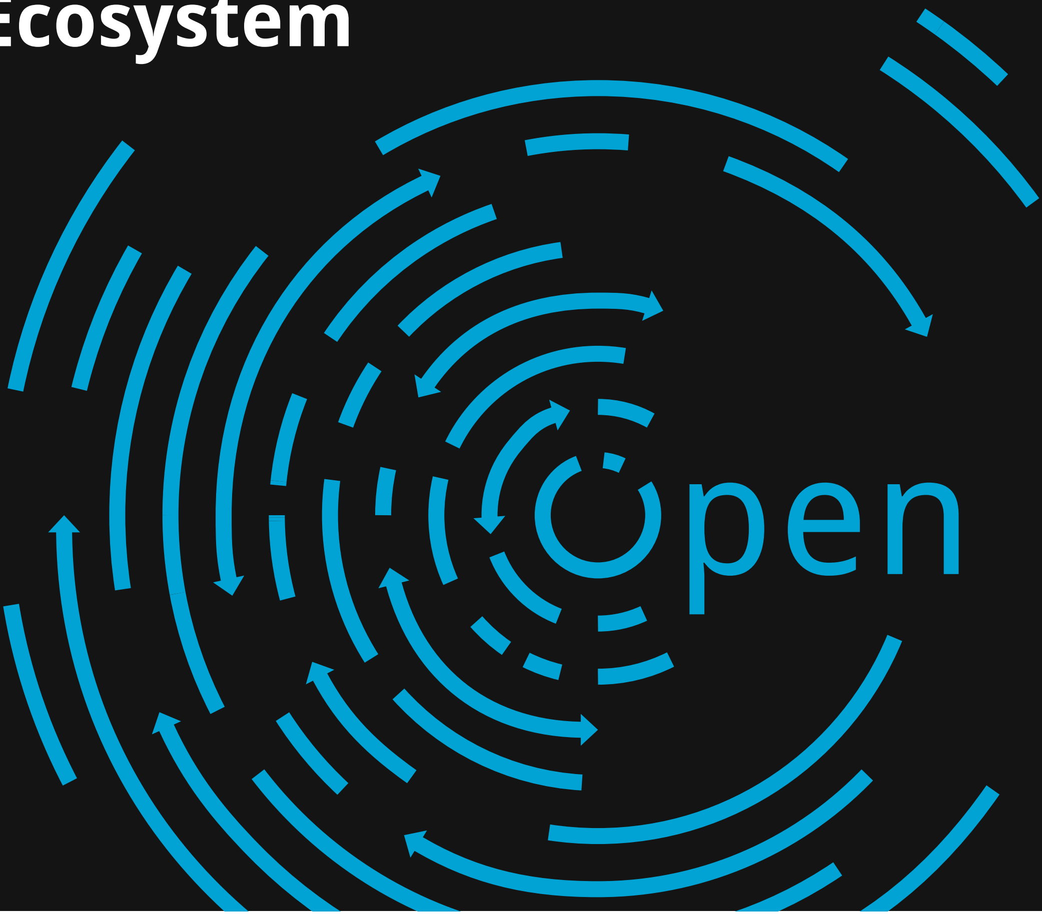
INTERNATIONAL DATA SPACES ASSOCIATION

Position Paper | Version 1.0 | December 2018

Open Data Spaces
Towards the IDS Open Data Ecosystem



- ● Position Paper of members of the IDS Association
- ○ Position Paper of bodies of the IDS Association
- ○ Position Paper of the IDS Association
- ○ White Paper of the IDS Association

Authors:
Vincent Bohlen, Lina Bruns, Nadja Menz,
Fabian Kirstein, Dr. Sonja Schimmler

Fraunhofer Institute for
Open Communication Systems FOKUS
www.fokus.fraunhofer.de

Version 1.0
December 2018

Contact:
Digital Public Services
Fraunhofer FOKUS
dpskontakt@fokus.fraunhofer.de

Cite as

Bohlen V., Bruns L., Menz N. et al. (2018): Open Data Spaces. Towards the IDS Open Data Ecosystem. International Data Spaces Association. https://doi.org/10.5281zenodo.5675977

1. Introduction

A relevant aspect of the International Data Spaces ecosystem is the allocation and application of Open Data. Typical examples for Open Data are weather data, traffic information or geographical data. Evidently, such data can be significant to enhance data-driven businesses and many IDS use cases. Especially, it concerns both core aspects of the IDS: the data provision and the data use. Although, traditionally Open Data is mostly provided by public institutions, it can be assumed that industries will become significant providers of Open Data in the future. The rapid growth of applying advanced data processing and data exchange in the industry sector will foster new strategies and business models for opening previously closed or yet undiscovered data. Open Data is widely accepted as a driver of innovation and an anchor for trust and transparency. Companies are already harnessing the benefits of using Open Data in order to improve decision-making, streamline business processes and increase efficiency. A better integration of Open Data and IDS data sources and specifications would increase efficiency and nurture both ecosystems. An easier and optimized integration of Open Data in industry processes and access to new industry data sources for the public and data enthusiasts is enabled.

This position paper gives an overview of this topic related to concepts, technologies and use cases of the International Data Spaces. In the first section **Open Data Overview,** the current organizational and technical state of Open Data is reported and its relevance for both, the industry and the public sector is characterized. In the second section **IDS and Open Data,** significant intersections and potentials of joining the principles and specifications of the International Data Spaces and the concept of Open Data are highlighted. The outcome of this section is transferred into a practical architecture and software artefacts in the third section **The IDS Open Data Ecosystem**. Hence, this position paper delivers practical guidelines and solutions to harness the full potential of combining Open Data and IDS in order to initiate a short-term adoption. Successively, the practical artifacts will be made available for actual application, reuse and further development.

2. Open Data Overview

This section provides a general overview and background information about Open Data, including relevant stakeholders, real-world examples and the technical state-of-the-art.

Definition

Due to widespread and still increasing digitization of all areas of life, the volume of data is steadily growing. A great part of this data is of public interest and can help create more transparency, participation or even novel business models. For this potential to be used, the data has to be made available to the public.

"Open Data" describes a concept in which machine-readable data can be freely used, reused and distributed by anyone. Access to the data is possible at any time without mandatory registration or justification. The data is offered free of charge for unrestricted reuse. However, Open Data must not contain personal data or cannot be subject to data protection. Typical Open Data is, for example, environmental and weather data, geodata, traffic data, household data, statistics, publications, protocols, laws, judgments and ordinances.

Open Data Principles

Openness is the key principle of Open Data and relates to two aspects: First, from a legal point of view, the data should be published under an open license with the possibility of reuse and disclosure. This also includes the change of the data by third-parties for commercial purposes. Second, the aspect of technical openness refers to providing the data in machine-readable and non-proprietary file formats so that it can be universally used and processed.

Next to the legal and technical openness, there are other principles, which have been set up by a group of Open Data leaders already back in 2007[1]. The principles have been expanded over the course of time and are still valid today[2]: According to the principles, data should be published as fully as possible and with the highest possible level of granularity (raw data) without delay right after data collection. The data should be accessible free of charge and without any discrimination (no registration needed) or legal restriction in open and machine-interpretable formats.

Open Data Stakeholders

Roughly speaking, there are two main groups in the context of Open Data: data publishers and data users. Data publishers are the persons or organizations that make content accessible to users, e.g. via a platform or a data portal. The data publisher is responsible for the decision on the publication, terms of use and formats. Data users are those natural or legal persons who use the data provided by data publishers in accordance with the intended terms of use.

Although the topic of Open Data is usually discussed in connection with the modernization of public administration, it is ultimately supported by very different actors. For example, data publishers may be municipalities, offices, public sector, non-governmental organizations, associations but also commercial companies, research institutes, non-governmental organizations or citizens. Therefore, Open Data includes a variety of types of data: ministerial data, (e.g. statistics, data from cadasters), company data (e.g. market data, data on consumer behavior), research data (measured data, laboratory values) and so on. The Open Data user group is also diverse, ranging from start-ups, data journalists and

[1] https://www.data.gov/blog/open-data-history

[2] https://sunlightfoundation.com/policy/documents/ten-open-data-principles/

hobby developers to businesses and governments.

Motivation and Benefits

Data is published for various reasons. On the one hand, Open Data creates transparency. This especially applies when data from decision-making processes is being published: In doing so, decisions and actions become more understandable for the public and thus, trust and acceptance potentially are increased. In public administration, Open Data can also contribute to participation and democracy. Open Data also holds enormous potential for innovation. If data is available as Open Data, it can be used, processed, refined and distributed by citizens, non-governmental organizations, universities and companies. This can result in new applications, products and services, business models and production chains. Therefore, Open Data is also considered a driver of innovation and economic development.

History

Over the past few years, the concept of Open Data has moved into the focus of politics. Yet, the origins go back further: In 1995, a document from an American scientific agency appeared, which dealt with the disclosure of geophysical and environmental data. In this document, the term "Open Data" appeared for the first time.[3]

Over the years, different standards, data formats and principles have been developed to facilitate data exchange and push forward the idea of Open Data. In 2007, a group of data enthusiasts discussed and set up basic principles for Open Data, as already mentioned above. Also, in 2007, the Infrastructure for Spatial Information in the European Community Regulation of the European Commission (INSPIRE)[4] was introduced. This includes a reference framework for a uniform spatial data infrastructure, which enables the consistent publication of geospatial data across Europe.

Open Data really got into the public eye in 2009, when the Memorandum on Transparency and Open Government was passed under President Barack Obama. From now on, administrative data in the United States should be machine-readable, freely available, and further processed. This directive has brought a new level of transparency and openness to government and administration in the US.[5]

Today, Open Data is on the political agenda of many institutions and organizations, especially in the field of public administrations. The brisk open data implementation in public administration is attributable to the open data laws that exist in many countries. In Germany, for example, the e-Government law[6] (EGovG) requires the opening of administrative data for federal authorities (§12a EGovG).

Examples

The most common channel of distribution for Open Data is the deployment and operation of a web portal. These portals usually constitute catalogs, which not only provide the data to users but also make it searchable. In total, more than 2,600 such data portals exist today[7], for example issued by local authorities, companies or higher-level institutions. In the following, some examples of Open Data Portals are presented.

Open Data Portals of Public Administrations

Open Data portals of public administrations can be found at different federal levels. In recent years, more and more German municipalities have published their data on

[3] https://www.bitkom.org/sites/default/files/pdf/noindex/Publikationen/2017/Leitfaden/Open-Data-Leitfaden/171103-Open-Data.pdf S. 15
[4] https://inspire.ec.europa.eu/inspire-directive/2
[5] https://www.data.gov/blog/open-data-history
[6] http://www.gesetze-im-internet.de/egovg/
[7] https://www.opendatasoft.com/a-comprehensive-list-of-all-open-data-portals-around-the-world/

their own municipal data portals. Some federal states also have their own state portals, for example North Rhine-Westphalia and Berlin. Above the country and municipal portals, there is the national Open Data portal for Germany, GovData. On international level, there is also a European data portal, which provides central access to the administrative data of European countries. The European Data Portal, GovData and the Open Data Portal of the city of Bonn, are briefly presented below.

Bonn[8]

Open Data has been introduced as a standard for all public information of the city of Bonn since 2014. Also, in 2014, the city's Open Data Portal went online, making it one of the first Open Data portals at the municipal level in Germany. The portal currently provides 259 datasets[9]. The greatest part of this data comes from the city of Bonn itself (220 datasets), but also municipal companies such as the "Stadtwerke Bonn" (municipal utilities) or commercial companies such as the company Cambio (car-sharing provider) publish their data via the portal.

GovData[10]

Since 2013, GovData is the central Open Data portal of Germany, which provides unified access to administrative data from the federal government, federal states and local authorities. The use of the portal by the federal government and several federal states is governed by an administrative agreement. The goal of the portal is to facilitate and improve the findability and usability of the data through a central access. According to the Federal Government's evaluation report on the government program "Verwaltung 2020", 18.500 datasets have been published so far via GovData[11].

European Data Portal[12]

Since 2015, the European Commission has been making data available from European countries on the European Data Portal in form of a one-stop-shop. The portal gathers open governmental data available in the different European countries. It currently covers 34 countries and 79 catalogs, links to 890,530 datasets[13] across Europe, and offers a variety of interactive learning resources and Open Data usage examples.

Open Data Portals of Companies

As already mentioned, companies can also publish Open Data. Two examples are presented below: The Open Data Portal of the Deutsche Bahn and the Open Data Portal of the distribution grid operator Stromnetz Berlin GmbH.

Deutsche Bahn[14]

At the end of 2015, Deutsche Bahn (German railway) launched its own Open Data Portal. Here, selected data of the DB Group that are suitable for Open Data and that are neither personally identifiable nor for internal use, only, are published. Included are, for example, data on stations, elevators or bus stops, as well as the list of offices and the route network. The portal is in beta and is further developed in parallel. Currently, 33 datasets are provided, but new data is continuously made available.[15] In addition, Deutsche Bahn holds hackathons and workshops on Open Data several times a year.

[8] https://opendata.bonn.de/
[9] As of December 2018.
[10] https://www.govdata.de/
[11] https://www.bmi.bund.de/SharedDocs/downloads/DE/publikationen/themen/moderne-verwaltung/evaluierungsbericht-digitale-verwaltung-2020.pdf?__blob=publicationFile&v=1, S. 81
[12] https://www.europeandataportal.eu/
[13] As of December 2018.
[14] http://data.deutschebahn.com/
[15] As of December 2018.

Stromnetz[16]

Since 2012, the distribution grid operator Stromnetz Berlin GmbH has been providing network data in open form on a distinct Open Data Portal. Currently, there are 231 datasets available.[17] These are typical structural data such as circuit length, number of withdrawal points or information about network charges and feed-in of renewable energies in the Berlin network. All interested parties can use the data freely. Stromnetz also supports the use of open data by holding hackathons in the context of energy data.

Impact

Economic Impact of Open Data

Open Data holds a high economic potential. The use, aggregation or combination of Open Data (with other data) can be the basis for new products and services. An everyday example for the reuse of Open Data can be found in vehicle navigation systems, weather forecasts or financial and insurance services. The potential of Open Data also becomes visible with the use of Smartphones: Every day, millions of apps that are based on Open Data are used, such as pollen forecasts or timetable information for local public transport.

According to a study of the European Commission[18], the total market size of Open Data is expected to grow from 55.3 bn EUR for 2016 to 75.7 bn EUR by 2020 in the EU28+ countries. The same study estimates that through Open Data about 25,000 new jobs are to be created in the Member States in the period of 2016 until 2020. For Germany, a study of the Konrad-Adenauer-Stiftung[19] estimates the value creation potential of Open Data very differently, depending heavily on further engagement of German politics and economic decision-makers: from a conservative point of view, the value-added potential of Open Data in Germany is estimated by 12.1 bn EUR p.a., from an optimistic point of view by 131.1 bn EUR p.a. All these numbers imply the high economic potential inherent in Open Data. Still, next to the economic potential, there is also a growing conviction that, by opening up their data, companies are gaining greater value through transparency and that openness brings greater prosperity in general.

Open Data Business Models

Open Data can help companies make better decisions, streamline business processes, and use their own infrastructure more efficiently. Various business models are conceivable based on Open Data[20]: for example, statistics and insights gained from Open Data can be sold as well as products that have been enriched with Open Data. Services can also be linked to Open Data: Calculations of, for example, customer data enhanced with Open Data can be offered as a service. Also conceivable are services in the field of data analysis: Data collection, aggregation, evaluation and visualization offer numerous possibilities for business models. The creation and operation of platforms for the publication and use of Open Data as a product or service is also a possible business model.

The Open Data Institute has analyzed 270 UK companies using Open Data[21]. The results show that these companies operate in various industries: mostly in the information and communication sector, but also

[16] http://www.netzdaten-berlin.de/

[17] As of December 2018.

[18] https://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf

[19] https://www.kas.de/c/document_library/get_file?uuid=3fbb9ec5-096c-076e-1cc4-473cd84784df&groupId=252038

[20] https://www.opendata.sachsen.de/835.htm

[21] https://theodi.org/article/open-data-means-business/

in education, research, agriculture, finance, services, health and other industries. Many of these companies use Open Data to develop new products, services and business models. Most frequently used by companies are open geospatial data, transport data and demographic data. According to the study, most of the companies that were surveyed, also publish Open Data themselves, which increases the data supply and attracts more customers.

Maturity Levels

If and to what extent companies can use Open Data depends on the degree of maturity of the data being published. In this context, for example, it is relevant whether the data can be automatically processed by machines or whether they need to be manually upgraded at first. Maturity models help assess the maturity of the data. A model for determining technical maturity is the Five Star Model[22] by Tim Berners-Lee. In five cascading maturity levels, this model classifies the maturity of data for automated processing. The individual levels are: (1) data is available under open license, (2) data is structured, (3) data is structured in open formats, (4) uniform resource identifiers are used, (5) data is linked to other data and thus, creates a new context.

Another maturity model, but on a different level of consideration, is the Open Data Barometer[23] of the World Wide Web Foundation. It uses various dimensions, such as economic, social and political impact to assess states in terms of their Open Data activities. Currently leading in the ranking are Canada, Great Britain and Australia.[24] Of the 30 countries listed, Germany ranks 10th.

[22] https://5stardata.info/de/
[23] https://opendatabarometer.org/
[24] As of December 2018.

Technologies

Established Open Data Software and relevant standards and specifications existing in the context of Open Data are described below.

Open Data Software

CKAN[25]

The open source solution CKAN is a basic software for building data catalogs, especially for Open Data. It is established as the de facto standard in the public sector, but is also increasingly used by companies. CKAN provides numerous functionalities for mapping the entire process of publishing data catalogs. In particular, a comprehensive range of extensions has been developed. The CKAN API is extensively documented and provides a simple machine-readable way to retrieve the metadata of the data catalog.

OpenDataSoft[26]

Open Data Soft is a complete and proprietary solution for collecting, processing and delivering Open Data. In France, in particular, OpenDataSoft is a de facto standard and has accompanied many public institutions in the implementation of Open Data. OpenDataSoft also allows for interaction and visualization through automated API generation.

Socrata[27]

Socrata is a proprietary SaaS cloud-based Open Data catalog platform that provides API tools and data manipulation. A distinctive feature of Socrata is that it allows users to create views and visualizations based on published data and save them for others to use. In addition, Socrata offers an open

[25] https://ckan.de/ckan/
[26] https://www.opendatasoft.de/
[27] https://socrata.com/

source version of its API, designed to facilitate the transition for customers who decide to migrate from the SaaS model.

Standards and Specifications

DCAT[28]

The Data Catalog Vocabulary (DCAT) is a popular vocabulary for describing data records meaningfully. The overall goal is to increase the interoperability between data catalogs and enable compound searches. DCAT makes no assumptions about the format of the actual data. The standard core consists of three main classes: catalog (dcat:catalog), dataset (dcat:dataset) and distribution (dcat:distribution), whereby one catalog consists of several datasets and one dataset can have multiple distributions. The latter refers to the actual data, so to the resources. DCAT is published as a so-called Linked Data specification, but its use in other serialization formats is possible and established. The basis for DCAT is the Dublin Core metadata vocabulary[29]. Dublin Core is a general standard for describing digital resources.

DCAT-AP[30]

The DCAT Application Profile for data portals in Europe (DCAT-AP) is an extension of DCAT for describing public sector records. DCAT-AP is developed by the European Commission (among others) and is intended to establish itself as the standard for the publication of Open Data across Europe. For this, DCAT has been extended by numerous metadata fields and concrete namespaces, e.g., for categories or languages. DCAT-AP is becoming increasingly popular throughout Europe. In particular, country-specific extensions have emerged. In June 2018, the German IT Planning Council defined DCAT-AP.de[31] as the formal exchange standard for open governmental data in Germany.

INSPIRE - Infrastructure for Spatial Information in the European Community[32]

In 2007, the European Parliament and the Council adopted a project for a common spatial information infrastructure in Europe - the so-called INSPIRE Directive (2007/2/EC). INSPIRE facilitates the use of geodata in Europe. INSPIRE came into force in May 2007 and has since been transposed into EU law by the EU Member States. The INSPIRE Directive defines the legal framework for the development of spatial data infrastructures. Technical details are regulated by the EU with implementing provisions that are directly binding on the Member States. In practice, INSPIRE demands a uniform description of geodata and their provision on the internet, including services for search, visualization and download. The data itself must also be in a uniform format. The measures mentioned concern the type of acquisition, storage and processing of geodata. The purpose is to collect geodata at a specific administrative level for use by other administrative authorities.

RDF

RDF (Resource Description Framework) is most likely to be understood as a model or system for storing data and metadata. The focus is less on good readability for people, but on the connection of related data. These links are called relations or triples and always follow the pattern "subject-predicate-object". Thus, the structure is based on the natural language. Each of the triples can be addressed by a globally unique Internet address (the so-called Uniform Resource Identifier of the Hypertext Transfer Protocol).

[28] https://www.w3.org/TR/vocab-dcat/
[29] http://dublincore.org/
[30] https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe
[31] https://www.dcat-ap.de/
[32] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32007L0002&qid=1543570495931&from=DE

The URIs point out where further information and facts about the designated concepts can be retrieved. RDF provides the basis for Linked Open Data (LOD), an approach that makes data understandable beyond its original context.

Frictionless Data

Frictionless Data is a set of open specifications and software artefacts to facilitate and simplify the publication and consumption of Open Data.[33] It was initiated by the Open Knowledge Foundation and is partly based on established standards. The main focus lies on a detailed, semantic and structured description of tabular data. It advocates the use of so-called "data packages", which contain both, a payload of data and structural information about the data. In general, the specification is domain-agnostic, but a specialization for fiscal data is available. A rich selection of apps, libraries and platforms is on hand in order to create and process data packages, especially in combination with the Open Data platform CKAN. A notable pilot use case is the Open Power System Data portal, which offers many of its data as Frictionless Data.[34]

Barriers

As already mentioned, Open Data is credited with great potential. More and more administrations open their data, and data portals are heavily expanded. However, there are some barriers in dealing with Open Data that occur on both the data publisher and the data user side. A study of the European Union conducted under the framework of the European Data Portal[35] concludes that the existing barriers can be of political, organizational, legal, financial and technical nature.

Political barriers occur when politicians as well as decision-makers in companies and public administrations classify the priority of Open Data as low and attach it little or no importance at all. Organizational barriers, for example, refer to the processes within an organization: If Open Data is not an integral part of the organization's internal processes and if not all relevant actors are being involved in opening up the data, organizational barriers may be present. A legal framework is the basis for Open Data: legal barriers arise when, for example, users are unclear about what to do with a dataset that has been published under a given license. Publishing Open Data costs resources on the data provider side, while reusing Open Data means resources and investment on the data user side. Therefore, Open Data can also create financial barriers.

Technical barriers relate to both the availability of data (as well as the necessary infrastructures for data provision) and data quality. For example, data that is available in proprietary or non-machine-readable formats is problematic: in these cases, automatic further processing is difficult or impossible for data users, which means that the data will lose value. Looking at the most widely used formats in which data is made available on the European Data Portal, it is fortunate to find that many of the most popular formats are both machine readable and open[36]. They include CSV, JSON, HTML and XML. However, some of the top 10 most widely used formats on the European Data Portal still include proprietary or non-machine readable formats such as XLS and PDF, indicating that this type of technical barrier still exists.

The lack of standardization is also a technical barrier: Open Data is diverse, various formats, languages and licenses make it difficult for data users to implement permanent solutions for the reuse of Open Data in

[33] https://frictionlessdata.io/
[34] https://open-power-system-data.org/
[35] https://www.europeandataportal.eu/de/highlights/barriers-working-open-data
[36] https://www.europeandataportal.eu/data/en/dataset. As of December 2018

their processes. Another problem of technical nature for data users is the quality of metadata: often, the metadata is incomplete, which makes the data more difficult to find. In addition, missing descriptions in the metadata lead to the data not being interpreted correctly by data users.

In general, the different barriers may apply to both sides, data publishers and data users, when working with Open Data. However, some barriers apply in particular to only one of the two groups. As part of the study about barriers in working with Open Data, a survey was conducted among data publishers and data users. The results reveal that data publishers face financial and legal barriers when opening up their data, while data users have to cope with technical barriers like low quality and availability issues.

In order to meet the technical barriers of data users, it is important that metadata is given particular attention when providing data: complete and easy-to-understand metadata is a prerequisite for data to be found, understood and, ultimately, used beneficially. The quality of the data also plays an important role: in particular, it must be ensured that the data (and metadata) are as up-to-date, accurate, consistent, transparent, trustworthy and reliable as possible.

Outlook

Open Data as a global phenomenon will continue to evolve in organizational and technical aspects. This is emphasized by a proposal of the European Commission for a revision of the Public Sector Information (PSI) Directive.[37] The revision already addresses many of the before-mentioned barriers of Open Data. In addition, it aims to expand the amount of available Open Data. This includes especially data generated by organizations, which are partly or fully funded by public money and still refrain from making relevant data available. Typical examples are transport sector data and public research data. Furthermore, it aims to enable the provision of real-time and dynamic data, which is considered the most valuable type of Open Data. Alongside with such organizational frameworks, Open Data stakeholders continuously improve the entire Open Data lifecycle. A key aspect is the improvement of data and metadata quality. Among other things, this is approached by standardizing metadata and data standards. Data providers will be better supported with integrated tools and easy workflows in order to ensure high-quality data from the beginning.[38] In addition, Open Data services and portals shift the actual data into focus, offering integrated preview and analysis features, allowing an immediate evaluation and reuse of Open Data. The foundation for such functionality will be an improved application of Linked Data principles and semantic annotation on metadata and data. Open Data with the highest quality will be more likely reused by all stakeholders, leading to a broader awareness and act as an incentive for new stakeholders.

[37] https://ec.europa.eu/digital-single-market/en/proposal-revision-public-sector-information-psi-directive

[38] https://theodi.org/article/to-improve-open-data-help-publishers/

3. IDS and Open Data

Section 2 elaborated on the current state of Open Data. While Open Data is a field that is well researched, government agencies and especially corporate Open Data is facing barriers. Section 3 provides an overview on the potential that a consolidation of Open Data and the IDS has.

Open Data as an Enabler

While promising vast possibilities for companies, International Data Spaces are in a developing state with a need for more practical real-world applications. Commercial implementations of IDS concepts are being released and are proving themselves in a market environment. However, the barriers to entry are still high. To foster commercial data exchange on a larger scale, offering a way to exchanging data without negotiating contracts lowers this barrier.

The foundation of the IDS architecture is the exchange of data. This includes consumption of data published by external creators, as well as distribution of data collected in business processes. Incorporating Open Data into International Data Spaces reveals new business models to participants. It also offers an approach to actually consume and publish such data in an IDS conforming way. Open Data is not only a use case for companies to integrate into their business processes. It also offers an easy access to exchanging data in International Data Spaces. Open Data consequently acts as an enabler to encourage companies to start sharing their data using the IDS architecture.

Secure data exchange is fundamental to the impact of International Data Spaces. Security is established by building trust between participants and by the option to negotiate binding contracts for data usage. These possibilities provide vast opportunities for the creation of new business processes. However, removing the legal hurdles from the setup process presents a demonstration use case to interested IDS participants. The development of International Data Spaces benefits from enabling companies to easily publish data. Additionally, efforts to increase the amount of corporate Open Data benefit from the incorporation into IDS.

While government agencies are encouraged to release their data as Open Data, private companies are traditionally on the consumer side. The perspective of opening internal data, which would allow insights into business processes, production capacities, or development stages to competitors is undesirable. On the other hand, investing into identifying data that would profit from an open availability is a step towards new business models. An example for the kind of data to publish as Open Data are catalogs of goods. These catalogs are usually already available to interested entities. The potential consumer group is therefore enlarged by publishing the data on a central platform. The data provided as such directly affects IDS related use cases such as the digital supply chain, which aims to enable dynamic restructuring of supply chains. The Connector structure reverses the widely used API infrastructure, which requires the data owner to implement the platform's API in order to deliver the data. Using the IDS, the data owner actively opens up the desired data using their own connector. The central Open Data platform eventually consumes the data by querying the owner's connector.

In conclusion, Open Data and the International Data Spaces accelerate each other by increasing the amount of data shared via the IDS infrastructure as well as increasing the overall amount of corporate Open Data published.

Using Open Data

Traditionally, Open Data use cases aim to process data to present them in an accessible way for the end user to consume. Examples include applications displaying information about road conditions, water quality or locations of specific service providers, e.g. day care centers. Open Data, however, also has use cases in real world industrial scenarios. For example, in the context of the Fraunhofer IDS research projects, three such use cases are developed. Possible applications for Open Data in said use cases are described below. In all of these use cases, using Open Data improves the performance of the system. Furthermore, data is created in each use case. Publishing this data again is directly beneficial for the system.

Digital Supply Chain

The Digital Supply Chain use case aims to enable companies to orchestrate their supply contracts. This allows them to effectively react to changes in demand, availability, etc. Open Data about traffic allows transport partners to effectively plan delivery routes, order capacities, etc. by dynamically calculating timetables, idle times and more. The use case provides means to dynamically change contractor networks without losing sovereignty about company insides. Not all data exchanged is sensitive, though. Opening up information from the supply chain can improve service quality beyond companies.

Smart Urban Mobility / Logistics

Transportation and delivery of goods pose a high risk to the capacities of urban transport infrastructure. Especially when not orchestrated in an intelligent and interconnected way. Open Data aids the decision making in route planning by providing Open Data about traffic. This data can be used as a baseline to anticipate traffic jams or by pointing out highly sensitive traffic hubs, prone to causing collapse of traffic when used extensively without strict planning.

Corporate logistics data promise to help further reduce overstrain of said sensitive traffic areas. By opening up corporate gathered data on traffic, orchestration of transport and delivery of goods can be refined substantially. The urban mobility / logistics use case is therefore a promising example of the impact of corporate Open Data on the participants in International Data Spaces.

Smart Urban Mobility / Humans

Digitization leads to enhanced human mobility through combination of public transport and shared vehicles. Open Data improves decision making and helps distribute passengers across the available mobility providers to effectively utilize capacities. By using Open Data about traffic, passengers can be redirected to street independent vehicles, e.g. subways. Open Data provided by the public transport organizations on the other hand can indicate technical difficulties and guide passengers to use the nearest bus, shared car, etc.

Cities like San Francisco or Beijing are well known for environmental problems related to traffic but German cities as well have taken action against increased fine dust loads. Open Environment Data helps in the reduction of exhaust gases by routing passengers around areas with high air pollution. On the one hand this reduces carbon dioxide concentration on strained areas. On the other hand, routing cyclists and pedestrians around said areas actively improves national health.

The use cases mentioned so far illustrate only a small portion of the potential benefit Open Data has as a component of the International Data Spaces.

IDS Scenarios

Projects in the context of International Data Spaces and related initiatives have already developed artefacts and platforms, which deal with Open Data.

Advaneo[39]

The Advaneo data marketplace offers different kinds of data to participants based on their subscription status. Company users access the data marketplace via an IDS certified component and exchange data under consideration of data usage policies attached to the datasets. Additionally, the data marketplace offers Open Data. This data can be consumed by certified company users, as well as private and personal users not affiliated with a certified company. The Open Data available on the Advaneo marketplace stem from common fields, e.g. health, government, population, environment, etc.

Urban Data Space[40]

The Urban Data Space (UDS) is a virtual room in which all urban data can be exchanged safely and linked between relevant stakeholders. It supports proactive provision, secure exchange and easy linking of data in urban areas. Urban data include freely available data (Open Data), commercially available data as well as internal urban data (e.g. emergency management). The Urban Data Space is integrated organizationally into the municipal processes and, as an infrastructure, establishes data sovereignty for local and regional actors. New business models can thus be initiated sustainably at the local level.

Both, the UDS and IDS have a number of similarities. For example, both concepts aim to create an open platform that is user-driven and collaboratively developed. In addition, IDS and the UDS both emphasize and guarantee data sovereignty of data publishers and data users. However, concerning data sovereignty, there is one exception in the UDS regarding government data: Based on the PSI-Directive (2003/98/EC), certain government data is required to be made available freely as Open Data.

Another common aspect of IDS and UDS is the ambition of creating greater trust in data exchange. For this purpose, the components used are checked by certification procedures and trust & identity management systems are being used as well. Scaling and network effects are further goals shared by both concepts: The ICT ecosystems of both IDS and UDS are supposed to be dynamic and scalable. In the long term, both ecosystems should contain a large number of different services and applications.

FOKUS Open Data Connector[41]

The Open Data Connector developed at Fraunhofer FOKUS presents a first solution to offering Open Data in International Data Spaces. The purpose of the connector is to offer Open Data via the IDS architecture. It therefore contains an adapter to a CKAN backend as well as a GUI for importing and publishing data. Data consumption in turn is performed by querying the connector's self-description.

IDS as Open Data Technology

International Data Spaces and Open Data ecosystems address different problems occurring with data sharing. However, the underlying technologies developed to accomplish the approached subject are similar in many aspects. They especially resemble each other in terms of building metadata repositories to spread information about available participants and data. However, to achieve the problem of broadcasting the information, both architectural styles apply different strategies: State of the art Open Data platforms use a pull approach, while

[39] https://mvp.advaneo.de/
[40] https://www.fokus.fraunhofer.de/de/fokus/projekte/urbane_datenraeume
[41] http://ids-connector.fokus.fraunhofer.de/

the IDS architecture applies a push approach.

The following section describes how the concepts of the IDS can be applied to the domain of Open Data, in which aspects they are alike and how approaches pursued by the IDS can be applied to Open Data.

Openness of IDS Components

The fundamental architecture of International Data Spaces inherently follows principles of Openness. The implementation base is accessible by everyone. However, it is to note that the concept of trust is established by certification done by an appointed authority. Some aspects of trusted data exchange are therefore closed.

The fundamental definition of components and communication in the International Data Spaces is given by the information model, which is published as an open ontology.[42] All connectors and brokers are implemented using said information model and openly publish their metadata. Communication, as well, is accomplished by describing the content in accordance to the model. Even data apps are described in this way and the description can potentially be published openly.

Data Sources

Open Data is made available for download via various Open Data portals. Currently, Open Data portals can be sorted hierarchically by degree of specialization. For example, city administrations publish their data on the corresponding Open Data portal of the city. The data of multiple city portals are then in turn merged and published on country wide Open Data portals. Those are then merged and published in continental portals, etc.

Harvester

The current state of the art process to gather information to be published in a specific portal is called "harvesting". During the process of harvesting, Open Data portals gather the metadata from other portals or directly from the producer. This allows portal operators to build up portals dedicated to special topics or use cases, e.g., gathering all Open Data published in a country or field.

Broker

In the IDS architecture, the broker component performs a similar task, namely making connectors locatable to participants interested in acquiring the connector's data. For this purpose, the broker can likewise aggregate connectors offering data of specific characteristics, e.g. logistics data, supply data, stock data, etc.

While the two technologies are similar in the objective they accomplish, the core principle is diametrically opposed. The IDS architecture takes a "push" approach where the data producer informs the central entity about their existence. The current Open Data state of the art follows a "pull" approach in which the central entity gathers the information it aims to assemble.

Data Storage

An important topic in the context of International Data Spaces is data storage and the associated topic data sovereignty. While the problem of data sovereignty does not play such an important role in the context of Open Data portals, availability of data is an issue.

Open Data Metadata Portals

In the field of Open Data and the interlinked topic Open Data distribution, the state of the art is to distribute metadata amongst different data portals, only. The actual data

[42] https://github.com/IndustrialDataSpace/InformationModel

is made accessible via a webserver by the data producer or at a data portal. The data is then directly hyperlinked to.

Since the portals take care of gathering the metadata from other portals, information about the availability of data may be lost over time. If the description of a dataset is removed from one portal, the change can only be adopted by other portals after additional harvesting. As a result, dead links may appear in Open Data portals.

Connector Resource Catalog at the Broker

In the IDS architecture, the distribution of information about available data is handled in a similar way. The connector's available data is depicted in the self-description of the connector. This self-description document in turn is communicated to a broker where it can be obtained by interested participants. The data itself remains at the producer until a data exchange is negotiated and initiated.

Benefits of the "Push" Approach

As described before, the IDS follows a "push" principle when publishing data. This approach, on the one hand, leads to increased management options for the data producer. One important option is the possibility to un-publish data by informing the portal operator about the unavailability. This, on the other hand, leads to higher availability in data portals, as the un-publishing of data is directly communicated to the portal operator. Affected metadata can then be removed from the portal's system, i.e., GUI, search index, etc.

Reversing the responsibility of communicating the availability of data, furthermore, yields standardized ways of distribution. Producers can easily register their Open Data Connector to the broker of a portal and thereby apply for the data to be published on said portal. It also releases portal operators from the responsibility of gathering information about the available datasets, published by the data provider or in Open Data portals.

As described above, this leads to higher availability of data in data portals, as well as an undiminished representation of the data catalog.

Standardization

In the first section, several standards and specifications for publishing Open Data are described. Particularly relevant is DCAT-AP as a standard for describing public sector datasets. DCAT-AP is heavily promoted by the European Commission, already widely adopted and is supposed to become the standard for Open Data interfaces. Like the IDS information model, DCAT-AP is based on the Data Catalog Vocabulary (DCAT). Therefore, IDS and DCAT-AP share the same core vocabulary for describing the available data. This allows straightforward interoperability and enables IDS to be fully compatible with DCAT-AP on a data description level.

Improving Industry-Public Data Integration

At the moment, Open Data and industry data oppose each other in multiple aspects. By bringing them together on a technological level, new possibilities for integration and reuse emerge. Open Data can then be created and processed with the same tools and applications, which are already available in the industry. This can lower the barriers for using and providing Open Data. Hence, by applying IDS specifications to Open Data publishing, the entire value creation process can be streamlined and optimized.

Increasing Trust

Trust and verifiability are not the most demanded properties for Open Data users, since the data is usually acquired from official web portals. Still, a clear provenance and check of integrity is desired. Especially, when Open Data is used to draw new conclusions and insights. Existing Open Data standards do cover these aspects only marginally. The IDS architecture emphasizes such aspects in its very core concept, e.g., by applying a certification scheme. Thus, the use of the IDS architecture for Open Data can upgrade the credibility of the data and consequently increase the value of derived products and data.

4. The IDS Open Data Ecosystem

Section 3 described how Open Data contributes to the success of International Data Spaces as well as its use cases and further illustrated how Open Data and International Data Spaces fit together. In this section, the concept, features and architecture of a practical IDS Open Data ecosystem are presented.

The Open Data Connector thereby embodies a toolset for easily opening existing data sources to the IDS Open Data ecosystem.

User Stories

The user stories that are eliciting the requirements for the IDS Open Data ecosystem are described below. The stories are structured based on the five user groups that were identified.

Data Steward

The Data Steward role contains participants of the ecosystem publishing data to the Open Data ecosystem. They are either affiliated with an IDS participant or implement IDS solutions for publicly available data.

Data Stewards want to publish data, stored in common internal data management systems, e.g., databases, file systems, etc. Further, they wish to publish data, which is available in common public data management systems, e.g., CKAN, Eurostat, etc. For convenience, an easy-to-use GUI for describing the content and type of the data should be provided. Data Stewards need to be able to supply origin, format, license, and other metadata. Additionally, they want to configure the scope of the data they are publishing. Furthermore, Data Stewards want to register the connector to one or multiple brokers to enhance the visibility of the data published. They want precise information about who, i.e., which connector or IP, has accessed the published data to infer information about the relevance of a given dataset. Analogous to publishing data via the connector, Data Stewards need to be able to un-publish data from the connector and the broker(s) to ensure data sovereignty inside the open data ecosystem. To ensure that the published data is up-to-date, a scheduling process for refreshing the offered data is necessary.

Administrator

The Administrator role covers a person undertaking administrative tasks for the Data Steward or their organization. Administrators want a generic "base distribution" of the connector that can be brought up without much configuration so that insensitive data can quickly be published on the IDS. "Convention over configuration" should be applied. They want to adapt to well-known data management solutions on behalf of the Data Steward to help them make data accessible. Additionally, detailed logs about data access and need to be able to configure the access frequency of a dataset, need to be accessible.

Provider-Side Developer

The Provider-Side Developer describes a developer affiliated with a Data Steward or their organization. The Provider-Side Developer wants to comfortably hook up not supported data management systems to the connector.

Data User

The Data User role describes participants of the ecosystem consuming Open Data using a connector. Data Users want a GUI which allows them to easily browse the broker for all the available open datasets and connectors. Instead of displaying the metadata in the form of the IDS Information model, a GUI should display the metadata in a human-readable way. Additionally, Data Users want to search for and find data fitting their needs and use cases. They need to be able to not only query the metadata, but directly query the actual data from the publishing connector.

Consumer-Side Developer

The Customer-Side Developer describes a developer affiliated with a Data User or their organization. Consumer-side developers want a RESTful interface so that they can programmatically query metadata of the offered datasets.

Applied IDS Artefacts

The Open Data Ecosystem is based on the IDS architecture and therefore encompasses a variety of concepts and artifacts stemming from said environment. Which IDS components are utilized during design and development and which roles they take in the resulting ecosystem is described below.

Information Model

The IDS information model describes the relations and attributes of the components described by the reference architecture[43] and the IDS handshake document[44]. The information model therefore defines how components have to be modeled and how communication needs to be performed in order to comply with IDS standards. In the implementation of the Open Data ecosystem, the information model will play a defining role in modeling the backend architecture and data exchange.

Connector

The connector component of the IDS reference architecture is the central entity acting as a gatekeeper and ensuring save and trusted data exchange between IDS participants. In the Open Data ecosystem, the connector plays the role of the local provider of data. Each data providing participant uses an instance of the connector to grant access to their data to the public. The connector then accesses the desired data on the backend data management systems and offers them to the ecosystem.

Data Apps

In the reference architecture, Data Apps are defined as applications running inside connectors to perform manipulation of the data being accessed, as well as acting as adapters to backend data storage systems. They are encapsulated to prohibit interference with other apps and the other processes being performed in the connector.

In the ecosystem, transformation Apps are an optional instance during data exchange since, under certain circumstances, a data manipulation, e.g., format transformation, may be necessary. Particularly with regards to extensibility, transformation apps become important. Data gathered from proprietary data sources are not necessarily available in common data formats and therefore need transformation. Additionally, data stemming from internal sources may have to undergo anonymization processes in order to be published openly.

Broker

The IDS Broker offers the possibilities for users to find connectors of other participants as well as making their own connector locatable. The broker offers means to register and unregister a connector, updating the connector's description and querying it for other connectors.

In the Open Data ecosystem, the connector of the participant registers at a broker and regularly updates its self-description in order to refresh the catalog of data offered. The broker is part of an Open Data portal, which processes the given information to make them available in a human readable format.

[43] https://www.internationaldataspaces.org/publications/ids-ram2-0/

[44] available to IDS members only

Data Sources

Since data is stored in diverse management systems, the Open Data connector needs to be able to enable publishing from a variety of data sources. As a base implementation, the Open Data connector supports three widely used data management systems. Additionally, it provides connections to two specific Open Data sources. These sources do not use CKAN and therefore provide an example for implementing specific adapters.

Internal Data Sources

Internal Data Sources describe the data management systems available inside an organization. Data in these systems is not usually open to the outside world.

PostgreSQL

In contrast to formatted data stored in files, databases centrally store the majority of raw data accumulating in organizations. Data stored in databases needs to undergo a transformation process before delivery to a data consumer. These transformations include but are not limited to formats such as CSV or XLS. Databases can either be queried regularly to create the files to be published or the querying and transformation can be applied in real-time when a data request is received. By default, the IDS Open Data ecosystem will support PostgreSQL.

File System

In addition to data stored in databases, large quantities of data is collected in single files on local or network machines. In contrast to data stored in databases, files can directly be accessed and transferred without undergoing further transformation. However, for the sake of data quality, readability and usability, it is desirable for the files to undergo a company internal review process. Files are uploaded to a separate server instance to make them available via the connector.

Public Data Sources

Public Data Sources describe data management systems in which data is already available to the public. In these cases, the connector's purpose is to ensure distribution and availability of the data.

CKAN

CKAN has been described in section 3 as the de facto standard for publishing open data. Metadata is stored inside a CKAN database while the actual data is accessible by providing hyperlinks to external data sources. The management responsibilities of CKAN and information brokers are therefore similar. The CKAN REST API offers methods to make the stored information available via the connector.

Eurostat

Eurostat is the statistical office of the European Union, publishing statistical data gathered in the EU. The comprised themes include transport, agriculture, finance, etc. Eurostat provides a REST API for convenient querying of the published data.

DWD-Geodata[45]

In addition to data stored in data management systems and published publicly, a large quantity of data is published on conventional HTML websites, i.e., in download sections, ftp servers or simple HTML directories. Publishing this data therefore requires a high degree of customization. The Open Data service of the German Meteorological Office publishes its near real time data as a hybrid of an ftp server and an html-based visualization of the data.

[45] https://opendata.dwd.de/

Concept and Architecture

The IDS Open Data Ecosystem is built on the concept of decentralized data storage to ensure the data sovereignty of the data publisher. The data publisher therefore remains in control of which data they want to share and is responsible for publishing the metadata to one or multiple data portals. The concept of Open Data in the International Data Spaces is described below.

Ecosystem Architecture

The architecture encompasses a combination of connectors and brokers. Every organization, which wishes to share data with the public, sets up a connector and configures the access mechanisms to the selected data. The connector is not limited to a single type of data management system, but is able to publish data from multiple data sources.

The broker acts as a central point for data consumers to collect information about the available data. The connector therefore registers to a desired broker by sending its self-description document. The self-description document includes the whole data catalog of the published data. Data consumers can query the broker by using an IDS connector and requesting the broker's self-description. Alternatively, information about available datasets can be found by using a GUI presenting the data catalogs in human readable form.

Scheduling Updates

The described architecture allows for various ways of providing data. Data can be stored on separate storage devices, e.g., to publish files, stored on a local machine, or can be accessed directly from the data management system, e.g., from a database. The latter case even allows for near real time data to be published, by making the connector query the latest dataset from a database and transforming it into the desired file format on demand.

Whatever the case is, the data undergoes additions and changes over time. The Open Data ecosystem allows for easy change of the data since the data is conveniently stored at the data infrastructure of the provider. Additionally, metadata registered at a broker can easily be updated by simply updating the self-description document of the connector at the broker. The connector allows for defining the scheduling interval of updating the data locally as well as scheduling the interval of updating the self-description at the broker(s).

Component Internals

As described before, the proposed IDS Open Data ecosystem consists of two main components: the broker and the connector. The components are deployed independently and implementation and extension of one component does not immediately affect the other.

Connector

The Connector is designed to be extensible. It innately supports a variety of data sources and additionally offers the possibility to extend it with adapters customized to support in-house data storage solutions, APIs, etc. The Connector furthermore supports Data Apps to perform manipulation of the data.

The Data Apps and adapters are encapsulated, running in separate containers and having their own personal storage. This ensures that the separate Apps do not interfere with each other. All administrative tasks of the connector are also performed by separate services. There exists a message routing component which receives the data request and routes it to the corresponding input adapter. Additionally, there exists a component, which generates and keeps the connector self-description up to date.

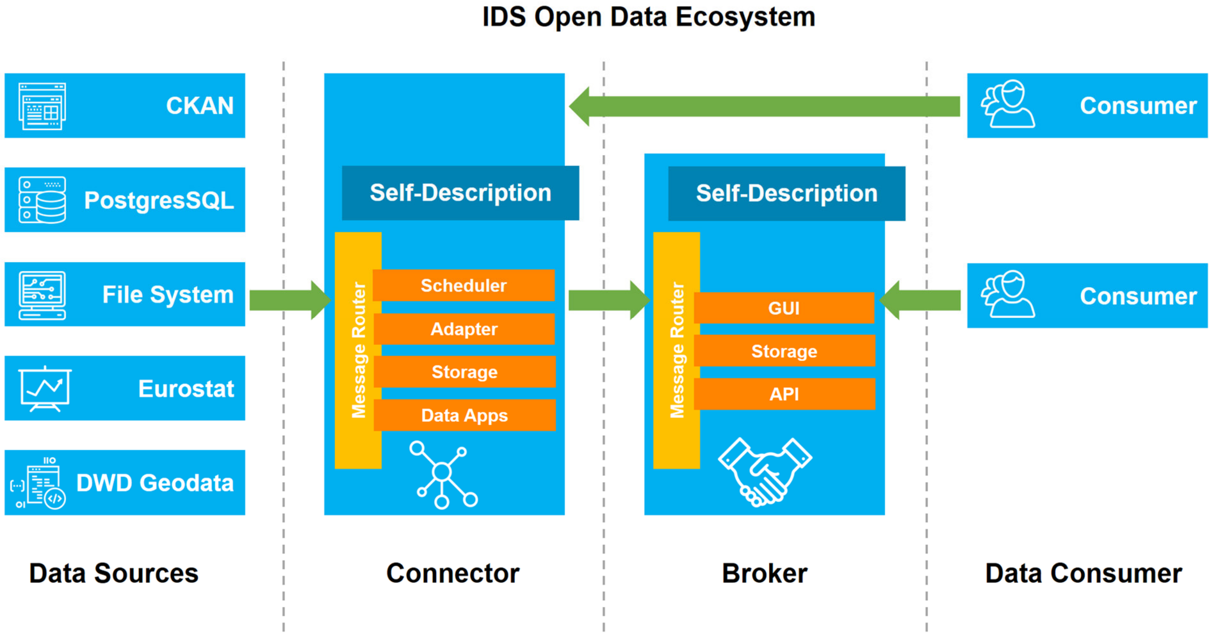


Figure 1 Overview and Architecture of the IDS Open Data Ecosystem

Broker

The purpose of the broker is to receive the registration messages from the connectors, storing the self-description information and exposing them for the GUI to display. The broker therefore consists of the following separate components:

A message routing component, similar to the one used in the connector, receiving the input messages and routing them to either the storage component or the self-description component. The storage component receives the self-description documents from the registering connectors and stores the metadata.

Eventually, there also exists a service for generating the self-description document for the broker. The broker's self-description is consumed by the GUI to display the registered datasets. Additionally, connectors can directly consume the self-description of the broker. This way, the manual step is bypassed and the collection of metadata can be performed fully automatically.

Technology

Details about implementation of the ecosystem components are out of the scope of this position paper. A brief overview over the architectural concepts as well as technological details are given below.

Microservices

Encapsulation and modularization have a large share in the establishment of trust and security in the IDS connector. Modularization does not only aim to ensure that different modules do not interfere with each other's processes, but also provides an easy way of extending the connector by injecting new modules to the system.

To achieve this modularization, the IDS Open Data ecosystem makes use of the architectural technique of microservices. Each module deployed in the connector or broker is implemented as a service of its own and follows a lightweight communication protocol. To achieve this modularity, the ecosystem is built on the Docker[46] containerization engine. Docker allows the connector to be deployed on a huge variety of machines by virtualizing the runtime environment of each container deployed. As a result, the connector can easily be brought up on the hardware of a data provider.

Certification

After implementation, IDS Open Data ecosystem will apply for its IDS-ready assessment within the IDS certification scheme.

Data security and data sovereignty are the fundamental characteristics of the International Data Spaces. As such, participants within the Industrial Data Space must use certified components[47] (e.g., the "International Data Spaces Connector") in order to securely exchange data in a sovereign way.

For core components to provide a sufficiently high level of security regarding the integrity, confidentiality, and availability of data exchanged, evaluating and certifying is vital for the functioning of the International Data Spaces. This necessitates a certification scheme defining all processes, rules and standards governing the certification of core components in order to ensure a consistent evaluation and certification process for all core components (i.e., Connectors and Brokers).

To build trust in the intended cross-industrial and cross-company information exchange, IDS core components must verifiably provide the required functionality and an appropriate level of security. As such, the core component certification is interoperability- and security-focused, while aiming to strengthen the development and maintenance process of all IDS core components.

The matrix certification approach as shown in Figure 2 was defined for the core components certification. The approach ensures on the one hand a low entry barrier specifically suitable for SMEs and on the other hand a scalable certification to meet high information security requirements.

The depth and rigor of a component evaluation consists of three assurance levels:

- Checklist Approach: The vendor of the component validates the claims made about the implementation via a security feature checklist. Additionally, an automated test suite will be used to verify the component's security features.
- Concept Review: An in-depth review is performed. The review includes an evaluation of the provided concept as well as practical functional and security tests.
- High Assurance Evaluation: In addition to the functional and security tests, an in-depth source code review will be performed. Furthermore, the development process will be evaluated, including an audit of the development site.

The second dimension of the approach defines three security profiles:

- Base Security Profile: This profile includes basic security requirements, like limited isolation of software components, secure communication, mutual authentication, and basic access control.
- Trust Security Profile: This profile includes, amongst others, strict isolation of software components, secure storage of cryptographic keys, secure communication, usage control and trusted update mechanisms.

[46] https://www.docker.com/

[47] While the IDS certification scheme encompasses certification both of participants and components, for the context of this Position Paper, the focus will be on the component certification.

Assurance →

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

Figure 2: Certification Approach for Core Components

- Trust+ Security Profile: This profile additionally requires hardware-based trust anchors and supports remote integrity verification (i.e., remote attestation).

For the Open Data ecosystem, the target level is verification of the requirements from the Base Security Profile via a Concept Review.

An in-depth description of the core component certification approach and how it applies to the key elements of IDS architecture can be found in Part 3 of the White Paper Certification[48]. The certification criteria catalog for Core Components is freely available for all IDSA members.

Availability and Documentation

The Open Data ecosystem, including software, source code, demos and documentation will be gradually made available here:

http://ids.fokus.fraunhofer.de

48 https://www.internationaldataspaces.org/publications/whitepaper-certification/

5. Summary and Outlook

This position paper has given a comprehensive overview about the combination of Industrial Data Spaces and Open Data. Open Data is mostly published by stakeholders from the public sector and currently only little by the industry. Several specifications and software standards exists for publishing Open Data, like CKAN or DCAT-AP. Although its great economic potential, Open Data has to face many barriers, mostly regarding data quality and availability. The IDS architecture can be used to overcome essential barriers and enable the industry to become an Open Data user and provider. Many IDS use cases can benefit from integrating public data, e.g. using traffic data in a digital supply chain. In addition, the engagement with data in the context of IDS can create awareness about Open Data, leading to the creation of industrial Open Data.

IDS offer a variety of concepts and artefacts for trustful data sharing. Although many features relate to confidential data exchange, the very core mechanisms are suitable matches for building Open Data ecosystems. This includes the openly available specifications, the concept of decentralization and the standardized interfaces. Furthermore, the IDS Connector-Broker architecture allows an inversion of control. The data providers actively push their data offering to a central platform while keeping sovereignty. This increases quality and timeliness in comparison to established pull mechanisms. The underlying data structure of IDS is the Data Catalog Vocabulary (DCAT), which is also fostered by the European Commission as standard for public sector Open Data.

A practical implementation and adoption of IDS concepts in the context of Open Data requires concrete guidelines and software artefacts. Therefore, a straight-forward architecture for an IDS Open Data ecosystem is proposed. In its core it applies and consolidates the IDS artefacts information model, connector, broker and data apps to an out-of-the-box solution for IDS-conform data publishing. It supports generic and exemplary data sources, like PostgreSQL, file system and Eurostat. The connector will be certified based on the requirements from the Base Security Profile of the IDS certification approach.

The proposed architecture will be successively implemented and published as open source software. It will act as a transparent implementation of IDS artefacts and represents an entry point for potential IDS Open Data providers.

This paper presents the current state, and findings and designs will be progressively refined as the IDS architecture evolves and practical implementations yield to new insights.

OUR MEMBERS

orbiter
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
pi-lar GmbH
POLITECNICO MILANO 1863
pwc
QuinScape

realworldsystems
RI. SE
REWE GROUP
RITTAL
SALZGITTER AG Stahl und Technologie
SAP

SCHAEFFLER
SETLOG The Smart Connective
SICK Sensor Intelligence.
SIEMENS Ingenuity for life
Signal Solutions
SINTEF

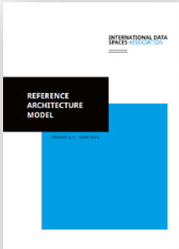
SMEV AG Smart Mobility Evolution
TARC
tecnalia Inspiring Business
IK4 TEKNIKER
T · ·
mtc Manufacturing Technology Centre

thyssenkrupp
TNO innovation for life
TÜV NORD GROUP
TÜV SÜD
UNIKLINIK RWTH AACHEN
UNiSCON Sealed Cloud Technologies

UNITY CONSULTING & INNOVATION
universität Witten/Herdecke
University of Amsterdam
UNIVERSITY OF TWENTE.
vastuu group
VDMA

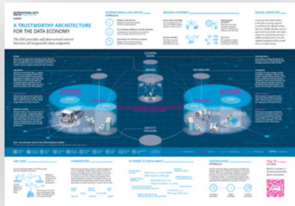
VOLKSWAGEN AKTIENGESELLSCHAFT
VTT
welotec a byte smarter
WeTech IDS-China
ZVEI: Die Elektroindustrie

OVERVIEW PUBLICATIONS


Reference Architecture Model


Executive Summary


Image Brochure


Infographic

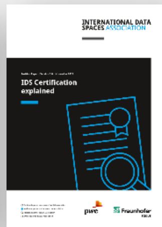
Use Case Brochures


Study on Data Exchange


Position Paper Implementing the European Data Strategy


Position Paper GDPR Require-ments and Re-commendations


Position Paper Usage Control in the IDS


Position Paper IDS Certification Explained


White Paper Certification


Sharing data while keeping data ownership







Magazine Data Spaces_Now!

For these and further downloads: www.internationaldataspaces.org/info-package

Code available at: https://github.com/industrial-data-space

CONTACT

Head Office

INTERNATIONAL DATA SPACES ASSOCIATION

Emil-Figge-Str. 80
44227 Dortmund | Germany

phone: +49 231 70096 501
mail: info@internationaldataspaces.org

WWW.INTERNATIONALDATASPACES.ORG

@ids_association

international-data-spaces-association