Towards a Federated Data and Service Catalogue

Position Paper

Johannes Theissen-Lipp[1,2] (Speaker at the Workshop), Ahmad Hemid[1],
Christoph Lange[1,2] and Christina Gillmann[1]

[1]*Fraunhofer Institute for Applied Information Technology FIT, Sankt Augustin, Germany*
[2]*RWTH Aachen University, Aachen, Germany*

Abstract

The integration and management of heterogeneous data pose challenges across various applications, requiring scalable solutions to handle large volumes of data, maintain compatibility, and ensure security, privacy, and regulatory compliance. This position paper presents a federated data and service catalogue based on the Eclipse XFSC framework. It presents enhancements such as individual application profiles, metadata service offerings, trusted management, interoperable data sharing and ecosystem integration. These contributions address the critical requirements of federated dataspaces, providing a robust, secure and scalable infrastructure for effective data integration and management across diverse applications.

1. Introduction and Background

Managing and integrating heterogeneous data is a significant challenge, involving diverse data formats, semantic inconsistencies and ensuring data quality. Scalable solutions are needed to handle large data volumes, maintain compatibility, and ensure security, privacy, and regulatory compliance. Effective integration facilitates insights and decision-making. Despite these challenges, many applications benefit from meaningful data integration. In healthcare, it enhances patient care and research. Scientific research is improved by combining data, fostering collaboration and results [1, 2]. Mobility and public transport benefit from unified interfaces, providing better services, travel plans and forecasts.

Dataspaces promise a solution by enabling the flexible correlation of disparate data sources, allowing the spontaneous discovery and exploration of relationships, patterns and trends that may not be apparent in isolation. In 2014, the Fraunhofer Society in Germany launched the *Industrial Data Space* [3] initiative to create a reference architecture that enables sovereign and secure data exchange between companies without them losing control of their data. Over the years, several dataspace initiatives have been launched with different scope (national or international), focus (industrial or research), technology readiness, communities, impact, reach and target domains. Among these, two initiatives have established themselves as de facto standards, because they meet the above-mentioned aspects particularly well: The *International*

Data Spaces Association[4] with the *International Data Spaces* (IDS) [5] and Gaia-X [6]. These two initiatives are linked and co-operate with each other [7] and are in line with the *Common European Data Spaces* [8]. Within the Gaia-X ecosystem, a set of *Cross Federation Services Components* (XFSC) [9] has been proposed as a toolbox for providing interoperability across federations in dataspaces. The XFSC services are now available as open source through the *Eclipse Foundation* [10].

As part of the Data Science & Artificial Intelligence department of the Fraunhofer Institute for Applied Information Technology FIT, the *Data Management* research group develops scalable solutions for small and large datasets in various domains. Our research areas include life sciences, healthcare, manufacturing, logistics, culture and tourism. This position paper aims to share our experiences and lessons learnt from the wide variety of projects we have conducted, resulting in a *federated data and services catalogue* based on the Eclipse XFSC.

2. A Federated Data and Service Catalogue

Based on the Eclipse XFSC [10] open source project, this position paper proposes a catalogue for the trusted management of domain-specific data and service metadata, enabling interoperable data sharing and exchange across digital ecosystems.

2.1. The Eclipse XFSC

Eclipse XFSC (Cross Federation Services Components) [10] develops software for creating federated systems that connect participants in a data and service infrastructure, aiming to create new data-driven services and innovative products through interconnected data ecosystems, or federations, orchestrated by Federation Services as part of Gaia-X. The project comprises microservice components that ensure interoperability within these federations. The Eclipse XFSC Toolbox offers services for self-sovereign identities, W3C credential management, trust services, knowledge graph databases, policy negotiation, and a core low-code engine to support decentralized ecosystems. It meets the technical requirements for federations, enabling organizations to engage in self-sovereign identity and data ecosystems while ensuring data sovereignty and trust in line with Gaia-X goals. As a snapshot of June 2024, at the time of writing of this position paper, the set of services is [10]:

- **Identity, Credential, Access Management and Trust:** These services authenticate and authorise users and systems in a decentralised, self-sovereign manner, maintaining trust without a central authority through credential validation.
- **Decentralised Catalogue and Contracting Service:** Acts as an inventory, allowing participants to discover, understand and use available data in an ecosystem. It serves as a repository for federations, enabling participants to find each other's information and services, negotiate contracts and track data transactions.
- **Orchestration & Monitoring:** Orchestration and monitoring services manage complex ecosystems to ensure seamless and compliant operations. This includes orchestrating tasks and managing infrastructure services. Continuous automated monitoring provides transparency through automated compliance monitoring.

- **Portal:** The Portal is an integration layer that highlights federation services and provides access to onboarding, accreditation, service discovery, orchestration and delivery. It includes functionality for searching and viewing content from the federated catalogue, onboarding new participants, and orchestrating services.

2.2. Towards A Federated Data and Services Catalogue

In this section, we propose our additions to the Eclipse XFSC. These findings are a result of our experience and lessons learned from the wide variety of projects we have conducted in dataspace initiatives such as the International Data Spaces [11], Gaia-X [12, 13], or the National Research Data Infrastructure (NFDI) [1, 14], or in domain-specific dataspaces such as the Mobility Data Space [15, 16], NAPCORE (National Access Point Coordination Organisation for Europe) [17], or the German *Datenraum Kultur* for culture [18]. An ongoing series of scientific workshops on *Semantic in Dataspaces* [19, 20] at international conferences is also contributing solutions, insights and future research areas. We propose the following additions to the Eclipse XFSC to pave the way for a federated data and service catalogue for dataspaces:

- **Individual Application Profiles:** Our solution introduces a set of metadata elements, policies and guidelines that represent different application profiles to meet the unique characteristics of different use cases. This addresses the need for customised data and service representations, thereby increasing the flexibility and specificity of the Eclipse XFSC.
- **Metadata Service Offering:** We ensure privacy by allowing data to be stored on the provider's servers, while defining metadata that is visible in dataspaces, allowing searchability without exposing the actual data. This extends the catalogue system by providing a secure way to share data insights while maintaining data sovereignty.
- **Trustable Management:** Our approach manages access control through the data provider, using digital signatures from certified organisations to ensure the integrity and trustworthiness of the entities accessing the data. This addition strengthens the trust and security framework and addresses the need for robust identity and access management.
- **Interoperable Data Sharing:** We facilitate data reuse through the adoption of interoperable data formats based on the IDS and Gaia-X standards, and provide extensions for enhanced semantic mapping to meet specific needs. This supports a requirement for interoperability, making the components of the Eclipse XFSC more versatile and capable of seamless data integration.
- **Towards Merging Ecosystems:** We are working to connect multiple ecosystems by continuously improving our service catalogue to include an increasing number of ecosystems and enable seamless transitions between them. This will extend the catalogue's ability to operate in a multi-ecosystem environment, fostering greater collaboration and integration.
- **Metadata Broker Provision:** An extension to the *Eclipse Dataspace Connector* facilitates the discovery of available metadata assets and automates the generation of queries. The broker streamlines the contracting process, encompassing the entire cycle from offering contract proposals and negotiating terms to finalising contracts between participants.

3. Conclusion

In conclusion, this position paper presents our contributions towards a federated data and service catalogue that extends the existing Eclipse XFSC framework. By introducing individual application profiles, we address the need for customised data and service representations, thereby increasing flexibility. Our metadata service offering ensures privacy and data sovereignty, extending the capabilities. Trusted governance is enhanced through the use of digital signatures from certified organisations, addressing the need for robust identity and access management. Interoperable data sharing is facilitated by the adoption of IDS and Gaia-X standards, making the system more versatile. Finally, our efforts to merge ecosystems extend the catalogue's ability to operate in a multi-ecosystem environment, fostering greater collaboration and integration. These additions address the critical requirements of federated dataspaces and provide a robust, secure and scalable infrastructure for managing and integrating heterogeneous data across multiple applications.

Acknowledgments

References

[1] N. Hartl, E. Wössner, Y. Sure-Vetter, Nationale Forschungsdateninfrastruktur (NFDI), Informatik Spektrum 44 (2021) 370–373. doi:10.1007/s00287-021-01392-6.

[2] National Research Data Infrastructure (NFDI) e.V., NFDI: The Association, 2022. URL: https://www.nfdi.de/?lang=en, accessed on 2024-05-23.

[3] B. Otto, J. Jürjens, J. Schon, S. Auer, N. Menz, S. Wenzel, J. Cirullies, Industrial Data Space - Digital Sovereignity Over Data, Technical Report, 2016. doi:10.13140/RG.2.1.2673.0649.

[4] IDSA, International Data Spaces Enabling Data Economy, 2021. URL: https://internationaldataspaces.org/download/16822/?tmstv=1670550685, accessed on 2024-05-23.

[5] B. Otto, M. Jarke, Designing a Multi-sided Data Platform: Findings from the International Data Spaces Case (2019) 561–580. doi:10.1007/s12525-019-00362-x.

[6] Plattform Industrie 4.0, Project GAIA-X: A Federated Data Infrastructure as the Cradle of a Vibrant European Ecosystem, Technical Report, BMWi, Berlin, Germany, 2019.

[7] International Data Spaces Association, IDSA Position Paper: GAIA-X and IDS, 2021. doi:10.5281/zenodo.5675897.

[8] European Commission, Staff Working Document on Common European Data Spaces, Technical Report, Brussels, Belgium, 2022.

[9] Gaia-X European Association for Data and Cloud AISBL, GXFS and the XFSC Toolbox, 2022. URL: https://www.gxfs.eu/set-of-services/, accessed on 2024-05-23.

[10] Eclipse Foundation, Eclipse XFSC (Cross Federation Services Components), 2023. URL: https://projects.eclipse.org/projects/technology.xfsc, accessed on 2024-05-23.

[11] S. Bader, J. Pullmann, C. Mader, S. Tramp, C. Quix, A. Müller, H. Akyürek, M. Böckmann,

B. Imbusch, J. Theissen-Lipp, S. Geisler, C. Lange, The International Data Spaces Information Model - An Ontology for Sovereign Exchange of Digital Content, Springer International Publishing, Cham, 2020, pp. 176–192. doi:10.1007/978-3-030-62466-8_12.

[12] GAIA-X European Association for Data and Cloud AISBL, Gaia-X Architecture Document, Technical Report, Brussels, Belgium, 2022. URL: https://docs.gaia-x.eu/technical-committee/architecture-document/22.10/, accessed on 2024-05-23.

[13] A. Strunk, C. Lange, Self-Description of Resources, Service Offerings and Participants within Gaia-X Ecosystems, Technical Report, Brussels, Belgium, 2022. URL: https://gaia-x.eu/wp-content/uploads/2022/08/SSI_Self_Description_EN_V3.pdf, accessed on 2024-05-23.

[14] S. Arndt, B. Farnbacher, M. Fuhrmans, et al., Metadata4Ing: An Ontology for Describing the Generation of Research Data Within a Scientific Activity, Technical Report, NFDI, 2022. doi:10.5281/zenodo.5957104.

[15] H. Drees, D. O. Kubitza, J. Lipp, S. Pretzsch, C. S. Langdon, Mobility data space–first implementation and business opportunities, in: 29th ITS World Congress, ERTICO – ITS Europe, Hamburg, Germany, 2021, pp. 1–12.

[16] Fraunhofer IVI, Mobility Data Space: Sovereign Mobility Data Ecosystem, 2019. URL: https://www.mobility-data-space.de/en.html, accessed on 2024-05-23.

[17] NAPCORE, National Access Point Coordination Organisation for Europe, 2022. URL: https://napcore.eu/metadata/, accessed on 2024-05-23.

[18] Fraunhofer FIT, Datenraum Kultur, 2022. URL: https://www.fit.fraunhofer.de/de/geschaeftsfelder/human-centered-engineering-and-design/HCED_Forschung/datenraum-kultur.html, accessed on 2024-05-23.

[19] J. Theissen-Lipp, S. Decker, E. Curry, The First International Workshop on Semantics in Dataspaces, co-located with the ACM Web Conference, April 30, 2023, Austin, TX, USA, Association for Computing Machinery, 2023. doi:10.1145/3543873.3589750.

[20] J. Theissen-Lipp, P. Colpaert, S. K. Sowe, E. Curry, S. Decker, The Second International Workshop on Semantics in Dataspaces, co-located with the Extended Semantic Web Conference, May 26–27, 2024, Hersonissos, Greece, CEUR-WS, 2024.

Speaker Bio

Johannes Theissen-Lipp is a researcher at the Chair for Databases and Information Systems at RWTH Aachen University and deputy leader of the Data Management team at Fraunhofer FIT. He studied computer science and conducted research at the Institute for Information Management in Mechanical Engineering for about two years. After joining Fraunhofer, Johannes continues his research in the areas of dataspaces, Semantic Web technologies and interoperability. Project highlights include the International Data Spaces, the National Research Data Infrastructure NFDI, the cluster of excellence Internet of Production and Gaia-X. Johannes recently submitted his PhD thesis on "Semantic Foundations of Dataspaces" to RWTH Aachen University.