

Sascha Meckler

Solid for Dataspaces

Gefördert durch

Bayerisches Staatsministerium für
Wirtschaft, Landesentwicklung und Energie



Solid for Dataspaces

Agenda

1. Dataspace concepts

- DataSpace Support Platform
- Web of Data
- International Data Spaces (IDS)
- GAIA-X
- Social Linked Data (Solid)

2. Solid Data Space (SDS)

3. Comparison between the SDS and IDS

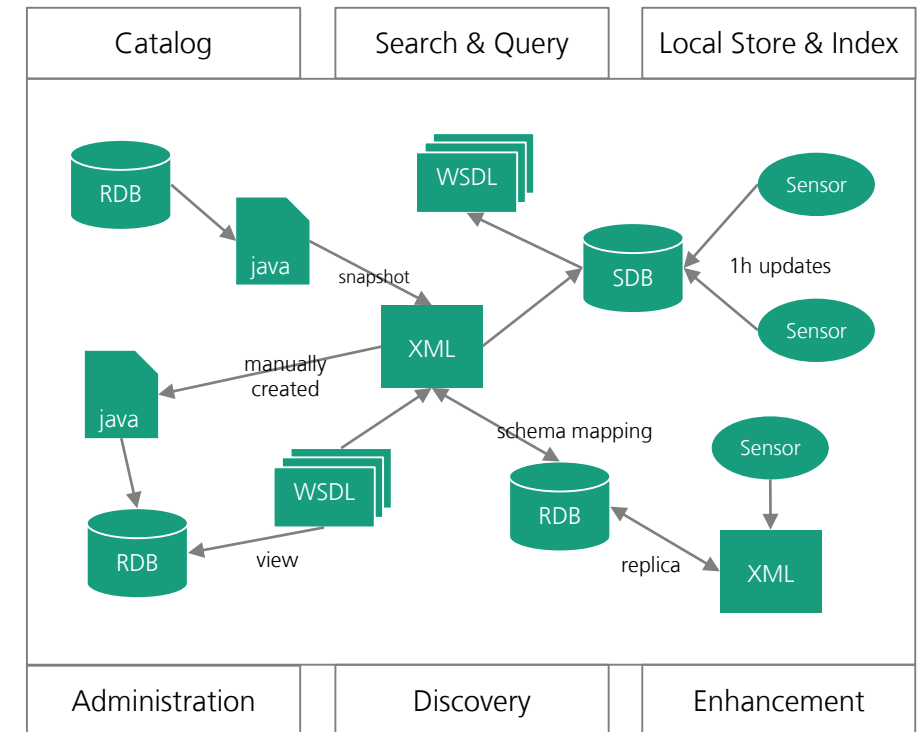
4. Conclusion and Outlook

Dataspace concepts

DataSpace Support Platform (DSSP)

Dataspace management system

- 2005, 2006
- Information integration system for multiple data sources of one person (PIM) or organization
- Co-existence of multiple schemas
- Best-effort answers and “pay-as-you-go” integration (Halevy, Franklin & Maier 2006)
- “*Dataspaces can be seen as an umbrella for much of the research that is already being actively pursued in the database community*” (Franklin, Halevy & Maier 2005)



Example dataspace and the components of a dataspace system, adapted from (Franklin, Halevy & Maier 2005)

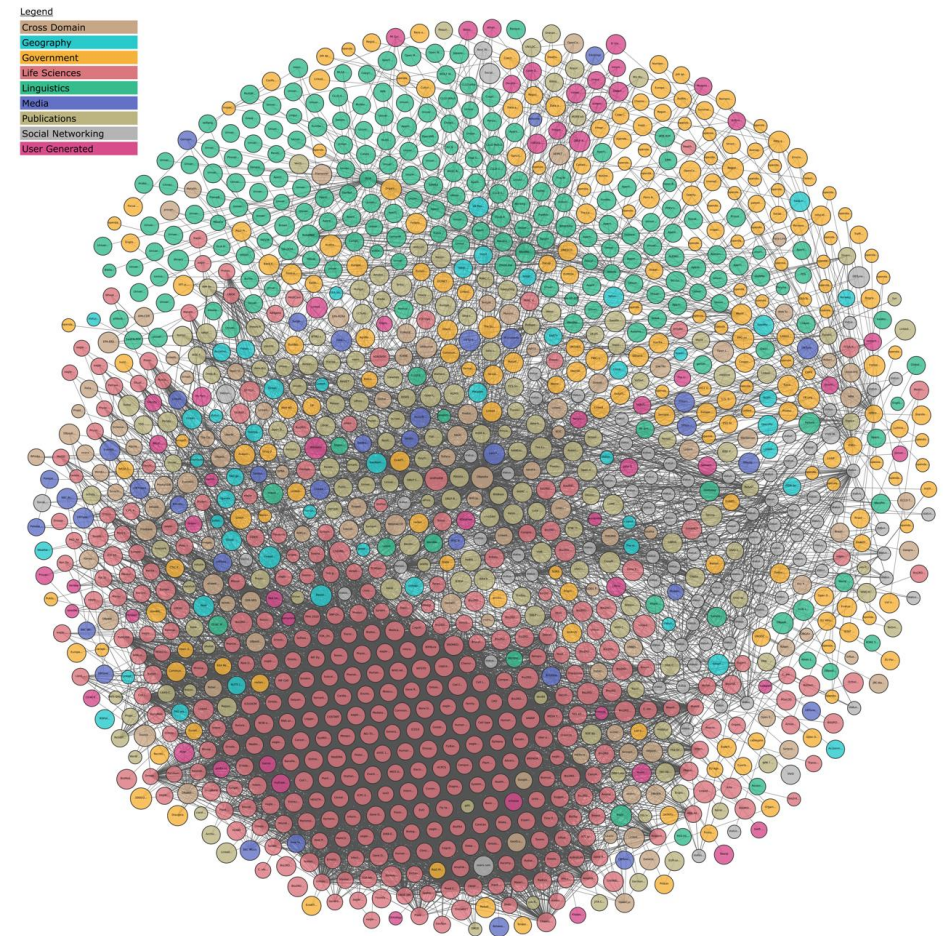
Dataspace concepts

Web of Data

Web of machine-understandable, interconnected data

(Berners-Lee, Hendler & Lassila 2001)

- 2001 – 2011
- Web of Data (as part of the Semantic Web) forms a global data space (Heath & Bizer 2011)
- Publishing Linked Data contributes to the building of a Web of Data
- *“Web of Data can therefore be seen as a realization of the dataspace concept [by Franklin, Halevy & Maier] on global scale, relying on a specific set of Web standards”* (Bizer, Heath & Berners-Lee 2009)



Linked Open Data Cloud (lod-cloud.net)

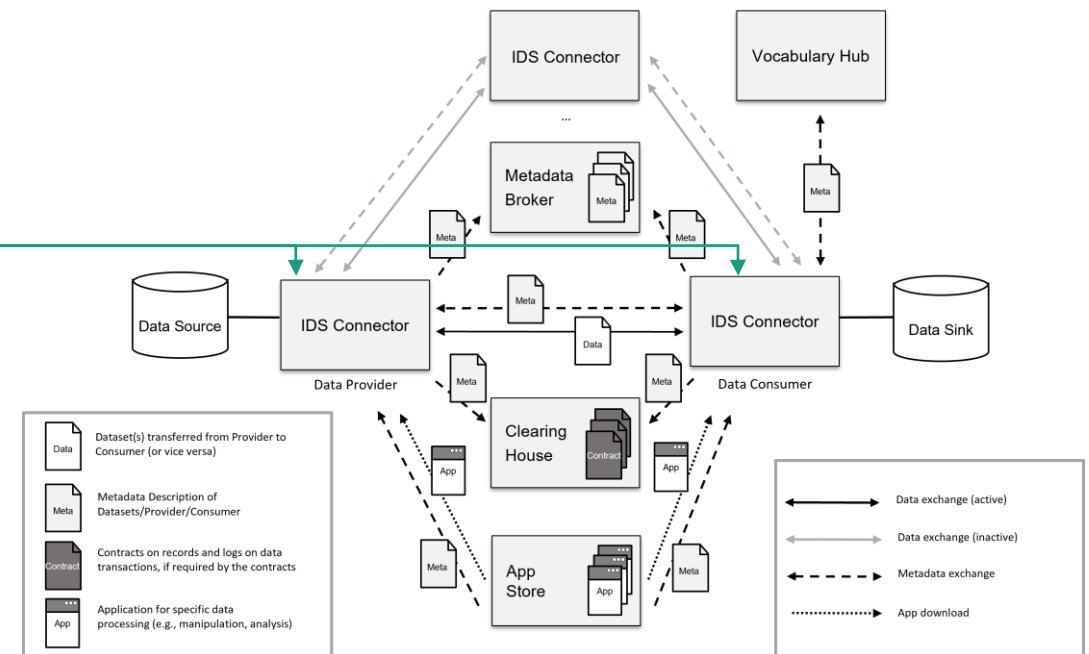
Dataspace concepts

International Data Spaces (IDS)

Architecture for sovereign data exchange

- 2016
- International Data Spaces Association with mostly European members from industry and research (Fraunhofer)
- Focus on data sharing between companies
- Definition of a reference architecture model (IDS-RAM) (Steinbuss et al. 2022) and specifications
- IDS (Data) Connector is the interface to the dataspace that handles data exchange, policy enforcement, app execution
- Certification of components and participants by IDSA and associates to create trust
- Dataspace functions are driven by metadata defined in the IDS Information Model (Bader et al. 2020)

IDS System Layer: Interaction of technical components
(Steinbuss et al. 2022, section 3.5)

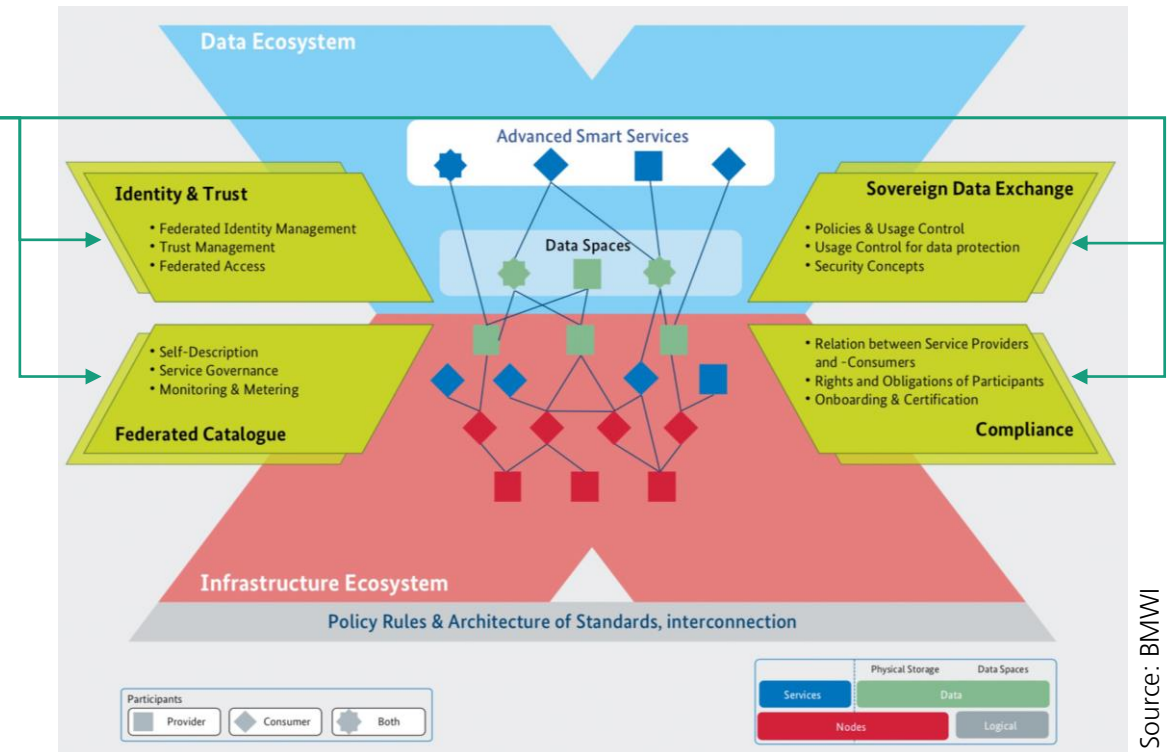


Dataspace concepts

Gaia-X

EU initiative for federated and trusted cloud infrastructure and dataspaces

- 2019
- Gaia-X European Association for Data and Cloud AISBL is an international non-profit association which develops the technical framework and operates the Gaia-X Federation services.
- Architecture Model for a data sharing architecture (GAIA-X 2022) contains many contributions from the IDS (Otto 2021)
- Central, trusted third-parties for the supervision of certificates, policies and data usage
- IDS-RAM provides more sophisticated concepts for dataspaces whereas the Gaia-X architecture additionally provides solutions for data storage and cloud components; IDS components align with the Gaia-X architecture (Otto 2021)



Dataspace concepts

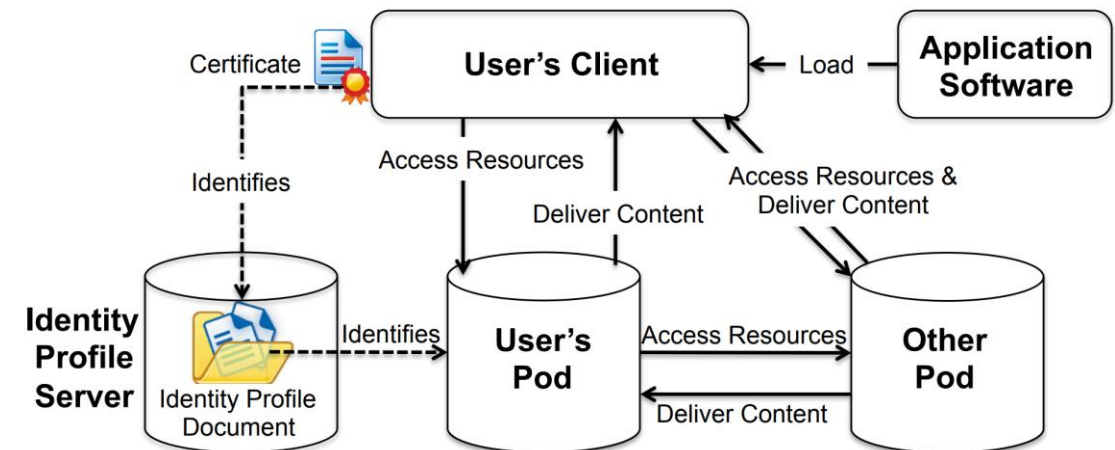
Social Linked Data (Solid)

Web standard that puts people in control of their data

- (2015), 2016
- Storage of data in Personal Online Data Stores (Pods) on the Web
- Paradigm of separating personal data from applications; users can control data access of apps and other users
- Decentralized architecture for identity, storage and applications
- Consolidation and extension of existing (Semantic) Web standards: W3C recommendations for WebID, Linked Data Platform, RDF and SPARQL
- Standardization of a RESTful, read-write Linked Data interface, identity (authentication) and access control (authorization)



Solid logo
(solidproject.org)



Solid architecture (Sambra et al. 2016)

The Web and Linked Data as a Solid Foundation for Dataspaces

Agenda

1. Dataspace concepts

2. Solid Data Space (SDS)

- Technology layers
- Components
- Data catalog example

3. Comparison between the SDS and IDS

4. Conclusion and Outlook

Solid Data Space

Concept

The Web of Data is the ideal foundation for dataspaces!

- Build on stable Web architecture
 - Reuse proven Web technologies and frameworks
 - Keep participation easy
- Gradually extend the agreements for communication and collaboration

Solid Data Space

Technology layers

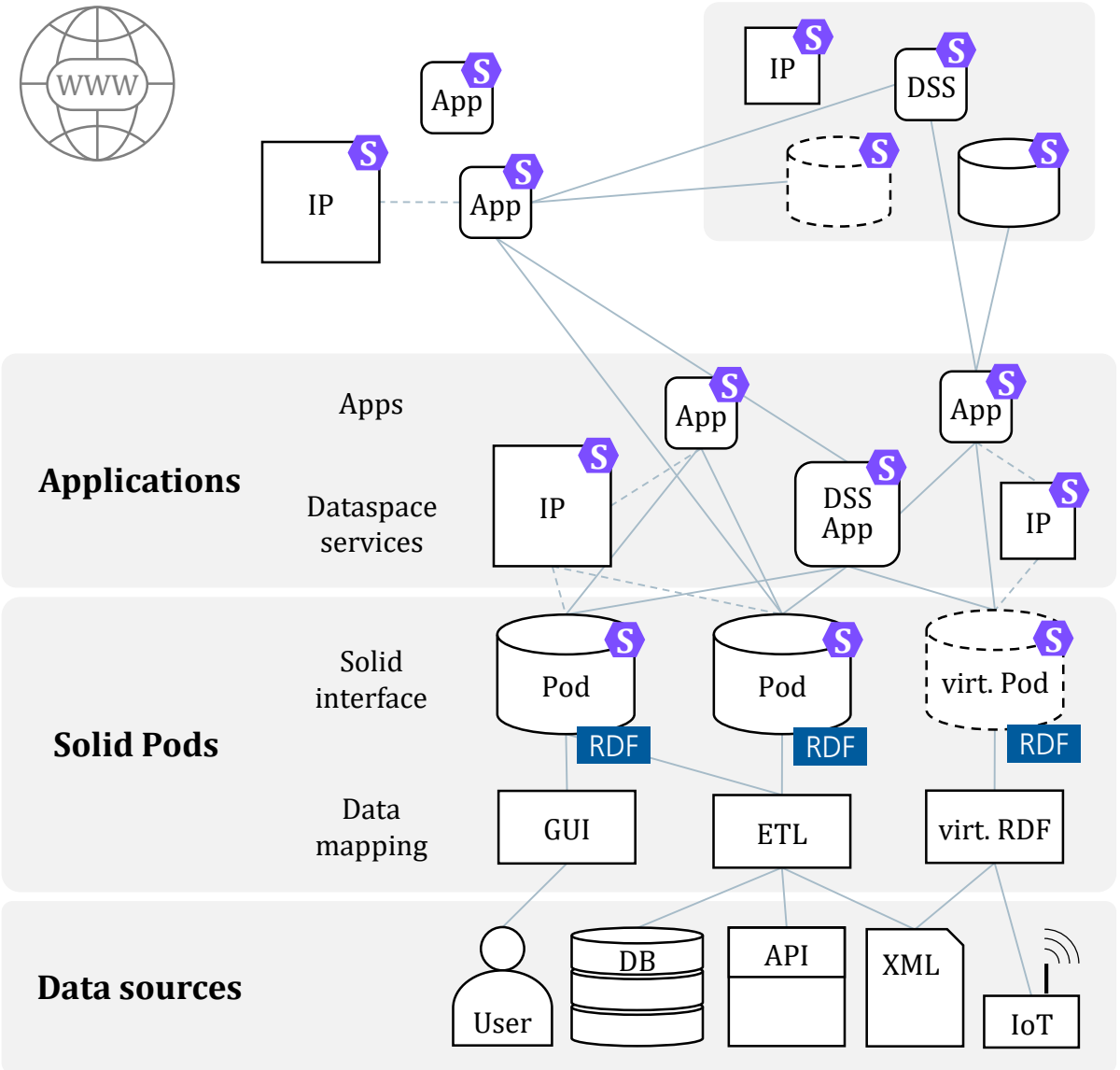
Layer	Interoperability	Agreements on standards and technologies
Dataspace	Application	<ul style="list-style-type: none">• Additional agreements between participants of a specific dataspace• Services for easier data sharing, e.g. data catalog• Authorities that establish trust or supervise data sharing processes
Solid (Web)	Access	<ul style="list-style-type: none">• Decentralized identity (WebID, Solid-OIDC)• Access control (WAC)• Read-write Linked Data API (LDP)
Linked Data	Data	<ul style="list-style-type: none">• Uniform data format (RDF)• Human- and machine-understandable data with semantics• Knowledge representation via ontologies and shared vocabularies (OWL, RDFS)• Linked Data concepts (LD principles and 5-star LD)
Web	Communication	<ul style="list-style-type: none">• Uniform communication protocol (HTTP, HTTPS)• Communication security and certificate authorities (TLS)• Common data formats (XML, JSON, HTML)• Device and browser compatibility

Solid Data Space

Components

Decentralized network of Web services, Solid Pods and identity providers

- **Data is mapped to RDF and stored in Solid Pods**
*if possible
- **Virtual Solid Pods**
 - Solid interface for non-RDF data sources
 - RDF mapping at runtime
 - Transformation of write operations to LDP documents into updates to the data sources (view update problem!)
- **Dataspace Service Apps (DSS)**
 - Solid apps that provide services for a dataspace, e.g. catalog and search
 - Solid apps that follow additional agreements/protocols



Solid Data Space

Example: Data catalog on Solid Pod

- **Dataspace layer**

- Agreement on catalog Pod(s), vocabulary and registration process
- Solid apps for crawling, searching, visualizing, etc.

- **Solid layer**

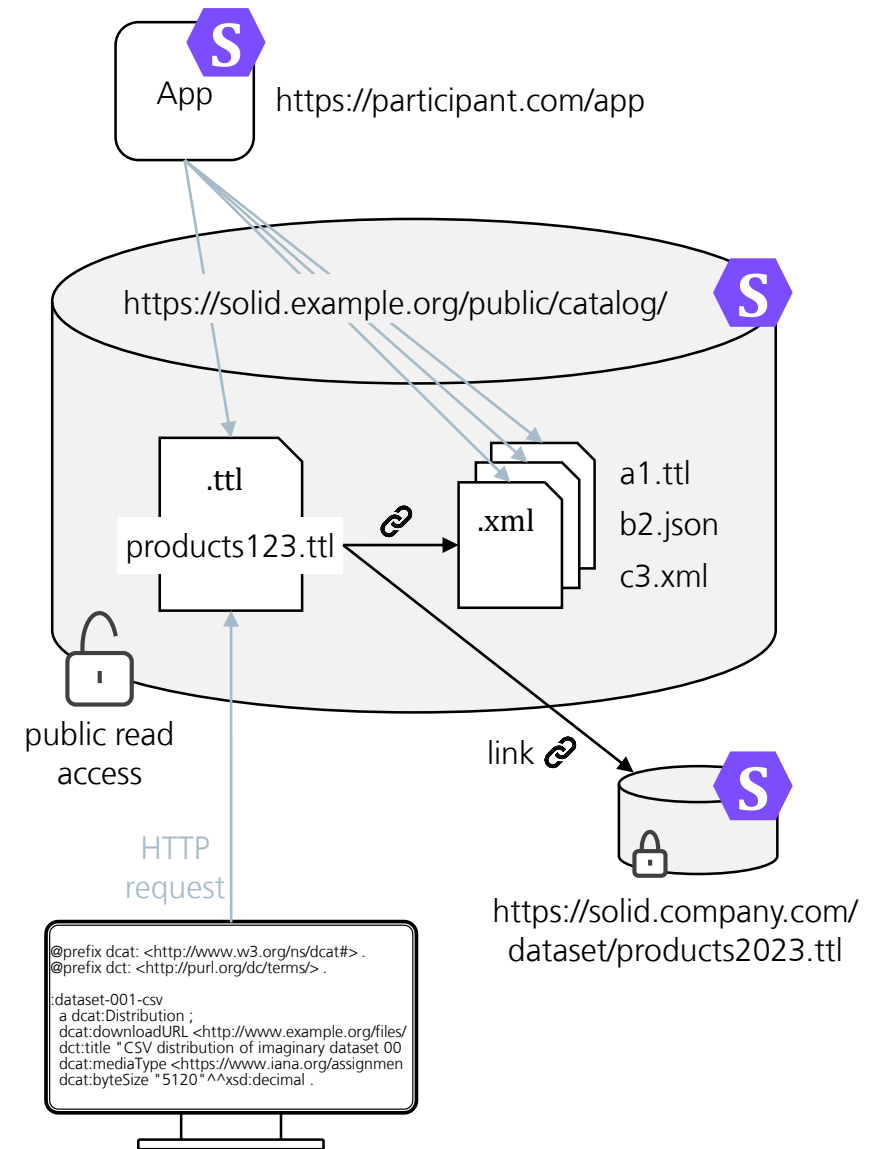
- Access rules (e.g. access for group of participants → *VCARD*)
- RDF data is organized in *LDP* containers with uniform interface

- **Linked Data layer**

- RDF data with links to the dataset, data types, responsible entities, etc.
- *DCAT* or *VoID* ontology

- **Web layer**

- Easy access with browser in case of public access
- Fallback to text content



The Web and Linked Data as a Solid Foundation for Dataspaces

Agenda

1. Dataspace concepts
2. Solid Data Space (SDS)
3. Comparison between the SDS and IDS
4. Conclusion and Outlook

Comparison between the SDS and IDS

International Data Spaces

- Architecture and blueprint for IDS dataspace
- Developed by IDSA
- Top-down approach; detailed prescription of components and processes
- Complex specification, connector and certification
- Marketplace for datasets and applications; many centralized components
- Certification processes; technical measures for access control and policy enforcement

Solid Data Space

- Concept for Solid-based dataspace
- Solid spec. developed in W3C groups
- Bottom-up approach; specification of the framework
- Low threshold for participation
- Web servers for data sharing; decentralized architecture
- (Currently) only access control

Comparison between the SDS and IDS

Dataspace services

IDS components

- **IDS Identity Provider**
 - identification and metadata about participants
 - certificate authority and token service
- **IDS (Metadata) Broker**
 - searchable catalog of available data (providers)
- **IDS Clearing House**
 - data exchange logging and billing
- **IDS App Store**
 - searchable repository for IDS app containers
 - certification for “trusted apps”

Equivalent features using Solid

- **decentralized identity providers**
 - metadata when dereferencing participants' WebIDs
 - Web PKI; Verifiably Credentials **TODO**
- **custom solution** **TODO**
 - e.g. catalog dataspace service
 - without a central catalog, decentralized data is discovered via link following
- **Solid server with access logging** **TODO**
 - only logs of access requests
- **Web apps**
 - any kind of Web application that supports the Solid protocol
 - trusted Solid apps? **TODO**

The Web and Linked Data as a Solid Foundation for Dataspaces

Agenda

1. Dataspace concepts
2. Solid Data Space (SDS)
3. Comparison between the SDS and IDS
4. Conclusion and Outlook

The Web and Linked Data as a Solid Foundation for Dataspaces

Conclusion

- **Linked Data and Solid already provide the most important features for sovereign data exchange**
- **Balance between complexity for guaranteed features and low participation hurdle**
 - E.g. automatic contract negotiation requires complex metadata about legal aspects
- **Dataspace participants must agree on trusted third parties**
 - For certification or supervision
 - Trust in the dataspace stems from trust to third party
- **Certain dataspace functions require centralized components**
 - Finable information services (e.g. online index / repository server for metadata)

Examples from the WWW

- DNS Security Extensions (DNSSEC) with a single global trust anchor (ICANN)
- Web PKI with more than 100 CAs that issue TLS certificates. Browsers vendors decide which CA is trusted (CA/Browser forum)
- BitTorrent network with repositories and trackers

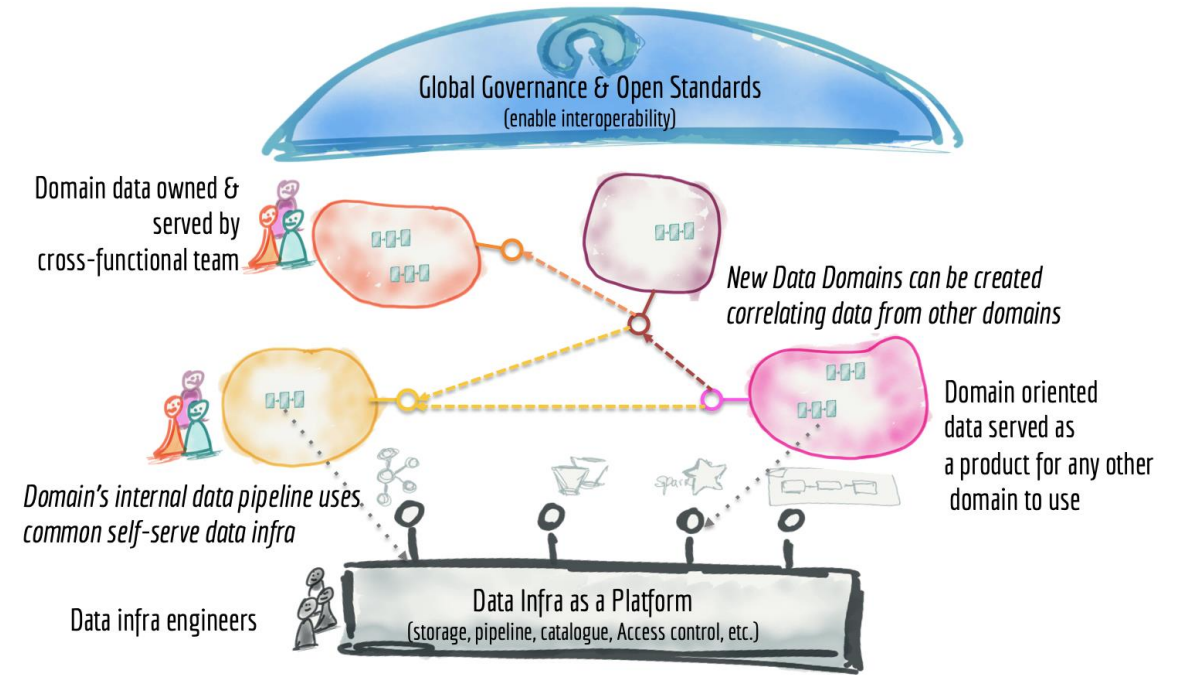
Outlook

Relation to the »Data Mesh«

“**Data mesh** is a decentralized sociotechnical approach to share, access, and manage analytical data in complex and large-scale environments within or across organizations.”
(Dehghani 2019)

- Domain-oriented decentralized data ownership and architecture
- Evolution of (centralized) data lakes or data warehouses

Solid dataspaces implement many technical parts of the Data Mesh principles



High-level Data Mesh architecture (Dehghani 2019)

Characteristics of domain datasets as product
(Dehghani 2019)

- DISCOVERABLE 
- ADDRESSABLE 
- TRUSTWORTHY (DEFINED & MONITORED SLOs) 
- SELF-DESCRIBING 
- INTER OPERABLE (GOVERNED BY OPEN STANDARDS) 
- SECURE 

References

DataSpace Support Platform

- Franklin, M., Halevy, A. & Maier, D. 2005. From Databases to Dataspaces: A New Abstraction for Information Management. ACM SIGMOD Record 34, 4 (Dec. 2005), 27–33.
- Halevy, A., Franklin, M. & Maier, D. 2006. Principles of Dataspace Systems. In Proceedings of the Twenty-Fifth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (Chicago, USA) (PODS '06). New York, USA, 1–9.

Web of Data

- Berners-Lee, T., Hendler, J. & Lassila, O. 2001. The Semantic Web: A New Form of Web Content That Is Meaningful to Computers Will Unleash a Revolution of New Possibilities. Scientific American 284 (May 2001), 5.
- Heath, T. & Bizer, C. 2011. Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool.
- Bizer, C. & Heath, T. & Berners-Lee, T. 2009. Linked Data - The Story So Far. IJSWIS 5, 3 (2009), 1–22.

International Data Spaces (IDS)

- Steinbuss, S. et al. 2022. IDS Reference Architecture Model 4. Technical Report. International Data Spaces Association. https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0
- Bader, S. et al. 2020. The International Data Spaces Information Model – An Ontology for Sovereign Exchange of Digital Content. In The Semantic Web – ISWC 2020. Springer International Publishing, Cham, 176–192.

GAIA-X

- GAIA-X European Association for Data and Cloud. 2022. Gaia-X: Architecture Document - 22.04 Release. Technical Report. Gaia-X European Association for Data and Cloud AISBL.
- Otto, B. 2021. GAIA-X and IDS. Technical Report. Zenodo. <https://doi.org/10.5281/ZENODO.5675897>

Solid

- Samba, A.V. et al. 2016. Solid: A Platform for Decentralized Social Applications Based on Linked Data. Technical Report. MIT CSAIL & Qatar Computing Research Institute.

Data Mesh

- Dehghani, Z. 2019. How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh, <https://martinfowler.com/articles/data-monolith-to-mesh.html>, accessed 2023-03-24

Contact

Sascha Meckler
Data Spaces & IoT Solutions
sascha.meckler@iis.fraunhofer.de

Arbeitsgruppe für Supply Chain Services SCS des Fraunhofer IIS
Nordostpark 93
90411 Nürnberg
www.scs.fraunhofer.de



Fraunhofer Institute for Integrated
Circuits IIS

Gefördert durch

Bayerisches Staatsministerium für
Wirtschaft, Landesentwicklung und Energie

