecosystems

Precision and usefulness of trained functions hinges on the quality of the data used for training (e.g., relevance to the task, distribution, coverage of relevant cases, etc.). Training data and data that can be analyzed by trained functions (e.g., about habits and preferences of consumers) has already become a critical factor for business success, making data the strategic resource for business success. A whole ecosystem is evolving around collecting, producing, labeling, refining, and trading data. In the past couple of years there has been a surge in start-ups that offer training data from various sources, tailor-made for individual training goals. First data market places exist in which such data sets are traded.

In strongly regulated and security- or safety-critical domains (e.g., autonomous vehicles), where machine learning is expected to have a big impact as well, these data ecosystems currently start to emerge. Figure 9 sketches the anticipated roles in such ecosystems: They will also comprise producers of raw data, curators of data sets with certain characteristics, data certification authorities, standardization bodies, providers of hardware and services for training functions as well as suppliers that offer readily trained functions.

4

DATA ECOSYSTEM CONSTITUENTS

4.1 Data Ecosystem Constructs

Though data ecosystems still lack conceptual clarity in research, they comprise different basic constructs that can be found in the presented cases. These constructs require consideration as important elements for understanding the dynamics of data ecosystems and enabling inter-organizational relationships (Oliveira and Lóscio 2018).
Data ecosystems may include but are not limited to the following constructs:

- **Actor:** A data ecosystems consists of multiple actors.
- **Capability:** Each actor has different capabilities that qualifies him for the engagement in data ecosystems.
- **Expectation:** Actors anticipate different views, goals and requirements of the interaction in data ecosystems.
- **Role:** A role fulfills a function that is performed by an actor.
- **Duty:** A role incorporates duties and tasks.
- **Activity:** A role fulfills duties that involve different activities.
- **Relationship:** Interaction between actors.
- **Transaction:** The relationship between actors involves the transaction of resources e.g. data sharing.
- **Infrastructure:** Technical environment that supports the interaction between actors.
- **Business Model:** Relationships can be reflected in a business model.
- **Resource:** Resources are tangible or intangible capabilities, products or services of use or value for or created by actors.
- **Governance:** Governance encourages relationships and assures desirable behavior of actors.
- **Standard:** Resources and transactions may regard standards.
- **License:** Licenses by restrict resources.
- **Quality Metric:** Quality metrics influence value of resources.
- **Value Creation:** Common goal among actors of a data ecosystem.

4.2 Value Co-Creation

In an ideal state of a data ecosystem, each actor contributes to the ecosystem in the form of data resources to fulfil common goals. The sharing of data enables new opportunities for synergies in intra-organizational relationships. Thus, the behavior of each single actor can have an impact on the value creation of others or the ecosystem as a whole. The goal is to create an environment in which all actors benefit. The formerly isolated organizational landscape therefore increasingly extends towards an inter-organizational decentralized network. Data sharing in cross-organization or even cross-domain scenarios enable more mature data analyses and derivation of efficiency-enhancing measures. In comparison to traditional value chains, data ecosystems are characterized by more dynamic value creation processes. The dynamics in data ecosystems are more complicated as data represent intangible resources, and the value-creation process is not linear as with tangible physical products. Relationships for the exchange of data can be executed in real-time. Unlike the traditional life cycle of data from an internal organizational that covers steps from the generation of data until its archived or deleted state, the data life cycle in data ecosystems is characterized by a complete coverage of the life cycle by different actors at each stage.

Collaboration in data ecosystems generate new opportunities of value creation for the participating organizations. It is a radical paradigm change from linear to multi-sided and scalable value creation. This type of collaboration and the ability to share data with other participants to enable digital value propositions challenges conventional practices and structures, which are characterized by separation without links beyond the single organization.

4.3
Data Ecosystem Roles

A variety of different actors are involved in a data ecosystem. They can differ in their contribution to the ecosystem. Depending on the engagement, an ecosystem participant can take on a certain role or functionality. For the achievement of common value propositions, these roles and functionalities are associated with tasks and activities. The setting of the roles differs in scientific publications and practice-oriented initiatives. However, a general view can be taken from the derivation of roles, which are necessary to enable the creation of an ecosystem (Oliveira and Loscio 2018).

In general, a data ecosystem requires actors that make data available. These actors are called data providers. A data provider publishes data that can be used by various other participants (data consumer). For instance, data consumers use data to enhance and enrich existing services with additional data e.g. for analytics. Depending on the scenario, often other actors are mentioned that are essential to enable a holistic and decentral network organization. For example, further instances may be necessary, one which supports building the connection between data providers and data users, a so-called "broker", providers of the technical infrastructure e.g. platforms or providers of services for the realization of data-related activities such as preparation, analysis and visualization of data. Further instances can include services for the exchange and monetary transactions of data. In this context, depending on the role and perspective, all actors are involved conducting various activities such as providing data or services in the data ecosystem. At its core, the focus lies on the end customer, who benefits from new and innovative individualized data-driven products and services enabled by the exchange of data from various partners (Otto et al. 2019).

4.4
Data Value Chain

The general concept of the value chain was introduced by Porter (2004) as an analytical instrument to describe all activities that a company carries out to bring a valuable product or service to market. In line with the transfer of the resource perspective from tangible goods to data and information, the application of value chains to data or information flows should enable the understanding of how data creates value (Curry 2016). Generally the data value chain is subdivided into four phases:

- Data generation
- Data acquisition
- Data storage
- Data analysis

Much of the existing data value chain literature has a company-internal focus. This reflects the evolution of big data concepts in general, as companies were concerned with leveraging potentials of various forms of data from their own perspective.

In contrast to this, the concept of data supply chains extends this view. Spanaki et al. (2018) define the data supply chain as an independent type of supply chain in which data and non-material objects are moved, shared, reconfigured and aggregated to offer new opportunities for competitive advantage and business model innovation.

In data ecosystems the value-added process of data is subject to dynamic change and can no longer be characterized by a linear structure (Attard et al. 2016). Individual activities may be skipped or repeated, if necessary. They are independent activities, each with the goal to form data to a product and sell it. Five general activities can be identified, namely:

- Data discovery
- Data curation
- Data interpretation
- Data distribution
- Data exploitation

Each activity consists out of a variation of actions like generation, procuring, gathering and selection and will be performed by different actors or even companies. Each activity forms a new version of the data product or a data product, which can be sold and used as input for the next activity.

4.5
Data Sharing and Data Exchange

Data exchange across company boundaries, using cross-organizational information systems has been evolving for some time, in the process various data exchange scenarios have been developed accompanied by associated technical standards. This development will necessarily progress in terms of exchanging and sharing data in the context of data ecosystems. Especially in order to foster the establishment of concepts like data sovereignty within the exchange of data in business ecosystems.

Due to this development, it is necessary to differentiate between data exchange and data sharing in the context of data ecosystems.

- Data exchange focuses on cooperation and collaboration between companies along value chains and supply chains, in order to support, enable or optimize on a vertical level.
- Data sharing, on the other hand, takes place in vertical but especially in horizontal cooperation and collaboration between companies. The achievement of a common goal is the focus here (e.g. predictive maintenance scenarios in manufacturing) or to enable new business models by generating additional value out of data (e.g. in data marketplaces). Furthermore, data sharing implies a mode of collaboration towards coopetition.

5

DATA ECOSYSTEM PERSPECTIVES

The architecture of a data ecosystem can be viewed from different perspectives, which are important for the individual participants in the ecosystem and for providing a top-down holistic view on the topic. We differentiate between three perspectives that differ in the level of granularity, namely ecosystem, company and asset perspective:

- **Ecosystem perspective:** Holistic view on whole ecosystem including interactions between actors.
- **Company perspective:** View from the perspective of a single entity interacting within and outside the organization including internal and external systems.
- **Asset perspective:** Detailed perspective from the point of view of a technical asset e.g. sensor or software component.

The differentiation of perspectives allows for a detailed view on all interactions beginning from the generation of data to its use in services in the ecosystem.

5.1 Case Description - Smart Production and Predictive Maintenance

The following case describes a scenario in the context of production and predictive maintenance. It is a practical example currently analyzed within the research project DEMAND, funded by the Federal Ministry for Economic Affairs and Energy. Smart Production stands for intelligent manufacturing operations and encompasses all measures for the increased demands placed on production systems by digitization and Industrie 4.0. The goal is to ensure the required degree of flexibility, availability and reliability of the production facilities from an economic point of view. Additionally, the aim is to advance the automation of processes and improve production performance by means of data analysis.

The application scenario involves multiple actors engaging in a data ecosystem. However, the arising ecosystem has not been made transparent before. The OEM in this case is an industrial manufacturer of cement machines. The portfolio of the OEM covers a wide range of machinery that provides technologies for the processing of raw materials, cement production and automation solutions. Its customers are plant operators. In the course of the organization's digital transformation, the OEM's pursuit is to include innovative data-driven solutions to its products in the cement industry. Various services are being developed to further expand the product portfolio and to add value adding services to the customers. These services include reports, condition monitoring and predictive maintenance to optimize machine usage, processes and the plant operation. This approach tackles the issue of creating a gap in the OEM's machines and their operations. The loss of valuable insights into the product life cycle and operations complicated the integration of knowledge retrieval into new stages of product development. The OEM and the plant operator realize that both actors can profit from data sharing. The plant operator is capable of reacting to upcoming failures in operations by improving maintenance cycles and the OEM retrieves valuable data on operations for product improvement and providing better services to customers. To better understand the involvement of further actors in this ecosystem, the ecosystem perspective in the next section provides insights into the dynamics of the case.

5.2
Ecosystem Perspective

Figure 10 illustrates the interactions within the data ecosystem in terms of data, cash and payment flow. Within the ecosystem perspective, a holistic view on the business relations is provided.

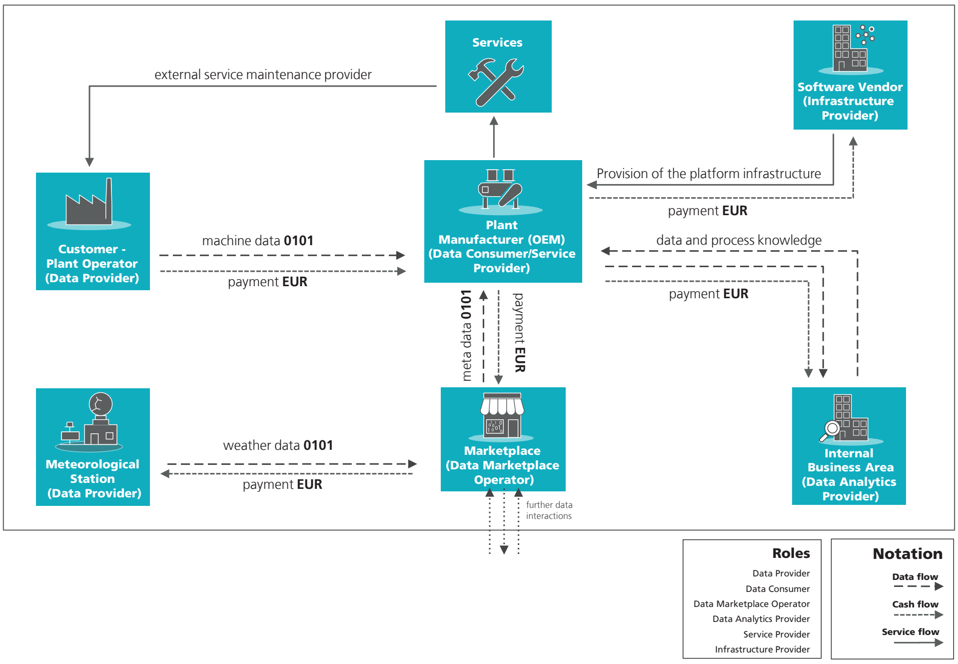


Figure 10:
Data Ecosystem Interactions

The digital ecosystem is enabled via multiple technical infrastructures. In this example two major infrastructure providers (data marketplace operator and software vendor) provide a network to connect various entities such as the physical machines, customers, service staff, and third party services e.g. for data analyses. The OEM collects data through specially installed edge devices on the customer's sites. The operational machine data is transferred to the platform-based cloud infrastructure provided by the OEM. The data is stored, processed, and analyzed on the OEM's platform. Part of the data analysis is performed by the OEM itself within separate organizational units, so that the role of data analytics provider is also taken internally. Therefore, additional historical and product development data is provided from internal resources to improve the analysis. Based on the analyses of the OEM, various services can be provided to the plant operators via the platform (e.g., dashboards on machine usage, specific operational reports, process optimization recommendations, maintenance planning). New external data sources (e.g., weather data) from meteorological stations can be considered as new parameters for analysis. This data is provided within a marketplace, which the OEM uses to enrich the existing data basis. The data marketplace acts as a gateway to other data ecosystems and promotes cross-industry data exchange. The ecosystem perspective provides an overview of the dynamics within a practical data ecosystem including the flow of data, payment and services.

5.3
Company Perspective

The Company Perspective is focused on the individual company, which participates in the data ecosystem. It is important for each individual organization to realize its position in the ecosystem to utilize existing capabilities and to reflect on how the participation occurs from a technical perspective. Each individual in this data ecosystem scenario is concerned with different business, organizational and technical aspects. There is also a limited transparency because an organization only considers its key activities in the data ecosystem and the activities of the direct connected partners.

The company perspective considers the system landscape of internal systems and shows, which external systems are linked to them. Hence, internal systems require adaption to be aligned for the exchange with the ecosystem.

Figure 11 shows several systems which can be found exemplarily in a company. In the predictive maintenance case, the plant operator primarily provides the process data from manufacturing execution systems (MES), supervisory control and data acquisition (SCADA) or programmable logic controller (PLC) systems. Which specific systems are used for the use case depends on the data required by the service provider. The systems within the company are often linked to each other via data lakes or enterprise service bus (ESB) systems. Depending on the company, the data lake or the ESB are connected to the ecosystem for the data ecosystem. Typically, the systems are not external provisioned directly for security reasons. So the companies provide external servers at the demilitarized zone (DMZ) which are gateways between the internal and the external systems. The external partners then get their data from the DMZ servers which is shown in the external area in the figure, where the service provider is placed.

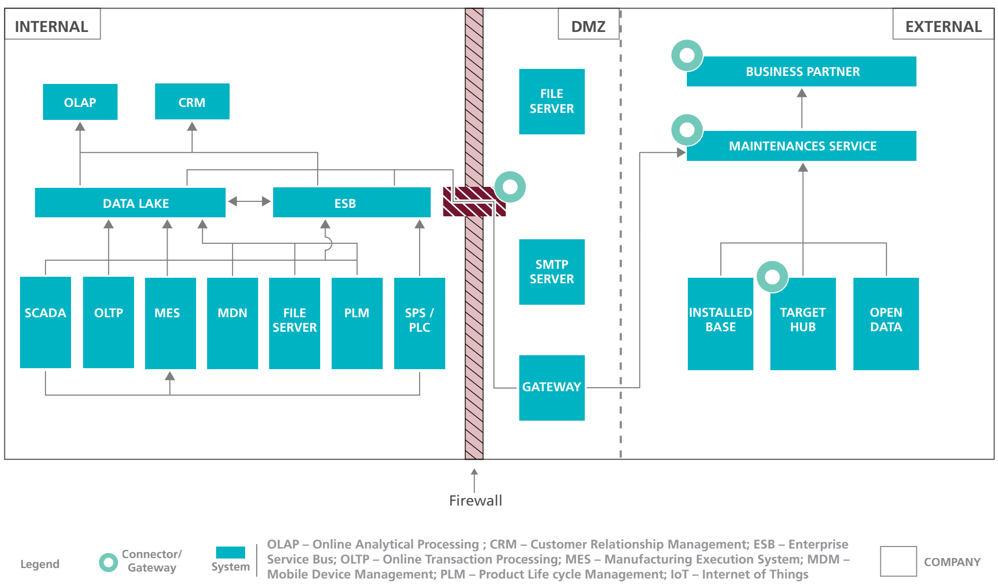


Figure 11:
Company Perspective

5.4
Asset Perspective

In comparison the overall holistic ecosystem perspective, the asset perspective regards the technical component itself. It shows how physical assets from the industrial environment can also participate directly in the ecosystem. In the example, the physical asset is a cement machine from the plant operator or a sensor from the meteorological station. These assets generate data and can be described using higher levels of abstraction e.g. using the administration shell by the Plattform Industrie 4.0 (2018). An administration shell encapsulates a physical asset, providing a standardized communication interface, for storing all data and information about the asset itself. To communicate with actors of the ecosystem, including non-industrial partners, a standard gateway for data is needed. A standardized technological gateway is capable of sharing data between the different ecosystem partners, while every data provider has the possibility to decide how other actors use the provided data assets. It provides the basis for sovereign data exchange, which is further described in the next chapter as a practical example of designing an ecosystem architecture. It is required to connect such gateway to the administration shell/asset to enable data sharing with the ecosystem. In the context of practical example of the initiative International Data Spaces (IDS) (chapter 6), the gateway is realized by the implementation of a so-called "IDS Connector". There are different options to design the implementation of the gateway/IDS Connector depending on the application scenario. Figure 12 shows the three options of implementation:

- IDS Connector as gateway at asset level (edge level)
- IDS Connector as central gateway for the company (fog level)
- IDS Connector as gateway in the cloud (cloud level), from where the data is made available

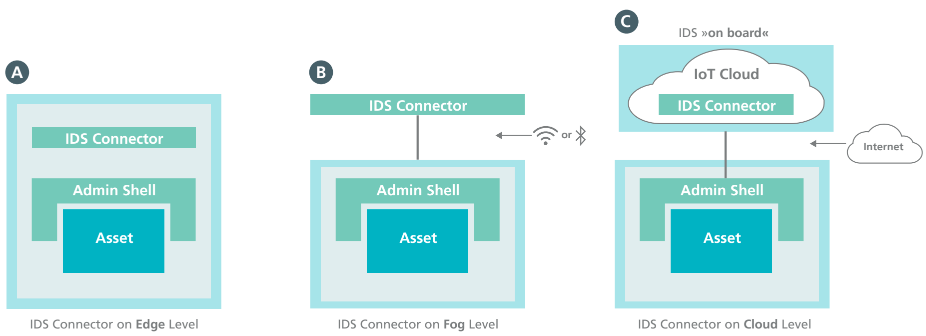


Figure 12: Asset Perspective[1]

[1]Legend: IDS – International Data Space; IoT – Internet of Things

6

A FEDERATED APPROACH FOR DATA ECOSYSTEMS - INTERNATIONAL DATA SPACES

6.1
Fair Data Ecosystem with Sovereign Data Sharing

In the course of digitalization, a number of initiatives have been launched to guide the way towards the design of a networked and digital world. To identify relevant partners for the exchange of data, domain-specific communities arise and promote a federated approach to share knowledge and develop technological solutions. International Data Spaces (IDS) is an initiative that promotes a virtual data space leveraging existing standards and technologies, as well as governance models to facilitate secure and standardized data exchange and data linkage in a trusted data ecosystem. It thereby provides a basis for creating smart-service scenarios and facilitating innovative cross-company business processes, while guaranteeing data sovereignty for data owners (Otto et al. 2019b).

The IDS initiative aims at data sovereignty for companies and citizens on an international level. More than 100 members from 20 countries are organized within the IDS Association to design and implement a reference architecture for data sovereignty as well as software components. An important software component is the IDS Connector, which acts both as a gateway for data exchange and as a secure execution environment for data analysis. The IDS Association is open and non-profit and sets standards for the secure and trustworthy use of data in the platform economy. Market participants - software companies, technology providers, etc. - take up the standards and develop innovative business models based on them.

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

The development of the IDS architecture is guided by the following principles and requirements:

- **Trust:** Trust is the basis of the IDS. Each participant is evaluated and certified before being granted access to the trusted business ecosystem.
- **Security and data sovereignty:** All components of the IDS rely on state-of-the-art security measures. Apart from architectural specifications, security is mainly ensured by the evaluation and certification of each technical component used in the IDS. In line with the central aspect of ensuring data sovereignty, a data owner in the IDS attaches usage restriction information to their data before it is transferred to a data consumer. To use the data, the data consumer must fully accept the data owner's usage policy.
- **Ecosystem of data:** The architecture of the International Data Space does not require central data storage capabilities. Instead, it pursues the idea of decentralization of data storage, which means that data physically remains with the respective data owner until it is transferred to a trusted party. This approach requires a comprehensive description of each data source and the value and usability of data for other companies, combined with the ability to integrate domain-specific data vocabularies. In addition, brokers in the ecosystem provide services for real-time data search.
- **Standardized interoperability:** The IDS Connector, being a central component of the architecture, is implemented in different variants and can be acquired from different vendors. Nevertheless, each Connector is able to communicate with any other Connector (or other technical component) in the ecosystem of the International Data Space.

- **Value adding apps:** The IDS allows to inject apps into the IDS Connectors in order to provide services on top of data exchange processes. This includes services for data processing, data format alignment, and data exchange protocols, for example. Furthermore, data analytics services can be provided by remote execution of algorithms.
- **Data markets:** The IDS enables the creation of novel, data-driven services that make use of data apps. It also fosters new business models for these services by providing clearing mechanisms and billing functions, and by creating domain-specific broker solutions and marketplaces. In addition, the IDS provides templates and other methodological support for participants to use when specifying usage restriction information and requesting legal information.

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

6.2
The Architecture for Enabling Sovereign Data Sharing

By proposing an architecture for secure data exchange and trusted data sharing, the IDS contributes to the design of enterprise architectures in commercial and industrial digitization scenarios. It does so by bridging the gaps between research, industrial stakeholders, political stakeholders, and standards bodies. The architecture is designed with the objective to overcome the differences between top-down approaches and bottom-up approaches. Figure 13 shows a typical architecture stack of the digital industrial enterprise. The IDS connects the lower-level architectures for communication and basic data services with more abstract architectures for smart data services. It therefore supports the establishment of secure data supply chains from data source to data use, while at the same time making sure data sovereignty is guaranteed for data owners.

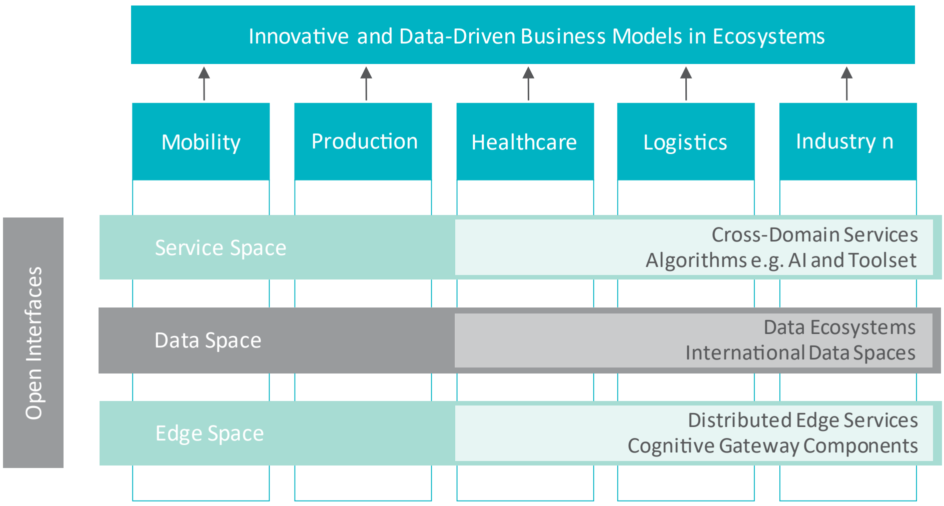


Figure 13: IDS Architecture Layers

In broadening the perspective from an individual use case scenario to a platform landscape view, the IDS positions itself as an architecture that links different cloud platforms through policies and mechanisms for secure data exchange and trusted data sharing (or, in other words, through the principle of data sovereignty).

Over the IDS Connector, the central component of the IDS, industrial data clouds, as well as individual enterprise clouds, on-premises applications and individual devices can be connected to the IDS ecosystem.

6.3 International Data Spaces Roles

Data ecosystems can comprise different roles. Figure 14 shows the interactions between the different roles in IDS.

- **Broker:** Intermediary that stores and manages information about data sources.
- **Clearing house:** Intermediary providing clearing and settlement services for all data exchange and sharing transactions.
- **Data consumer:** Legal entity or natural person receiving data from a data provider.
- **Data owner:** Legal entity or natural person which has complete control over its data.
- **Data provider:** Legal entity or natural person which makes data available for being exchanged and/or shared between a data owner and a data consumer.
- **Data user:** Legal entity or natural person that has the legal right to use the data of a data owner as specified by a usage policy.

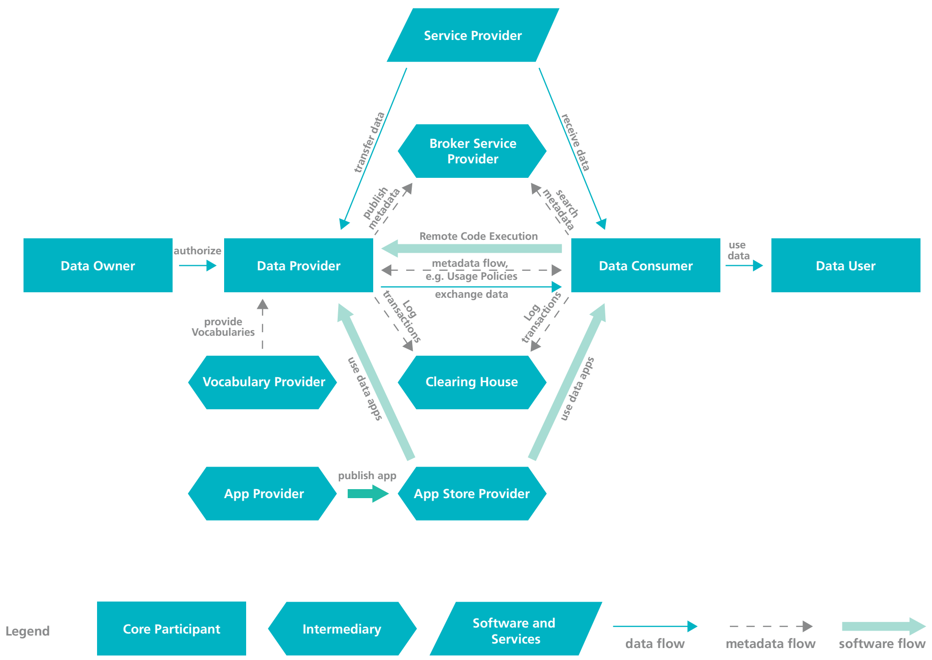


Figure 14: Data Ecosystem Roles based on IDS Reference Architecture

7

RECOMMENDATIONS FOR ACTION

7.1
Data Provider Perspective

Lay the foundations to identify business relevant and profitable data and make it available: A data provider makes data available for being exchanged and/or shared between participants in a data ecosystem. Data providers lay the foundation to successfully engage in data ecosystems internally within the organization. The data provider requires a precise overview of existing data assets and which business-models can be realized with these assets. Ideally, data providers are capable of specifying their data resources and perform evaluation techniques regarding there value proposition. Elaborating on pricing models for data requires a high maturity in the management and governance of data. The whole data life cycle from the generation until the provision on data marketplaces depend upon the support of adequate governance structures that allow to have transparency over the relevant data assets. From an organizational perspective, these prerequisites should be reflected in a data strategy. Fraunhofer ISST conducts multiple projects to support partners in utilizing their data resources for the creation of innovative business models. The research project DEMAND developed a maturity model[1] to evaluate the data maturity of an organization and to give guidance which issues to tackle in the future to successfully engage in data ecosystems.

Though research on the nature of data ecosystems and platforms is still in its infancy, it can be expected that the diversity and number of data platforms will rise. From a market perspective, there is a necessity for the data provider to analyze which platforms are suitable for its data products. It is possible that data providers allocate resources and use different platforms for addressing the needs of certain data consumers. As more platforms will penetrate the market, data providers need to choose trusted platforms with the correlating technical infrastructure for the adequate data domain. The IDS initiative supports necessary measures to provide the ability of natural and legal persons to determine the usage of their own data resources.

7.2
Data Consumer Perspective

Re-evaluate business models and identify data needs: In the sphere of data ecosystems, the transparency of existing datasets in platforms can be limited. It is relevant to build on a technical infrastructure that enables potential data consumers to make inquiries for the search of existing data assets. Therefore, a data consumer must be able to search datasets provided by various data providers. Once the data consumer identifies suitable data for its purpose, a link between the data provider and consumer needs to be established. The IDS initiative is one example to enable this data transaction on a secure basis.

There are multifold scenarios when data consumers can benefit from data sharing in data ecosystems. Organizations need to re-evaluate existing business models in terms of its digital capability. It encompasses knowing which data is existent on the one hand, and understanding what data is required to extend and increase the value of products or services. However, all parties involved need to overcome the trust barrier by building on a trusted and agreed technical infrastructure where a data consumer respects the terms of use defined by the data owner. Fraunhofer ISST employs proven methods such as Digital Business Engineering or Smart Data Engineering to enrich and transform existing products and services to sustainable digital business models.

[1]www.data-economy-benchmark.de/

7.3
Platform Provider Perspective

Define platform strategy and governance structures to establish trusted platform: Many digital platforms emerged in recent years and had a considerable societal and economic impact. Especially B2C-platforms became prominent examples for success in the digital economy. To sustain a vital position as a trusted platform in this dynamic and turbulent development, platform providers need to take various design principles into consideration. Various design principles concerning the technical infrastructure, architecture or openness have an impact on the level of control and transparency. Ultimately, they provide different governance mechanisms for platform providers on how to manage users on the platform. The design of the platform is elementary to deal with complexity, control and growth due to the involvement of multiple parties from different business domains.

It is necessary to find the adequate platform architecture that resolves governance issues between all stakeholders. On the one hand, platform providers need to be able to motivate data providers to share data and on the other hand data consumers need to find the right and qualitative data on the platform. All these aspects are reflected in the design and functionality of the platform. The goal is to create value-generating connections between all stakeholders within the platform.

Politics

Policy Maker and Legal Perspective

- Establish and promote digital sovereignty as unique selling point for global data exchange
- Support the development of a trustworthy digital ecosystem based on European values
- Institute fairness and protection of interests as a trademark
- Support technical environment and incentives to enable fast and uncomplicated cooperations

Industry

Data Provider

- Build-up data capabilities
- Identify business-relevant data resources
- Elaborate on data-driven business models
- Establish data governance on organizational and technical level
- Find trustworthy plat-forms for providing data
- Identify relevant partners in your own ecosystem for data sharing

Data Consumer

- Identify relevant data resources to enhance existing business models
- Combine various data sources to enrich data-driven services
- Identify suitable providers of qualitative data
- Find trustworthy plat-forms for acquiring data
- Identify relevant partners in your own ecosystem for data sharing

Infrastructure Provider (e.g. platform)

- Establish trusted platform to enable engagement of multiple actors
- Know your governance mechanisms to manage engagements
- Find balance between openness and control in the platform design
- Build trust by respecting security standards and sovereign data exchange

Figure 15: Recommendations Overview

7.4
Policy Maker and Legal Perspective

Provide data infrastructure and define common legislation system to enable open and sovereign data exchange: The global competition for data is uprising and leads to the development of data strategies in the economic and political landscape. Existing approaches do not take the idea of data democracy or data sovereignty into consideration and therefore need more attention. These concepts would allow data generators and owners to have control over their data resources and protect their interests in terms of data protection and privacy.

These challenges reinforce the need for developing trustworthy cloud and data infrastructures and strengthen the consideration of European values in its design such as self-determination, participation, federalism and openness.

The development of such a data infrastructure creates the basis for a sovereign use of data and data-enabled services with AI. Small, medium-sized and large companies, data producers and owners as well as data users from all domains count on the provision of an adequate legislation system and technical infrastructure for the use and sharing of data.

8

LITERATURE

ACCENTURE, 2018. Value of Data. The Dawn of the Data marketplace [online]. Accenture.com, 2018 [viewed 14 October 2019]. Available from: http://www.accenture.com/t20180904T113809Z_w_/us-en/_acnmedia/PDF-85/Accenture-Western-Digital-Value-of-Data-of-the-Data-Marketplace.pdf

ALHASSAN, I., SAMMON, D., DALY, M., 2016. Data governance activities: an analysis of the literature. Journal of Decision systems, 25(1), pp. 64-75.

ALLEMANG, D. and TEEGARDEN, B., 2016. A Global Data Ecosystem for Agriculture and Food. In: Global Open Data for Agriculture and Nutrition (GODAN). Godan.info, 2016 [viewed 12 August 2019]. Wallingford, UK. Available from: https://www.godan.info/sites/default/files/documents/Godan_Global_Data_Ecosystem_Publication_lowres.pdf

AL-RUITHE, M., BENKHELIFA, E., HAMEED, K., 2018. A systematic literature review of data governance and cloud data governance. Personal and Ubiquitous Computing, pp. 1-21.

ATTARD, J., ORLANDI, F., AUER, S., 2017. Data Value Networks: Enabling a New Data Ecosystem. 2016 IEEE/WIC/ACM International Conference on Web Intellegence.

BÄRENFÄNGER, R., 2017. Managing Information Services in the Digital Economy (doctoral dissertation). University St. Gallen, 2017.

BHARADWAJ, A. et al., 2013. Digital business strategy: toward a next generation of insights. MIS quarterly, pp. 471-482.

BOUDREAU, K., 2010. Open platform strategies and innovation: Granting access vs. devolving control. Management science, 56(10), pp. 1849-1872.

CHANG, E., WEST, M., 2006. Digital Ecosystems A Next Generation of the Collaborative Environment. In: iiWAS 2006, pp. 3-24.

CHEN, P., 1976. The entity-relationship model. Toward a unified view of data. ACM Transactions on Database Systems (TODS), 1(1), pp. 9-36.

CONSTANTINIDES, P., HENFRIDSSON, O., PARKER, G. G., 2018. Introduction - Platforms and infrastructures in the digital age. In: Information Systems Research 29(2), pp. 381–400.

CURRY, E., 2016. The big data value chain: definitions, concepts, and theoretical approaches. In: New horizons for a data-driven economy. Springer, Cham, pp. 29-37.

DAMA, 2017. DAMA-DMBOK. Data management Body of Knowledge (2nd Edition).

DE REUVER, M, SØRENSEN, C., BASOLE, R., 2018. The digital platform: a research agenda. Journal of Information Technology, 33(2), pp. 124-135.

DEMCHENKO, Y., DE LAAT, C., MEMBREY, P., 2014. Defining architecture components of the Big Data Ecosystem. In: International Conference on Collaboration Technologies and Systems (CTS), Minneapolis, USA, pp. 104-112.

DING, L. et al., 2011. A portal for linked open government data ecosystems. Web Semantics: Science, Services and Agents on the World Wide Web, 9(3), pp. 325–333.

EATON, B. et al., 2015. Distributed Tuning of Boundary Resources: The Case of Apple's iOS Service System. In: MIS Quarterly, 39(1), pp. 217–243.

EVANS, D. S., 2013. Economics of vertical restraints for multi-sided platforms. University of Chicago Institute for Law & Economics Olin Research Paper, Nr. 626.

EVANS, D. S. and SCHMALENSEE, R., 2013. The antitrust analysis of multi-sided platform businesses. National Bureau of Economic Research.

EVANS, P. C. and GAWER, A., 2016. The rise of the platform enterprise: a global survey.

LANEY, D. B., 2017. Infonomics: how to monetize, manage, and measure information as an asset for competitive advantage.

GAWER, A., 2009. Platform Dynamics and Strategies: From Products to Services, ch. 3. Platforms, Markets and Innovation. Edward Elgar Publishing, Cheltenham UK, Massachusetts, USA 2009.

GAWER, A. and CUSUMANO, M. A., 2014. Industry Platforms and Ecosystem Innovation. In: Journal of Product Innovation Management 31(3), pp. 417–433.

GHAZAWNEH, A. and HENFRIDSSON, O., 2013. Balancing Platform Control and External Contribution in Third-Party Development: The Boundary Resources Model. In: Information Systems Journal 23(2), pp. 173–192.

GHAZAWNEH, A. and HENFRIDSSON, O., 2015. A Paradigmatic Analysis of Digital Application Marketplaces. In: Journal of Information Technology 30(3), pp. 198–208.

GHAZAWNEH, A. and MANSOUR, O., 2015. Value Creation in Digital Application Marketplaces: A Developers' Perspective. In: 2015 International Conference on Information Systems: Exploring the Information Frontier, ICIS 2015.

GLAZER, R., 1993. Measuring the Value of Information. The Information-Intensive Organization. In: IBM Systems Journal 32(1), pp. 99-110.

GOODHUE, D. L., QUILLARD, J. A., ROCKART, J. F., 1998. Managing the Data Resource. A Contingency Perspective, MIS Quarterly. Management Information Systems Research Center, University of Minnesota, 12(3), p. 373.

HAGIU, A. and WRIGHT, J., 2015. Multi-sided platforms. In: International Journal of Industrial Organization 43, pp. 162–174.

HINDERER, H. and OTTO, B., 2009. Datenqualitätsmanagement im Lieferanten-Controlling. In: Controlling & Management Review, 53(1), pp. 2-10.

ISO/IEC, 1993. Information technology: Vocabulary Part 1: Fundamental terms (ISO/IEC 2382-1:1993). ISO/IEC, Genf 1993.

KAZAN, E., DAMSGAARD, J., 2013. A Framework for Analyzing Digital Payment as a Multi-Sided Platform: A Study of Three European NFC Solutions. In: ECIS 2013 - Proceedings of the 21st European Conference on Information Systems.

KING, S. P., 2013. Two-Sided Markets. In: Australian Economic Review 46(2), pp. 247–258.

KOUTROUMPIS, P., LEIPONEN, A., THOMAS, L. D. W., 2017. The (unfulfilled) potential of data marketplaces. ETLA Working Papers.

LEVITIN, A. and REDMAN, T., 1995. Quality dimensions of a conceptual view. Information Processing & Management, 31(1), pp. 81-88.

MADNICK, S., WANG, R. Y., Zhang, W., 2002. A Framework for Corporate Householding. In: Proceedings of the Seventh International Conference on Information Quality. Cambridge, MA, USA, pp. 36-40.

MOORE, J. F., 2006. Business Ecosystems and the View from the Firm. In: The Antitrust Bulletin 51(1), pp. 31–75.

MOORE, J. F., 1993. Predators and prey: a new ecology of competition. In: Harvard Business Review 71(3), pp. 75–86.

MUSCHALLE, A. et al., 2012. Pricing approaches for data markets. In: international workshop on business intelligence for the real-time enterprise. Springer, Berlin, Heidelberg, pp. 129-144.

NACHIRA, F., 2002. Towards a network of digital business ecosystems fostering the local development. European Commission, Bruxelles.

OPPENHEIM, C., STENSON, J., WILSON, R. M. S., 2003. Studies on information as an asset I: definitions. In: Journal of Information Science, 29(3), pp. 159-166.

OTTO, B., 2015. Quality and value of the data resource in large enterprises. In: Information Systems Management 32(3), pp. 234-251.

OTTO, B. et al., 2019. Data Economy. Status Quo der Deutschen Wirtschaft & Handlungsempfehlungen in der Data Economy. Whitepaper. Fraunhofer ISST, Dortmund 2019.

OTTO, B. et al, 2019b. International Data Spaces. Reference Architecture Model Version 3.0, April 2019. International Data Spaces Association, Dortmund 2019.

OTTO, B. and ÖSTERLE, H., 2016. Corporate Data Quality. Springer Berlin Heidelberg, 2016.

PLATTFORM INDUSTRIE 4.0, 2018. Specification Details of the Asset Administration Shell, Part 1 – The exchange of information between partners in the value chain of Industrie 4.0, Version 1.0, BMWi, Berlin Germany, November 2018.

POIKOLA, A., KOLA, P., HINTIKKA, K. A., 2011. Public data, an introduction to opening information resources. Finnish Ministry of Transport and Communications.

POLLOCK, R., 2011. Building the (open) data ecosystem. Open Knowledge Foundation Blog, 31, 2011.

PORTER, M., 2004. Competitive Advantage: Creating and Sustaining Superior Performance, original version 1985.

ROWLEY, J., 2007. The wisdom hierarchy. Representations of the DIKW hierarchy. Journal of information science, 33, pp. 163-180.

RUUTU, S., CASEY, T., KOTOVIRTA, V., 2017. Development and competition of digital service platforms: A system dynamics approach. In: Technological Forecasting and Social Change, 117, pp. 119-130.

SCHREIECK, M., WIESCHE, M., KRCMAR, H., 2016. Design and Governance of Platform Ecosystems-Key Concepts and Issues for Future Research. In: Twenty-Fourth European Conference on Information Systems, ECIS 2016.

SPANAKI, K. et al., 2018. Data supply chain (DSC). Research synthesis and future directions. In: International Journal of Production Research, 56(13), pp. 4447-4466.

STAYKOVA, K. S. and DAMSGAARD, J., 2015. A Typology of Multi-Sided Platforms. The Core and the Periphery. In: 23rd European Conference on Information Systems, ECIS 2015.

TEECE, D. J., 2007. Explicating dynamic capabilities. The nature and microfoundations of (sustainable) enterprise performance. In: Strategic Management Journal 28(13), pp. 1319–1350.

TIWANA, A., 2013. Platform Ecosystems. Aligning Architecture, Governance, and Strategy. Watham, USA.

WANG, R. and STRONG, D., 1996. Beyond Accuracy. What Data Quality Means to Data Consumers. In: Journal of Management Information Systems, 12(4), pp. 5-33.

YABLONSKY, S., 2018. Multi-Sided Platforms (MSPs) and Sharing Strategies in the Digital Economy: Emerging Research and Opportunities: Emerging Research and Opportunities. IGI Global.

YOO, Y., HENFRIDSSON, O., LYYTINEN, K., 2010. The New Organizing Logic of Digital Innovation. An Agenda for Information Systems Research. In: Information Systems research 21(4), pp. 724-735.

YOON, V. Y., AIKEN, P, GUIMARAES, T., 2000. Managing organizational data resources. Quality dimensions. Information Resources Management Journal (IRMJ), 13(3), pp. 5-13.

ZECHMANN, A., OTTO, B., MÖLLER, K., 2017. Nutzungsbasierte Datenbewertung. In: Controlling 29(5), pp. 57-66.

ZUBCOFF, J. J. et al., 2016. The university as an open data ecosystem. In: International Journal of Design & Nature and Ecodynamics, 11(3) pp. 250–257.

ZUIDERWIJK, A. et al., 2016. The wicked problem of commercial value creation in open data ecosystems: Policy guidelines for governments. Information polity.