Dataspaces for Collaborative Research

Soo-Yon Kim*, Liam Tirpitz*, Max Wagels*, Benedikt T. Arnold†,
Christian Rennert‡, István Koren‡, Janik Rapp§, Mario Moser¶,
Wil van der Aalst‡, Bernhard Rumpe§, Robert H. Schmitt¶, Jan Pennekamp‖, Sandra Geisler*
*Data Stream Management and Analysis, RWTH Aachen University, Germany
{kim, tirpitz, geisler}@dbis.rwth-aachen.de · {max.wagels}@rwth-aachen.de
†Fraunhofer FIT, Germany · {benedikt.arnold}@fit.fraunhofer.de
‡Process and Data Science, RWTH Aachen University, Germany · {rennert, koren, wvdaalst}@pads.rwth-aachen.de
§Software Engineering, RWTH Aachen University, Germany · {rapp, rumpe}@se-rwth.de
¶WZL-IQS, RWTH Aachen University, Germany · {mario.moser, robert.schmitt}@wzl-iqs.rwth-aachen.de
‖Communication and Distributed Systems, RWTH Aachen University, Germany · {pennekamp}@comsys.rwth-aachen.de

***Abstract*—Data-driven collaborative research is a key driver of innovation. However, effective data exchange across institutions remains difficult in practice, often relying on ad hoc mechanisms that provide little support for discoverability, sovereignty, or the dynamic sharing of datasets during ongoing collaboration. Dataspaces have been proposed as a method to address these shortcomings, but their suitability for collaborative research remains largely unexplored in terms of the needs that academic collaborations impose, of practical deployment experiences, and of the potential for concrete use cases.**

Addressing this gap, we derive the requirements for infrastructures in data-driven collaborative research, providing the basis for assessing dataspaces. We further report on the deployment of a pilot dataspace for a real-world, large-scale research project, focusing on the onboarding of four institutes with diverse data types, interests, and disciplinary backgrounds. The deployment highlights the practical steps required to select and prepare a dataspace technology stack, establish connectors, and further assesses challenges posed by heterogeneous environments and the level of effort involved in integration. Beyond deployment, we explore the dual role of research dataspaces, serving as both a generic data sharing infrastructure and as a testbed for practical research on data sharing technologies. A federated process mining use case for data-driven production demonstrates the latter, where distributed process data are analyzed collaboratively.

Our findings indicate that dataspaces are indeed a viable option for collaborative research if supported by adequate expertise. By deriving requirements, reporting deployment experiences, and demonstrating use cases, we contribute guidance for research practitioners. Next, future work should focus on sustainability and scalability needs, such as lowering entry barriers, developing trust mechanisms, and extending use case scenarios.

***Index Terms*—dataspaces, data ecosystems, data exchange, collaborative research, research data management.**

I. Introduction

Research is increasingly characterized by its dependence on large amounts of heterogeneous data from diverse sources, as exemplified in the context of crucial societal challenges, including pandemics [1], climate change [2], and industrial transformation [3], [4]. This need for data poses challenges for underlying research infrastructures, which must support data exchange across institutional and disciplinary boundaries.

Dataspaces offer a promising approach to address this need. A dataspace provides a federated architecture where data remains with the owners, or is only shared under specific conditions, and where interoperability, discoverability, and governance are ensured through common standards, protocols, and agreements. In this paper, we explore dataspaces as an enabler for data-driven collaboration in research.

The handling of data in research is the focus of the field of Research Data Management (RDM) [5], with sharing and reuse emphasized as essential steps in frameworks such as the RDM lifecycle [6]. Guidelines for research data dissemination [7] encourage the use of repositories such as Zenodo[1], which provide a way to make data publicly available, preserve it, and reliably cite it, thereby enabling global reuse and contributing to the reproducibility of scientific results [8].

However, the needs of collaborative research go beyond publication-based approaches to data sharing. Only a fraction of relevant data can be considered publication-ready, as ongoing collaborations are inherently work in progress and often involve data that is dynamically evolving [9]. Moreover, a timely exchange of data is crucial for collaboration progress, but preparing data for publication in repositories is time-intensive, and further delays can occur when release is tied to written publications. Such characteristics limit the usefulness of data publications for dynamic collaborative research.

In practice, research data is often produced and retained within organizations, with limited actions taken toward making it available for immediate reuse [10]. When data exchange does occur, it is often facilitated through ad hoc mechanisms such as emails [11], rather than through structured, sustainable approaches. This underlines the need to better understand the requirements of data-driven research collaborations, and to develop approaches that adequately support them.

The scale and complexity of modern collaborative research environments are well observable in large-scale research projects. They typically span multiple institutes, disciplines, and infrastructures while producing highly heterogeneous data. To meet the demands of agile collaboration in such projects, infrastructure must provide mechanisms to support the lifecycle of data shared across project institutes [12]. Researchers

[1]https://zenodo.org/

 DOI: 10.1109/BigData66926.2025.11402648

must be able to discover, share and reuse data, while retaining sovereignty and control over what is shared and with whom. Further, the infrastructure needs to accommodate heterogeneous local environments and assets. Rather than centralizing data, these requirements point toward federated, decentralized solutions that keep data at the source while orchestrating controlled but flexible exchange between participants.

The Internet of Production (IoP)[2] is a long-lasting large-scale research project with a focus on data-driven, industrial collaboration, which exemplifies such a setting. While the IoP has developed infrastructures to support individual aspects of joint data management, such as discovery across institutes, the need for a comprehensive solution for structured collaborative research remains unresolved.

We propose to address this gap with dataspace technologies. While many existing works on dataspaces focus on a conceptual level [13], or build on industrial use cases such as manufacturing, smart cities, or agriculture [14], reports on practical deployments in academic research settings, which may differ essentially from industrial ones [15], are scarce. As a result, the potential of dataspaces to improve collaborative research remains largely unexplored.

In this paper, we investigate this potential by exploring how to deploy and use industry-grade dataspace technologies as a basis for federated infrastructures in cross-institutional, data-driven collaborative research environments, on the example of the IoP as a large-scale, cross-domain, and multi-institutional research project.

Contributions: (i) *Requirements:* We analyze and derive the requirements for data infrastructures in cross-institutional, collaborative research. (ii) *Deployment:* We report on the deployment of a pilot dataspace in the IoP research project, including the onboarding of four institutes with heterogeneous infrastructures and disciplinary backgrounds. (iii) *Use cases:* We demonstrate how dataspaces can concretely support collaborative research, based on realistic use cases from the project: (a) as an integrated tool for project-wide sharing of research data, and (b) as a testbed for research on data sharing itself, exemplified through federated process mining on a data–driven production scenario. Our findings can thus contribute to shaping future data infrastructures for collaborative, cross-organizational research in large-scale projects and beyond.

Structure: The remainder of this paper is structured as follows. In Section II, we first present the setting of our large-scale research project, the IoP. We discuss current practices of sharing data in cross-institutional environments in Section III. In Section IV, we derive a structured overview of data-sharing-related requirements in such settings. In Section V, we then report on the practical setup of a dataspace and our experiences with onboarding institutes. We demonstrate the capabilities of dataspaces as infrastructure to support collaborative research use cases in Section VI, and conclude this paper with future research directions in Section VII.

[2]https://www.iop.rwth-aachen.de/

II. A Collaborative Research Environment: The Internet of Production

The IoP [16] is a large-scale research project running for seven years, which investigates topics connected to Industry 4.0 and the Industrial Internet of Things (IIoT), i.e., the use of data in industrial manufacturing. A core research objective of the IoP is to enable cross-organizational collaboration throughout the entire production life cycle, e.g., along and across supply chains [3], [17]. A prerequisite for this vision is to enable sovereign and dynamic data exchange, which allows collaborators to reliably share, reuse, and cooperate on data across organizational boundaries.

A. Scale and Diversity of the IoP

The IoP involves 200+ researchers across 30 institutes [18], spanning disciplines such as engineering, computer science, and the social sciences. The participating institutes contribute a large variety of assets, ranging from raw machine and experimental data to curated datasets, algorithms, and code artifacts. With this scale and diversity, the project itself provides a collaborative research environment in which data sharing challenges and opportunities frequently emerge in practice.

Data-driven collaboration is central to the vision of the IoP in two dimensions. On the industrial side, joint data use across organizations is intended to enable improvement in production processes. On the scientific side, it is essential to combine the expertise and data of researchers from different disciplines and institutes, with the goal of advancing knowledge across domains. In both dimensions, effective mechanisms for data sharing are a cornerstone for the success of the IoP.

B. Existing Data Infrastructure

Centrally, data-driven collaboration is supported in several forms. A GitLab instance is available to all participants for versioning and sharing code and other small artifacts. In addition, the research data management platform Coscine[3] offers a long-term storage solution for datasets of various sizes with integrated metadata management. The project also makes use of RDMO[4], a tool for the creation and maintenance of data management plans, which encourages researchers to manage their research artifacts in a structured and responsible way. While these centralized tools exist, data is still largely managed in decentralized systems maintained by different institutes, primarily due to the heterogeneous nature of their respective local infrastructures and the need for sovereign data management. Centralized measures therefore provide a basis for coordination in the IoP, but do not sufficiently address the challenge of making research data discoverable and accessible across institutional boundaries.

To address the need for discoverability, the IoP developed and established the Dataset Finder [19]. The Dataset Finder is a project-wide data catalog that enables researchers in the IoP to add descriptions of their artifacts, thus creating a central

[3]https://coscine.rwth-aachen.de/

[4]https://rdmo.rwth-aachen.de/

list of dataset descriptions while the data itself remains stored within the institutes. Searching and browsing functionalities provide researchers with the ability to discover relevant data, e.g., by keywords or project domain. The catalog thus greatly improves the discoverability of project assets. Yet, access to the data itself still has to be organized ad hoc, prompting for advances in the direction of interoperable data-sharing extensions.

C. Collaborative Research in the IoP

Collaborative research can take different forms. In some scenarios, collaboration consists of very basic data exchanges in which researchers access data generated by other institutes to reuse it for their own purposes. In others, by contrast, data remains with the original holders while analyses are coordinated across sites through shared protocols or algorithms. Here, collaboration often focuses on developing and validating methods that are specifically designed for distributed data. A typical case is research conducted on personal or industrial data, where confidentiality and privacy concerns prohibit the exchange of raw datasets, despite the need for collaborative analysis. They highlight requirements that extend beyond simple data exchange. To concretize these aspects, we discuss the concrete use case from the IoP project of conducting federated process mining in the following.

Federated process mining. Process mining analyzes data on processes, where each data entry contains at least: (i) an identifier for the process it is related to, (ii) an activity that took place within the process, (iii) and a timestamp of the time at which that activity took place. Based on such data, it is possible to automatically discover process models, check their conformance, and extend them with perspectives on Key Performance Indicators (KPIs) [20].

For the past decade, inter-organizational process mining has been a focus of interest. More recently, attention has shifted toward the inclusion of so-called confidentiality domains [21]. The objective is to design process mining algorithms that satisfy the constraints imposed by privacy regulation laws such as GDPR and HIPAA, or by process owners that are interested in gaining insights without disclosing complete datasets, or exposing corporate secrets. These algorithms aim to reproduce results identical to those obtained with shared data, without requiring actual data exchange. Consequently, federated process mining is particularly relevant to domains such as healthcare and agriculture, but also supply chain management and distributed manufacturing, which are prevalent in the IoP.

As a simple use case, we consider abstraction-based approaches [22]–[25]. Here, an abstraction of the input data is first computed and then shared, only requiring for the transfer of aggregations to be shared, leading to a low number of data exchanges. Often, the abstractions are aggregations of the event data. In the IoP, applying these federated, abstraction-based methods constitutes a collaborative research case, with the goal of assessing the feasibility of these methods in practice, and exploring how to refine them for future use [26].

By presenting the IoP as an example, we have introduced both some of the opportunities and challenges in collaborative research environments. To better understand the technological foundations that enable such scenarios, we next review the state of the art in data-driven collaboration methods.

III. State of the Art in Technologies for Data-Driven Collaboration

Data-driven collaborations rely on infrastructures that make data findable and usable across institutions, with researchers consistently emphasizing control over their data and the availability of reliable, easy-to-use services as key enablers of sharing [27]. A range of solutions exists for the support of data sharing and collaboration within and beyond academia.

A. Collaboration Methods in Research and Academia

To give an overview of the state of the art in this context, we now survey different conceptual directions that promise to enable and accelerate data-driven collaborative research.

Data catalogs. Data catalogs provide structured metadata and search functionalities to improve the discoverability of datasets, without storing or preserving the data themselves. Catalogs cover different scales: project-specific catalogs, e.g., IoP's Dataset Finder, national and international services such as the general purpose catalog OpenAIRE Explore[5], and disciplinary catalogs, such as the CESSDA Data Catalogue[6] for the social sciences. These systems are valuable for increasing transparency and visibility. However, only a fraction of catalogs cover datasets that are not yet publication-ready, and catalogs remain limited to describing data rather than managing access or enabling direct exchange.

General and domain-specific repositories. General purpose repositories, such as Zenodo or Dryad[7], and discipline-specific platforms, such as EBRAINS[8] or PANGAEA[9], are well-suited for accessibility and discoverability [8], [11]. However, these platforms primarily target finalized, curated datasets and are therefore less suitable for sharing dynamically evolving data that arises during collaboration [28].

RDM platforms. RDM platforms such as iRODS[10] or Coscine go beyond simple repositories by typically supporting researchers to integrate the Findable, Accessible, Interoperable, Reusable (FAIR) principles [29], a best practice set of guidelines for scientific data management, more directly during the research process. Depending on the platform, functionalities may include organizing data in project spaces, defining policies for access and reuse, metadata management, and integration with heterogeneous storage backends. However, their adoption in large-scale collaborations is often hindered by the technical effort needed for setup and maintenance, the complexity of aligning metadata and policies across institutes,

[5]https://explore.openaire.eu/
[6]https://datacatalogue.cessda.eu/
[7]https://datadryad.org/
[8]https://www.ebrains.eu/
[9]https://www.pangaea.de/
[10]https://irods.org/

which in some platforms is a prerequisite for their usage, and usability challenges for researchers [9].

Collaborative workflow environments. Solutions such as GitHub[11], JupyterHub[12], or electronic lab notebooks address a different need. They facilitate the sharing of code and methods within labs or projects. While they support collaborative workflows, they do not inherently provide mechanisms for cross-institutional discovery and are primarily suited for sharing code rather than for sharing different kinds of data.

Academic and commercial cloud storage. Academic cloud-based services such as Sciebo[13] provide convenient file exchange across institutions. Their well-known commercial counterparts, e.g., Dropbox or Google Drive, are also widely adopted for scalability and ease of use. However, their centralized nature reduces data sovereignty, and, in the case of commercial services, accessibility and scalability to larger volumes of data is dependent on subscription payment.

Ad hoc sharing. Email and file transfer services (e.g., Gigamove[14]) are still widely used for individual exchanges of small numbers of files. While convenient, they can only be used when data has already been discovered, and they lack any form of update mechanisms, therefore not scaling to long-lasting or dynamic collaborations.

Overall, these solutions form a fragmented landscape. While the technologies address specific aspects of data sharing, no solution emerges as a comprehensive fit to the challenges of collaborative, cross-organizational research as a whole.

B. Dataspaces: An Introduction and Overview

The need for comprehensive data sharing solutions in collaborative settings is not unique to research. Most prominently, dataspaces [30], [31] have recently emerged as a paradigm for sovereign and federated data sharing in different domains, such as healthcare and manufacturing [14], where they are increasingly researched and deployed as productive systems.

Dataspaces are an abstract concept to enable data sharing between independent organizations in an easy-to-use way, while maintaining data sovereignty [14], i.e., data owners can choose how and when data leaves their control. Dataspaces follow a federated approach for decentralized data exchange, based on shared principles, interoperable software components, and protocols [32].

Conceptual work highlights several core requirements for the realization of dataspaces, such as the facilitation of standardized and interoperable data exchange [33], [34], strong federation with minimal centralized components, as well as strong security and trust. Gleim et al. [33] investigate fully decentralized data sharing in the context of industrial collaboration with a focus on versioning and persistent identification [35] and develop an interoperable data sharing framework based on web standards [36], but without considering trust. Similarly, the Solid project [37] builds on semantic web technologies to decouple applications from personal data and enables sovereign data sharing between independent entities. Meckler et al. [38] propose Solid as a lightweight foundation for dataspaces. However, so far, their approach remains conceptual without practical applications.

On the industrial side, several large-scale initiatives have strongly influenced the evolution of this field, such as Gaia-X[15], a European initiative for federated data infrastructures, and Catena-X[16], which adapts dataspace principles to the requirements of the automotive domain. The International Data Spaces (IDS) initiative provides a mature reference architecture [32] and feature-rich technology stack for standardized communication in dataspaces. In IDS, participation is realized through a *connector*, a software component maintained by each organization. Connectors link internal data sources and sinks to the dataspace by establishing connections to other connectors, and interact with central services such as catalogs and identity providers. While IDS requires more centralized components for managing the dataspace compared, e.g., to Solid-based dataspaces, it brings components for identity management and usage control, strengthening trust.

Together, these research and industrial efforts illustrate the breadth of dataspace technologies and their potential to support federated applications that rely on cross-organizational exchange. While the idea of using dataspaces as collaborative research infrastructures has already been envisioned [4], [39], to the best of our knowledge, there has not yet been a comprehensive report on a practical deployment in an ongoing research project. To assess the applicability of dataspaces for collaborative research environments, we must first define the requirements that infrastructures need to meet for research data exchange across independent institutions.

IV. Requirements for Infrastructures for Data-Driven Collaborative Research

Building on the description of the IoP as a collaborative research environment and the review of state of the art methods, we consolidate the needs and recurring challenges for collaborative research into requirements for data infrastructures. Analogously to the technical and operational dimensions of the TELOS framework [40], we distinguish between *functional requirements*, which describe the necessary functionalities, and *operational requirements*, which concern aspects of deployment and operation.

A. Functional Requirements

With regard to required functionalities that infrastructures for collaborative, data-driven research must provide, we derive six key aspects.

Discoverability. Researchers must be able to find out what data exists across institutes. This requires an index of appropriate range, e.g., project-wide, with a search functionality, and integration of adequate, descriptive metadata.

[11] https://github.com/
[12] https://jupyter.org/hub
[13] https://hochschulcloud.nrw/en/index.html
[14] https://gigamove.rwth-aachen.de/de
[15] https://gaia-x.eu/
[16] https://catena-x.net/

TABLE I
COMPARISON OF TECHNOLOGIES AGAINST REQUIREMENTS.

Method	Disc.	Dyn.	Decent.	Sov.	Workf.	Heter.
Data catalogs	✓	△	✓	–	–	✓
Repositories	✓	–	–	△	–	✓
RDM platforms	✓	△	△	△	△	△
Workflow env.	–	✓	–	△	✓	△
Cloud storage	–	✓	–	△	–	△
Ad hoc sharing	–	–	✓	✓	–	✓
Dataspaces	✓	✓	✓	✓	△	✓

Disc. = discoverability, Dyn. = dynamic data, Decent. = decentralized storage, Sov. = sovereign exchange, Workf. = workflow coordination, Heter. = heterogeneous assets. ✓= suitable, – = not suitable, △ = partial/case-by-case.

Dynamic data. In collaborative projects, often intermediate and evolving data is generated. Infrastructures must therefore accommodate updates or iterative contributions during ongoing research and enable timely data exchange.

Decentralized storage. Data in collaborative projects is rarely consolidated in one location. Infrastructures must enable data to remain with the original holders while still being accessible for exchange or analysis.

Sovereign exchange. Institutes must be provided with a function to exchange data while retaining control over who accesses which data. This includes mechanisms for defining usage conditions, and for granting and restricting access.

Workflow coordination. In federated scenarios, collaboration often extends to the joint execution of workflows across sites. Infrastructures must therefore provide mechanisms to deploy and orchestrate distributed workflows.

Heterogeneous assets. Collaborative research typically produces a wide variety of data types. Infrastructures need to be able to concur such heterogeneity and enable meaningful integration of diverse assets.

A comparison of the existing approaches along these requirements is given in Table I. The comparison indicates that dataspaces align well with the functional requirements of infrastructures for collaborative research.

B. Operational Requirements

Beyond functional capabilities, infrastructures for collaborative research should also satisfy operational requirements that determine their practical feasibility in real projects. We derive seven key operational aspects.

Ease of setup and onboarding. Deployment should be feasible even without highly specialized technical expertise.

Fast setup and onboarding time. The time needed for setup and for onboarding of a new institute should be manageable, as lengthy procedures hinder adoption in collaborations.

Compatibility with heterogeneous environments. Participating institutes have their own, established systems. The infrastructure must interface with different storage backends, authentication systems, and metadata practices without forcing fundamental local changes.

Usability. Collaborative research projects usually involve not only heterogeneous systems, but also actors of various backgrounds. Researchers should be able to work with the infrastructure without requiring extensive familiarization. Intuitive interfaces, straightforward workflows, and adequate documentation lower barriers to everyday use.

Maintenance. Keeping the system operational requires updates, monitoring, and troubleshooting. Ongoing workload should remain manageable within available resources.

Scalability. As collaborations grow, the infrastructure should accommodate more participants, datasets, and applications without disproportionate increases in effort or complexity.

Support and governance. Long-term viability depends on available support and clear governance. This requirement includes help resources, sustainability of the technology stack, and agreed responsibilities among participating institutes.

These operational aspects complement the functional requirements, and together allow for evaluating infrastructure solutions for collaborative research in practice.

While functional requirements can often be assessed conceptually, operational aspects demand actual deployment for meaningful evaluation. To assess the accessibility of dataspace technologies for large-scale collaborative projects, we thus detail our approach to deploy a dataspace infrastructure for the IoP project in the next section.

V. DEPLOYMENT OF A DATASPACE FOR THE IOP

To understand the practical feasibility of dataspaces as a solution for collaborative research settings, we examined how to orchestrate a deployment within the context of the IoP. This step involved selecting an appropriate technology stack, realizing a project-wide dataspace, and onboarding institutes with diverse disciplinary and technical backgrounds, as we report on in the following. Building on these steps, we then assessed the feasibility of dataspaces for collaborative research in light of the identified requirements.

We structured the deployment of the IoP dataspace along the lifecycle model provided by the Data Space Support Centre (DSSC) Starter Kit[17]. This model distinguishes between the phases of *exploration*, *preparation*, *implementation*, *operation*, and *scaling*. Although originally intended for long-term industrial dataspaces, it offered a useful reference for systematically considering the steps in our research context as well.

A. Exploration

In the exploration phase, we clarified the motivation and scope of the dataspace. The central motivation was to develop a research data exchange system for sovereign data exchange across IoP institutes that integrates with and extends established infrastructures, which we described in Section II-B. We also selected a diverse subset of IoP institutes to participate.

The participating institutes were selected to represent different types of data, collaboration use cases, and disciplinary background from the IoP project, and are depicted in Figure 1.

The **Data Stream Management and Analysis (DSMA)** group focuses on managing and analyzing data in cross-organizational settings. It typically works with streaming or

[17]https://dssc.eu/space/SK/759234564/Starter+Kit+for+Data+Space+Designers+|+Version+1.5+|+September+2024

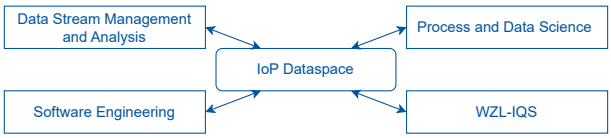

Fig. 1. Institutes connected to the IoP dataspace.

static data in application domains such as production and research. The institute is particularly interested in dataspaces as a means to support distributed analytics in production, e.g., in supply chains, and for research data sharing.

The Chair for **Process and Data Science (PADS)** conducts research in process mining, business intelligence, and workflow management, typically working with event data and process models. The institute is interested in dataspaces as possible enablers of use cases such as federated process mining, as discussed in Section II-C.

The Chair for **Software Engineering (SE)** focuses on improving software and systems development through methods, tools, and infrastructures for efficient and flexible incremental development of high-quality systems. Within the context of dataspaces, their interest extends to research-related assets, including research software. A potential use case lies in the exchange between SE and DSMA, where methods and tools for handling research software and research data could be shared and reused across institutes.

The Chair for **Intelligence in Quality Sensing (WZL-IQS)** conducts research in engineering sciences, with a focus on data about production processes, products, and quality. Their interest in dataspaces centers on supporting the exchange of heterogeneous and large-scale research assets. As an example, they considered x-ray computer tomography (xCT) data, which typically includes a parameterization file, hundreds of generated images, and an object model derived from these images [41]. Such datasets are heterogeneous in structure and can reach volumes of around 150 GB, making sharing challenging despite the transferability of the data once produced.

Overall, these institutes illustrate the range of possible roles within a dataspace—some acting as providers of data, others as consumers interested in accessing or testing data, and others as participants to research the dataspace infrastructure itself.

B. Preparation

In the preparation phase, we specified the functional goals of the dataspace, as previously compiled in Section IV. We decided on two core use cases to pursue within this work: First, we chose the RDM use case of integrating the Dataset Finder with the dataspace to cover both the discovery and exchange of research data in general. Complementing the first, we chose a federated process mining use case to explore data spaces as a testbed for federated analysis pipelines in industrial collaborations beyond research projects.

C. Implementation

Creating a new dataspace requires configuring, deploying, and connecting multiple components. While common standards and protocols are readily available, there are various implementations and technology stacks for dataspace deployment [42]. Depending on the requirements of a given dataspace project, some components can be reused, whereas others must be adapted or developed from scratch.

The Eclipse Dataspace Components (EDC) connector is a widely used dataspace base component, employed in different versions, e.g., in the automotive sector in Catena-X, the Mobility Dataspace[18] and the German Culture Dataspace[19].

For the IoP, we chose the EDC-based software stack of the German Culture Dataspace[20], which is built upon sovity's EDC stack[21], for its ready-to-use preconfiguration that only requires one file per connector for configuration, and its overall strong emphasis on usability through its extended connector frontend.

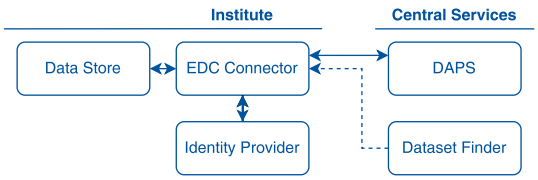

Fig. 2. Architecture of the IoP dataspace.

We first focused on deploying and validating the core components within one institute before rolling out the implementation to the other participating institutes. Central to this setup was the Dynamic Attribute Provisioning Service (DAPS) component, which contributes to trust by issuing and validating tokens conveying the identity and access rights of participants, while also acting as the central intermediary that enables participants to discover and recognize each other within the dataspace.

First, we deployed the DAPS and two connectors on three virtual machines to explore and verify the complete setup. Two of the machines run connectors, which enabled us to establish connections between them and test the full data exchange workflow under realistic conditions. The third machine hosts the DAPS, providing authentication and authorization, just as in a real-world dataspace deployment. This local deployment allowed us to validate the system's functionality, identify potential issues early, and gain practical experience before the eventual roll-out to the participating institutes.

Based on the experiences gained during these tests, we prepared streamlined, step-by-step guides for the setup of the DAPS and the connectors. We further prepared general instructions for using the connector interface to offer, negotiate, and receive data with other connectors, which we provided as part of the onboarding process. The instruction files are available in the supplementary material [43].

[18]https://mobility-dataspace.eu/
[19]https://datenraumkultur.de/
[20]https://github.com/Fraunhofer-FIT-DSAI/drkultur-edc
[21]https://github.com/sovity/edc-ce

TABLE II
ASSESSMENT OF THE IOP DATASPACE IN ITS PILOT STAGE AGAINST OPERATIONAL REQUIREMENTS.

Operational Requirement	Rating
Ease of setup and onboarding	* / **
Fast setup and onboarding time	* / **
Compatibility with heterogeneous environments	***
Usability	***
Maintenance	**
Scalability	**
Support and governance	**

Ratings: * = limited, ** = moderate, *** = good.

D. Operation

Following the deployment of the core components, we onboarded four institutes. The objective was to evaluate the feasibility of connecting heterogeneous environments to the dataspace, to identify recurring challenges, and to assess how well the deployment aligned with the requirements derived earlier. We streamlined connector deployment as much as possible, e.g., by preregistering the connectors with the DAPS. By handling this part of the configuration in advance, we lowered the participation effort and allowed participants to focus on completing the simplified local deployment. We provided each institute with a reference implementation of the connector. Authentication in the connectors' web application frontend was linked with existing institutional identity providers via Keycloak, while the integration with local storage systems remained the responsibility of each institute. During the rollout, we supported the institutes in resolving connector deployment challenges and iteratively improved the deployment documentation based on their feedback. This process helped preventing typical errors and improved the clarity of the instructions for subsequent deployments.

E. Scaling and Continuation

While ideally, all institutes of the IoP would be connected to the dataspace, such a deployment spanning all institutes of the IoP was out of scope for this work. However, we specifically focused on a straightforward deployment process across diverse infrastructures, and the gained insights, developed processes, and written documentation provide a foundation for scaling out the deployment in the future.

Considering the continuation raises the question of long-term dataspace formats. As research projects have a finite duration, this could be realized through dataspaces that are maintained independently of individual projects, while allowing project-specific subspaces to be added and removed without disrupting the core infrastructure of the dataspace.

F. Feasibility Assessment

With the deployment and onboarding experiences described above, we conducted an assessment of the practical feasibility of the IoP dataspace. Table II summarizes our assessment along the operational dimensions described in Section IV-B.

Regarding the *ease of setup*, our exploratory deployment showed that selecting an appropriate dataspace technology stack and setting up its central components had a steep learning curve and required substantial technical expertise. However, the *ease of onboarding* of the institutes was of moderate complexity: With the provided guide, the connector setup could be completed in a straightforward manner by researchers with a computer science background, or by technical administration staff. While the *setup time* of the central components was prolonged due to several iteration and adaptation steps, documented *onboarding times* for the technical onboarding of institutes were moderate, ranging from one to four hours. This included firewall configuration and integration to differing local *environments*. The chosen stack offers a connector frontend which provided good *usability*; only the asset creation and negotiation concepts were not entirely self-explanatory, and researchers relied on step-by-step instructions for first-time use. While an elevated amount of troubleshooting was necessary during the initial phase, *maintenance* tasks remained modest once recurring issues were identified and the system was running. For long-term operation of the dataspace however, resource constraints should be addressed; e.g., by sourcing funding for a designated maintainer role. For *scalability* considerations, scaling the dataspace to all 30 institutes in the IoP project would be challenging to realize in the current state with the initial cost of onboarding in terms of resources. However, preregistration and refined documentation were found to streamline subsequent onboardings, indicating potential for a significant enhancement of the process. Finally, while the step-by-step guides were helpful for onboarding the institutes, timely expert *support* remained important for troubleshooting, indicating a level of required support not fully covered in currently existing documentation.

Overall, while the system operates smoothly once deployed, onboarding and especially setup currently demand substantial technical expertise, highlighting the need for further work on lowering the entry barrier for broader adoption.

VI. DATASPACES AS ENABLERS FOR COLLABORATION

To challenge our derived dataspace design with respect to real-world needs, we consider two orthogonal applications. First, we source the IoP project in light of the required data exchange that arises in complex, collaborative research settings. Second, we go beyond the research project and focus on the vision of the IoP, which considers cross-organizational collaboration in industrial manufacturing supply chains. This way, our assessment captures two future deployment areas.

A. Integrating the Data Catalog with Dynamic and Sovereign Research Data Exchange

As a large-scale research project, the IoP maintains its own research data infrastructure, including the Dataset Finder (Section II-B), which provides a centralized catalog for discovering data resources. As part of the metadata for a dataset, data providers can describe how the data can be accessed. Currently, links to external data repositories, e.g., Github or Zenodo are supported, as well as e-mail addresses for contacting the data provider for exchanging data on a separate

channel. In this use case, we extend the capability of the Dataset Finder by linking its metadata entries to data resources available in our dataspace.

We realized the following scenario from our project work. In a previous study, interviews had been conducted on established research data management practices inside the IoP. The interviews were transcribed and anonymized. While this dataset does not contain personally identifiable information, due to its internal nature, it is not to be published or stored out of the project's immediate control. However, to make the insights accessible to interested IoP researchers and provide them with a basis for improved practices, there is a need to share the dataset, while keeping track of the actual users of the data, and without the overhead of negotiating the data transfer via email for every requesting researcher.

We store the interview transcripts within SE's infrastructure and grant access through their dataspace connector. We further extend the Dataset Finder such that each dataset can reference a resource inside the dataspace as part of the metadata. This approach enables the following process: (i) Users can search for specific datasets. (ii) If the resource is available through the dataspace, a corresponding button is displayed, which forwards the potential data user to the web interface of their local dataspace connector. (iii) In our modified version of the connector interface, this link redirects the user directly to the specific listing of the interviews where they can negotiate the actual data exchange to their local infrastructure.

This way, we augment the existing research data management infrastructure in the IoP with dataspace technology and enable easy, sovereign, and traceable exchange of research data inside the project. The Dataset Finder repository, the modification script, and recordings of the exchange are available in the supplementary material [43].

B. Dataspaces as Testbeds: Federated Process Mining

Our second use case is motivated by the vision of the IoP, as we explore a simplified scenario of cross-organizational collaboration in industry. Given is a supply chain comprising a carrier, a manufacturer, and a supplier. They jointly manage purchase, payment, and transportation within an order. For their process, the organizations want to identify which of the activities within the supply chain follow each other. This sensitive information should be exploited without sharing detailed event or order data. Using partial logs of each organization and their handover relations, i.e., which activities occurred at the sender and receiver of a handover-of-work, a joint Directly-Follows Graph (DFG) is computed using abstractions as introduced in Section II-C.

Figure 3 details the corresponding protocol that is utilized as part of the federated process mining algorithm:

1) Each organization maintains their event logs in their local data store, which is connected to the dataspace through the local EDC connector. All three organizations internally compute their partial DFGs as their abstractions.
2) The partial DFGs and handover relations are provided as assets through the respective connectors.
3) One organization acts as the server for the joint computation; in our case, the manufacturer takes on that role. The server requests the necessary assets from the other two organizations by initiating negotiation via the connectors.
4) The collected abstractions and handover relations are aggregated into the resulting DFG by the server.
5) The resulting DFG is provided as an asset through the manufacturer's connector, ready to be requested by the carrier and supplier.

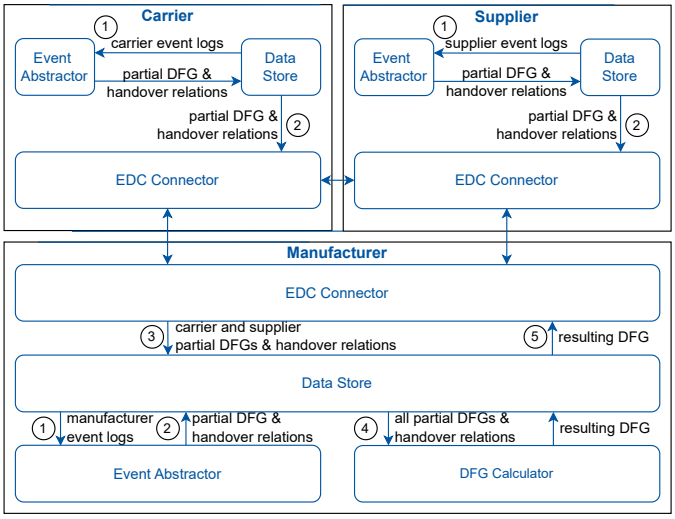

Fig. 3. Dataspace perspective on the federated process mining use case.

Figure 4 shows a snippet of the resulting DFG. In the snippet, the directly-follows relations between activities of manufacturer, supplier, and carrier are given with their counts, and the interplay of activities for shipping of goods is shown.

In this use case, our dataspace testbed enables us to research ad hoc collaborative computations across organizational boundaries through standardized interfaces and highlights the interplay between federated computations and cross-organizational data exchange. The event data used for our use case is created artificially and can be simulated using CPN Tools[22]. For our assessment, we adapt a CPN Tools model from prior work [25]. The event data, the Python scripts for preprocessing, the model to create the event data, the full resulting DFG, and screen recordings are available in the supplementary material [43].

Overall, the use cases illustrate possible ways of realizing and enhancing collaborative research through dataspaces. Although a broader evaluation would be valuable to further assess impact in practice, the two exemplary yet realistic scenarios suggest potential for various applications.

VII. Conclusion and Future Directions

This paper derived steps necessary for developing and deploying a dataspace for a large-scale research project, as exemplified through the IoP, thereby highlighting three main

[22]https://cpntools.org/

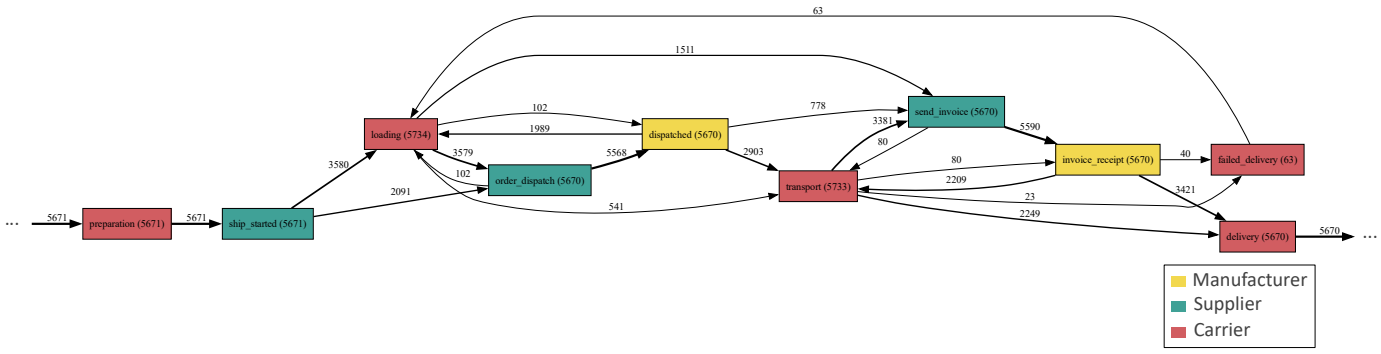

Fig. 4. Snippet from the resulting directly-follows graph of the supply chain that shows the counts of activities (with their counts) and the count of relations indicating activities that directly followed each other. The activities are mapped to organizations through color coding.

contributions: (i) deriving requirements for collaborative research infrastructures, (ii) demonstrating the deployment and onboarding of institutes using an EDC-based stack, (iii) and illustrating through two use cases the potential of dataspaces as infrastructure for research data exchange, and as a practical research testbed. Together, these contributions demonstrate that dataspaces are indeed a viable approach for cross-institutional collaboration in large-scale research projects.

There are a few limitations to note. Onboarding was limited to four institutes, with most of them having strong technical expertise, and data assets were either synthetic, or selected for exploratory purposes. The effort required for onboarding institutes from other disciplines, or for handling sensitive or high-volume data, remains to be assessed. Moreover, there is a need to examine additional use cases beyond the two presented examples to capture further forms of collaborative research.

As our findings highlight the need for work on broader practical adaptation, we derive the following research directions toward sustainable, production-grade research dataspaces.

- *Governance frameworks.* Existing governance concepts from industrial dataspaces need to be evaluated for their adaptability to research collaborations, where organizational structures and incentives may differ.
- *Incentive structures.* To support sustainability and quality of collaborative research through dataspaces, measurements should be incorporated to incentivize and motivate researchers in data sharing and collaboration.
- *Trust and quality.* Research requires not only trust between participants, but also confidence in the quality, provenance, and reproducibility of shared data.
- *Privacy and security.* Especially for sensitive data, dataspaces must provide dependable guarantees for privacy protection and secure data handling [30], [44]. Moreover, the technical capabilities and organizational perceptions of security building blocks must be better communicated to close the gap between promise and practice [45].
- *Lowering entry barriers.* To enable broad participation, technical hurdles must be reduced through improved tooling, documentation, and managed services.
- *Integration with existing tools.* As exemplified through the Dataset Finder, the integration of dataspaces with other RDM platforms and repositories in use may be promising to improve collaborative workflows.
- *Working with varied data.* Future research should explore the handling of a greater variety of data. Streaming data, for instance, poses different requirements, and brings other use cases, than static datasets: Through collaborative pipelines that combine preprocessing and edge-based reduction with dataspace exchange capabilities, practical research on massive data streams could be enabled.
- *Networks of dataspaces.* Moving beyond individual projects, future research may explore how research dataspaces can interconnect across projects and domains.

With our findings suggesting that dataspaces hold potential as an infrastructure for collaborative research, this paper provides a contribution toward and foundation for their systematic adoption and broader use in academic settings.

Finally, our deployment of an industry-grade dataspace in research demonstrates potential to enhance collaboration not only within academia, but also to extend seamlessly to data-driven cooperation with industry.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2023 Internet of Production – 390621612 and by the Federal Ministry of Research, Technology and Space under the grant number 16IS25009. We would like to thank Dr. Johannes Theissen-Lipp for his valuable input.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work, the authors used OpenAI's GPT-5 to improve the readability of the paper. After using this service, the authors reviewed and edited the content. They take full responsibility for the content of the publication.

References

[1] E. Tacconelli, A. Gorska, E. Carrara, R. J. Davis, M. Bonten, A. W. Friedrich, C. Glasner, H. Goossens, J. Hasenauer *et al.*, "Challenges of data sharing in european covid-19 projects: A learning opportunity for advancing pandemic preparedness and response," *The Lancet Regional Health - Europe*, vol. 21, 2022.

[2] W. K. Michener, "Ecological data sharing," *Ecol. Informatics*, vol. 29, pp. 33–44, 2015.

[3] J. Pennekamp, M. Henze, S. Schmidt, P. Niemietz, M. Fey, D. Trauth, T. Bergs, C. Brecher, and K. Wehrle, "Dataflow challenges in an internet of production: A security & privacy perspective," in *CPS-SPC@CCS*. ACM, 2019, pp. 27–38.

[4] Y. Demchenko, C. de Laat, W. Los, and L. Gommans, "Defining platform research infrastructure as a service (priaas) for future scientific data infrastructure," in *Designing Data Spaces*, 2022, pp. 241–260.

[5] D. Patel, "Research data management: a conceptual framework," *Library Review*, vol. 65, no. 4-5, pp. 226–241, 07 2016.

[6] L. Corti, M. Woollard, L. Bishop, and V. Van den Eynden, *Managing and Sharing Research Data : A Guide to Good Practice*. SAGE, 2019.

[7] T. Ross-Hellauer, J. P. Tennant, V. Banelytė, E. Gorogh, D. Luzi, P. Kraker, L. Pisacane, R. Ruggieri, E. Sifacaki, and M. Vignoli, "Ten simple rules for innovative dissemination of research," *PLoS Comput. Biol.*, vol. 16, no. 4, p. e1007704, Apr. 2020.

[8] M. Sicilia, E. García-Barriocanal, and S. Sánchez-Alonso, "Community curation in open dataset repositories: Insights from zenodo," in *CRIS*, ser. Procedia Computer Science, vol. 106. Elsevier, 2016, pp. 54–60.

[9] T. E. Ferrin, C. C. Huang, D. M. Greenblatt, D. Stryke, K. M. Giacomini, and J. H. Morris, "Enhancing data sharing in collaborative research projects with dash," in *Biocomputing 2005*, 2005, pp. 260–271.

[10] C. Tenopir, S. Allard, K. Douglass, A. U. Aydinoglu, L. Wu, E. Read, M. Manoff, and M. Frame, "Data sharing by scientists: Practices and perceptions," *PLOS ONE*, vol. 6, no. 6, pp. 1–21, 06 2011.

[11] L. Tedersoo, R. Küngas, E. Oras, K. Köster, H. Eenmaa, A. Leijen, M. Pedaste, M. Raju, A. Astapova *et al.*, "Data sharing practices and data availability upon request differ across scientific disciplines," *Scientific Data*, vol. 8, no. 1, p. 192, Jul. 2021.

[12] R. Abdul, A. Kumar, M. Mohana, K. Dandu, P. Goel, A. Jain, and E. Shrivastav, "The role of agile methodologies in product lifecycle management (plm) optimization," *International Journal of Computer Science Engineering*, vol. 11, no. 2, pp. 363–390, 06 2022.

[13] T. Dam, L. D. Klausner, S. Neumaier, and T. Priebe, "A survey of dataspace connector implementations," in *itaDATA*, ser. CEUR Workshop Proceedings, vol. 3606. CEUR-WS.org, 2023.

[14] M. Bacco, A. Kocian, S. Chessa, A. Crivello, and P. Barsocchi, "What are data spaces? systematic survey and future outlook," *Data in Brief*, vol. 57, p. 110969, 2024.

[15] N. Glombiewski, Z. Boukhers, C. Beilschmidt, J. Drönner, M. Mattig, A. Piet, R. Pietrzynski, M. Jaberansary, M. Maia *et al.*, "From theory to practice: Demonstrators of FAIR data spaces across different sectors," in *SAC*. ACM, 2025, pp. 509–513.

[16] P. Brauner, M. Dalibor, M. Jarke, I. Kunze, I. Koren, G. Lakemeyer, M. Liebenberg, J. Michael, J. Pennekamp *et al.*, "A computer science perspective on digital transformation in production," *ACM Trans. Internet Things*, vol. 3, no. 2, pp. 15:1–15:32, 2022.

[17] J. Pennekamp, R. Glebke, M. Henze, T. Meisen, C. Quix, R. Hai, L. C. Gleim, P. Niemietz, M. Rudack *et al.*, "Towards an infrastructure enabling the internet of production," in *ICPS*. IEEE, 2019, pp. 31–37.

[18] S.-Y. Kim, S. Hillemacher, S. Decker, B. Rumpe, and S. Geisler, "Designing and implementing practicable data management plans in large-scale projects," *Bausteine Forschungsdatenmanagement*, no. 3, pp. 1–12, 2023.

[19] S. Kim, S. Hillemacher, M. Kocher, B. Rumpe, and S. Geisler, "The dataset finder: A tool utilizing data management plans as a key to data discoverability," *Data Sci. J.*, vol. 23, 2024.

[20] W. M. P. van der Aalst, *Process Mining - Data Science in Action, Second Edition*. Springer, 2016.

[21] J. Rott, M. Böhm, and H. Krcmar, "Laying the ground for future cross-organizational process mining research and application: a literature review," *Bus. Process. Manag. J.*, vol. 30, no. 8, pp. 144–206, 2024.

[22] M. Rafiei and W. M. P. van der Aalst, "An abstraction-based approach for privacy-aware federated process mining," *IEEE Access*, vol. 11, pp. 33 697–33 714, 2023.

[23] L. Nucciarelli, R. Gatta, A. M. Tudor, E. Tavazzi, G. Arcuri, M. Vallati, G. Ibáñez-Sánchez, Z. Valero-Ramon, C. Fernández-Llatas, and A. Damiani, "Towards distributed process discovery in healthcare: Testing and proving the feasibility of the federated alpha+ algorithm," in *AIME (2)*, ser. Lecture Notes in Computer Science, vol. 15735. Springer, 2025, pp. 294–299.

[24] J. D. Hernandez-Resendiz, E. Tello-Leal, H. M. Marin-Castro, U. M. Ramirez-Alcocer, and J. A. Mata-Torres, "Merging event logs for inter-organizational process mining," in *New Perspectives on Enterprise Decision-Making Applying Artificial Intelligence Techniques*. Springer, 2021, pp. 3–26.

[25] M. Rafiei, M. Pourbafrani, and W. M. P. van der Aalst, "Federated conformance checking," *Inf. Syst.*, vol. 131, p. 102525, 2025.

[26] W. M. P. van der Aalst, "Federated process mining: Exploiting event data across organizational boundaries," in *SMDS*. IEEE, 2021, pp. 1–7.

[27] W. D. Chawinga and S. Zinn, "Global perspectives of research data sharing: A systematic literature review," *Library & Information Science Research*, vol. 41, no. 2, pp. 109–122, Apr. 2019.

[28] F. Urbano, F. Cagnacci, and Euromammals Collaborative Initiative, "Data Management and Sharing for Collaborative Science: Lessons Learnt From the Euromammals Initiative," *Frontiers in Ecology and Evolution*, vol. 9, p. 727023, Sep. 2021.

[29] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos *et al.*, "The fair guiding principles for scientific data management and stewardship," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[30] J. Lohmöller, J. Pennekamp, R. Matzutt, and K. Wehrle, "On the need for strong sovereignty in data ecosystems," in *DEco@VLDB*, ser. CEUR Workshop Proceedings, vol. 3306. CEUR-WS.org, 2022, pp. 51–63.

[31] S. Geisler, M. Vidal, C. Cappiello, B. F. Lóscio, A. Gal, M. Jarke, M. Lenzerini, P. Missier, B. Otto *et al.*, "Knowledge-driven data ecosystems toward data transparency," *ACM J. Data Inf. Qual.*, vol. 14, no. 1, pp. 3:1–3:12, 2022.

[32] S. R. Bader, J. Pullmann, C. Mader, S. Tramp, C. Quix, A. W. Müller, H. Akyürek, M. Böckmann, B. T. Imbusch *et al.*, "The international data spaces information model - an ontology for sovereign exchange of digital content," in *ISWC (2)*, ser. Lecture Notes in Computer Science, vol. 12507. Springer, 2020, pp. 176–192.

[33] L. Gleim, J. Pennekamp, M. Liebenberg, M. Buchsbaum, P. Niemietz, S. Knape, A. Epple, S. Storms, D. Trauth *et al.*, "Factdag: Formalizing data interoperability in an internet of production," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 3243–3253, 2020.

[34] J. Lohmöller, R. Matzutt, J. Loos, E. Vlad, J. Pennekamp, and K. Wehrle, "Complementing organizational security in data ecosystems with technical guarantees," in *The 1st Conference on Building a Secure and Empowered Cyberspace (BuildSEC '24)*. IEEE, 2024, pp. 49–56.

[35] L. Gleim, L. Tirpitz, and S. Decker, "HTTP extensions for the management of highly dynamic data resources," in *ESWC*, ser. Lecture Notes in Computer Science, vol. 12731. Springer, 2021, pp. 212–229.

[36] L. Gleim, J. Pennekamp, L. Tirpitz, S. Welten, F. Brillowski, and S. Decker, "Factstack," in *BTW*, ser. LNI, vol. P-311. Gesellschaft für Informatik, Bonn, 2021, pp. 371–395.

[37] A. V. Sambra, E. Mansour, S. Hawke, M. Zereba, N. Greco, A. Ghanem, D. Zagidulin, A. Aboulnaga, and T. Berners-Lee, "Solid: a platform for decentralized social applications based on linked data," 2016.

[38] S. Meckler, R. Dorsch, D. Henselmann, and A. Harth, "The web and linked data as a solid foundation for dataspaces," in *WWW (Companion Volume)*. ACM, 2023, pp. 1440–1446.

[39] M. Jaberansary, M. Maia, Y. U. Yediel, O. Beyan, and T. Kirsten, "Analyzing distributed medical data in FAIR data spaces," in *WWW (Companion Volume)*. ACM, 2023, pp. 1480–1484.

[40] J. K. Ssegawa and M. Muzinda, "Feasibility assessment framework (faf): A systematic and objective approach for assessing the viability of a project," *Procedia Computer Science*, vol. 181, p. 377–385, 2021. [Online]. Available: http://dx.doi.org/10.1016/j.procs.2021.01.180

[41] S. Carmignato, W. Dewulf, R. Leach *et al.*, *Industrial X-ray computed tomography*. Springer, 2018, vol. 10.

[42] G. Giussani, S. Steinbuss, N. Gras, and T. Prasse, "Data connector report," Sep. 2024.

[43] Data Stream Management and Analysis, "GitHub - dsma-org/dataspace-demo: Supplementary material of the project "Dataspaces for Collaborative Research". — github.com," https://github.com/dsma-org/dataspace-demo, [Accessed 28-09-2025].

[44] J. Lohmöller, J. Pennekamp, R. Matzutt, C. V. Schneider, E. Vlad, C. Trautwein, and K. Wehrle, "The unresolved need for dependable guarantees on security, sovereignty, and trust in data ecosystems," *Data Knowl. Eng.*, vol. 151, p. 102301, 2024.

[45] J. Lohmöller, H. Jeon, J. Hentschel, K. Wehrle, and J. Pennekamp, "Between Promise and Practice: Challenges and Misperceptions of Applying Privacy Enhancing Technologies in Business Contexts," in *The 59th Hawaii International Conference on System Sciences (HICSS '26)*, 2026.