11

Federated Data Infrastructures for Scientific Use

NFDI, EOSC, Gaia-X, and the European Data Spaces:

Comparison and Recommendations for a Committed Engagement to Shape the European Research Data Ecosystem

rfii
Rat für Informations-infrastrukturen

Federated Data Infrastructures for Scientific Use

NFDI, EOSC, Gaia-X, and the European Data Spaces: Comparison and Recommendations for a Committed Engagement to Shape the European Research Data Ecosystem

IMPRINT

Adopted in June 2024. Editorial deadline: 29 May 2024.

German Council for Scientific Information Infrastructures (RfII) – Head Office
Papendiek 16, 37073 Göttingen, Germany

Phone +49 551-392 70 50
Email info@rfii.de
Web www.rfii.de/en/home

SUGGESTED CITATION
RfII – German Council for Scientific Information Infrastructures: Federated Data Infrastructures for Scientific Use. NFDI, EOSC, Gaia-X, and the European Data Spaces: Comparison and Recommendations for a Committed Engagement to Shape the European Research Data Ecosystem, Göttingen 2024, 103 p.

Photos (Conference Report): Philip Bartz, Vera Münch

The German National Library lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available online: urn:nbn:de:101:1-2023021772.

CONTENTS

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AISBL	Association internationale sans but lucratif (international non-profit association)
DFG	Deutsche Forschungsgemeinschaft (German Research Foundation)
DG CNECT	Directorate-General for Communications Networks, Content and Technology
DG RTD	Directorate-General for Research and Innovation
DGA	Data Governance Act
EDIB	European Data Innovation Board
EDIC	European Digital Infrastructure Consortium
EOSC	European Open Science Cloud
ERA	European Research Area
ESFRI	European Strategy Forum on Research Infrastructures
e. V.	eingetragener Verein (registered association)
FAIR	Findable, Accessible, Interoperable, and Reusable
GDPR	General Data Protection Regulation
GHGA	German Human Genome-Phenome Archive
GWK	Gemeinsame Wissenschaftskonferenz (Joint Science Conference)
GXDCH	Gaia-X Digital Clearing House
HEI	Higher Education Institution
IDSA	International Data Spaces Association
IoT	Internet of Things
IPCEI-CIS	Important Project of Common European Interest for Next Generation Cloud Infrastructure and Services
NFDI	Nationale Forschungsdateninfrastruktur (National Research Data Infrastructure)
PID	Persistent Identifier
RDM	Research Data Management
RfII	Rat für Informationsinfrastrukturen (German Council for Scientific Information Infrastructures)
SIMPL	Smart open-source Middleware Platform
SRIA	Strategic Research and Innovation Agenda
WR	Wissenschaftsrat (German Science and Humanities Council)

EXECUTIVE SUMMARY

This comparative report **Federated Data Infrastructures for Scientific Use***, adopted by the German Council for Scientific Information Infrastructures (RfII – Rat für Informationsinfrastrukturen) in June 2024, focuses on three different initiatives – the National Research Data Infrastructure (NFDI – Nationale Forschungsdateninfrastruktur) in Germany, the European Open Science Cloud (EOSC), the transnational initiative Gaia-X – and on the European Data Spaces as a complex and ambitious European framework for better accessibility and unobstructed flow of all kinds of data within the Single European Market.

The report

- looks at different aspects of the above-mentioned initiatives regarding structure, fundamental objectives, and genesis up to the present day (29 May 2024),
- compares the different types of governance and funding,
- describes commonalities and differences in terms of technical structure and existing services, as well as involved stakeholders and their cooperation with each other and further actors from academia,
- identifies pending questions for the long-term perspectives of the different initiatives and the need for joint efforts to strengthen coherence, and interoperability in certain aspects such as basic technical solutions, services and standards.

At first glance, the initiatives and the even more complex European Data Spaces differ primarily in their embedding in the multi-level system of policy making and research institutions with regard to their founding purposes. With NFDI, an infrastructure is currently being developed for the challenges of research data management (RDM) at national level and the high demand for FAIR (Findable, Accessible, Interoperable, and Reusable) research data and corresponding services in Germany. EOSC, as a European initiative, implements a complex multi-actor structure for the shared and Europe-wide use of research data and related services, especially for scientific actors, but in the future potentially also for commercial users and providers. Gaia-X, as a transnational, industry-driven initiative, aims to strengthen European sovereignty in terms of data provision and to improve the competitiveness of European companies through a commitment to interoperability and the introduction of common standards. The European Data Spaces, finally, are a multifaceted instrument in the European Data Strategy, aimed at fostering a single market for data within the EU to facilitate secure, free-flowing data across sectors and borders, while ensuring high standards of data protection, privacy, and interoperability.

* In 2023 the RfII published a prior version as a working paper in German. However, the present release as a policy paper is a fundamentally revised, amended and updated document with partially new recommendations and differs significantly from the German publication. See for this prior version RfII (2023a) – Föderierte Dateninfrastrukturen für die wissenschaftliche Nutzung.

Key part of the European Data Spaces is the embedding of the data spaces in an ambitious and supporting legal framework and updated funding conditions for data-related European programmes. In the realm of this framework, the EOSC is commonly considered to be a genuine data space devoted to (public) research.

The report derives impulses for users in science and humanities, operating entities and providers, funding organisations, higher education and research institutions as well as research policy actors for the further development and shaping of scientific and science-related data infrastructures. Among other things, the RfII recommends:

- to strengthen the active engagement of researchers in the development and maintenance of data infrastructures and consider this activity as an integral part of scientific practice,
- to implement measures for networking the various data infrastructures to enable the mutual connectivity of the different initiatives,
- to further transform project-oriented funding programmes in a way that project costs which relate to RDM and according qualification of research personnel can be covered. Data use from reliable data infrastructure sources should be generally encouraged by funding organisations,
- to drive the digital transition in research institutions by promoting the development of competencies of research and support personnel for the (re)use, processing and storage of (research) data as well as the confident usage of data services and infrastructures,
- to emphasise the understanding that access to genuine high-quality research data is a matter of basic supply – especially when NFDI and the sources linked by the EOSC Federation are concerned. Basic supply should be free of costs for individual researchers. Costs for additional services or access to other research-relevant but not genuine scientific data spaces should be eligible for research funding,
- to continually strengthen EOSC to create a European Research Data Space through in-kind contribution via a strong and qualitative outstanding NFDI,
- to initiate a coordinated process for developing research data infrastructures aiming for sustainable ownership and maintenance structures.

After all, the RfII points out that the blueprints for NFDI, EOSC and Gaia-X and, to a certain extent, the European Data Spaces as well, have been designed before the actual geopolitical crises and the latest leaps in development and application of generative artificial intelligence occurred. The Council will consider and discuss potential effects of these recent events on strategies and further developments of (research) data infrastructures in Germany and Europe as well as for the Open Science debate in general in future recommendations.

1 INTRODUCTION AND MOTIVATION

Need for suitable services for research data

The exponential growth of digital data in all aspects of life and work, along with the increasing monetary and societal value placed on this data, has resulted in significant changes in the management of digital infrastructures in recent years. This trend is accompanied by a growing demand for suitable services for storing and archiving as well as analysing and reusing data. Specifically, the conditions under which researchers work with research data, act as producers of data themselves or actively utilise commercial data for research are subject to unique considerations. These encompass the entire data life cycle, i.e. collection, processing, analysis, and archiving. Individual adaptations are required in terms of discoverability, accessibility, interoperability and reusability, as well as compliance with (European) data protection and legal frameworks for barrier-free data utilisation. In a rapidly changing system, questions of data quality and sovereignty also play a significant role. In recent years, the high demand for suitable data infrastructures have been met at various levels – institutional, regional, national, transnational, European, and international – with different approaches, initiatives, and concepts for institutionalisation and standardisation.

The RfII contributes with its advice to the establishment and expansion of federated data infrastructures[1] in science and humanities and at the interfaces between research, politics and other areas of society. The RfII has accompanied the development of data infrastructures since its founding in 2014 – by involvement of individual Council members in the various infrastructures and committees, the exchange with other stakeholders and the intensive monitoring of developments in this area in Germany and abroad. Consequently, the Council has a lasting impact on the digital transformation of the scientific system, especially with its recommendation for establishing a National Research Data Infrastructure (NFDI) in Germany.[2] While the initial analyses of European, transnational, and international developments aimed to provide crucial insights for establishing NFDI, since 2018 the Council's focus has increasingly shifted towards research policy activities to facilitate high-quality digital and data-intensive science and humanities by establishing federated data infrastructures across Europe.[3]

1 In this publication, the RfII consistently uses the term 'Data Infrastructures' as it is currently common in the international debate and is predominantly used by the initiatives being examined. However, the RfII generally prefers the slightly broader defined term 'Information Infrastructures': 'Information infrastructures are technically and organisationally networked services and facilities for accessing and maintaining databases, information bases, and knowledge bases. In the context of the RfII's counselling work, they primarily serve research purposes, are often objects of research, and always function as an enabler. […]', RfII (2016) – Enhancing Research Data Management, p. 78.

2 RfII (2016) – Enhancing Research Data Management

3 RfII (2018) – Statement on the EOSC; RfII (2022) – Datenpolitik, Open Science und Dateninfrastrukturen

Infrastructural requirements of the scientific system

The scientific system as a whole, as well as researchers, have specific requirements towards data infrastructures, primarily aiming at easily accessible infrastructures, tailored to their disciplinary or domain-related needs. To counteract fragmentation and non-interoperable infrastructures, and to efficiently and effectively use public resources – especially in the current challenging budgetary periods – it is crucial that existing and new infrastructures interconnect and ideally allow for seamless integration. It would be undesirable if various initiatives or data spaces developed with different technical conditions and maturity levels, would (again) lead to (new) insular infrastructural solutions.

Structure and objectives of this report

As a first step, this paper therefore traces the genesis of three different data infrastructure initiatives and the data spaces framework and describes their basic objectives and structures based on publicly available information:

- The NFDI as a national initiative in Germany,
- the European Open Science Cloud (EOSC) as an approach initiated at European level,
- Gaia-X as a transnational initiative focussing on business data and users as well as providers from industry and
- the European Data Spaces as an ambitious supranational framework to establish a secure and interoperable single market for data with several interconnected spaces for different types of data – using for instance Gaia-X-solutions for standardisation and harmonisation.

While NFDI and EOSC primarily focus on scientific data and users, Gaia-X adopts a policy-driven approach. It is not primarily aimed at the scientific community but focuses on the issue of European data sovereignty and incentivising especially business data sharing. Gaia-X does not build an independent cloud but primarily defines standards and technical requirements for a sovereign data exchange. It therefore enables the creation of common data spaces. Gaia-X is currently just one example of numerous initiatives driven by both public and private sector stakeholders to build European data infrastructures. Furthermore, the report turns a spotlight on the European Data Spaces as part of the broader European Data Strategy and the associated EU legislation aimed at transforming the EU into a leading data-driven market.

The paper concentrates on these four subjects for the following reasons:

All four efforts have been developed in Europe over the past decade with a high level of financial commitment from responsible political institutions. Research as well as policy actors and the scientific communities consider these efforts as valuable foundations for innovative research and closer ties between the scientific system and other societal sectors such as the economy or public health. With NFDI, the scientific actors and institutions in Germany in particular are contributing an ambitious, research-driven data infrastructure to the European and global context. However, it is of great importance for the functionality and international

connectivity of NFDI that the data stored and maintained in NFDI or NFDI-related repositories can also be found, used and linked with other data and data sources by scientific and economic users abroad – especially via the EOSC structures currently being established. Technical solutions, standards and applications developed as part of the business-driven Gaia-X initiative also play a crucial role in making this happen – particularly when it comes to the interoperability of the research data space with all the other emerging data spaces.

In addition to NFDI and EOSC, which can currently be described as the most important initiatives in the field of accessing research data in Germany and Europe, Gaia-X is an initiative that harbours great potential for the interlinking of scientific and business data which can be beneficial for public research. The high relevance of Gaia-X for the scientific system can also be demonstrated by the participation of numerous research organisations and the existing collaboration between NFDI and Gaia-X (e.g. the 'FAIR Data Spaces' project)[4] and in some cases also between EOSC and Gaia-X.

After all, the RfII sees NFDI, EOSC and Gaia-X as well as the emerging European Data Spaces as interlocking and partially overlapping core elements for a viable and coherent federated (research) data infrastructure in Europe and beyond.

Following the descriptive overview in chapter two, the third chapter is devoted to an analysis and comparison in terms of the technical and financial constitution, governance structure, stakeholders and participating organisations as well as available or planned services and standards of the aforementioned infrastructures.

Recommendations to develop data infrastructures

In chapter four a demand-oriented development and scientific use of data infrastructures, which also considers the intended interaction between the various initiatives is recommended by the RfII. In addition to the users and providers of data infrastructures, the recommendations are aimed in particular at research performing organisations (higher education institutions and non-university research institutions), funding organisations as well as research policymakers. The recommendations relate not only to the further development of NFDI, EOSC, Gaia-X, and the European Data Spaces but also to other infrastructures and initiatives that may become relevant for scientific users. Nevertheless, the aim of the recommendations is also to ensure that the initiatives continue to interlock with each other in the future and to further expand structural, technical and personnel links. The central point of orientation is the identification of scientific and target-group requirements. In turn, aspects of research and funding policy play a vital role with regard to the legal design, technical structure and long-term availability (e.g. project funding vs. institutionalisation) of the infrastructures. The establishment of a single market for data, as envisaged by the European Commission, also has far-reaching implications for the entire scientific system. Crucial questions

4 https://www.nfdi.de/fair-data-spaces/

include the availability of data, conditions for data usage and data management in different sectors and disciplines, as well as the increasing regulation through European legal acts (Data Governance Act, Data Act, etc.), which also encompass scientific data use. All these factors influence the further development of federated data infrastructures.

Insights from the Herrenhausen Conference on Data Spaces in 2023

The recommendations given in chapter four also arise from discussions at the Herrenhausen Conference *Shaping Data Spaces in Germany and Europe – The Impulse of Science* (Datenräume in Deutschland und Europa gestalten – Impulse der Wissenschaft), organised by the RfII with support of the Volkswagen Foundation in Hanover, Germany, in April 2023. The conference, attended by over 220 participants from research, higher education, science funding and research policy, as well as members from the data infrastructure institutions, facilitated discussions on the current challenges of building new scientific and science-related data infrastructures. Also, ideas were developed to address different stakeholders of the scientific system. The RfII has taken up some of these ideas within its working groups and consultations and rolled them out in this report.[5]

The final chapter gives an outlook on the future topics and areas of interest for the RfII, especially in light of current challenges and crises. These challenges span geopolitical issues as well as technological progress, such as developments in the field of generative artificial intelligence.

5 See the conference report in the annex (pp. 88), also available in German language under: https://rfii.de/de/tagungsdokumentation-herrenhaeuser-konferenz/.

2 NFDI, EOSC, GAIA-X, AND THE EUROPEAN DATA SPACES

2.1 NATIONAL RESEARCH DATA INFRASTRUCTURE (NFDI)

> 'The aim of the National Research Data Infrastructure (NFDI) is to systematically manage scientific and research data, provide long-term data storage, backup and accessibility, and network the data both nationally and internationally. The NFDI will bring multiple stakeholders together in a coordinated network of consortia tasked with providing science-driven data services to research communities.'[6]

2.1.1 History and Objectives

Federal Government-Länder agreement to establish NFDI

The establishment and funding of NFDI in Germany was determined on 26 November 2018 through a Federal Government-Länder agreement by the Joint Science Conference (GWK – Gemeinsame Wissenschaftskonferenz).[7] This decision traces back to a recommendation by the RfII, which had suggested the establishment of NFDI in the policy paper **Enhancing Research Data Management: Performance through Diversity** published in 2016. The recommendation aimed to ensure the continuation of successful infrastructure projects, primarily by facilitating self-organisation and establishing an appropriate and durable supporting structure.[8]

Guiding principles of NFDI

NFDI's primary objective is to systematically unlock data resources for research, establish standards in research data management (RDM), and, as a digital, regionally distributed and networked repository of knowledge, facilitate the sustainable storage and continued utilisation of research data. The central element of NFDI are the consortia, where various scientific institutions and providers of data infrastructures collaborate to establish a sustainable range of services aligned with principles such as FAIR. Originally, NFDI was to comprise up to 30 consortia in its final stage of development. These consortia were selected in a science-led process organised by the German Research Foundation (DFG – Deutsche Forschungsgemeinschaft)[9] and accompanied by an NFDI expert committee of the DFG. This process should cover the various scientific fields and domains. Following three calls for proposals, NFDI currently comprises 26 subject-specific consortia, along with an overarching network of all consortia dedicated to basic services (see section 2.1.4).

6 https://www.dfg.de/en/research_funding/funding_initiative/nfdi/index.html

7 GWK (2018) – Agreement between the Federal Government and the Länder. In the GWK representatives of the Federal Government and the 16 Länder Governments prepare and decide on science policy matters of common interest and national scope.

8 RfII (2016) – Enhancing Research Data Management, Recommendation 4.2.1.

9 The DFG is the most prominent funding institution for basic research and research related services in universities in Germany. Its funding committees and procedures are based on the principle of self-organized governance by science and humanities.

Structural evaluation of NFDI conducted by the WR

The current funding phase of NFDI is scheduled to conclude in 2028. Preparations for the second funding phase are underway involving various stakeholders, including the structural evaluation by the German Science and Humanities Council (WR – Wissenschaftsrat)[10], which will assess the governance and effectiveness of NFDI. The individual consortia will be evaluated by the DFG. The GWK's decision on the further design of NFDI beyond 2028 is expected in 2026 (see section 2.1.7).

2.1.2 Governance

Networked Data Infrastructure

To coordinate the activities regarding the establishment of NFDI, the non-profit association NFDI e. V. was founded in October 2020 by the Federal Government and all 16 Länder, with its headquarters in Karlsruhe, Germany. The association manages the development of NFDI as a networked data infrastructure.[11]

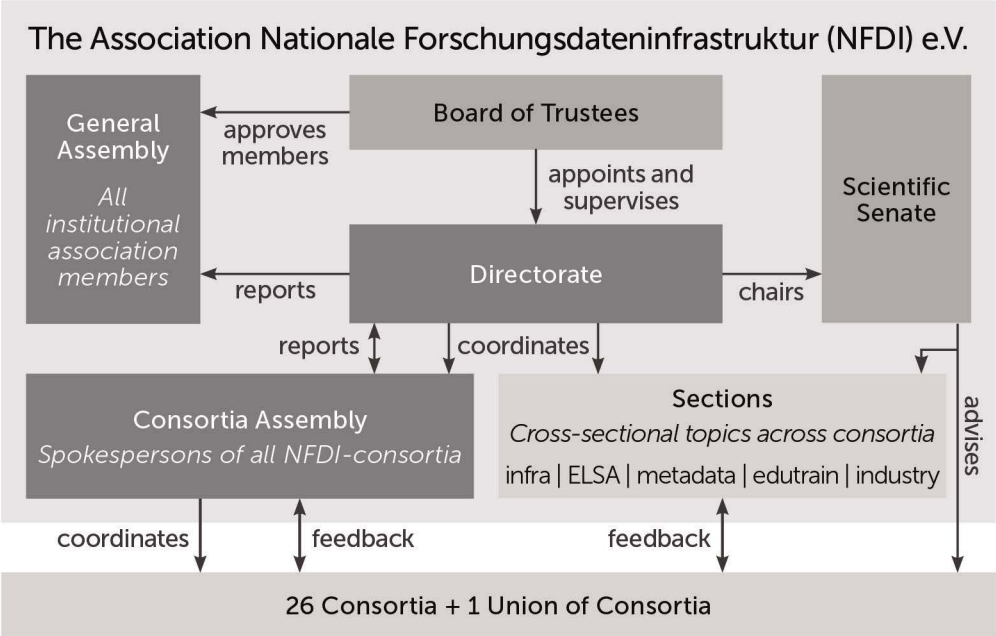


Figure 1: NFDI e. V. Governance
Source: Own illustration based on NFDI e. V. (2020) – Satzung.

Directorate: Executive Board

The governance of the NFDI e. V. consists of five bodies: Scientific Senate, Consortia Assembly, Directorate, General Assembly, and Board of Trustees.[12] The Directorate consists of a scientific head (Director) and an administrative manager. It acts as the Executive Board of the Association and is supported by a head office. The task of the directorate is to coordinate the association and to promote cooperation on content and strategy within the various bodies of the NFDI e. V.

10 The WR is an advisory body for all matters of science and higher education related policy making. It consists of representatives from science and humanities as well from Federal Government and Länder governments.
11 On the basic idea of the NFDI, see also: Brünger-Weilandt, Gehring (2022) – NFDI.
12 NFDI e. V. (2020) – Satzung.

In the NFDI Consortia Assembly, the spokespersons of each consortium meet and determine in particular the content-related and technical principles for the work of the consortia. The NFDI Consortia Assembly is thus the central coordination body for and between the consortia. It also submits decision proposals to the Scientific Senate on overarching standards and defines framework conditions for their implementation.[13]

Consortia Assembly: central coordination body between the consortia

The Scientific Senate assumes a content-related and strategic advisory function, in particular for the consortia and their projects, as well as on issues of cross-disciplinary standards and services. Comprising twelve experts nominated by the GWK, the Alliance of Science Organisations[14], and the Consortia Assembly, the Scientific Senate is chaired by the Director of the Association as its 13th member.

Scientific Senate: strategic advice

All members of the NFDI Association are entitled to participate in the General Assembly. The assembly assumes classical tasks according to German association law. More than 270 organisations belong to the association (as of May 2024). Association membership is particularly intended for organisations funded as part of an NFDI consortium by the Federal Government and the Länder or seeking such funding. However, any scientific institution can become a member of the Association, even if it does not seek funding within the scope of NFDI.

General Assembly: more than 270 institutional members

The Board of Trustees is the administrative-strategic control body of the Association. Among its responsibilities, it decides on the admission or exclusion of members to or from the Association and monitors the work of the Directorate. The Board of Trustees consists of a total of nine members, three of whom are delegated by the Federal Government and three by the Länder, with an additional three academic members elected by the General Assembly.

Board of Trustees: administrative control

According to the statutes, sections are meant to serve 'cross-disciplinary cooperation in terms of content' within the NFDI Association.[15] Their task is, among other things, to 'develop decision proposals on cross-consortium standards, metadata standards and formats'[16]. After a negotiation process among the consortia and a public consultation, four sections were established by the Scientific Senate in autumn 2021, and a fifth section was added in March 2023. These sections aim to work on common cross-sectional topics:[17]

Sections to foster cross-disciplinary cooperation

- (Meta)data, terminologies and provenance: Developing common standards, best practices for modelling terminologies, vocabularies, and ontologies, as well as 'documentation procedures of technical and cultural aspects of the context of creation of (meta)data'.[18]

13 NFDI e. V. (2022a) – Leitfaden für Konsortien im NFDI-Verein
14 https://www.allianz-der-wissenschaftsorganisationen.de/en/
15 NFDI e. V. (2020) – Satzung, § 23.
16 Ibid., p. 17.
17 NFDI e. V. (2022b) – Leitfaden für Sektionen im NFDI-Verein
18 Koepler et al. (2021) – Sektionskonzept Metadaten

- Common Infrastructures: Enhancing interoperability of subject-specific data infrastructures by embedding them in a common infrastructure, conceptualising and realising a multi-cloud-based basic infrastructure, and ensuring sustainable usability and long-term archiving.[19]
- Training & Education: Strengthening data literacy, developing target group-oriented training modules, teaching materials, and designing a certificate course for data stewards.[20]
- Ethical, Legal & Social Aspects: Serving as an exchange forum for legal, social / socio-cultural, and research ethics experiences, and developing of common guidelines and legal standards for RDM in science.[21]
- Industry Engagement: Promoting cooperation between NFDI consortia and industry, aiming to facilitate opportunities for greater exchange and the development of joint solutions and standardisations.[22]

The first three NFDI sections (Meta)data, Terminologies and Provenance as well as Common Infrastructures and Training & Education have also formed numerous working groups on specific topics or issues. The Ethical, Legal and Social Aspects section plans to work in a topic-oriented manner, for example by commenting on EU and national legislation processes.[23]

2.1.3 Financing

Funding ratio by Federal Government and the Länder: 90:10

NFDI is co-funded by the Federal Government and the Länder, providing up to EUR 90 million per year from 2019 to 2028.[24] The Federal Government contributes the absolute majority of funding with 90 per cent, while the Länder provide the remaining 10 per cent based on the Königsteiner Schlüssel[25]. With up to EUR 85 million annually, the main amount of funding goes to the consortia. Individual consortia are allocated between two to five million euro per year. Furthermore, partner institutions within consortia contribute additional funding. The NFDI e. V. does not charge membership fees. Additional activities and projects are also financed through the acquisition of third-party funds.

19 Diepenbroek et al. (2021) – Sektionskonzept Common Infrastructures

20 Herres-Pawlis et al. (2021) – Sektionskonzept Training & Education

21 Boehm, Buchner et al. (2021) – Sektionskonzept ELSA

22 Stahl, Hamann (2023) – Sektionskonzept Industry Engagement; also see Stahl, Hamann et al. (2023) – Collaboration Models Between Industry and Academia

23 NFDI-Sektion ELSA (2022) – Stellungnahme zum EU Data Act and NFDI e. V. (2022c) – Stellungnahme EU Open Data

24 Annually, up to EUR 85 million are marked for the consortia (to be distributed by the DFG), while up to EUR 2.5 million go to the NFDI directorate and office to coordinate the work of the Association and another up to EUR 2.5 million go to the DFG to cover administration costs for the selection, evaluation and overall management procedures of the consortia.

25 The Königsteiner Schlüssel regulates the distribution of the Länder share in joint financing with the Federal Government. In research policy it is used primarily for the distribution of the Länder share of the grants for the extra-university research organisations and their institutes such as the Max-Planck Society, Leibniz Association, Fraunhofer Association and for the DFG. The legal foundation for this practice is Art. 91b of the German Constitution (Grundgesetz).

2.1.4 Stakeholders and Participating Organisations

27 consortia established since 2020

The structure of NFDI, based on subject-specific consortia, enables the development of customised solutions for individual scientific fields and domains. In addition, these consortia are expected to develop cross-disciplinary standards in intensive collaboration with each other (see section 2.1.2). In the first phase of the selection process, nine consortia were selected, and their funding commenced in October 2020. Subsequently, an additional ten consortia were approved the following year, starting in October 2021. In November 2022, the GWK approved funding for eight more initiatives in the third and final phase, commencing in March 2023, aligning with the DFG's recommendations in each case.[26] In addition to seven subject-specific consortia, funding was also approved for an overarching network of NFDI consortia, Base4NFDI, which will offer services for all consortia (see section 2.1.5). For an overview of the consortia and their subject-specific tasks, see Table 1 in the appendix on pages 76 to 80.

Broad response in the German scientific landscape

Over 200 (co-)applicant organisations and more than 400 other participants are involved in the 27 consortia, including universities and research institutions, professional associations, public institutions, scientific libraries, data centres, archives and other non-profit infrastructure providers. Some consortia build on established infrastructures, such as KonsortSWD in the social sciences and economics, German Human Genome-Phenome Archive (GHGA) in human omics research and NFDI4Biodiversity in biodiversity and environmental research.

2.1.5 Technical Structure and Planned Services

NFDI will also use existing services

In the NFDI consortia, partner institutions are already providing generic services for the respective scientific communities (such as repositories and discovery systems, ontologies, knowledge bases), in addition to numerous scientific and methodological analyses and exploration tools. Some of these tools and services have been well-established in scientific communities for an extended period, with certain ones already achieving institutional sustainability. Others are actively being developed within consortia where such services are required. The term 'service' is understood within the NFDI consortia as a 'technical-organisational solution', encompassing storage and computing services, software, processes and workflows, as well as the necessary personnel support for different service desks.[27]

Reliable and sustainable basic services

In consortium collaboration, NFDI-wide basic services arising from cross-cutting themes are to be selected and provided reliably and sustainably. In particular, they should ensure interoperability and thus prevent the development of

26 GWK (2022) – Aufnahme von acht Konsortien in die NFDI. Detailed information on NFDI consortia: https://www.nfdi.de/consortia/?lang=en

27 Consortium meeting of the National Research Data Infrastructure Association, NFDI e. V. (2022d) – Stellungnahme der NFDI-Konsortien zu Basisdiensten. The definition of the term 'service' was developed by Base4NFDI, https://base4nfdi.de/process/basic-services.

divergent and incompatible solutions. In coordination with the consortia, services for collaborative work, authentication and authorisation, as well as cloud-based services and persistent identifiers, are prioritized for development.

Conditions of use of basic services

Participating institutions in individual consortia are expected to contribute subject-specific and generic services to NFDI. These services realise the implementation of the FAIR principles, address a proven need for the entire (specialist) community and should be available to them – regardless of their membership in the NFDI e. V. or in an NFDI consortium. However, this does not mean that all services can necessarily be used free of charge. Costs may arise, for example, in the form of licence fees for operational services, which may be provided by associations or a non-profit organisation, or in form of user-specific, cost-covering fees. In certain domains, particularly those closely tied to the economy, services that cannot be opened up for broad use may also be operated in cooperation with commercial partners. Various forms of an access regime are also conceivable – for example, a division into a public, non-pay area and an area subject to licensing.

Initialisation and integration phase of three projects has started

Six projects for the development of basic services are currently being funded as part of the NFDI-wide consortium Base4NFDI, which has been active since 1 March 2023 (see Table 2, p. 81). Five projects are currently in the initialisation phase. During this phase, the projects intend to develop necessary strategies and identify existing solutions and requirements to prepare the implementation of NFDI-wide basic services. The sixth project IAM4NFDI is currently in the integration phase.

2.1.6 Cooperation

The NFDI e. V. acts as mandated organisation in the EOSC Association

NFDI, specifically the NFDI e. V., cooperates with Gaia-X and EOSC on different levels and projects. For instance, the NFDI e. V. is a member of the EOSC Association, and acts as mandated organisation according to a provision of the Federal Ministry of Education and Research (BMBF). Furthermore, numerous organisations involved in the NFDI consortia are also stakeholders in the EOSC process (see section 2.2. EOSC).

Cooperation between Gaia-X and NFDI

Collaborations between Gaia-X and NFDI exist in particular in the FAIR Data Spaces project, which has been funded by the BMBF from May 2021 until May 2024 and is intended to develop synergies between the two infrastructures. The aim is to establish a joint cloud-based data space for industry and research in compliance with the FAIR principles. To this end, ethical and legal framework conditions for data exchange are being advised, technical foundations are being developed and the advantages of Gaia-X for the storage and use of research data are being demonstrated.[28] In the context of various events, the NFDI e. V. also exchanges information with other players from industry. For example, the potential and conditions for data-related cooperation between public research and industry were

28 https://www.nfdi.de/fair-data-spaces/

discussed with representatives of the BMBF and the Federation of German Industries (BDI) during an event in September 2022.[29] The exchange organised jointly by the NFDI e. V. and the BDI is to be continued.

Link to the RDM initiatives of the Länder

In addition, the NFDI e. V. and the consortia increasingly collaborate with various RDM initiatives of the Länder governments (including those from North Rhine-Westphalia – fdm.nrw, from Saxony – SaxFDM, and from Hesse – HeFDI), aiming to establish closer contact with specialist researchers and facilitate direct exchange on community requirements. Through the *Data Competence in NFDI* project, initiated in mid-2022, the NFDI e. V. aims 'to support and make visible the holistic and sustainable development of data literacy for researchers through the targeted networking of different actors.'[30] The project is funded by the BMBF and coordinated by the NFDI Directorate.

Perception of NFDI in HEIs

The Higher Education Barometer 2023 of the Stifterverband and the Heinz Nixdorf Foundation offers insights into the current usage of NFDI at German universities and the significance universities attribute to it.[31] According to the survey, distinct differences exist, particularly among different types of higher education institutions (HEIs). Universities show a significantly higher proportion (80.4 per cent) actively using NFDI compared to Universities of Applied Sciences (Fachhochschulen) (17.3 per cent) and private HEIs (6.3 per cent). Of the surveyed HEIs, 29.4 per cent state that they provide data via NFDI or benefit from NFDI (35.6 per cent). However, around 30 per cent of the HEIs also state that the provision of data via NFDI is not relevant to their institution. The survey also reveals that the majority (76.4 per cent) of higher education management staff consider further development of NFDI to be 'rather relevant' or 'relevant' for the long-term development of the higher education system.

2.1.7 Outlook

In 2026, the GWK decides on the future of NFDI

The agreement between the Federal Government and the Länder from 2018 frames the development of NFDI up to and including 2028. Decisions regarding NFDI's continuation beyond that date will be made by science policymakers after a structural evaluation by the WR, scheduled to conclude by the end of 2025.[32] In early 2023, the WR formed the according working group tasked with submitting an evaluation report to the GWK. Based on this report, the GWK will decide in 2026 on the conditions for continuing the funding of NFDI as a comprehensive system from 2029 on. A separate timetable exists for the evaluation of the consortia conducted

29 https://www.nfdi.de/forschungsdaten-fair-fuer-wirtschaft-und-wissenschaft-nutzbar-machen/

30 https://www.nfdi.de/data-literacy/?lang=en

31 Stifterverband, Heinz Nixdorf Stiftung (2023) – Hochschul-Barometer 2023, p. 6-8.

32 The RfII recommended the development of a concept for regular peer review processes that assesses not only scientific quality and user orientation but also social relevance and the economic use potential of the data sets, see RfII (2016) – Enhancing Research Data Management, Recommendation 4.1.3.

by the DFG, which is based on the first five-year funding period of the consortia. The second funding phase for consortia selected in the first round is set to commence on 1 October 2025 with the GWK's decision on continued funding anticipated in the summer of 2025. The deadline for applications for further funding is accordingly in August 2024.[33]

Long-term financial security required by stakeholders

In an impulse paper for the 20th legislative period of the German Federal Parliament (Bundestag), the DFG advocated for the permanent financial security of NFDI, urging its transition from a project status to a stable, enduring infrastructure. The DFG also called for additional funding to establish and operate essential basic services within NFDI, as well as for expanding existing consortia to maximise NFDI's utility across (sub)communities.[34] In its 2021 report, the Commission of Experts for Research and Innovation (EFI) echoed the call for early assurance of long-term financial security for NFDI, emphasising its connectivity and adaptability across scientific fields, transnational interoperability, and global consolidation.[35]

The NFDI Association identifies the formation of consortia, network creation for an overarching NFDI structure, subject coverage, and funding issues as challenges in the start-up phase. For a later stage, the development of NFDI is considered as a process of mutual adjustment with scientific subjects and domains.[36]

Criteria for assessing the development status of consortia

From the perspective of the NFDI expert committee[37], the second funding phase of NFDI from 1 October 2025 should serve in particular to consolidate the consortia. This means the integration of new expectations regarding data management and services into the existing consortia, and thus the adaptation to the changing requirements of the scientific users. To achieve this consolidation, the Expert Committee outlines eight key tasks – these include for instance the further development of RDM and the identification of the core tasks for which long-term funding is necessary. Expanding the target communities and increasing the use of offerings beyond the consortia,as well as expanding information, training and education offerings are further objectives.[38] The Expert Committee recognizes the differing developmental stages and prerequisites of consortia at their inception and proposes three criteria for assessing their status at the time of evaluation, along with quantitative information on work results. For a balanced assessment of the development status of a consortium, the following three criteria should therefore be taken into account: (1) the specific framework conditions of the consortium at the start of funding; (2) the progress made by a consortium in the measures justified in the (initial) application and (3) the objectives formulated in a follow-up application.[39]

33 Amelung, Bodenschatz et al. (2023) – Umgang mit Zielen der BLV, p. 5.
34 DFG (2021) – Erkenntnisgeleitete Forschung stärken
35 EFI (2021) – Gutachten 2021
36 Lipp, Sure-Vetter (2020) – NFDI
37 https://www.dfg.de/de/ueber-uns/gremien/nfdi/liste
38 DFG (2023) – Eckpunkte für die zweite Förderphase der NFDI, p. 2.
39 Ibid. p. 3.

The Consortia Assembly of the NFDI e. V. has expressed its opinion in a white paper on how to deal with the objectives of the Federal Government-Länder agreement as a basis for the structural evaluation of the WR. The white paper suggests establishing a common understanding of the goals, developing appropriate interpretations or definitions and clarifying the time dimension and long-term nature of building NFDI.[40] With regard to the measurement of success, the Assembly advocates a combination of impact case studies to assess community-level impacts and key figures and metrics to evaluate NFDI's influence on shaping the data infrastructure in Germany overall. Proposals for both impact case studies and individual metrics have already been put forth.[41]

40 GWK (2018) – Agreement between the Federal Government and the Länder, § 1 Objectives of Funding.

41 Amelung, Bodenschatz et al. (2023) – Umgang mit Zielen der BLV

2.2 EUROPEAN OPEN SCIENCE CLOUD (EOSC)

'The ambition of the European Open Science Cloud (EOSC) is to provide European researchers, innovators, companies and citizens with a federated and open multi-disciplinary environment where they can publish, find and reuse data, tools and services for research, innovation and educational purposes.'[42]

2.2.1 History and Objectives

Enabling an open and cross-disciplinary exchange of research data

The idea of an EOSC as a federated data infrastructure first appeared on the European Commission's agenda in 2015. In a Communication on the Digital Single Market[43], the Commission announced the launch of a cloud for research data. This project was concretised one year later with the Commission's communication *European Cloud Initiative – Building a Competitive Data and Knowledge Economy in Europe*[44]. The aim of EOSC is to facilitate an open and cross-disciplinary exchange of research data for at least 1.7 million researchers and 70 million other employees in the science and technology sector in Europe. From the beginning, EOSC was understood as an initiative to bring together the existing scientific data infrastructures in Europe. EOSC was officially launched with the Vienna Declaration on the European Open Science Cloud[45] on 23 November 2018. During the pilot phase until the end of 2020, the European Commission allocated funding of approximately EUR 320 million to over 50 projects under Horizon 2020, the EU's research and innovation funding programme from 2014-2020. This stage saw the development of initial proposals on governance, structure, and funding of EOSC, laying essential foundations for its further development from 2021 on. These developments were encapsulated in the 2020 declaration *Opening the Doors to a World of FAIR Digital Objects*[46]. A group of experts, active since 2015, also provided concrete recommendations for implementing EOSC as a federated data infrastructure.[47] Their focus was on enabling trustworthy access to services, systems, and the reuse of shared scientific data across disciplinary, social, and geographical boundaries, building on existing structures in the Member States.[48]

42 https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en

43 EC (2015) – A Digital Single Market Strategy.

44 EC (2016a) – European Cloud Initiative.

45 https://eosc-launch.eu/declaration/

46 o. A. 2021 – Threshold to the European Open Science Cloud.

47 EC (2016b) – Realising the European Open Science Cloud; EC (2018a) – Prompting an EOSC in Practice.

48 For a further insight into the origin of EOSC, see also the following article by the former head of unit at the Directorate General for Research and Innovation (DG RTD): Burgelman (2021) – How the EOSC Became Reality; and the reply by Wittenburg (2021) – Reply to Jean-Claude Burgelman.

Tripartite governance

The central steering structure, known as the tripartite governance, comprising the EOSC Association, EOSC Steering Board, and European Commission, was established with the creation of the EOSC Association as a non-profit organisation under Belgian law (AISBL – Association internationale sans but lucratif) in the summer of 2020. The Memorandum of Understanding (MoU)[49] with the European Commission followed in early 2021, formalising the co-programmed partnership.

Developing a web of FAIR data and services

EOSC is intended to provide access to research data, related services and scientific publications – guided by the principle: as open as possible and as restricted as necessary. Its objective is to develop a web of FAIR data and services[50] primarily for science and humanities in Europe, encompassing a wide range of services, from the long-term safeguarding of research information to the visualisation and analysis of data, as well as the tracking of Open Science practices, all aligned with the FAIR principles. The envisaged characteristics of EOSC extend beyond its name; it is not exclusively European, not inherently open (as protected data, like patient data, will play a role), not exclusive for public research (as commercial or economic data will also be used and shared), and not a central cloud (since not all data will be stored centrally, but will be found on federated infrastructures).

By the end of 2021, the Council of the European Union defined EOSC as one of 20 pilot measures to deepen and strengthen the new European Research Area (ERA).[51] As a data space for science, research, and innovation, EOSC, alongside other sectoral data spaces, is intended to form part of the European Data Spaces outlined in the European Data Strategy (see section 2.4).[52]

2.2.2 Governance

Member states are involved via the Steering Board

For the management of EOSC, a tripartite governance has been established, consisting of the EU represented by the Commission and its Directorates-General (DG RTD – Directorate-General for Research and Innovation and DG CNECT – Directorate-General for Communications Networks, Content and Technology), the participating countries (EU Member States and Associated Countries) represented through the EOSC Steering Board, and the research community represented by the EOSC Association. The Steering Board plays a crucial role in integrating European Member States and Associated Countries into the EOSC setup process. Additionally, it provides strategic advice to the Commission, particularly on coordinating national and European guidelines for RDM and other EOSC-relevant aspects, as well as funding issues through in-kind contributions and other means.

The Association's member organisations convene regularly in General

49 EOSC AISBL, EC (2021) – Memorandum of Understanding.

50 https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en

51 Council of the European Union (2021) – Future Governance of the ERA, p. 15.

52 EC (2020) – A European Strategy for Data.

Assemblies, where key decisions, including the election of the Board of Directors (BoD), are made. The nine members (including the President) of the BoD work on a voluntary basis and receive no remuneration from the EOSC Association. The main task of the Board is to realise the purpose of the Association and to manage its activities by implementing the decisions, instructions and recommendations adopted by the General Assembly.[53] The Board is also responsible for involving the member organisations and other EOSC stakeholders. To facilitate this, various Advisory Groups have been established, addressing different aspects and issues of EOSC, such as implementation, technical challenges, metadata, monitoring, research careers, and sustainability. Several Task Forces were assigned to each Advisory Group, which provided strategic advice on specific topics. A total of 13 Task Forces were set up and active in the period from 2021 to the end of 2023. However, the aim of the Task Forces was not to develop specific services or service components; this task is taken on by the projects funded under Horizon Europe or by the successful consortia in the procurement procedure. In mid-2024, however, four new Task Forces will be launched, for which the EOSC Association published a call for experts in March 2024.[54] These new Task Forces will focus on the following topics: FAIR Metrics and Digital Objects; EOSC Technical and Semantic Interoperability; Long-term Data Retention; and Health Data. Not a Task Force but an advisory group on *Data Spaces and Industry* has been set up at the 7th General Assembly at the end of 2023. The objective here is apparently to establish a stronger thematic link to Gaia-X and the European Data Spaces activities.

The primary guiding document for steering EOSC and defining the development steps is the Strategic Research and Innovation Agenda (SRIA),[55] which also defines the general framework for future research, development, and innovation activities. The Multi-Annual Roadmap (MAR)[56], a part of the SRIA, outlines interim goals and a concrete work plan for two-year periods.

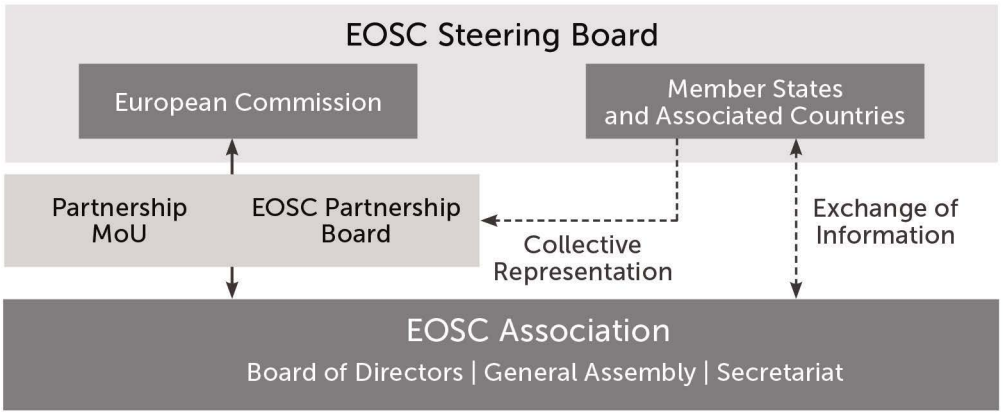


Figure 2: EOSC Governance (CC BY)
Source: Own illustration based on Beckmann (2022) – Monitoring EOSC Readiness.

53 EOSC Association (2022) – Articles of Association, p. 9.
54 https://eosc.eu/open-call-for-eosc-association-task-force-members/
55 EOSC Association (2023a) – SRIA
56 EOSC Association (2023b) – EOSC Multi-Annual Roadmap

EOSC Federation and the role of nodes

As part of the ongoing planning for the actual implementation of EOSC, the concept of the EOSC Nodes was proposed by the Commission in spring 2023. This concept outlines the integration of European, national, regional, institutional, and thematic initiatives into the EOSC Federation. The EOSC Federation is defined as a federation of distributed systems ('system of systems'), comprised of multiple nodes that can be scaled by adding further nodes. In turn, EOSC Nodes contain resources that add value to the federation, act as legal representatives that interact with the EOSC governance structure, provide interfaces consistent with the EOSC interoperability framework, control their own operations and resources and ensure adherence to policies within the node. The planned EOSC EU Node is intended to be the central entity of EOSC at the European level and will include core services and infrastructure developed as part of the *Managed Services for the EOSC Platform* procurement (see section 2.2.5). The EOSC EU Node will enable the federation of existing data, research, and e-infrastructures operating at institutional, national, regional, pan-European and thematic levels.[57]

Overarching tasks for further development

The long-term governance and financing of EOSC beyond 2027 are currently under intensive discussion within the tripartite governance, as well as with the members of the Association and other stakeholders. The European Commission has identified five overarching tasks for the further development of EOSC:

Task 1: Deploying and operating the EOSC EU Node (Core, Exchange, FAIR Data Federation). The EOSC EU Node will then represent the fully functional infrastructure for EOSC[58].

Task 2: Maintaining and updating the EOSC EU Node and expanding the EOSC Federation (with elements that are close to the market).

Task 3: Enabling a web of FAIR data and services for science and humanities.

Task 4: Developing, prototyping and testing new elements supporting the evolution of the EOSC Core and Exchange and the tools enabling the federation (elements that can be made ready for the market).

Task 5: Enabling Open Science policies and the uptake of Open Science practices.[59]

Possible governance models

For the fulfilment of these tasks and the continuation of EOSC beyond 2027, possible governance and financing models are being discussed in the EOSC Steering Board, especially regarding the further operation of the EOSC EU Node. They range from a very flexible model where these tasks may be taken over by an existing institution to a more complex model in which the establishment of a new institution a) as an Art. 185 or 187 measure ('Joint Undertaking') or b) under private law is envisaged.

57 https://eosc.eu/wp-content/uploads/2024/01/20240116-17_EOSC-A_Webinars-on-EOSC-Federation-and-the-role-of-Nodes.pdf

58 See section 2.2.5 for details.

59 Dumouchel (2023) – EOSC Building, p. 8.

In order to continue a coordinated development of the EOSC Federation a 'Tripartite Group' was formed by EOSC Tripartite Governance stakeholders in April 2024, consisting of representatives of the Commission, the EOSC Steering Board and the EOSC Association. The main task of this group is to prepare joint positions on the development of the EOSC Federation and its further operation – particularly regarding the continuation of the EOSC beyond 2027.

2.2.3 Financing

Project funding and in-kind contributions

The financing of EOSC is secured from two sides within the tripartite governance: the European Commission has committed itself to investments through the current research and innovation funding programme until 2027 and within the framework of a co-programmed European partnership amounting to EUR 490 million. Concurrently, the Member States and Associated Countries contribute to EOSC via in-kind contributions amounting to around EUR 500 million. The EOSC Association costs are covered, among other sources, by membership fees from the participating organisations, set at EUR 10 000 per year, and EUR 2 000 per year for organisations with observer status.

The work programmes for research infrastructures within Horizon Europe, notably *Destination – Enabling an Operational, Open and FAIR EOSC Ecosystem (INFRAEOSC)*, provide funding for numerous projects aimed at contributing to building EOSC and achieving the goal of a web of FAIR data and services. For 2021 and 2022, approximately EUR 90 million were allocated for this purpose.[60] Calls for tenders totalling EUR 130 million have been planned for 2023 and 2024.[61] In addition, an EOSC Procurement[62] with an estimated total value of EUR 32 million, will finance the development of the EOSC Core Platform and the EOSC Exchange, thereby developing the basic functions of the future fully operational EOSC.

The financing of EOSC after 2027, as outlined in the previous section, remains unresolved. It is anticipated that ongoing operational funding for basic services within the EOSC EU Node (see section 2.2.5) will necessitate permanent contributions from participating states.[63]

2.2.4 Stakeholders and Participating Organisations

Key stakeholders in EOSC are the member organisations of the EOSC Association and the partners in the EOSC-related projects funded by the European Commission under the Horizon Europe research programme (and its predecessor Horizon 2020).

60 EC (2022a) – Work Programme Research Infrastructures 2021-2022.
61 EC (2022b) – Work Programme Research Infrastructures 2023-2024.
62 https://digital-strategy.ec.europa.eu/en/funding/managed-services-european-open-science-cloud-platform
63 EOSC – Financial Sustainability Task Force (2022) – Towards Sustainable Funding Models

As of May 2024, the EOSC Association consists of more than 250 members, with 87 organisations having observer status, and 29 organisations designated as Mandated Organisations. The latter are member organisations mandated by the respective Member States to represent their scientific landscape in the Association. For Germany, the NFDI e. V. has assumed this role after the DFG acted as a mandated organisation during a transitional period. However, not all Member States and Associated Countries have appointed mandated organisations so far, even though they represent national interests in accordance with the Association's statutes: Each Member State or Associated Country with one or more organisations as association members may appoint one member as its Mandated Organisation to represent national interests.[64]

EOSC Association consists of more than 250 members

Prerequisites for admission as a member organisation in the EOSC Association include legal entity status or status as an intergovernmental organisation (e.g. European University Association), based in a European Member State or an Associated Country. For observers, the latter is not necessary. Furthermore, member organisations may not be national ministries or their departments but are primarily research organisations, research funders, and service providers.

Other important EOSC actors are the various projects funded under the EU programmes for research and innovation, addressing different EOSC-relevant topics in successive funding periods and involving numerous research organisations from the Member States. Alongside completed projects that initiated crucial specifications of EOSC (EOSCpilot, EOSC Hub, etc.), several projects launched since the beginning of 2021 focus on the actual implementation of EOSC and its federated infrastructure and associated services. With EOSC Future[65], for example, a project was launched that is intended to introduce the actual implementation of EOSC. Other projects increasingly explore interfaces with various scientific disciplines (e.g. EOSC-Life) and to the roadmap of the European Strategy Forum on Research Infrastructures (ESFRI), which provides essential data resources for EOSC, enabling cooperation with the involved disciplines and infrastructures. In addition, various projects are funded that involve different scientific disciplines and stakeholders and develop concrete services.[66]

Several projects foster the implementation of EOSC

In 2024, seven new projects start (funded under Horizon Europe) or have already started. These include, for instance, the EOSC Beyond project, which aims to provide new EOSC core functions that enable scientific applications to find, compile and access Open Science resources and offer them to researchers as integrated functions. The project also supports the new EOSC EU Node concept (see 2.2.5), which aims to create a federated and integrated network of nodes operating at different levels – national,

64 EOSC Association (2022) – Articles of Association
65 https://eoscfuture.eu/
66 E.g. https://eosc.eu/news/six-new-he-infraeosc-projects

regional, international and thematic.[67] As another example, the *Open Science Cluster Action for Research & Society (OSCARS)*[68] project aims to unite European research infrastructures listed in the ESFRI Roadmap and beyond, thereby advancing the adoption of Open Science in Europe. These research infrastructures belong to five science clusters (Humanities and Social Sciences, Life Sciences, Environmental Sciences, Photon and Neutron Science, Astronomy, Nuclear and Particle Physics) that strive to make open data easily accessible to users and the public by implementing FAIR data management policies and practices to enable EOSC Open Science. The project, and thus the science clusters, are an integral part of the development of EOSC and are expected to help link ESFRI and other world-class research infrastructures to EOSC.[69]

2.2.5 Technical Structure and Planned Services

Development in various iterations

The technical structure and functionalities of EOSC are to be developed through various iterations, building on each other and subsequently forming the core of the federated data infrastructure. It is planned to build the minimum required functions and services for EOSC based on scientific demand. This follows the principle of a 'Minimum Viable EOSC' (MVE), consisting of a governance structure, the EOSC Core, EOSC Exchange, and the actual federated services and is primarily available for scientific users. To realise its full potential, EOSC will be extended beyond the European research community to the wider public and private sector at a later stage (see Figure 3, p. 24).

Public procurement to develop the platform and core services

The EOSC Core will provide the basic functional framework (including authentication and authorisation rules for access, persistent identifiers, metadata standards, catalogues, etc.) of EOSC for discovery, sharing, access, and reuse of scientific data and objects. These will be processed by the federated services. With EOSC-Exchange, a marketplace is to be established in a second and partly parallel process, which will make it possible to offer and use various indepth services and tools. These are to be oriented towards an EOSC interoperability framework and follow various Open Science principles.

In November 2023, the European Commission announced the results of the EOSC Procurement *Managed Services for the European Open Science Cloud (EOSC)*.[70] With this tender, the Commission and the responsible DG CNECT intend to build and deploy a fully functional infrastructure for EOSC, the so-called EOSC EU Node. The requirements for participating in the tender were challenging research performing institutions in Germany, which are accustomed to grant allocating procedures rather than to com-

67 See https://eosc.eu/news/2024/01/the-year-2024-welcomes-seven-new-he-infraeosc-projects/.

68 https://oscars-project.eu/about-oscars

69 https://oscars-project.eu/science-clusters

70 https://digital-strategy.ec.europa.eu/en/news/commission-announces-winners-eosc-procurement

mercial procurement laws. The latter requires application documents and proofs which are generally geared towards commercial enterprises.

The public tender was organised in three lots:

Lot 1 – Core Federation Services for the EOSC EU Node
The winner of the lot is the Open Science Agora Consortium. The consortium will offer professionally managed services for the essential components of the EOSC EU Node. This includes functions like the Web Portal Front Office, Resource Catalogues, Registry Services, Application Workflow Management Engine, Federated Identity Management and Single-Sign-On solution, Monitoring and Accounting, and an overall Service Management System and Service Integration. The consortium is coordinated by the Athena Research Center (ARC). Alongside other research and infrastructure organisations from Europe, the Scientific Compute and Competence Centre of the University of Göttingen and the Max Planck Society (GWDG mbH), for example, is also involved in the consortium.

Lot 2 – Exchange Infrastructure Services for the EOSC EU Node
The winner of the lot is the Poznan Supercomputing and Networking Center (PSNC) which is affiliated to the Institute of Bioorganic Chemistry of the Polish Academy of Science. The consortium will deliver fully managed services for the infrastructure services component of the EOSC EU Node. This includes Managed Container Service, Managed Compute (Virtual Machine) Service, and Managed Bulk Data Transfer Service. The GÉANT Association serves as the underlying network provider.

Lot 3 – Exchange Application Services for the EOSC EU Node
The winner of the lot is also PSNC, which will offer fully managed services for the application services component of the EOSC EU Node. This includes Managed File Synchronization and Sharing Service, Interactive Notebooks Service, and Managed Large File Transfer Service for end-users.

Over 36 months (started in January 2024), the Consortium and PSNC will build, set up and operate the Core Federation Services, the Exchange Infrastructure Services, and the Exchange Application Services for the EOSC EU Node. The Commission is investing EUR 35 million in these projects (Lot 1, 2, and 3) through the Horizon Europe Research Infrastructures Work Programme 2022. The EOSC EU Node should be supplemented by national nodes in the participating countries, reflecting the principle of the federated data infrastructure of EOSC.[71]

71 EOSC Association (2023c) – Position Paper on the EOSC Federation and the Role of EOSC Nodes

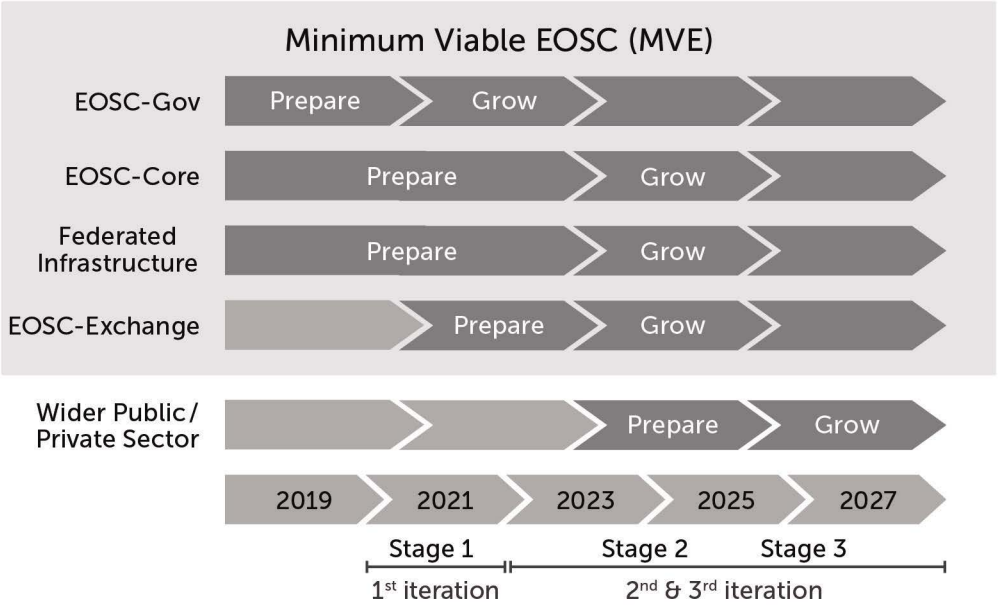


Figure 3: EOSC Timetable for Implementation (CC BY)
Source: EOSC Association (2023a) – Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud, p. 75.

First point of contact for researchers

The EOSC Portal has been a central point of contact for RDM services within the EOSC ecosystem in recent years.[72] It has been jointly developed by various projects since 2018 (eInfra-Central, EOSCpilot, EOSC-hub, OpenAIRE-Advance, EOSCsecretariat.eu, EOSC Enhance) and was most recently operated by the EOSC Future project. The portal currently contains about 400 generic and disciplinary services and tools but serves primarily as a reference system. In addition, countless publications, datasets, software components, and other resources can be found by the search function. Due to the end of the EOSC Future project on 31 March 2024 and the planned transition to an EOSC EU Node concept, the portal is offline since April 2024. However, in the spring of 2024, the European Commission published a new website about EOSC and the EOSC EU Node. The functionalities of the portal were also transferred to this new website.[73] Nevertheless, the fully comprehensive EOSC with a broad-based and cross-disciplinary data federation based on agreed standards – such as minimum metadata schemes and the FAIR principles – will be established through the EOSC EU Node, which will be owned by the European Commission and governed by the EOSC Tripartite Governance. The infrastructures and services currently being developed by the winners of the procurement (lots 1 to 3) will form the basis for the establishment of the EOSC EU Node.

72 https://eosc-portal.eu/
73 https://open-science-cloud.ec.europa.eu/

2.2.6 Cooperation

National Tripartite Events to discuss national priorities

EOSC collaborates with numerous national, European and international initiatives, networks and infrastructures in various ways. Apart from integration through project funding within Horizon 2020 and Horizon Europe, exchanges primarily occur through personal and institutional connections within the governance structure and by membership in the EOSC Association, including organisations such as CESAER, GÉANT, Science Europe, OpenAIRE, PRACE, and the Research Data Alliance (RDA). Close cooperation also exists between the EOSC Association and the EOSC projects funded under the Horizon Europe research framework programme, contributing significantly to the development of EOSC. Since mid-2022 National Tripartite Events in the Member States have facilitated stakeholder exchanges at the national level, discussing priorities and specifics.[74]

Cooperation with commercial actors is currently limited, with EOSC only planning to open to the private sector in a later stage of the implementation. Discussions are ongoing among relevant bodies regarding the appropriate involvement of commercial actors, with Gaia-X being referenced in this context.[75]

EOSC-ESFRI cooperation

The ESFRI-EOSC Task Force[76], established in 2023, brings together representatives from the EOSC Steering Board, the Commission, ESFRI, and observers from other organisations (in particular the EOSC Association, e-Infrastructure Reflection Group, European Intergovernmental Research Organisation Forum) to exchange views on cooperation between EOSC and ESFRI. Their goal is to collaboratively build an integrated ecosystem of research infrastructures, creating a web of FAIR data and services for science. The Task Force aims to reduce the fragmentation of the research data landscape in Europe by connecting ESFRI RIs and European Research Infrastructure Consortia (ERICs) to the EOSC network. It is expected to operate until the end of 2024 and may continue thereafter.

European Data Spaces and other initiatives

The pan-European Data Spaces proposed by the European Commission in early 2020 as part of the *European Strategy for Data*[77] are designed to cover various areas (e.g. Health, Green Deal, Industrial and Manufacturing), with EOSC serving as a sectoral data space for research data (see section 2.4). The development and implementation of the Data Spaces is coordinated by the Data Spaces Support Centre (DSSC)[78]. The several EOSC initiatives will need to interact with the DSSC in the future to achieve the highest possible interoperability between the various Data Spaces.

74 https://eosc.eu/news/series-tripartite-events-are-coming
75 See EC, EOSC Steering Board Expert Group (2022) – EOSC and Commercial Partners.
76 https://www.esfri.eu/working-groups/esfri-eosc-task-force
77 EC (2020) – A European Strategy for Data
78 https://dssc.eu/

2.2.7 Outlook

EOSC as sectoral data space for science, research and innovation

The aim of the EOSC is to contribute to the development of a web of FAIR data and services for science in Europe. In addition to the federation of national data infrastructures, international connections and cooperation will undoubtedly play a crucial role in the future. Similar structures are emerging on other continents and in various countries and regions, such as the African Open Science Platform (AOSP), the National Research Infrastructure for Australia (ARDC) or the China Science and Technology Cloud (CSTCloud). Initiatives are also underway to establish a Global Open Science Cloud[79]. EOSC was designated by the Council of the European Union at the end of 2020 as one of the pilot measures for consolidating and expanding the ERA and will represent the sectoral data space for science, research, and innovation in the European Union.[80]

As outlined in the previous sections, various stakeholders of EOSC are presently involved in in-depth discussions regarding the design of EOSC from 2028 onwards, especially concerning governance and funding issues. These discussions will continue and are crucial for the future success of EOSC. Furthermore, the EOSC Association is working on an update of the SRIA, which defines the foundations for the further development as well as research and innovation activities of the EOSC in the coming years. It is also a central source for the Call for Proposals in the INFRAEOSC programme part of Horizon Europe.[81]

79 https://codata.org/initiatives/decadal-programme2/global-open-science-cloud/

80 Council of the European Union (2021) – Future Governance of the ERA.

81 https://rea.ec.europa.eu/funding-and-grants/horizon-europe-research-infrastructures/enabling-operational-open-and-fair-eosc-ecosystem-infraeosc_en

2.3 GAIA-X ASSOCIATION FOR DATA AND CLOUD

> 'Gaia-X is a project initiated by Europe for Europe and beyond. Representatives from business, politics, and science from Europe and around the globe are working together, hand in hand, to create a federated and secure data infrastructure. Companies and citizens will collate and share data – in such a way that they keep control over them. They should decide what happens to their data, where it is stored, and always retain data sovereignty.'[82]

2.3.1 History and Objectives

Initiated by Germany and France

Gaia-X is a not-for-profit project that originated from a joint initiative of the German and French governments[83], first unveiled to the public in 2019 during the Digital Summit of the German government. Initially, coordination on the German side was undertaken in particular by the Federal Ministry for Economic Affairs and Energy (in December 2021 renamed in Federal Ministry for Economic Affairs and Climate Action, BMWK).[84] In their joint declaration *Building the next generation cloud for businesses and the public sector in the EU*, the 27 EU Member States and the European Commission also emphasised the role of Gaia-X as an important example of a public-private initiative for the European cloud federation and for constructing European Data Spaces.[85] Gaia-X's primary objective is to enhance European digital and data sovereignty, thereby constructing a reliable and competitive data infrastructure for Europe. To this end, Gaia-X develops cloud and data standards as well as open-source software code for cloud solutions designed to operate interoperable and independently from individual providers and business models. It is important to note that Gaia-X does not offer services itself; rather, it defines standards for a sovereign cloud infrastructure, similar to DIN standards, for instance. Although Gaia-X is primarily geared towards companies, the standards and compliant services defined by Gaia-X could also be relevant to scientific users. Similar to EOSC, Gaia-X is constructing a federated infrastructure based on common standards. Like EOSC, Gaia-X does not aim to function as an independent cloud provider. Instead, it enables the establishment of self-determined federations, allowing market players to collaborate and offer their services as certified providers. Gaia-X's overarching guiding principle is a commitment to European values and standards, particularly European regulations on data protection, to which all participating organisations must commit. Furthermore, important guiding principles of the initiative include interoperability, free market access, openness and transparency, as well as a federated and decentralised structure.[86]

Gaia-X develops cloud and data standards

82 https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html

83 Federal Ministry of Economic Affairs and Energy, French Ministry of the Economy, Finance and the Recovery (2020): Franco-German Position on GAIA-X.

84 For a detailed description, see e. g. Person, Schütrumpf (2023) – Das Projekt Gaia-X.

85 European Member States, EC (2020) – Building the Next Generation Cloud for Businesses and the Public Sector in the EU, p. 4.

86 Bonfiglio (2021) – Vision and Strategy, p. 3.

2.3.2 Governance

Joint establishment of an AISBL by organisations each from Germany and France

The organisational structure of Gaia-X consists of three pillars: the Gaia-X Association, the national Gaia-X Hubs, and the Gaia-X Community. Founded in January 2021 as a non-profit organisation under Belgian law (AISBL), the Gaia-X Association operates as an international umbrella organisation. Currently (as of May 2024), more than 300 companies and organisations have become members.[87] The Association's general direction is managed by a Board of Directors, while operational control falls under the responsibility of the Management Board. The member organisations also meet in General Assemblies. The governance structure includes two Advisory Boards and various Committees, which, with the assistance of working groups, handle operational tasks and address substantive issues.

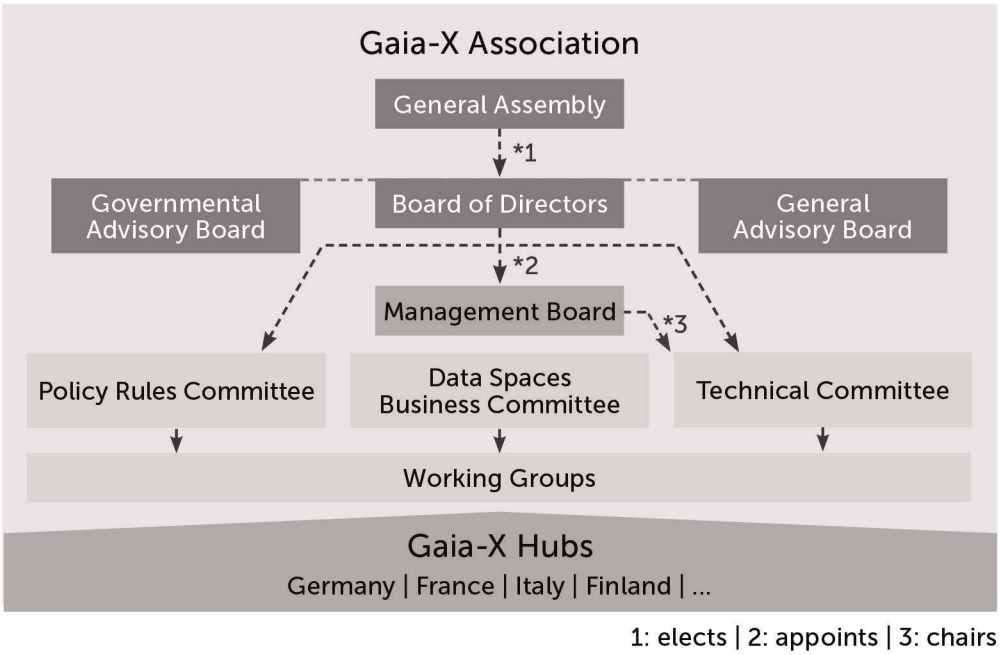


Figure 4: Gaia-X Governance (CC BY)

Source: Own illustration created in May 2024 based on https://gaia-x.eu/who-we-are/association/.

National hubs to provide central and country-specific contact points

This structure is complemented by the national Gaia-X Hubs. These exist in France, the Netherlands, Germany, and Finland, among others, and are to be established in all European countries, if possible.[88] There are currently 16 European and five non-European hubs. The national hubs provide central and country-specific contact points for companies, initiatives, public institutions and other stakeholders and have, among other things, the objective of identifying the concrete requirements of the various companies and stakeholders for the implementation of their projects within Gaia-X. In addition, the national hubs are also supposed to communicate with each other and with other national initiatives, both from the public and private sectors. In order to ensure

87 https://gaia-x.eu/what-is-gaia-x/about-gaia-x/
88 https://www.gaia-x.eu/who-we-are/hubs

a European orientation of the project, the board of the Gaia-X AISBL consists exclusively of representatives of companies based in Europe.

2.3.3 Financing

The Gaia-X Association's costs are financed through annual membership fees paid by the participating organisations. For commercial companies, these fees range from EUR 5 000 to EUR 75 000, while non-profit organisations contribute a lower amount. It is not yet determined whether the membership fees will also cover funding for technical services.

Project funding from the German and French governments

In addition, individual consortia or so-called lighthouse projects receive project-based funding, particularly in Germany and France. Since the beginning of 2022, the BMWK has been funding eleven consortia within the framework of the funding competition *Innovative und praxisnahe Anwendungen und Datenräume im digitalen Ökosystem Gaia-X* until the end of 2024, with a total of EUR 122 million.[89] The funding aims to develop practical applications based on Gaia-X. Numerous scientific organisations are involved in the various consortia of the first funding round. In a second call in February 2022, five additional projects were selected and initially planned to commence in 2022. However, due to constraints in the federal budget, they have not been funded. Nevertheless, the 2023 federal budget included an increase of EUR 4 million for Gaia-X, with a total increase of EUR 67 million for the upcoming years. The French government is also investing at least EUR 150 million in the development of Gaia-X-compliant data rooms as part of its national cloud strategy.[90]

2.3.4 Stakeholders and Participating Organisations

Commitment to European laws and data protection

Membership in the Gaia-X AISBL is not restricted to European companies or organisations. However, organisations must commit to European laws and requirements, especially in the areas of data protection, security, openness, and transparency. The provision of services within the Gaia-X ecosystem must comply with European legislation.[91]

German scientific organisations involved

The members of the German Gaia-X Hub are organised in the following domains: Education, Energy, Finance, Geoinformation, Health, Industry 4.0, Culture / Creative Industries, Agriculture, Mobility, Public Sector, Planning / Building / Operating, Smart City / Smart Region, Smart Living. The aim of these domains is to strengthen networking between the participating actors and to develop various use cases. Numerous scientific organisations and researchers from Germany are engaged in the respective domains. In addition, more than 80 organisations from Germany are already members of Gaia-X AISBL, including research organisations and universities such as

89 https://www.bmwk.de/Redaktion/DE/Dossier/gaia-x.html
90 Le Maire et al. (2021) – Stratégie nationale pour le cloud.
91 Gaia-X AISBL (2022a) – Policy Rules Document, p. 1.

German Research Centre for Artificial Intelligence (DFKI), Charité Berlin, German Aerospace Center (DLR), TU Darmstadt, Fraunhofer Society, Karlsruhe Institute of Technology, Gauss Centre for Supercomputing e. V. and University of Passau.[92]

2.3.5 Technical Structure and Planned Services

Architecture that enables sovereign data sharing

With the assistance of the so-called Federation Services (i.e. a series of required technical services, such as authentication mechanisms, catalogues for offers and services as well as possibilities for data exchange and compliance mechanisms), Gaia-X aims to establish an architecture facilitating sovereign data sharing. This architecture will be developed based on use cases across different domains or data spaces, identifying existing needs and testing the advantages of an independent European data infrastructure. The basic technical structure of Gaia-X is described in the Gaia-X Architecture Document.[93]

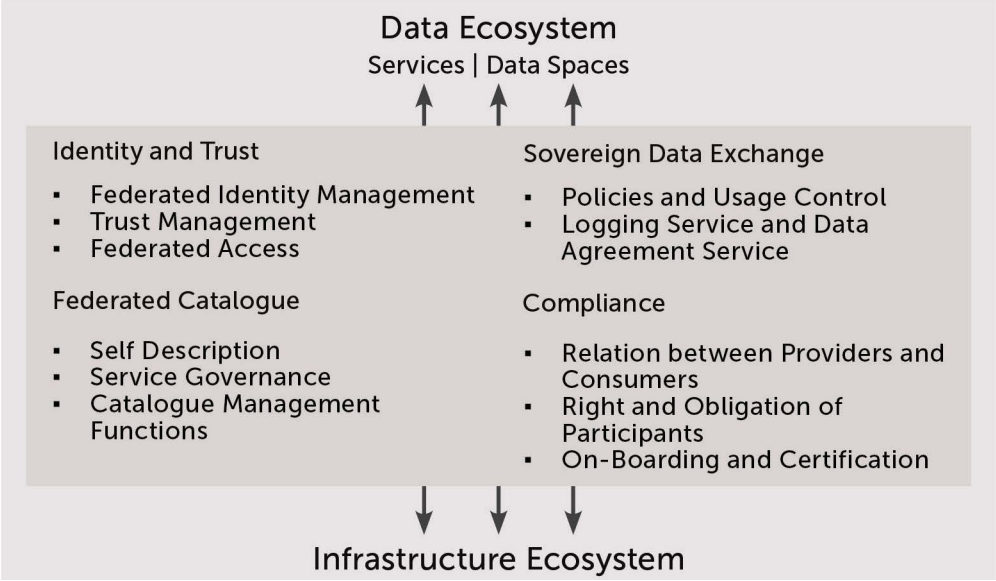


Figure 5: Gaia-X Ecosystem (CC BY)
Source: Own illustration based on Gaia-X AISBL (2022b) – Gaia-X-Architecture Document.

Gaia-X Federation Services (GXFS)

The core services for constructing the Gaia-X ecosystem are jointly funded by the German and French governments. Two projects are being funded by the BMWK including the Sovereign Cloud Stack (SCS) project, with overall EUR 14.9 million, and the Gaia-X Federation Services (GXFS-DE) project, receiving around EUR 13.5 million, with the latter being coordinated by eco – Association of the Internet Industry. Both projects aim to create an open-source toolbox for a sovereign, standardised, and federated cloud infrastructure, providing basic functions for organisations to join the Gaia-X ecosystem and implement their projects. The French government is investing EUR 15 million in a consortium to design part of the

92 https://gaia-x.eu/membership/members-directory/
93 Gaia-X AISBL (2022b) – Gaia-X Architecture Document.

core Gaia-X services. In November 2023, CISPE, part of the French project GXFS-FR (coordinated by the Institut Mines-Télécom) and member of the Gaia-X AISBL, presented the first Federated Cloud Services Catalogue[94] containing over 500 cloud services meeting Gaia-X requirements.[95]

GXFS provides technical basis for the planned European data ecosystem

The open source-based Gaia-X Federation Services are intended to provide the technical foundation for the envisaged European data ecosystem. They enable the creation of data spaces for different sectors and purposes. The Federation Services serve as a toolbox containing basic technical requirements, categorised into Identity and Trust Management, Federated Catalogue, Sovereign Data Exchange, and Compliance (see Figure 5, p. 30). Each area encompasses various core services:[96]

- Identity and Trust Management: Facilitates decentralised identity management, ensuring transparency for both customers and service providers.
- Federated Catalogue: Ensures uniform metadata and standards by requiring all organisations and providers in the Gaia-X ecosystem to present themselves through self-description.[97]
- Sovereign Data Exchange: Essential for the emergence of data ecosystems, enabling transparency and control over the use and application of data and data exchange.
- Compliance: Part of the Gaia-X Trust Framework, includes the Code of Conduct, the Policy Rules, and the agreement to the Terms and Conditions of the participants. Specific services, such as monitoring self-description or issuing digital and legally binding certificates, are intended.

Eclipse Foundation moderates the developer community

The federation services developed within the framework of the GXFS-DE were transferred to the Eclipse Foundation in mid-2023, which will moderate the developer community under the new and neutral name Eclipse XFSC, following the principles of the Eclipse Foundation.[98]

The Gaia-X Digital Clearing House (GXDCH) was created in March 2023 as a central point of contact for the automated verification of conformity with the Gaia-X rules. It is described as a necessary element to operationalise Gaia-X in the market.[99] This digital clearing house enables the Gaia-X Framework-compliant connection to the services of the developed compliance components. This will ensure a decentralised hub that is not operated by the Association and to which anyone can turn to obtain compliance. In this way, the Association aims to enable a Gaia-X market. The first two infrastructure nodes of the GXDCH are operated by T-Systems and Aruba.

94 https://catalogue.cispe-cde.cloud/
95 https://cispe.cloud/gaia-x-builds-momentum-with-first-federated-cloud-services-catalogue/
96 Gaia-X AISBL (2022b) – Gaia-X Architecture Document, pp. 41-52.
97 https://www.eco.de/news/httpswww-gxfs-eu/
98 https://projects.eclipse.org/projects/technology.xfsc
99 https://gaia-x.eu/gxdch/

2.3.6 Cooperation

Gaia-X closely collaborates with various national and European initiatives and projects, drawing on the activities and outcomes of different organisations. Examples include the International Data Spaces Association (IDSA) and the Big Data Value Association (BDVA), which, in 2021, changed its name to DAIRO – Data, AI and Robotics AISBL. While the IDSA primarily focuses on developing reference architectures for sovereign data exchange based on open standards, DAIRO aims to establish Europe's leadership role in researching, developing, and applying value-based and trustworthy technologies.

Project-based link between Gaia-X and NFDI

The connection between Gaia-X and NFDI was established under the BMBF-funded project *FAIR Data Spaces* (see section 2.1.6). There are also links and collaborations with the field of scientific high-performance computing (HPC). Since May 2021, the Gauss Centre for Supercomputing (GCS) has been a member of the Gaia-X Association, evaluating services for the use of HPC resources via Gaia-X as part of the InHPC-DE project.[100]

Loose links between Gaia-X and EOSC

Currently, there are no concrete or even institutionalised links between Gaia-X and EOSC. Despite the EOSC Association and the Gaia-X Association holding joint working meetings in 2021 to explore synergies and opportunities for cooperation, there has not been a regular exchange or tangible cooperation since. However, in July 2021, the EGI Foundation published a comprehensive paper on Governance, Architecture, and Business Models for Data and Cloud Federations, including case studies on Gaia-X and EOSC. This paper contains recommendations for future collaboration between the two initiatives.[101]

Numerous projects related to Gaia-X

In parallel with the establishment of the Gaia-X governance, several projects have been set up to develop Gaia-X-compliant infrastructures and services for specific industries and sectors. In addition to the automotive sector (Catena-X)[102], corresponding initiatives have emerged in the agricultural sector (AgriGaia)[103] and the area of infrastructure providers (Structura-X). In Structura X, for example, more than 20 European cloud and infrastructure providers, including the German internet node operator DE-CIX, have collaborated to build joint Gaia-X-compliant infrastructure services. Together with six other projects, Catena-X and Structura-X belong to the Lighthouse Projects, which act as pioneers for the implementation of the Gaia-X Framework[104] and follow the objective to build industry-specific data exchange platforms.[105] In addition to the official Lighthouse Projects of the Gaia-X association, the eleven consortia funded by the BMWK have created further projects that are testing practical applications within the Gaia-X structure.[106] For an overview of the Gaia-X Lighthouse Projects, see Table 3 in the appendix on pages 82 to 84.

100 https://www.hlrs.de/projects/detail/inhpc-de
101 Dietrich, Ferrari (2021) – EOSC and Gaia-X Case Studies
102 https://catena-x.net/de/
103 https://www.agri-gaia.de/
104 https://gaia-x.eu/gaia-x-framework/
105 https://gaia-x.eu/who-we-are/lighthouse-projects/
106 https://gaia-x-hub.de/gaia-x-foerdervorhaben/

With Manufacturing-X, the BMWK is promoting the development of another cross-sector digital ecosystem for data exchange in industry. The call for proposals for the associated funding programme ended on 31 December 2023 and includes up to EUR 150 million for projects that will advance data-based collaboration between industrial companies in industrial value networks by the end of 2026.[107]

Further projects with a similar focus expected

Under the German Presidency of the Council of the European Union in the second half of 2020, it was also decided to establish the Important Project of Common European Interest (IPCEI)[108] for Next Generation Cloud Infrastructure and Services (CIS). The objectives of IPCEI-CIS have been set out by twelve European states in a joint manifesto at the beginning of 2022.[109] This project also aims to strengthen digital sovereignty in Europe and to align data and cloud infrastructures with European rules and standards. In May 2023, the German BMWK published a funding guideline for cloud and edge infrastructure and services as part of the IPCEI-CIS.[110]

The IPCEI-CIS 'is to be compatible with Gaia-X services and standards, but [is] to go significantly beyond Gaia-X through the use of next-generation cloud and edge technologies and the funding opportunity for initial industrial use cases.' [111] The European Commission approved the IPCEI-CIS on 5 December 2023, enabling financial support with public funds from seven Member States of up to EUR 1.2 billion.[112] In addition, private investments of EUR 1.4 billion is expected. As part of this IPCEI, 19 companies, including Small and Medium-sized Enterprises (SMEs), will undertake 19 projects with highly innovative objectives. The IPCEI concerns the development of the first interoperable and openly accessible European data processing ecosystem, the multi-provider cloud-to-edge

107 https://www.bmwk.de/Redaktion/EN/Pressemitteilungen/2023/08/20230821-manufacturing-x-funding-call-for-the-construction-of-the-digital-ecosystem-launched.html

108 An Important Project of Common European Interest (IPCEI) allows for targeted state aid to facilitate large-scale, transnational projects that have the potential to make a substantial impact on the EU economy, environment, and society. IPCEIs are identified in sectors deemed critical for the EU's strategic interests. They aim to bolster the EU's technological leadership, enhance its industrial capacity, and address societal challenges while ensuring sustainability and adherence to European values. The European Commission's role involves assessing and approving these projects to ensure they comply with EU state aid rules, particularly that the benefits of such projects outweigh any distortion of competition that the aid may cause. See https://competition-policy.ec.europa.eu/state-aid/ipcei_en.

109 https://www.bmwk.de/Redaktion/DE/Downloads/M-O/manifesto-for-the-development-of-the-next-generation-cloud-infrastructure-services-capabilities-in-2022.pdf?__blob=publicationFile&v=1

110 BMWK (2023) – Förderrichtlinie im Rahmen des IPCEI-CIS, p. 1. See also the amendment of the funding guideline: https://www.bmwk.de/Redaktion/DE/Downloads/B/aenderung-forderrichtlinie-fcloud-edge-ipcei-cis.pdf?__blob=-publicationFile&v=2

111 BMWK (2023) – Förderrichtlinie im Rahmen des IPCEI-CIS, p. 1. [translated by the author].

112 https://competition-policy.ec.europa.eu/state-aid/ipcei/approved-ipceis/cloud_en

continuum. It will develop data processing capabilities, and software as well as data sharing tools that should enable federated, energy-efficient and trustworthy cloud and edge-distributed data processing technologies and related services. The participating companies have declared to develop an open-source software that will allow for real-time and low-latency (i.e. a few milliseconds) services by distributed computing resources close to the user, thereby reducing the need to transmit large volumes of data to centralised cloud servers. The research, development and first industrial deployment phases will run between 2023 and 2031, with timelines varying depending on the project and the companies involved. First novel result of the IPCEI – an open-source reference infrastructure – will probably be available around the end of 2027.

The Commission has also announced the development of Smart open-source Middleware Platform (SIMPL [113]), a new platform for cloud-to-edge federations to which the various sectoral data rooms will connect. SIMPL will ensure seamless inter-connection and interoperability of data sets and infrastructures. The consortium to build and operate SIMPL has been selected by the Commission in December 2023. It will run for three years and will be financed with EUR 41 million.[114]

2.3.7 Outlook

The 5-year outlook outlined 2021 by the Gaia-X Association is ambitious. The expectation is that Gaia-X should be disseminated in over 40 per cent of SMEs in Europe by 2024. It is anticipated that the dissemination of Gaia-X will continue to grow steadily beyond the initial phase (until the end of 2023) due to the increasing market demand. Starting from 2025, the European platform economy is expected to achieve a market share of more than 10 per cent.[115]

Harmonisation of funding by the EU and its Member States

In autumn 2023, the executive director of the Gaia-X Association articulated future objectives and challenges for the initiative. These include, among other things, greater harmonisation of funding by the EU and its Member States, as well as increasing market acceptance for Gaia-X solutions.

The number and selection of member organisations in the Gaia-X AISBL suggest that there is fundamental support for the initiative, although the media's assessment of the actual prospects of success is mixed in some cases. In particular, the involvement of non-European cloud providers such as Google, Microsoft, and Amazon, as well as the bureaucracy associated with the project, are core aspects of the criticism.[116]

113 https://digital-strategy.ec.europa.eu/en/news/simpl-cloud-edge-federations-and-data-spaces-made-simple-updated-august-2023

114 https://digital-strategy.ec.europa.eu/en/news/commission-awards-eu41-million-contract-develop-infrastructure-common-european-data-spaces

115 Bonfiglio (2021) – Vision and Strategy, p. 13f.

116 E.g. https://www.theregister.com/2024/01/08/gaiax_future/.

Long-term financing and political support still unclear

The question of whether and to what extent Gaia-X will achieve its own goals remains open and an answer depends much on individual perspectives and expectations (e. g. SME, Hyperscaler, scientific partner, kind of domain). Equally, it cannot yet be predicted which importance individual players will assign to Gaia-X in the near future and, in particular, how cooperation and interoperability with familiar cloud initiatives will develop. While some of the Lighthouse Projects may not expire until or even beyond 2025, the current German Gaia-X projects funded by the BMWK will actually expire at the end of 2024. It is still unclear whether and to what extent political ambitions sustain and resources in Germany and beyond will be allocated to fund further Gaia-X-related projects later than 2024. How the overall long-term funding schemes, objectives, and synergies between Gaia-X and related cloud initiatives – as the IPCEI-CIS in particular – may look like, is also uncertain yet.

2.4 EUROPEAN DATA SPACES

> 'The European Data Spaces will help unleash the enormous potential of data-driven innovation. They will allow data from across the EU to be made available and exchanged in a trustworthy and secure manner. EU Businesses, public administrations, and individuals will control the data they generate. At the same time, these data holders will benefit from a safe and reliable framework to share their data for innovation purposes.'[117]

2.4.1 History and Objectives

Data spaces in sectors and domains of public interest

NFDI, EOSC, Gaia-X and other initiatives for building a comprehensive federated data infrastructure in Europe – including research data – go alongside overarching strategies that aim at unleashing the enormous potential of data-driven innovation. Since the beginning of this decade, the European Commission has set in motion a series of strategies and regulatory proposals for data, technologies and infrastructures. The European strategy for data includes the evolution of common European Data Spaces in specific sectors and domains of public interest. As users within each distinct sector contribute to shaping these spaces, each data space should have its own unique characteristics. However, they should have in common to be built upon joint data infrastructures and governance frameworks that provide a trustworthy and reliable environment and facilitate data pooling, access, processing, sharing, use, or reuse.

Creating a genuine single market for data

The inception of the European Data Spaces initiative is linked to the European Data Strategy, set forth in 2020.[118] In this strategy, the Commission highlights the experience gained in connection with EOSC and the necessity of synergies between the work on European cloud federation and Member States' initiatives such as Gaia-X. The Data Strategy aims to strengthen the European Union as a relevant player in the global digital society, leveraging data as a pivotal asset for innovation, economic growth, and societal progress. At the heart of this strategy is the objective of creating a genuine single market for data where data can flow seamlessly across borders and sectors in a safe and trustworthy manner, in line with European rules and values.[119] As a pillar of the single market for data, the Commission argues for the establishment of sectoral and domain-specific common European Data Spaces that should be progressively interconnected.[120]

117 https://digital-strategy.ec.europa.eu/en/policies/data-spaces; for a recent update see EC (2024) – SWD on Common European Data Spaces.

118 EC (2020) – A European Strategy for Data

119 https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en

120 EC (2024) – SWD on Common European Data Spaces.

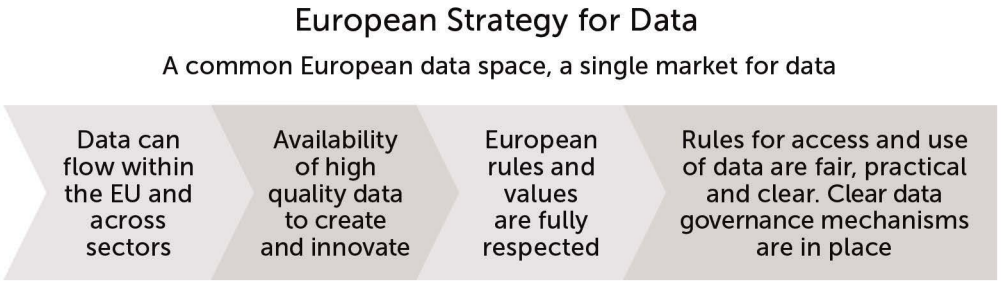


Figure 6: The European Strategy for Data – Key Objectives (CC BY)
Source: Siemen (2022) – Presentation on EOSC and related Horizon Europe Calls, p. 2.

A complex regulatory framework as foundation for the European Data Spaces

The regulatory framework enabling the development of common European Data Spaces includes the General Data Protection Regulation (GDPR)[121] as a first fundamental pillar intended to provide a 'framework for trust in the digital environment'[122] insofar as personal data is involved. Of crucial importance – not only for the European Data Spaces in general but also for the aforementioned and already established data initiatives NFDI, EOSC, Gaia-X, etc. – are the following further regulations:

- The Open Data Directive[123] aims to ensure that certain data sets and documents in the public sector are made available in open, machine-readable, accessible, findable and reusable formats, and may be re-used in the private sector. In particular, the reuse of 'high-value data-sets', which are defined under categories such as geospatial, earth observation and environment, meteorological, or mobility, is supposed to lead to significant benefits for society and the economy.[124]
- The Data Governance Act[125] (DGA) complements the Open Data Directive by covering specifically protected types of public sector data. It aims to promote safe and wide sharing of these data, while respecting privacy and confidentiality, and to facilitate reuse under certain conditions – first of all through the implementation of technological data protection and security concepts. The DGA initiates the establishment of novel data intermediaries and the sharing of data for altruistic purposes, for example in the field of medical research. Data intermediation services and recognised data altruism organisations are supposed to contribute to the bilateral or multilateral sharing of data or to create platforms or databases enabling

121 European Union (2016) – General Data Protection Regulation
122 EC (2018b) – Towards a Common European Data Space, p. 1
123 EC (2019) – Open Data and the Re-use of Public Sector Information
124 EC (2023) – Laying Down a List of Specific High-Value Datasets
125 European Union (2022) – Data Governance Act; see https://digital-strategy.ec.europa.eu/en/policies/data-governance-act. In its statement on the proposal for a Data Governance Act the RfII has emphasised the need to create incentives for the development of new data intermediaries and to make data access for research easier by means of a research clause, RfII (2021a) – Statement on the Proposal for a DGA. See also: RfII (2023b) – Stellungnahme zum Entwurf eines Gesetzes zur Durchführung der Verordnung über europäische Daten-Governance.

the sharing or joint use of data. The provisions of the DGA are also intended to promote the establishment of sector-specific data spaces, such as the already outlined European Health Data Space.[126]

- The Data Act[127] (DA) is focused on ensuring that personal and non-personal data generated in the context of the Internet of Things (IoT) is made available for use by various stakeholders. Among other things, this regulation clarifies the rights and obligations of the different parties involved in the use of data from connected products, facilitates data sharing and negotiating contracts, improves portability rights, and defines essential requirements with regard to the interoperability of data spaces. Its objective is to foster the exploitation of the knowledge and value-creation potential of data for the benefit of the European economy and society at large.[128]

Further regulations

The overarching European Data and Digital Strategy for Data includes further regulations adopted since May 2022, such as the Digital Services Act, the Digital Markets Act, the NIS2 Directive (on cybersecurity), or the Commission´s proposal for laying down harmonized rules on artificial intelligence.[129] All of these acts shape the legal framework not only for data spaces still under construction but also for the future strategies of the already established or initiated federated infrastructures for research (NFDI and EOSC) and economic data (Gaia-X).

Implications for the R&I ecosystem in Europe

While all these legislative acts have not been developed with research as the sole or even primary focus, they have important implications for the research and innovation ecosystem in Europe. For example, the DSA provides vetted researchers with the right to request data from very large online platforms and search engines to conduct research on systemic risks in the EU. The Open Data Directive encourages the principles of Open Science and data sharing among researchers and institutions. By providing a framework for data governance and clarifying rights on data access and use, the DGA and the DA facilitate the sharing of data, thus also increasing access to data, collaboration and transparency in scientific research. The DGA specifically introduces the concept of data altruism, encouraging organisations and individuals to share data for the common good, including scientific research, and provides the regulatory framework for sector-specific common European Data Spaces.

126 EC (2022c) – Proposal for a Regulation on the European Health Data Space.

127 European Union (2023) – Data Act; see https://digital-strategy.ec.europa.eu/en/policies/data-act. The RfII has also addressed the Data Act in its work: RfII (2021b) – Statement on the Proposed Data Act of the European Union, RfII (2022) – Statement on the Proposal for a Data Act by the European Commission.

128 In contrast, according to Kerber and Specht-Riemenschneider, it is questionable to what extent the Data Act will actually lead to innovations in the area of data sharing and improve the position of users; Kerber (2023) – Governance of IoT Data, Specht-Riemenschneider (2022) – Der Entwurf des Data Act (in German only). With regard to the discussed effects of the DGA on the development of data intermediaries, see, among others: Blankertz, Specht (2021) – What Regulation for Data Trusts Should Look Like.

129 EC (2021) – Proposal for a Regulation on Artificial Intelligence Act.

Common European Data Spaces are intended to promote data sovereignty and trust. They are envisaged to:[130]

Key aspects of the European Data Spaces are data sovereignty and trust

- deploy data-sharing tools and services for the pooling, processing and sharing of data by an open number of organisations, as well as federate energy-efficient and trustworthy cloud capacities and related services,
- include data governance structures, compatible with relevant EU legislation, which determine, in a transparent and fair way, the rights concerning access to and processing of the data,
- improve the availability, quality and interoperability of data – both in domain-specific settings and across sectors.

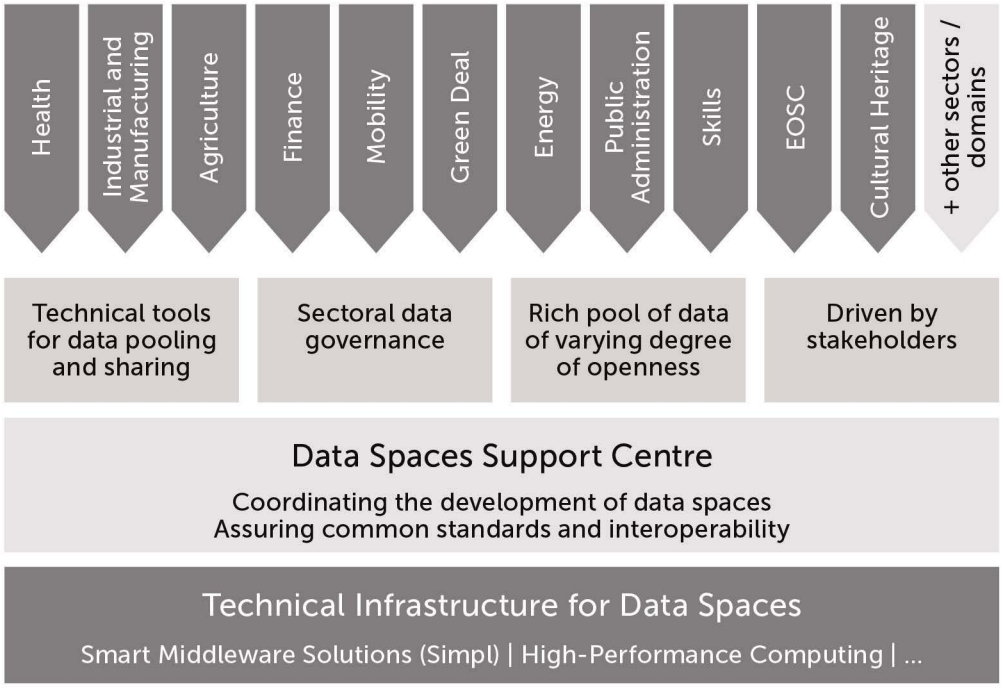


Figure 7: Overview of the European Data Spaces Sectoral Coverage (CC BY)
Source: Coduti (2023) – Presentation on European Data Spaces, p. 10.

Related initiatives and developments

The EU is currently funding several initiatives related to common European Data Spaces for sector and domain-specific initiatives, notably the Data Spaces Support Centre (DSSC) and SIMPL (see section 2.3.6) under the Digital Europe Programme and Horizon Europe. In addition, enabling institutions and instruments will be developed further. In particular, the European Data Innovation Board (EDIB) – established under the Data Governance Act as an advisory body – is supposed to provide guidelines. Recognised data altruism organisations, data intermediaries, and the European Register for Protected Data held by the Public Sector need to be established.

The Commission also works with the sectoral stakeholders and offers support for developing common European Data Spaces in the form of

130 https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces

reference architecture, building blocks, semantics, interoperability specifications and data models, accompanied by related advisory services, provided by the DSSC.[131]

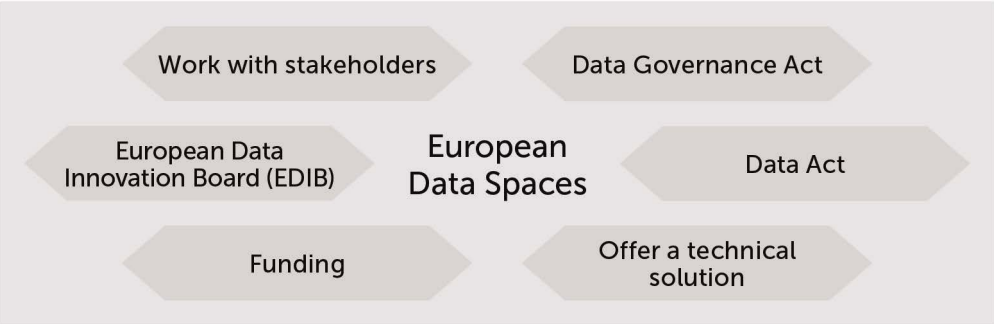


Figure 8: The Role of the EC in Supporting the Creation of Data Spaces (CC BY)
Source: Coduti (2023) – Presentation on European Data Spaces, p. 12.

2.4.2 Governance

Support structure of the European Data Spaces

Each data space will be tailored to the respective sector with sector-specific governance structures[132], including the active engagement of stakeholders through the use of existing or establishment of new networks, forums, and collaborations. Operational management is facilitated through initiatives like the DSSC[133] and specific actions under the Digital Europe Programme.[134] These should provide the necessary infrastructure, governance models, and standards to support the deployment and operation of data spaces. In addition, the responsibilities of the European Data Innovation Board (EDIB) include advising and assisting the European Commission in developing guidelines to foster the development of common European Data Spaces, identifying standards, ensuring that interoperability requirements for cross-sector data sharing are met, and facilitating the consistent application of rules for accessing and re-using public-sector data.

131 https://joinup.ec.europa.eu/collection/semic-support-centre/data-spaces
132 On the state of play see European Commission (2024) – SWD on Common European Data Spaces.
133 https://dssc.eu/
134 For example: https://digital-strategy.ec.europa.eu/en/news/digital-europe-programme-commission-opens-calls-invest-data-spaces-tourism-and-cultural-heritage.

2.4.3 Financing

Funding via different EU programmes

The financing of the European Data Spaces encompasses a strategic blend of EU funding programmes, focused on fostering digital transformation and innovation across various sectors:

- Digital Europe Programme: This programme is instrumental in financing coordination and support actions, deployment actions, and specific grants aimed at enhancing data sharing infrastructure, governance models, and interoperability across various data spaces. The Digital Europe Programme's work schedule for 2023 has allocated a budget of EUR 113 million specifically for the deployment of common data spaces and cloud-to-edge infrastructure and services.
- Horizon Europe: Focused on research and innovation, Horizon Europe funds projects that contribute to the technical and operational advancements of data spaces, including the development of open-source middleware and support for sector-specific initiatives.
- Connecting Europe Facility: Provides funding for digital infrastructure projects that enhance connectivity and interoperability across data spaces, supporting the EU's digital and data strategy.
- Recovery and Resilience Facility: Some Member States' recovery and resilience plans include actions related to common European Data Spaces, showcasing national commitment to the data economy.

Furthermore, a new instrument for multi-country projects, the European Digital Infrastructure Consortia (EDICs)[135], has been set up to enable Member States to collaborate on strategic digital infrastructure projects, including those related to data spaces. EDICs can access funding from centrally managed EU programmes.[136] With the Alliance for Language Technologies European Digital Infrastructure Consortium (ALT-EDIC) and CitiVERSE the first EDICs have already been established.

135 EDIC is an instrument made available to Member States under the Digital Decade Policy Programme 2030 to speed up and simplify the setup and implementation of multi-country projects. EDICs will enable the achievement of the Digital Decade general objectives and targets. Each EDIC is a legal person established by a Commission decision upon the application of at least three Member States and Commission approval. The founding Member States define the EDIC´s governance structure and other functioning rules in the Statutes. Its budget will be based on its members' contributions complemented by other sources of revenues, which may include EU and national grants. The seat of an EDIC is in a participating Member State and its legal personality must be recognised by all Member States. An EDIC may implement a multi-country project by deploying joint infrastructure, delivering services and bringing together – as considered appropriate by the founding Member States – public entities, private entities, final users and industry. See https://digital-strategy.ec.europa.eu/en/policies/edic.

136 https://digital-strategy.ec.europa.eu/en/policies/edic

2.4.4 Stakeholders, Participating Organisations and Cooperations

New instrument for multi-country projects

The development and implementation of European Data Spaces involves a wide array of stakeholders, participating organisations and cooperation partners, reflecting the initiative's comprehensive and multi-sectoral approach. Because potentially all kinds of data can be covered in the realm of the European Data Spaces a list of prospective stakeholders would be endless. For this reason, only a few stakeholders and participating organisations are mentioned here:

The European Commission (lead: DG CNECT) coordinates the overarching strategy, while various EU bodies, including specific directorates-general and agencies, provide sector-specific expertise and funding. The European Parliament and the Council of the EU are mainly involved in the legislative processes that provide the legal framework for data spaces and align them with other EU policies and regulations. The EU Member States contribute to the governance, implementation, and adaptation of data spaces within their national contexts as well as with their own funding, e.g. in the form of setting up EDICs (see 2.4.3). They ensure compliance with EU regulations and adapt their national policies to facilitate the development of data spaces. Local, regional, and national public authorities and public sector organisations contribute data and use data spaces for enhancing public services and policy-making. Also, sector-specific industries should participate actively, contributing data and leveraging data spaces for innovation and service enhancement. SMEs and startups are particularly requested to play a dynamic role in the new European data economy.[137] Universities and research organisations contribute to the development of data spaces through research, innovation, and the provision of scientific data. They should also benefit from access to data sets from different economic and societal spaces for research purposes. Civil society organisations, including consumer protection groups, and advocacy organisations, are envisaged to play a role in shaping the development of data spaces by ensuring that data spaces cater to societal requirements and uphold ethical standards, contributing to discussions on privacy, data governance, and societal impact. Further stakeholders respectively cooperating partners will come from a wide array of technical providers, standardisation committees and consortia (e.g. the European Interoperability Network – EIF) and – for sure – from the manifold consortia and networks of the NFDI, EOSC, and Gaia-X world.

137 In this context the Data Spaces Business Alliance is seen as an important actor. See https://data-spaces-business-alliance.eu.

2.4.5 Technical Structure and Planned Services

Accessibility for use across sectors and borders

The European Data Spaces are being designed on foundations which would facilitate secure, efficient, and seamless data sharing across sectors within the EU. These foundations consist of:

- Federated Architecture: European Data Spaces are built on a federated architecture model, allowing data to remain with its original owners while being accessible for use across sectors and borders. This approach respects data sovereignty and privacy regulations.
- Interoperability Frameworks: A key component of the technical structure is the establishment of interoperability frameworks that ensure data from different sectors and sources can be easily integrated and used together. This includes semantic interoperability for understanding data across different contexts and syntactic interoperability for the technical aspects of data exchange.
- Data Governance Mechanisms: Governance mechanisms have to be embedded within the technical infrastructure to ensure data sharing is compliant with EU regulations, including GDPR, and aligns with ethical standards. This includes access controls, data usage agreements, and mechanisms for data tracking and auditing.
- Secure Data Sharing Infrastructure: Security protocols and infrastructure must include encryption, anonymisation, and other cybersecurity measures to protect data during exchange and storage.
- Cloud and Edge Computing: The infrastructure must leverage cloud and edge computing technologies to provide scalable, efficient, and flexible data processing and storage solutions. This should support the real-time processing requirements of various applications, from AI to IoT.

Depending on the sector, specific applications and services have to be developed to meet the unique requirements of each data space, such as health data analysis tools in the Health Data Space or real-time traffic management systems in the Mobility Data Space.

2.4.6 Outlook

The roll out of the European Data Spaces is still at an early stage – however, with large parts of the legal framework already adopted by the EU and in implementation processes of the Member States. One of the most advanced developments is connected with the European Health Data Space. Here, the corresponding legal foundation is currently in the trialogue process of EU institutions. Consequently, European Data Spaces are expected to expand in scope and number, covering many sectors and integrating existing and new data spaces into a cohesive ecosystem for digital data.

Challenges of implementation

While the European Data Spaces hold great promise, their implementation poses several challenges, including ensuring interoperability among different data platforms, addressing legal and regulatory complexities, and overcoming technical and infrastructural barriers. Furthermore, it is essential to build trust among data providers and users, safeguarding data privacy and security while fostering a culture of data sharing and collaboration. The success of the European Data Spaces depends on collaborative efforts among EU institutions, Member States, industry stakeholders, and the research community to overcome obstacles, realise the full potential of Europe's data economy and increase databased innovations driven by scientific research.

European Data Spaces are expected to significantly drive innovation in various sectors

The development of the European Data Spaces will continue to be influenced by technological advancements, particularly in areas like AI, blockchain, cloud and edge computing as well as progress in the development of better cybersecurity. The feasibility of European Data Spaces will also depend on the availability of digital skills and capabilities among the workforce and data users. Efforts to enhance digital literacy, technical skills, and data science expertise will be critical to maximising the benefits of data spaces for the economy and society.

Concerns about the complexity of the regulatory framework

However, there are also concerns about the complexity of the regulatory framework constituting the European Data Spaces. The so-called Regulation Tsunami sparked by EU Institutions has created a lot of relatively open legal terms – still to be interpreted in application and by future court decisions. Quite a few observers fear that this will initially further increase legal uncertainty in data access and data exchange. Also, some stakeholders express caution regarding the technical challenges of ensuring interoperability across different data spaces and sectors, which could impede the seamless sharing and utilisation of data.

3 ANALYSIS AND COMPARISON

Key objectives

While NFDI and EOSC specifically aim to improve the access to research data and provide services to the scientific community, Gaia-X's primary goal is to introduce structures and standards for sharing economic and industrial data, allowing the development of common data spaces. Scientific users and research organisations can also participate in this process (which is already happening in practice) but are not the primary target group. To establish sector-specific data spaces under a common European legislation is the objective of the European Data Spaces. As a result, data from the entire EU should be made more available and exchangeable in a trustworthy and secure manner – making these data also better accessible for research purposes. For a brief comparative overview of the infrastructures and initiatives presented in Chapter 2, see also Table 4, pages 85 to 87.

At first glance, the considered infrastructures differ mainly in their primary target groups and underlying main objectives. The key aspect of NFDI is the development of a national solution for the challenges of RDM and the high demand for systematic cataloguing, networking and sustainable usability of research data. With EOSC as a European initiative, a complex multi-level structure is being implemented for accessing and sharing research data in Europe in particular (but not exclusively) by linking data, datasets and data sources combined with the provision of various services (Core Federation Services, Exchange Infrastructure etc.) for RDM.[138] In the future, however, EOSC also plans to make its functionalities available to the wider public and private sectors, thus enabling the exchange of data from different sources (scientific, public / administrative, and economic data). Gaia-X, as a transnational initiative supported by industry participation, aims to strengthen European sovereignty in the cloud sector and improve the competitiveness of European companies through a commitment to greater interoperability and standardisation. Last but not least the European Data Spaces are envisaged as sector-specific frameworks that facilitate the safe and fair sharing and use of data among a large number of stakeholders including businesses, researchers, and public administrations.

3.1 GOVERNANCE STRUCTURES

Different governance approaches

There are significant differences among the data infrastructures in terms of their organisational logic, specifically regarding the question whether they follow a user, political or market-driven approach – and in some cases a mixture of it all. NFDI comes with several elements of a user-driven initiative, constructed by scientific communities and domains based on recommendations from a research policy advisory body (the RfII) with the support of research policy actors, such as the GWK (task: legal framework for governance and funding)

138 The structure is complex, especially with regard to its governance and financing, which is carried out at supranational (EU Commission), national (EU Member States) and subsidiary (member organisations of the Association) levels.

and the DFG (task: science-led selection process for consortia). Services are (mostly) built inhouse via the NFDI-entity Base4NFDI and for NFDI's own purposes and demands. In contrast, the EOSC is a politically driven and administratively steered project, initiated by the European Commission, with continuous guidance by the Commission through a European Partnership involving Member States (represented in the Steering Board). This process takes place with the participation of expert committees and a comparatively small but steadily growing number of highly committed persons from the scientific system. Recently, commercial tenders of the Commission for the provision of the EOSC service core elements played a significant role in the infrastructure's further development. Some scientific players complained that – as public institutions – they could not meet the market economy-related requirements of the tenders and for that reason could not participate. However, in contrast to NFDI, the EOSC-Services will be built up and provided more via a top-down-approach and probably with less participation by the scientific communities and domains. Gaia-X started as a governmental initiative by the ministries of economics in Germany and France and is currently evolving quite independently from direct national government influence. It is being primarily constructed by consortia of economic actors, with varying degrees of participation by scientific organisations, depending on the consortium. Here, while in the beginning being closer to the more politically driven EOSC approach, a strongly user-driven commitment and self-governance is now in place – with more similarities compared to the NFDI governance. However, probably due to Gaia-X's strictly decentralised governance, a concept is still missing, how far the actually developed use cases of the different Gaia-X consortia could be transferred into regular and continuously provided services.

How far these three initiatives fit to and interact with and within the European Data Spaces framework remains to be seen in the near future. The latter is still in its very beginnings with just parts of an ambitious legal framework and the proposal for a European Health Data Space already laid out.

Integration of scientific requirements and founding purpose

While temporary or interim governance structures still prevailed in NFDI, EOSC and Gaia-X in the early development phase, legally secure structures in the form of non-profit associations and registered societies have been established in the meantime. Particularly in the case of NFDI, and in parts also in the case of EOSC, governance is strongly oriented towards a needs-based structure that is primarily intended to meet the differentiated requirements of researchers and scientific institutions. NFDI achieves this through the Consortia Assembly and the Scientific Senate. In case of EOSC, active researchers are primarily involved in the Task Forces. However, the latter do not have a specific position in the steering structure, but rather an advisory character. With the establishment of national hubs, Gaia-X takes its transnational character into account and thus enables central and country-specific contact points, which, however, with a view to the ultimately industrial policy orientation of the targeted solutions, do not per se aim at a science-led governance and towards structures that correspond with specifically scientific

demands. After all, we see varying degrees of scientific orientations in the construction and purposes of the infrastructures:

- NFDI: Serving scientific requirements is the founding purpose,
- EOSC: Serving scientific requirements is the main purpose, serving economic requirements is also an indirect purpose in the final stage of development,
- Gaia-X: serving the economic requirements is the founding purpose,
- European Data Spaces: serving scientific requirements is not a priority but one out of many potential benefits of a single market for data in Europe.

Federated architecture for data infrastructures

NFDI and EOSC rely on the federation or networking of existing infrastructures – which does not rule out the possibility that new infrastructure components and services will also be built up, such as in the context of the basic services for NFDI and the planned core components for storage, processing, etc. for EOSC. European Data Spaces are also built on a federated architecture model, allowing data to remain with its original owners while being accessible for use across sectors and borders. Gaia-X decisively sets up new structures (by means of rules and regulations to be developed for this purpose) among the participating cooperation partners.

The extent to which the different approaches or the selected architecture for the development of data infrastructures will favour a demand-oriented development is difficult to assess at the present time. However, the question of complementarities between the different structures and the acceptance within the communities play a decisive role for successful use in the scientific system.

3.2 FUNDING AND QUALITY ASSURANCE

Long-term financing has not yet been clarified

While NFDI is funded by joint mid- to long-term funding from German Federal Government and the Länder, the funding of Gaia-X and EOSC is based primarily on short- to mid-term public project funding provided by the European governments directly or via the European Commission, and through membership fees from the participating organisations. Similarly, the resources for establishing the data spaces come from a blend of European funding programmes. At the moment, none of the infrastructures offers any concrete information on sustained long-term funding perspectives beyond the implementation phases that are in progress. In the case of NFDI, however, there is a robust declaration of intent by the German Federal Government and the Länder to 'continue funding' beyond the already agreed development and expansion period.[139] Claims for early clar-

139 The Federal Government-Länder agreement states: 'Taking account of the results of the structural evaluation by the German Council of Science and Humanities in accordance with Section 13, the GWK shall decide in 2026 on the further organisation of NFDI and the details of further funding from 2029.', GWK (2018) – Agreement between the Federal Government and the Länder, Section 14 (2).

ification of the continuation of NFDI can also be heard from the ranks of research policymakers and the DFG. The question of the governance structure and financing of EOSC from 2028 onwards is also subject of intensive discussions in the various bodies and among associated stakeholders.

Involvement of research organisations

In all initiatives and also in the European Data Spaces, the participation of organisations and companies in the respective governance structure is a central point in the formation and implementation process. This is particularly true for NFDI, which is intended to represent the wide range of the German scientific system in its structure and is explicitly designed to be community- and user-driven. The involvement of research organisations and universities can to some extent be hindered by high membership fees payable to the respective entity. Membership in the NFDI e. V. is a requirement for the members of the NFDI consortia, but is also open to other stakeholders with an interest in research data. Membership in NFDI does not require a fee. EOSC and Gaia-X, in contrast, ask for financial contributions for their central membership structures. Considering the size of the institution or financial resources and in light of increasing cost pressures, it is conceivable that research organisations, in particular, will carefully weigh the decision to become a member and, especially, in how many of the initiatives.

Evaluation of the data infrastructures

It is not equally clear for all initiatives whether and in what form a (science-led) evaluation will be carried out. In order to review the effectiveness of NFDI, the WR will conduct a structural evaluation by the end of 2025, and the consortia will also be evaluated by the DFG 'at regular intervals'.[140] The timetables for these reviews are already known and corresponding evaluation structures and responsibilities have been defined. The RfII welcomes the transparency of this process. In its view, NFDI-related evaluation procedures will greatly benefit from evaluation criteria that are as unified and overarching as possible. Where this seems scientifically appropriate, the subject-specific and domain-related characteristics of the structures and offerings of the individual consortia should of course also be considered. At the European level, a comprehensive evaluation of all European Partnerships[141] is planned within the European Research and Innovation Programme. This therefore also applies to EOSC. The European partnerships are an important instrument for the implementation of the Horizon Europe Programme and the ERA. This implementation is continuously accompanied, for example, by the Biennial Monitoring Report (BMR).[142] However, it is not yet apparent how EOSC will be properly and professionally evaluated in the long term and in detail. Gaia-X presents its work publicly at summits. A scientific respectively science-led evaluation does not seem to be planned (due to its different target groups). Here, the extent of dissemination

140 See ibid. Section 13 (1 to 3).

141 More detailed information on the partnerships: Leibniz-Gemeinschaft (2021) – Europäische Partnerschaften in Horizon Europe

142 https://ec.europa.eu/assets/rtd/bmr/2022/

of standards and shared applications among private companies seems to be the key factor to measure functionality and success.

3.3 EMERGING NETWORKS AND INTERFACES

Institutionalised links between NFDI and EOSC

There are various networks, cooperation and commonalities between the initiatives – at the personnel, technical and policy levels. The NFDI e. V., for example, acts as the mandated organisation for Germany within the EOSC AISBL, with the aim of representing the interests of the German scientific landscape. Through the membership of the NFDI e. V., there is a direct link to EOSC. Connections between NFDI and Gaia-X exist in particular through the *FAIR Data Spaces* project[143] and the NFDI Section on Industry Engagement. However, other scientific actors are involved in Gaia-X within the framework of projects funded by the BMWK and as members in the Gaia-X AISBL – for example, the German Aerospace Centre, the German Research Centre for Artificial Intelligence, the Charité Berlin as well as various universities and non-university research institutions.[144] Furthermore, Gaia-X and, to a lesser extent, EOSC, are also referenced in the context of the European Data Spaces.

Application of the FAIR principles

With regard to the technical interfaces and prerequisites as well as the orientation towards common standards and schemes, there are initial similarities and overlaps, especially between NFDI and EOSC. In particular, both initiatives want to implement a consistent reference to and application of the FAIR principles. The Gaia-X architecture, on the other hand, does not explicitly refer to the FAIR principles but is intended to be in line with them.[145] Since Gaia-X primarily brings together commercially oriented actors, issues of trade secrecy and freedom of contract as well as aspects of antitrust law play a more prominent role in the context of data sharing and use.[146]

3.4 DEVELOPMENT AND PROVISION OF SERVICES

Comparable basic services are projected

Even though all initiatives are fundamentally based on existing infrastructures, services and organisations and thus follow the principle of federated data infrastructures, new services are also being set up or are projected within all initiatives. In the case of NFDI, generic or basic services will be collaboratively developed within Base4NFDI, the network of consortia that started by developing services in the fields of Identity & Access Management, Terminology Services and Persistent Identifiers. Within EOSC, basic functions are also being introduced as part of the core infrastructure and Gaia-X also requires basic services for the exchange of data. In terms of the approach to setting up these services, however, the initiatives differ fundamentally in some aspects.

143 https://www.nfdi.de/fair-data-spaces/
144 https://www.bmwk.de/Redaktion/DE/Dossier/gaia-x.html
145 See Dietrich, Ferrari (2021) – EOSC and Gaia-X Case Studies, p. 28.
146 See FRAND Principles: Fair, Reasonable and Non-Discriminatory.

While the majority of the NFDI services are offered and operated directly by scientific organisations, the European Commission has opened a procurement procedure for the EOSC core services in which commercial players can also participate. The extent to which later services will actually be offered by scientific actors and thus from the scientific community has not yet been conclusively answered (see also 3.1). With Gaia-X, the commercial provider side is system-immanent. However, the extent to which the large cloud service providers are also involved in the further development of basic systems is of particular interest here.

3.5 ACCEPTANCE IN THE SCIENTIFIC COMMUNITIES

Requirements for permanent and trustworthy availability of data infrastructures

The development and linking of different data infrastructures offers various advantages and opportunities for scientific research – including improved availability of research data from different sectors or domains (registry data, economic data, etc.), in different formats and, especially, the chance to reduce time and costs in the research process. In addition, however, there are also several challenges for this very development and the continued availability of infrastructures. For long-term availability to succeed, there must be a willingness within the scientific communities to actually use the data infrastructures that are being developed. For permanent and trusted use criteria such as the long-term suitability for different fields of research and possibly changing requirements of specific scientific communities, constant funding of the infrastructures, the guarantee of security and data protection as well as the preservation of research freedom are essential.

Public awareness of the different initiatives

As far as the awareness of federated data infrastructures in the broad academic community is concerned, NFDI has gained significant visibility in Germany since the agreement between the Federal Government and the Länder of 2018. The consortia already underway are also well-perceived within the specialist communities. With regard to the acceptance and expected benefits of NFDI, greater discipline-specific differences can be expected, which should not be surprising depending on the community and the associated affinity for the use of digital data infrastructures. On the other hand, the activities of EOSC and also of Gaia-X are still perceived by a rather small circle of specialists within the German scientific landscape. In the case of EOSC, this is naturally the case for researchers and organisations with strong European ties, as well as science policy actors and parts of research management (especially in the area of research funding and third-party funding management in the fields specialised in EU funding programmes). Participation in Gaia-X seems to be on the agenda of research organisations with already established cooperation with industry; these are in particular Fraunhofer institutes and other application-oriented scientific actors. With regard to the data spaces, sector-specific differences seem to exist. For example, the Health Data Space has garnered considerable attention in the health community, while this does not seem to be equally the case for other sector-specific

data spaces. However, it is difficult to make general assumptions here because empirical data on the state of collaborations is missing. The rudimentary mutual relationships that already exist through joint projects and formats (e.g. the national tripartite events of the EOSC Association [147]) increase awareness of the common challenges and the need for cooperation between the different levels. The RfII sees NFDI with its Directorate as an important actor that can provide the organisational interface between the different initiatives in the future and has already started doing so – for example in its role as a mandated organisation within the EOSC Association or through the *FAIR Data Spaces* project in cooperation with Gaia-X. However, as the *FAIR Data Spaces* project is financed on a temporary basis, the cooperation will prospectively come to an end in 2024. An institutionalised and long-term cooperation, possibly also integrated into governance structures, would be desirable.

147 https://eosc.eu/events/national-tripartite-event-germany

4 RECOMMENDATIONS

Recommendations for different stakeholders

In the following, the RfII formulates recommendations for the further design of the scientific (e.g. NFDI, EOSC) and science-related (e. g. Gaia-X, European Data Spaces) data infrastructures for five different groups of addressees:

- individual researchers and their academic communities (also professional associations),
- providers of data infrastructures,
- research performing organisations (HEIs and non-university research institutions), especially their management at various levels,
- research funders (including federal ministries or EU-directorates as grant-giving institutions),
- research policymakers at organisational, state (Länder), federal and European level.

As any institution within the scientific system might carry multiple roles at the same time – e.g. the role of a policymaker as well as the role of a funding and/or grant-giving institution, or the role of a research-performing organisation and of an infrastructure provider – the following recommendations may address the same actors in multiple ways.

Researchers and research support staff

Researchers and research support staff at universities and non-university research institutions are the primary target group regarding the benefits of a functioning system of federated (research) data infrastructures. They and their concrete research requirements – on both the supply and demand side – play an essential role in shaping the successful long-term operation of data infrastructures being appropriate for science and humanities. They are not only the most important source of information for service providers, but also play a decisive role in shaping the data infrastructures through their active use and feedback as well as with regard to design and implementation of new services and provision of new data products. Here, also professional associations with their respective guidelines for scientific practice and their rewarding of community services (should) have an important role. Within scientific communities, professional associations could provide orientation by enforcing the further development of professional cultures in the digital transformation. An understanding of technical possibilities (and limitations) of working with data infrastructures is also necessary for – yet – less data-intensive subjects and fields – especially adjustments in the design of subject curricula and methodological requirements need to be based on developed skills in research data analytics and data management.

Providers of data infrastructures

The second addressees are the institutional providers of data infrastructures, who must ensure the effective participation of researchers in order to be able to further develop their offerings in line with the demand. A participatory process of balancing interests becomes even more important when market and economic interests in operating models come into play, as envisaged for the future operation of EOSC, Gaia-X, and the European Data Spaces.

Research performing organisations

The third important addressee are the research performing organisations (HEIs and non-university research institutions) and their management (presidencies or rectorates, faculties and dean's offices). They must take up the new challenges by advancing study curricula and structured doctoral programmes as well as by enforcing recognition of scientific contributions to the further development of research data infrastructures in appointment procedures. The development and use of federated data infrastructures will have direct impact on many issues of research, teaching and transfer – presumably also in areas where research at universities is not yet data-intensive. Today, HEIs and non-university research institutions need to develop strategies on how they can incorporate the use of data infrastructure offerings into all of their performance dimensions or take them into account appropriately in research, teaching and transfer.

Research funders

The fourth addressee for recommendations is the group of research funders, who need to rethink the incentives they set with their funding programmes. Researchers' participation in establishing both, large-scale federated data infrastructures and good RDM practices in funded projects are personnel- and time-intensive. These investments require recognition, which must also be reflected in review and evaluation procedures. Research funders should therefore promote participation in consolidating, developing, expanding, and innovating research data infrastructures as a relevant component of scientific practice. Research funders, and in particular ministries at the Länder and the federal level should strive for coordinated approaches, building on and connecting existing research data infrastructure initiatives where ever possible. Consequently, all funding initiatives in the Länder and on federal level should be assessed to see whether they make sufficient reference to NFDI, EOSC and if appropriate to other relevant infrastructures as Gaia-X, IPCEI-CIS or emerging structures in the upcoming European Data Spaces.

Research policymakers at different levels

The fifth addressees are the research policymakers at the organisational, state (Länder), federal, and European level but also those inside the research performing and funding institutions. The task of research policy is to promote the initiated projects and infrastructures through constant support, innovative impulses and to legally and financially secure their long-term provision for scientific use. The focus here is on forward-looking linking and interoperability of the various data infrastructures – e.g. through suitable networking- and interface projects. Indispensable for a sustainable boost to innovation through federated data infrastructures in and outside of the scientific system is the permanent and reliable funding. Only public funding can create incentives for the active participation of as many researchers as possible in the development process. Every successful entry into an infrastructure solution is based on trust in the durability of this infrastructure and in the long-term return of investment (reputation, recognition) for commitment. This goes hand in hand with high demands on the design and mutual referencing of the evaluations of federated data infrastructures. In the context of evaluation, they should not be handled as insular entities, but rather as different interrelated elements of a larger data

ecosystem for research, innovation and overall prosperity in Germany and Europe. It is also necessary that the leaders of the research performing organisations – HEIs and the various non-university research organisations in Germany – overcome institution-centred and site-related paths and act together for the development of federated research data infrastructures as scientific commons.

Already in the context of its publication **Datenpolitik, Open Science und Dateninfrastrukturen: Aktuelle Entwicklungen im europäischen Raum** (Data Policy, Open Science and Data Infrastructures: Current Developments in Europe)[148], the RfII has dealt intensively with the similarities and differences between NFDI, EOSC and Gaia-X and made a number of suggestions for the further development and convergence requirements of the infrastructures in 2022. The following recommendations build up on these initial suggestions, summarising them in part, but also set new priorities with a focus on a demand- and resource-oriented construction, taking also into account ongoing developments in the European Data Spaces.

4.1 FURTHER DEVELOPMENT OF SCIENTIFIC PRACTICE IN DATA-INTENSIVE RESEARCH

Challenges of data integration

New research methods increasingly related to the use of digital data are changing the processes and practices within science. They create new challenges for the prerequisites and framework conditions for the generation and dissemination of scientific knowledge. In particular, challenges of data integration (collection, linking and long-term preservation) confront researchers with a number of issues that have so far only been partially solved. Federated data infrastructures that respond flexibly to scientific requirements can take up these challenges and become part of the solution. This requires a critical mass of experts from the entire range of disciplines and research fields who are actively involved in the development, long-term use and constant monitoring to closely link further developments in the field with technical innovations in the infrastructure and vice versa. The RfII aims to enable researchers at all career levels to actively communicate their requirements and experiences from everyday research both within their subject communities, in interdisciplinary contexts and to infrastructure providers respectively the support staff in the service-providing entities. The active communication of scientific demands – including the ongoing challenges of adaptation that become necessary in the context of progression in theories, methods and research tools – to the various data infrastructures is critical for their successful development and operation. Science-led demand and demand-oriented supply drive each other in a reciprocal process. This process is not linear; it does not reach an end point even after the institutional setup phase of data infrastructures. Instead, advisory support from research-active personnel for scientific data infrastructures has to

Communication of scientific requirements is necessary

148 RfII (2022) – Datenpolitik, Open Science und Dateninfrastrukturen

become a permanent scientific task. The RfII therefore suggests that both the use and the advisory support of scientific data infrastructures should be understood as a further dimension of scientific practice in relation to the respective professional standards. Broad acceptance within science for the use of infrastructures in the research process increases the connectivity, transparency and reproducibility or replicability of research. The sustainable handling of research data and the mutual further development and improvement of infrastructures and research methods must be promoted as early as possible in the education of students, especially in the stage of further academic qualification (Master and Doctoral studies). Working in this direction is also a task for the professional associations, which could reward participation in the development and expansion of data infrastructures and the teaching of related research data infrastructure literacy by activating their members across the board and visibly rewarding the commitment of individuals (e.g. through prizes). The digital transformation in science and humanities should be accompanied by scientific communities in a critical and reflective manner. It is important to analyse the extent to which research with the new data infrastructures will support scientific progress in the long-term. It is also crucial to avoid irreversible dependencies – especially on commercial offers – in the creation of tools and services for data-intensive science, and to uphold diversity in selecting research methodologies.

Use of data infrastructures is part of the scientific practice

Ensuring the independence of scientific infrastructures

4.1: RFII-RECOMMENDATIONS ON FURTHER DEVELOPMENT OF SCIENTIFIC PRACTICE IN DATA-INTENSIVE RESEARCH

Addressee: Researchers and their academic communities (including professional associations)

- **Prioritise data infrastructure utilisation and feedback for scientific advancement.**
  The use of data infrastructures – e.g. NFDI and EOSC and, in particular for applied, industry-related research, Gaia-X – and the provision of constructive feedback for their ongoing improvement must be an important aspect of scientific practice.
- **Excellent data-intensive research must go hand in hand with the advancement of research data infrastructures.**
  Researchers engaged in data-intensive research should also be committed to the design and validation of scientific data infrastructures and to the creation of next generation data products for use in the scientific system and beyond.
- **Reward and recognise researchers' contributions to data infrastructure enhancements.**
  It is imperative for professional associations and learned societies to recognise and encourage the active involvement of individual researchers and scientific institutions in enhancing data infrastructure services. This should also be reflected in recruitment and research performance evaluation processes.

4.2 ESTABLISHMENT OF RESEARCH-RELATED OPERATING MODELS FOR DATA INFRASTRUCTURES

Incorporating researchers' expertise into the development process

For the successful and sustainable development and expansion of data infrastructures in the full range of scientific disciplines and research fields, it is necessary for researchers to be willing to actively participate in their development and expansion. Rather, infrastructure providers (respectively those responsible for the build-up process in infrastructure institutions and data spaces) should be committed to make offerings for enabling researchers to participate in the process of designing and implementing structures and services. In addition to considering the recommendations of scientific expert committees, the expertise of researchers should be permanently incorporated into data infrastructures with the help of low-barrier forms of participation and suitable feedback structures. In a time of upheavals and crises that are having a considerable impact on society and the scientific system, the aforementioned reciprocal process of science-led demand and demand-oriented supply can also be understood as an opportunity to develop long-term and structured strategies. These strategies should drive a permanent mutual adjustment between the scientific and technical dynamics of the demand side and the information technology dynamics of the supply side – e.g. in the NFDI consortia. From this continuous and incremental feedback of demand and supply, the RfII expects an additional innovation impulse for the scientific system.

Connectivity of the various data spaces and infrastructures

Infrastructure providers face fundamental challenges in building federated and demand-oriented data infrastructures. In order to facilitate the creation of suitable offerings, a continuous and structured dialogue with infrastructure experts as well as researchers of universities and non-university research institutions is fundamental. Reliable and long-term access to basic services is also critical to the success of use and acceptance. For this reason, providers of data infrastructures and of services which are critical for the functioning of science and humanities should have a background as a reliable and publicly financed scientific or at least research related institution. The RfII suggests to strengthen the measures for appropriately linking the various data infrastructures on the operating level and to establish technical interoperability of the individual infrastructure systems. In this context the connectivity and interoperability of research data infrastructures such as NFDI and EOSC to the other sector-specific European Data Spaces is of great importance. Gaia-X should take action to play a prominent role in shaping the interfaces between the different data spaces or share at least Best Practices (derived from their Use Cases) on how to manage them. The interlocking of infrastructures should be enabled through:

- exchange and cooperation between NFDI and EOSC as well as with Gaia-X and the European Data Spaces through financed projects with the aim of mutual connectivity (e.g. *FAIR Data Spaces* project),
- cross-infrastructure governance for certain tasks with the possibility of

joint action Task Forces to ensure interoperability where this is feasible,
- technical interoperability, including agreed hardware protocols, shared vocabulary, and data models to enable shared data spaces.

This also means that scientific data infrastructures should work towards connectivity of data from science and humanities to other data spaces in business and society.

Operating models for research data infrastructures are needed

Until now the establishment of data infrastructures does not provide information on whether the existing services are actually utilized, and thus deliver added value to researchers. Therefore, success should be measured at least through appropriate metrics such as usage rates and intensity of use of the infrastructures and services, to make informed assessments regarding the relevance of the infrastructures. This is also important for establishing sustainable services in the long term.

In order to provide better evidence for institutional and project-based funding decisions, infrastructure providers should also develop comprehensible and sustainable operating models for research data infrastructures, which are financially viable and are being continuously updated to address evolving user needs and technological trends. These need to be based on the technical and the scientific requirements and need to consider best fit into the overall framework of an existing ecosystem of federated data infrastructures and services.

4.2: RFII-RECOMMENDATIONS ON THE ESTABLISHMENT OF RESEARCH-RELATED OPERATING MODELS FOR DATA INFRASTRUCTURES

Addressee: Providers of data infrastructures

- **Empower users for optimal performance of research data infrastructures by integrating user feedback.**
  Integrating scientific experts' requirements and enabling feedback in appropriate participation formats should be the core of a good operating model for research data infrastructures.
- **Customise computing and storage services to meet researchers' specific requirements for seamless engagement.**
  The provided computing and storage infrastructures, data management services, and applications must be aligned with the specific requirements of the scientific community and must offer user-friendly access options to researchers.
- **Foster compatibility among data infrastructures to enhance collaboration and efficiency.**
  Interconnection of all relevant data infrastructures – specifically NFDI and EOSC – on both, the operating and technology levels, should be implemented to enable shared access and integrated use by researchers across disciplines and domains in a multitude of research contexts.

- **Evaluate infrastructure success through appropriate usage metrics.** Providers should measure the usage, the intensity of usage and should – as far as possible – offer a basis for research-communities and science policy experts to reflect the impact on research productivity and quality to assess the success of data infrastructures.
- **Develop adequate operating models for sustainable data infrastructures and services.** Policymakers and research funders must rely on comprehensible and sustainable operating models for research data infrastructures to facilitate their institutional and project-related funding decisions. Infrastructure providers are responsible for all dimensions of sustainability of the offered services. They also have to consider (and monitor) the fit of their offerings into the (evolving) overall infrastructure framework on the national and European level.

4.3 RESEARCH, TEACHING AND TRANSFER IN THE DIGITAL TRANSFORMATION: FEDERATED DATA INFRASTRUCTURES AS DRIVERS OF INSTITUTIONAL STRATEGIES

In recent years, numerous universities and non-university research institutions in Germany have coped intensively with the challenges of the digital transformation in science and humanities and, for example, developed dedicated strategies for Open Science and RDM. In order to further improve the conditions for data-intensive research at innovative universities and non-university research institutions, the RfII suggests creating policies for the use of data infrastructures, the application of the FAIR principles and overarching as well as subject-related quality standards. Furthermore, corresponding competences should already be built into the subject curricula of the entire spectrum of existing Bachelor's and Master's degree programmes.

Developing expertise in dealing with data infrastructures

As already suggested by the RfII in various recommendations, the introduction of study programmes should be continued to address the demand for novel academic and science-supporting qualification profiles of e.g. data librarians, data scientists, science-supporting data stewards, etc. Universities and non-university research institutions also have an important task in supporting the development of data infrastructures. As established institutions of the scientific system, they are in a position to recognise undesirable developments at an early stage and to address them clearly. Furthermore, they represent the link between researchers, lecturers and the research-enabling institutions of the scientific system (e.g. archives, libraries, laboratories, computing centres and collections). After all, HEIs and research institutions can best assess the functions and consequences of the 'new' data infrastructures for the interplay of the different institutional levels involved.

Recruitment requirements for qualified personnel

The RfII presented recommendations on the challenges of the digital transformation for the academic labour market in 2019. These have only been implemented to some extent.[149] A major problem continues to be the classification and, closely related to this, the salary-related grouping of research supporting staff under the framework of the general public sector salary scheme. This staff is needed with increasing digitalisation, but – in terms of job description and salary category – can neither be clearly attributed to the scientific core area of research personnel nor to the higher education administration or research management staff. A greater permeability of the academic labour market at universities and non-university research institutions in both directions is also not yet apparent and ultimately also prevents reputation building through successful support services in infrastructure operations.

Scientific organisations compete with industrial employers for highly qualified personnel

The establishment of NFDI has created an institutional network that also addresses the challenge of creating a new academic workforce. Generally speaking, awareness for skill development and novel research related jobs is rising in the course of digital transformation. Nevertheless, the management level of universities and non-university research institutions is asked to create attractive job descriptions and incentives for personnel development which enforce the further expansion of scientific data infrastructures. This is necessary, to keep the scientific system competitive in recruiting digitally skilled personnel.

Enabling data-intensive research

Demand-oriented data infrastructures that are (co-)financed by public funds should in principle be firmly integrated into the daily research routine of universities and non-university research institutions. In this way, they generate a high intensity of data use and can achieve long-term running or operating times. The latter offers planning security and continuity for research lines and reusability of data for researchers across institutional boundaries. Last but not least, this helps to invest in profile building and focusing on alliances of universities and non-university research institutions. Data-intensive research at universities with a claim to excellence should be actively involved in the further development of scientific data infrastructures – which should also be reflected in science-led selection procedures for projects and priority areas (peer review, institutional strategic development) as well as for academic personnel (fellowships, appointment management).

Cooperation as a prerequisite for long-term data infrastructures

Universities and non-university research institutions must be aware that the path towards the long-term operation of data infrastructures is a joint transformation process. The necessary cultural change requires cooperation. Excessive competition thwarts the community-driven character of data infrastructures. It is also a task of research policy and funding to shift the system as a whole more towards cooperation and maintaining the commons – especially in the field of building and maintaining data infrastructures. The management of higher education and non-university research institutions should take bold steps here. Instead of setting up local

149 RfII (2019) – Digital Competencies

(and eventually: unconnected) data infrastructures, academic leaders are encouraged to promote cooperation in networked and cross-located data services. In general, there must be a greater awareness that infrastructure topics are ultimately relevant for success in research.

4.3: RFII-RECOMMENDATIONS ON RESEARCH, TEACHING AND TRANSFER IN THE DIGITAL TRANSFORMATION

Addressee: Management Level of Research Performing Organisations

- **Align education programmes with the creation and utilisation of data infrastructures to facilitate improved collaboration.**
  Curricula, qualification programmes, and guidelines at universities and non-university research institutions should be aligned and matched with the co-design, co-development and utilisation of data infrastructures such as NFDI and the EOSC Federation.
- **Enhance skills in data utilisation, data security, and legal considerations to empower researchers.**
  Competency development in data reuse, security, dissemination, and proficient management of data services, including legal aspects, should be encouraged. This effort should be incorporated into fundamental study curricula and ongoing training programmes for students, researchers, and support staff.
- **Support institutional policies adhering to FAIR principles and quality standards for effective infrastructure utilisation.**
  Institutional policies for the use of data infrastructures, the application of FAIR principles, and overarching as well as domain-specific quality standards should be supported.
- **Adapt personnel structures and job descriptions to meet the demands of digital transformation in academia.**
  Strict separations between the administrative and researcher work levels must be replaced by more fluid options to cover the greater complexity and variety in current academic professional roles in higher education and research institutions. Personnel structures, job descriptions and salary schemes should be adapted to the requirements of digital transformation in all academic ranks.
- **Emphasise cooperation over competition in infrastructure development – conceive research data infrastructures as common goods.**
  Data infrastructure construction and maintenance should not be seen as a field of competition between individuals, faculties or scientific centres or even between individual research performing organisations. It is a common good that requires cooperation. Higher education and research institutions should focus on cooperative transformation processes that enable the development of joint data infrastructures as commons.

4.4 INCENTIVES FOR INFRASTRUCTURE-BASED SCIENCE AND HUMANITIES THROUGH RESEARCH FUNDING

Promoting the active use of data infrastructures

Funding organisations and programmes by governments and supranational institutions can have an important function for the success of building research data infrastructures. They influence the scale and scope of the digital transformation of science and humanities by (further) re-adjusting their funding conditions. Therefore, it would be helpful when they promote the preferred use of high-quality infrastructure offerings such as data, data sources and services offered by NFDI and – in the near future – sources which are linked via EOSC (or equivalent data infrastructures in the realm of the European Data Spaces) to support the research process for the sake of higher productivity and better quality. The funding guidelines should also make it more attractive for researchers to contribute their own project data to the mentioned infrastructures and thus expand the scope of data reuse. Secondly, funding organisations, agencies and ministries should facilitate low-threshold and reliable access to the infrastructures in the long term – for example, by refunding licenses or access fees for specialised services – even if they are charged by publicly funded data infrastructure providers. However, access to the basic research data, data sources and data-related services provided by NFDI and via the EOSC Federation must be free of charge for researchers from higher education and public research institutions. In this sense they must be considered as a common good (basic supply) for science and humanities in general. For services that significantly exceed basic requirements, moderate fees for researchers or their institutions can be considered to be eligible for funding. This also applies to access fees to exploit data and data sets from other sector-specific (European) data spaces, e.g. business or health data, in a publicly funded research project. Additional funding is also needed to finance necessary further training to develop and expand digital skills and qualifications of staff in research projects – especially for early career researchers.[150] Thirdly, researchers should be more broadly involved in evaluations of the data infrastructures they use.

Linking infrastructure funding and evaluation processes

If these aspects are considered in the funding conditions and review processes, they will be generating incentives to use scientific data infrastructures that already exist or are currently being established. Particularly in the case of long-term transition processes, the RfII considers it advisable to gradually adapt the framework conditions of research funding and review – ultimately: of quality assurance in the scientific system. It is helpful here to interlock the review and evaluation processes of subject-specific research funding with infrastructure funding. Through greater coordination of the objectives, funding criteria and time horizons of infrastructure funding, and with permanent consideration of further developments in research-related funding of national (e.g. DFG) and Europe-wide (e.g. European Research Council)

150 The WR sees a need to critically review the existing regulations and the practice of recognising direct project costs and – within the framework of the possibilities under funding law – to adapt them if necessary. Wissenschaftsrat (2023a) – Strukturen der Forschungsfinanzierung an deutschen Hochschulen, p. 8.

organisations, parallel developments can be avoided and long-term synergies developed. The same holds true for the funding by governmental agencies, ministries or directorates.

In particular, when it comes to the establishment and operation of basic service structures research performing organisations and public research infrastructure providers are in competition with commercial enterprises. In order to secure long-term sovereignty over services and infrastructures used within the scientific community, it is important that scientific institutions are encouraged but also enabled to participate in tenders – such as pre-commercial procurements – which are unknown territory for most research performing organisations in Germany. Usually, they can hardly cope with mixed requirements for liability, guarantees, taxes or specific certifications for quality assurance and cyber security. Research funders and policy-makers need to react and make German research institutions fit to actively meet these new challenges in funding conditions. In parallel to an enabling-strategy RfII recommends that the German Federal Government advocates a research-friendly design of the tendering conditions for research data infrastructures and related services by the European institutions.

4.4: RFII-RECOMMENDATIONS ON INCENTIVES FOR INFRASTRUCTURE-BASED SCIENCE AND HUMANITIES THROUGH RESEARCH FUNDING

Addressee: Funders for Research Projects and Scientific Third-Party Funding

- **Adjust own funding activities for research and data infrastructure development in the light of other institutions' funding programmes.**
  Funders should coordinate with other grant or fund-giving institutions to contribute to a coherent development of research data infrastructures. In this sense, mutual adjustment of funding strategies from funding organisations, agencies and ministries would be helpful. German research performing organisations should also be encouraged and enabled to participate in public procurements of European institutions.
- **Provide incentives for utilisation of existing infrastructures as well as for contribution of project data and services to promote scientific advancement.**
  Funding conditions and review processes should a) enforce an appropriate utilisation of existing infrastructure offerings to effectively drive scientific progress; b) make it more attractive to give own research data and services into the infrastructures and widen their scope.
- **Support access to a wide scope of research-relevant data and data bases.**
  While access to genuine high-quality research data is a task of basic supply by NFDI or the EOSC Federation and should be free of costs for individual researchers, additional services or access to other research-relevant but non-scientific data spaces might be associated with additional costs for researchers from public and private HEIs. These costs should be eligible for research funding.

- **Invest in development of digital skills for early and mid-career researchers.**
  Funding organisations should co-finance digital skill development for personnel in third-party-funded research projects, particularly in the early career stage of researchers as well as for research supporting personnel.
- **Synchronise research reviews and infrastructure funding by recognising and utilising thematic interdependencies to amplify impact.**
  Better alignment of review and evaluation processes in research and infrastructure funding should be established. In particular, interdependencies between thematic research and infrastructure-related funding programmes should be identified and utilised.

4.5 COHERENT AND LONG-TERM RESEARCH POLICY AS A SUCCESS FACTOR FOR FEDERATED DATA INFRASTRUCTURES

Early discussion of long-term solutions for scientific data infrastructures

With NFDI, a suitable support structure is emerging in Germany that will make it possible to establish research-oriented and quality-assured infrastructure solutions beyond project funding. Data infrastructures financed through timely limited project funding run counter to the goal of a sustainable and reliable working structure in the sense of a common good. Temporary structures generate little trust among users and thus encourage the development of insular local solutions that, at least at first glance, promise a higher degree of long-term control over the fate of one's own data. For many years, the scientific system has successfully countered the negative effects of long-term fixed or institutional funding (eternity guarantees), such as structural rigidity and declining innovation dynamics, by incorporating quality assurance mechanisms like accreditations and evaluations at appropriate time intervals into the institutional funding of research or research-related institutions. The RfII assumes that such experiences can also be transferred to the field of quality-assured and demand-oriented basic provision of data and services by research data infrastructures.

The RfII regards NFDI as the main German data infrastructure for sharing research data and the necessary services for RDM in accordance with the FAIR principles. NFDI should therefore be prioritised by the Federal Government and the Länder, particularly in terms of funding. Germany should continue to contribute to EOSC in-kind by further advancing a strong NFDI – as a guarantor for high-quality research data from Germany for the (re-)use in the overall European Research Data Space. The basic services of EOSC should also be financed in such a way that researchers from Europe can use them free of charge. Services that go well beyond these basic requirements can also be accompanied by moderate fees for the users or their institutions (see 4.3).

In Germany, joint science policy and research funding to support institutions or scientific actions are tasks for different legislation authorities and there-

Initiate a process to establish sustainable ownership structures and convergence

fore need special political arrangements between the Federal Government and the Länder. This especially holds true when the long-term financing and permanent implementation and operation of spatially dispersed network-like structures for federated data infrastructures is to be supported. Construction and maintenance of such infrastructures represent a particular challenge for research policymakers. Appropriate solutions need to be developed now and continuously. In this sense, the RfII considers the system evaluation of NFDI by the WR as an important first step. Building on the system evaluation results, the RfII recommends further coordinated processes among policymakers, academia, and the scientific community to establish sustainable ownership and maintenance structures for scientific data infrastructures in and beyond the NFDI-context. Subject of the recommended coordinated process would also be the identification of interfaces and joint policies that enable a convergent development of NFDI, EOSC, Gaia-X, and other relevant European Data Spaces towards a European ecosystem for research data. Of vital importance is in this respect, a seamless further development and perpetuation of NFDI – not only as a single entity but also as Germany's ambitious in-kind-contribution to EOSC (and currently the biggest and most sophisticated contribution of any EU Member State to the EOSC Federation).

Close integration with European, international and interdisciplinary initiatives

Since research does not end at institutional and national borders, close integration with European and international as well as cross-domain initiatives is necessary – also with regard to the respective funding horizons. Research policy, funding organisations and the scientific communities are called upon to tackle these challenges together and to develop the necessary conditions for the long-term availability of research data and corresponding services in a demand-oriented and resource-saving manner. Research policy actors can support this process by providing incentives for further cooperation projects between the various infrastructure initiatives and by ensuring (co-) financing for these projects. Agreement on shared standards and technical interfaces should be defined by the scientific communities and at the same time in cooperation with economic and civil society actors, based on existing and proven solutions, in order to enable the greatest possible connectivity across different data spaces. An important question for both scientific actors and industry is the design of operating models, which are of central importance for the long-term and quality-assured operation of (scientific) data infrastructures. Last but not least, this also raises questions about the institutional allocation of and long-term sponsorship for the various infrastructures outside the realm of NFDI, which must be decided upon by research policy. The early and subsequently usable development of corresponding operating models in the form of best practice examples would contribute significantly to the success of the data infrastructures.

Operation of data infrastructures as a common good

In the course of establishing various sector-specific – and in the case of science and humanities, domain-specific – data spaces in Germany and Europe, particular care must be taken to prevent the establishment of new insular solutions. If possible, existing data infrastructures should be

integrated into NFDI and the EOSC Federation if they are suitable. On the other hand, the parallel creation of regionally or locally limited services and repositories should be avoided. Here, too, the following applies: local solutions must remain possible – also in the sense of experimental fields for new approaches – as long as this can be well justified from a professional and technical point of view. Purely local solutions should nevertheless become exceptions to the rule for the sake of collaborative research and – where this is feasible – a shared culture of open data. Rather, incentives for cooperation should be created that lead to the operation of data infrastructures in the sense of scientific commons. Such commons are used beyond the boundaries and development strategies of individual research performing organisations or single governments. The different control logics and the resulting mechanisms in the German scientific system (e.g. competition vs. cooperation in joint project applications) are often in conflict with each other – as can be seen, for example, in the competitive incentives of the Excellence Strategy. In the view of the RfII, a pure competitive logic is inappropriate for the establishment of research and data infrastructures. It hinders necessary and long-term innovations in this field and delays the progress of the scientific system as a whole.

Consolidation of evaluation criteria

The operational convergence desired by the RfII and the interplay of the various infrastructures require from research and innovation policy to adopt an early and coordinated approach with regard to the evaluation of initiatives and structures, considering the intensity of use and the respective degree of maturity. In recent years, proven evaluation criteria have already been developed in other contexts – especially in the context of institutional evaluations of infrastructure-bearing institutions (e.g. in the realm of the Leibniz Gemeinschaft). The RfII now sees the challenge in bringing these criteria together at a superordinate level and developing a common understanding of quality for the design of data infrastructures that does justice to federated data infrastructures as novel scientific entities sui generis. In particular, the networked and location-independent commons function that these infrastructures provide for scientific research in the 21st century could prove not only to be novel but one of the most outstanding innovations in building scientific institutions for a long time.

4.5: RFII-RECOMMENDATIONS ON COHERENT AND LONG-TERM RESEARCH POLICY AS A SUCCESS FACTOR FOR FEDERATED DATA INFRASTRUCTURES

Addressee: Research Policymakers at Organisational, State (Länder), Federal and European Level

- **Strengthen NFDI within the EOSC Federation as a major (in-kind) contribution for a European Research Data Space.**
  The lasting existence of NFDI is a necessary prerequisite and a driver for the European Research Data Space – and at least for a prospering European Single Market for Data. NFDI should be retained as the permanent German in-kind contribution to EOSC.
- **Carry on with coordinated processes to establish sustainable ownership and maintenance perspectives for research data infrastructures, building on the results of the NFDI system evaluation.**
  Institutional solutions, long-term ownership and maintenance perspectives must be discussed continuously between policymakers, research organisations and the scientific community. A continued coordinated process, building on the results of the NFDI system evaluation, should help to create convergence in the further co-development of NFDI, other national data infrastructures and the EOSC Federation.
- **Conduct coordinated evaluations of data initiatives and federated data infrastructures, assessing user demand, usage intensity, research impact, and interoperability.**
  Evaluations at different levels (local, regional, national and European) should be carried out in a coordinated manner as far as this is possible. User demand, usage intensity, potential impact on research quality and productivity, and maturity must be considered. Evaluations should also assess the alignment of the governance between data infrastructures.
- **Foster operational convergence and networking of federated data infrastructures.**
  The operational convergence and networking of a federated data infrastructure at the European level and in the EU Member States should be facilitated by appropriate and effective governance and viable funding mechanisms. In this respect, a harmonisation of funding conditions for research data infrastructures in the European Research Area (ERA) would be desirable.
- **Establish federated data infrastructures and collaboration networks as a common good to sustain broad access to research data in Germany and beyond.**
  The current governance logic within the German and European scientific landscape should be refined to provide basic supply of a high-quality research data infrastructure such as NFDI. Within the EU, access to infrastructures should be handled as commons, not as an object of competition. Broad access to research data is a crucial source for scientific advancement and economic prosperity.

5 OUTLOOK

Global challenges change our view on openness

Data infrastructures and thus digital services are a more and more important part of current research practice in all disciplines and therefore often an indispensable prerequisite for scientific progress and the acquisition of knowledge. However, a potentially global and open scientific system together with its supporting infrastructures is always embedded in societal transformations that can arise from politics, economy, technological innovations and, in particular, geopolitical challenges or even disruptions. Today global pandemics, climate and biodiversity change as well as new (and: renewed) lines of global and regional conflicts constitute threats that have impact on research policy, the openness debate in academia and the establishment of infrastructures for easier access to data of all kind. This holds especially true for questions of trustworthiness, fairness of mutual data exchange, and data security.

AI developments: Threat or opportunity?

Furthermore, we witness rapid developments in the field of (generative) AI, which also have the potential to challenge politics, society and scientific practice in their everyday dimensions – for example in form of algorithmic distortions, normative bias, copyright infringements or in the restriction of data protection and informational self-determination. The same applies to debates concerning digital and technological sovereignty.[151] On the other hand, the responsible use of generative AI for research purposes holds enormous opportunities for the whole scientific system and might become a key factor in the global competition for the best research-based solutions and economic innovations. Thus, data infrastructures' support in the merging and integration of data takes on a new significance in view of the possibilities of generative AI, which was hardly foreseeable just a few years ago. The blueprints for the construction of data infrastructures such as NFDI, EOSC, Gaia-X and many others were all developed before the rise of the new AI-driven possibilities of data collecting and merging in Large Language Models. Now, the already started infrastructures must adapt to this newly established and fast developing setting and coordinate with the evolving European Data Spaces. This includes attention to new regulatory frameworks on the European and national level (e.g. Data Governance, Data Act) which define new roles and obligations and prepare the ground for new intermediary actors such as data trustees.

Time pressure for responding to these trends

Until now, the German and, in some cases, European approach of initiating scientific data infrastructures from the bottom and implementing them in line with community demand seemed to create a proper ground to handle upcoming challenges – as far as these challenges were anticipated in the second decade of the 2000s. However, in international comparison – particularly with the big global competitors USA and China – we actually see that this approach bears the risk of not achieving sufficient speed to react

151 Wissenschaftsrat (2023) – Souveränität und Sicherheit der Wissenschaft im digitalen Raum

to disruptive technological innovations and geopolitical shifts in the power balance as fast as it becomes necessary to catch up, and at least to stay on equal par, with the developments in these regions. The often too complex governance structures and the level and complexity of financial support for scientific data infrastructures in Germany and Europe will also be an obstacle to future success in a global context.

Future tasks for the RfII

The RfII plans to fundamentally analyse these changing global conditions and challenges as well as their impact on the further development of the aforementioned infrastructures from 2025 on. The focus will be on identifying parallel developments and approaches to integrate generative AI as well as new data security concerns in digital policies and scientific institutions – including data infrastructures – of other countries and regions, particularly in the Americas and Asia. For this project, the RfII must focus on identifying ways and instruments that enable research policymakers, funders, research performing organisations as well as scientific communities to react more quickly than before to new technological and crisis-related challenges, to anticipate them in the best case and to adapt the further development of German and European data infrastructures accordingly.

REFERENCES

Amelung, Lisa et al. (2023): Umgang mit Zielen der BLV als Grundlage für die Strukturevaluation. White Paper. Verabschiedet durch die Konsortialversammlung des Vereins Nationale Forschungsdateninfrastruktur (NFDI) e. V. am 07.07.2023, https://doi.org/10.5281/zenodo.8192533.

Beckmann, Volker (2022): Monitoring EOSC Readiness on the National Level, https://fairsfair.eu/sites/default/files/Monitoring_national_level_FAIRsFAIR_Beckmann.pdf.

Blankertz, Aline; Specht, Louisa (2021): What Regulation for Data Trusts Should Look Like, Berlin, https://www.stiftung-nv.de/sites/default/files/regulation_for_data_trusts_0.pdf.

BMWK – Bundesministerium für Wirtschaft und Klimaschutz (2023): Förderrichtlinie für die Bundesförderung im Bereich Cloud- und Edge-Infrastruktur und -Services im Rahmen des IPCEI-CIS, Bundesanzeiger AT 16.06.2023 B1, https://www.bundesanzeiger.de/pub/de/amtliche-veroeffentlichung?2.

Boehm, Franziska et al. (2021): Sektionskonzept 'Ethical, Legal & Social Aspects' (section-ELSA). Konzept zur Einrichtung einer Sektion im Verein Nationale Forschungsdateninfrastruktur (NFDI) e. V., https://doi.org/10.5281/zenodo.5675972.

Bonfiglio, Francesco (2021): Vision and Strategy, https://gaia-x.eu/wp-content/uploads/2021/12/Vision-Strategy.pdf.

Brünger-Weilandt, Sabine; Gehring, Petra (2022): NFDI – wie sie wurde, was sie ist. Sabine Brünger-Weilandt und Petra Gehring im Gespräch, ZfBB 69, No. 1-2, pp. 6-9, https://doi.org/10.3196/1864295020691226.

Burgelman, Jean-Claude (2021): Politics and Open Science: How the European Open Science Cloud Became Reality (the Untold Story), Data Intelligence 3, No. 1, pp. 5-19, https://doi.org/10.1162/dint_a_00069.

Coduti, Maria Rosaria (2023): The Common European Data Spaces: Harnessing Data to Benefit the Society, the Planet and the Economy, European Commission, Directorate-General for Communications Networks, Content and Technology, https://data.europa.eu/sites/default/files/course/European%20data%20spaces%20for%20public%20administrations%20and%20data.europa.eu_.pdf

Council of the European Union (2021): Conclusions on the Future Governance of the European Research Area (ERA), https://data.consilium.europa.eu/doc/document/ST-14308-2021-INIT/en/pdf.

DFG – Deutsche Forschungsgemeinschaft (2021): Erkenntnisgeleitete Forschung stärken, von Wissensspeichern profitieren. Impulse der Deutschen Forschungsgemeinschaft für die 20. Legislaturperiode des Deutschen Bundestags, https://www.dfg.de/resource/blob/174878/35a379dba48731ecaaae-c31d1ed87329/20210505-pm-impulspapier-legislaturperiode-data.pdf.

DFG – Deutsche Forschungsgemeinschaft; NFDI-Expertengremium (2023): Eckpunkte für die zweite Förderphase der NFDI-Konsortien, https://www.dfg.de/resource/blob/204400/0d1b64172bf7e2f57ed-cd4b10115524e/eckpunkte-zweite-foerderphase-de-data.pdf.

Diepenbroek, Michael; Schimmler, Sonja; Ebert, Barbara (2021): Sektionskonzept Common Infrastructures zur Einrichtung einer Sektion im Verein Nationale Forschungsdateninfrastruktur (NFDI) e. V., https://doi.org/10.5281/ZENODO.5607490.

Dietrich, Mark; Ferrari, Tiziana (2021): Governance, Architectures and Business Models for Data and Cloud Federations: the EOSC and Gaia-X Case Studies, https://doi.org/10.5281/zenodo.4929021.

Dumouchel, Suzanne (2023): EOSC Building – Sharing Perspectives and Thoughts for the Future, https://eosc.eu/sites/default/files/2023-06/20230613_EOSCbuilding-Suzanne-Dumouchel.pdf.

EC – European Commission; EOSC Steering Board Expert Group (2022): Opinion Paper on EOSC and Commercial Partners, https://doi.org/10.2777/04436.

EC – European Commission (2015): A Digital Single Market Strategy for Europe. COM (2015) 192 final, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52015DC0192&from=EN.

EC – European Commission (2016a): European Cloud Initiative – Building a Competitive Data and Knowledge Economy in Europe. COM (2016) 178 final, Brussels, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=15266.

EC – European Commission (2016b): Realising the European Open Science Cloud. First report and recommendations of the

Commission High Level Expert Group on the European Open Science Cloud, High Level Expert Group, Luxembourg, https://data.europa.eu/doi/10.2777/940154.

EC – European Commission (2018a): Prompting an EOSC in Practice. Final Report and Recommendations of the Commission 2nd High Level Expert Group [2017-2018] on the European Open Science Cloud (EOSC), Brussels, https://data.europa.eu/doi/10.2777/112658.

EC – European Commission (2018b): Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 'Towards a Common European Data Space'. COM (2018) 232 final, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0232.

EC – European Commission (2019): Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on Open Data and the Re-use of Public Sector Information (Recast). Directive (EU) 2019/1024, 20.06.2019, in: Official Journal of the European Union, L 172/56-L 172/83, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=uriserv:OJ.L_.2019.172.01.0056.01.ENG.

EC – European Commission (2020): A European Strategy for Data. COM (2020) 66 final, https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN.

EC – European Commission (2021): Proposal for a Regulation of the European Parliament and of the Council Laying Down Measures Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. COM (2021) 206 final. 2021/0106 (COD), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

EC – European Commission (2022a): Horizon Europe Work Programme 2021-2022 – Research Infrastructures, https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/wp-call/2021-2022/wp-3-research-infrastructures_horizon-2021-2022_en.pdf.

EC – European Commission (2022b): Horizon Europe Work Programme 2023-2024 – Research Infrastructures, https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/wp-call/2023-2024/wp-3-research-infrastructures_horizon-2023-2024_en.pdf.

EC – European Commission (2022c): Proposal for a Regulation of the European Parliament and of the Council on the European Health Data Space. COM (2022) 197 final. 2022/0140 (COD), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX-%3A52022PC0197.

EC – European Commission (2023): Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 Laying Down a List of Specific High-Value Datasets and the Arrangements for their Publication and Re-Use, in: Official Journal of the European Union, L 19/43-L 19/75, https://eur-lex.europa.eu/eli/reg_impl/2023/138.

EC – European Commission (2024): Commission Staff Working Document on Common European Data Spaces. SWD (2024) 21 final, Brussels, https://ec.europa.eu/newsroom/dae/redirection/document/101623.

EFI – Expertenkommission Forschung und Innovation (2021): Gutachten 2021. Gutachten zu Forschung, Innovation und technologischer Leistungsfähigkeit Deutschlands, Berlin, https://www.e-fi.de/fileadmin/Assets/Gutachten/2021/EFI_Gutachten_2021.pdf.

EOSC Association; EC – European Commission (2021): Memorandum of Understanding for the Co-programmed European Partnership on the European Open Science Cloud, https://www.eosc.eu/sites/default/files/EOSC_Memorandum_30_July_2021.pdf.

EOSC Association (2022): EOSC AISBL Articles of Association, https://eosc.eu/wp-content/uploads/2023/08/20220524_EOSC-A_Revised-Articles-of-Association_PDF.pdf.

EOSC Association (2023a): Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC), Version 1.2, https://eosc.eu/wp-content/uploads/2023/12/20231114_SRIA_1.2_final2.pdf.

EOSC Association (2023b): EOSC Multi-Annual Roadmap (MAR) 2025 and 2026-2027, https://eosc.eu/sites/default/files/2023-01/MAR_2025-27_draft.pdf.

EOSC Association (2023c): Board Position Paper on the EOSC Federation and the Role of EOSC Nodes, https://eosc.eu/wp-content/uploads/2023/11/20231112-Short-paper-on-the-EOSC-Federation-draft-v3.pdf.

EOSC – Financial Sustainability Task Force (2022): Towards Sustainable Funding Models for the European Open Science Cloud. Financial Sustainability Task Force Progress report November 2022, https://www.eosc.eu/sites/default/files/2022-11/financial-sustainability-tf-progress-report-nov-2022.pdf.

European Member States; EC – European Commission (2020): Declaration – Building the Next Generation Cloud for Businesses and the Public Sector in the EU, https://ec.europa.eu/newsroom/dae/redirection/document/70089.

European Union (2016): Regulation (EU) 2016/779 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data and Repealing Directive 95/46/EC (General Data Protection Regulation), in: Official Journal of the European Union, L 119/1-L 119/88, https://eur-lex.europa.eu/eli/reg/2016/679/oj.

European Union (2022): Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European Data Governance and Amending Regulation (EU) 2018/1724 (Data Governance Act), in: Official Journal of the European Union, L 152/1-L 152/44, https://eur-lex.europa.eu/eli/reg/2022/868/oj.

European Union (2023): Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on Harmonised Rules on Fair Access to and Use of Data and Amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act), in: Official Journal of the European Union, https://eur-lex.europa.eu/eli/reg/2023/2854.

Gaia-X AISBL (2022a): Policy Rules Document, Brussels, https://gaia-x.eu/wp-content/uploads/2022/05/Gaia-X_Policy-Rules_Document_v22.04_Final.pdf.

Gaia-X AISBL (2022b): Gaia-X-Architecture Document – 22.04 Release, https://gaia-x.eu/wp-content/uploads/2022/06/Gaia-x-Architecture-Document-22.04-Release.pdf.

GWK – Gemeinsame Wissenschaftskonferenz (2022): Nationale Forschungsdateninfrastruktur (NFDI) mit der Aufnahme von acht Konsortien in die Förderung komplett. Press release 13/2022, München/Bonn, https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Pressemitteilungen/pm2022-13.pdf.

GWK – Joint Science Conference (2018): Agreement between the Federal Government and the Länder concerning the Establishment and Funding of a National Research Data Infrastructure (NFDI) of 26 November 2018, https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/BLV_english.pdf.

Herres-Pawlis, Sonja et al. (2022): Sektionskonzept Training & Education zur Einrichtung einer Sektion im Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V, https://doi.org/10.5281/zenodo.6475541.

Kerber, Wolfgang (2023): Governance of IoT Data: Why the EU Data Act Will not Fulfill Its Objectives, in: GRUR International 72, No. 2, pp. 120-135, DOI: https://doi.org/10.1093/grurint/ikac107.

Koepler, Oliver et al. (2021): Sektionskonzept Meta(daten), Terminologien und Provenienz zur Einrichtung einer Sektion im Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V, https://doi.org/10.5281/ZENODO.5619089.

Le Maire, Bruno; O, Cédric; Vidal, Frédérique (2021): Stratégie Nationale pour le Cloud, https://www.entreprises.gouv.fr/files/files/secteurs-d-activite/numerique/dp-strategie-cloud.pdf.

Leibniz-Gemeinschaft (2021): Europäische Partnerschaften in Horizon Europe 2021-2027, https://www.ipfdd.de/fileadmin/user_upload/ax/HEU_Partnerschaften_Broschuere_2021.pdf.

Lipp, Anne; Sure-Vetter, York (2020): NFDI – Akteure und Prozesse, Erfolgsfaktoren und Herausforderungen, ZfBB 69, No. 1-2, S. 10-17, https://doi.org/10.3196/1864295020691230.

NFDI e. V. (2020): Nationale Forschungsdateninfrastruktur (NFDI) e. V. – Satzung, https://www.nfdi.de/wp-content/uploads/2021/05/Satzung-NFDI-eV.pdf.

NFDI e. V. (2022a): Leitfaden für Konsortien im NFDI-Verein. Verabschiedet durch den Wissenschaftlichen Senat des Vereins Nationale Forschungsdateninfrastruktur (NFDI) e. V. am 22.02.2022, https://www.nfdi.de/wp-content/uploads/2022/03/Konsortialleitfaden.pdf.

NFDI e. V. (2022b): Leitfaden für Sektionen im NFDI-Verein. Verabschiedet durch den Wissenschaftlichen Senat des Vereins Nationale Forschungsdateninfrastruktur (NFDI) e. V. am 22.02.2022, https://www.nfdi.de/wp-content/uploads/2022/03/Sektionsleitfaden.pdf.

NFDI e. V. (2022c): Stellungnahme zum Entwurf einer EU Open Data Durchführungsverordnung der Sektionen Ethical, Legal & Social Aspects; Common Infrastructures und (Meta)daten, Terminologien und Provenienz des Verein Nationale Forschungsdateninfrastruktur (NFDI) e. V., https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12111-Open-data-availability-of-public-datasets/F3316361_en.

NFDI e. V. (2022d): Stellungnahme der NFDI-

Konsortien zu Basisdiensten, https://doi.org/10.5281/zenodo.6091656.

NFDI – Sektion ELSA (2022): Stellungnahme zum EU Data Act Proposal der Sektion ELSA (Ethical, Legal & Social Aspects) des Verein Nationale Forschungsdateninfrastruktur (NFDI) e. V., Verein Nationale Forschungsdateninfrastruktur (NFDI) e. V. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13045-Datengesetz-und-geanderte-Vorschriften-uber-den-rechtlichen-Schutz-von-Datenbanken/F3258672_de.

Federal Ministry of Economic Affairs and Energy, French Ministry of the Economy, Finance and the Recovery (2020): Franco-German Position on Gaia-X, https://www.bmwk.de/Redaktion/DE/Downloads/F/franco-german-position-on-gaia-x.pdf?__blob=publicationFile&v=4.

o. A. (2021): Opening the Door to a World of FAIR Research Digital Objects – On the Threshold to the European Open Science Cloud, https://www.bildung-forschung.digital/digitalezukunft/shareddocs/Downloads/files/eosc-declaration-opening-the-doors-to-a-fair-world.pdf?__blob=publicationFile&v=1.

Person, Christian; Schütrumpf, Moritz (2023): Das Projekt Gaia-X. Next Generation einer föderierten Dateninfrastruktur. Version 2.0, https://zevedi.de/wp-content/uploads/2023/03/ZEVEDI_Handreichung-Gaia-X_V2.0.pdf.

RfII – German Council for Scientific Information Infrastructures (2016): Enhancing Research Data Management: Performance through Diversity. Recommendations Regarding Structures, Processes, and Financing for Research in Data Management in Germany, Göttingen, https://rfii.de/?p=2075.

RfII – German Council for Scientific Information Infrastructures (2018): Statement of the Council for Scientific Information Infrastructures (RfII) on the proposals for a European Open Science Cloud (EOSC), Göttingen, https://rfii.de/?p=2792.

RfII – German Council for Scientific Information Infrastructures (2019): Digital Competencies – Urgently Needed! Recommendations on Career and Training Prospects for the Scientific Labour Market, Göttingen, https://rfii.de/?p=4015.

RfII – German Council for Scientific Information Infrastructures (2021): Statement on the Proposal for a Data Governance Act (DGA) by the EU Commission, Göttingen, https://nbn-resolving.org/urn:nbn:de:101:1-2020052666.

RfII – German Council for Scientific Information Infrastructures (2021): Statement on the Proposed Data Act of the European Union in View of the Inception Impact Assessment Published on this Subject, Göttingen, https://nbn-resolving.org/urn:nbn:de:101:1-2020052689.

RfII – German Council for Scientific Information Infrastructures (2022): Statement on the Proposal for a Data Act by the European Commission, Göttingen, https://d-nb.info/1258994119/34.

RfII – Rat für Informationsinfrastrukturen (2022): Datenpolitik, Open Science und Dateninfrastrukturen: Aktuelle Entwicklungen im europäischen Raum, https://rfii.de/?p=7743.

RfII – Rat für Informationsinfrastrukturen (2023a): Föderierte Dateninfrastrukturen für die wissenschaftliche Nutzung. NFDI, EOSC und Gaia-X: Vergleich und Anregungen für eine engagierte Mitgestaltung des Ausbaus und der Weiterentwicklung (RfII Berichte, 4), Göttingen, https://rfii.de/?p=8533.

RfII – Rat für Informationsinfrastrukturen (2023b): Stellungnahme zum Entwurf eines Gesetzes zur Durchführung der Verordnung (EU) 2022/868 des Europäischen Parlaments und des Rates vom 30. Mai 2022 über europäische Daten-Governance und zur Änderung der Verordnung (EU) 2018/1724 (Daten-Governance-Rechtsakt), Göttingen, https://nbn-resolving.org/urn:nbn:de:101:1-2021090897.

Siemen, Stephan (2022): EOSC Platform and Upcoming Related HE Calls, European Commission, Directorate-General for Communications Networks, Content and Technology, https://www.horizon-europe.gouv.fr/sites/default/files/2022-12/2-2022-12-14-ec-he-infraeosc-stephan-siemen-pdf-7726.pdf.

Specht-Riemenschneider, Louisa (2022): Der Entwurf des Data Act – Eine Analyse der vorgesehenen Datenzugangsansprüche im Verhältnis B2B, B2C und B2G. In: MMR-Beilage (09), pp. 809–826. https://www.jura.uni-bonn.de/fileadmin/Fachbereich_Rechtswissenschaft/Einrichtungen/Lehrstuehle/Specht/Dateien/MMR-Beilage.pdf.

Stahl, Florian; Hamann, Andreas (2023): Sektionskonzept Industry Engagement zur Einrichtung einer Sektion im Verein Nationale Forschungsdateninfrastruktur (NFDI) e.V, https://doi.org/10.5281/zenodo.7900079.

Stahl, Florian et al. (2023): Collaboration Models Between Industry and Academia.

Whitepaper on behalf of the Section Industry Engagement of the National Research Data Infrastructure (NFDI), https://stifterverband.sharefile.eu/share/view/s516afb4e907b-419485c1f053e8b3fcff.

Stifterverband für die Deutsche Wissenschaft; Heinz Nixdorf Stiftung (2023): Forschungsdaten aus der Wirtschaft: Die Sicht der Hochschulen. Ergebnisse aus dem Hochschul-Barometer 2023 (Discussion Paper, 6), Essen, https://www.stifterverband.org/sites/default/files/2023-08/forschungsdaten_aus_der_wirtschaft_die_sicht_der_hochschulen_2023.pdf.

Wittenburg, Peter (2021): Comments to Jean-Claude Burgelman's article Politics and Open Science: How the European Open Science Cloud Became Reality (the Untold Story), in: Data Intelligence 3, No. 1, 47-51, https://doi.org/10.1162/dint_a_00076.

WR – Wissenschaftsrat (2023a): Strukturen der Forschungsfinanzierung an deutschen Hochschulen. Positionspapier, Cologne, https://doi.org/10.57674/pms3-pr05.

WR – Wissenschaftsrat (2023b): Empfehlungen zur Souveränität und Sicherheit der Wissenschaft im digitalen Raum. Drs. 1580-23, Cologne, https://doi.org/10.57674/m6pk-dt95.

All URLs were last accessed on 18.09.2024.

APPENDIX

TABLE 1: NFDI CONSORTIA

Table 1 refers to section 2.1.4 in this publication (p. 11).

The project descriptions were taken from the DFG project database GEPRIS in an abridged and slightly modified version: https://gepris.dfg.de/gepris/OCTOPUS; the assignment to the scientific disciplines is based on the classification on the website: https://www.nfdi.de/consortia/?lang=en. Please note that the following descriptions are based exclusively on the consortia's descriptions of their tasks and objectives. They do not contain any assessment by the RfII regarding the appropriateness of the tasks or the chances of achieving the stated objectives.

	Consortium (Prior Scientific Disciplines Covered) and Description
First Selection Round 2020	**DataPLANT (Life Sciences)** The main goal of the Data in Plant Research consortium is to enhance added value in basic plant research by contextualizing research data according to FAIR principles with minimal effort. DataPLANT will establish a service environment supporting the entire research cycle in modern plant biology, offering both technical-digital and on-site personnel assistance. This initiative aims to create a central entry point and a valuable resource for plant-specific data and knowledge.
	GHGA (Life Sciences) The German Human Genome Phenome Archive will establish a national infrastructure for secure storage, access management, and analysis of human omics data. It will integrate existing German omics data providers to create a harmonised distributed infrastructure. GHGA will streamline data deposition and be part of European infrastructures as a federated German European Genome Archive node. This will shape international standards for data exchange and facilitate participation in international research consortia. GHGA will democratise access to large datasets with secure, cloud-based access and enable compute-intensive applications.
	KonsortSWD (Humanities and Social Sciences) The Consortium for the Social, Educational, Behavioural and Economic Sciences aims to enhance RDM for its interdisciplinary communities, leveraging its expertise in operating community-driven data infrastructures. KonsortSWD's mission is to develop a user-driven data infrastructure for studying human society, providing tools and services to manage and share data in compliance with FAIR principles. KonsortSWD develops a range of services that address the requirements of researchers and research data centres – e.g. the project *Open Data Format*, which offers a solution for data processing and data exchange in research.
	NFDI4Biodiversity (Life Sciences) The consortium Biodiversity, Ecology and Environmental Data aims to enhance access to interoperable data through six key objectives: integrating RDM into biodiversity research, ensuring FAIR data principles, consolidating FAIRness with quality, embedding into national and international landscapes, addressing cross-cutting topics, and promoting collaborative governance and sustainability.
	NFDI4Cat (Natural Sciences) The consortium NFDI for Science Around Catalysis follows the objectives to bring together the various disciplines of catalysis. Catalysis, a vital field enabling efficient production across industries and scales, addresses major challenges like sustainable energy and climate change. However, data management in catalysis is fragmented, lacking overarching repositories and standards. NFDI4Cat aims to revolutionise catalysis research by unifying data management across disciplines, fostering a digital future for catalysis.
	NFDI4Chem (Natural Sciences) The Chemistry Consortium for NFDI aims to digitise essential processes in chemical research, facilitating data collection, storage, analysis, sharing, and reuse. It promotes Open Science and RDM in line with FAIR principles, offering a comprehensive approach to data access for the chemistry community. The primary goal is to establish a data infrastructure for chemistry in Germany, fostering innovation and enabling easy access to data for novel scientific approaches.

	Consortium (Prior Scientific Disciplines Covered) and Description
First Selection Round 2020	**NFDI4Culture (Humanities and Social Sciences)** The consortium for Research Data on Tangible and Intangible Cultural Heritage strives for establishing a demand-orientated, research-led infrastructure for research data on cultural assets based on the FAIR and CARE principles. The consortium covers a broad spectrum of academic disciplines and specific research areas: Architecture, Art History, Musicology, Dance and Theatre Studies, Film and Media Studies. NFDI4Culture seeks to promote a cultural shift towards more sustainable RDM and enhance qualification and professionalisation in data-driven cultural heritage research.
	NFDI4Health (Life Sciences) The consortium NFDI for Personal Health Data seeks to compile a comprehensive inventory of German epidemiological, public health, and clinical trial data. It will establish a centralised data catalogue with advanced search features, robust data access management, and a data analysis toolkit, prioritising privacy protection for personal health data. Standardisation services will ensure interoperability, and prototypical study types will demonstrate the feasibility of harmonising all infrastructure, tools, and services in alignment with user needs.
	NFDI4Ing (Engineering Sciences) The consortium NFDI for the Engineering Sciences unites engineering communities to enhance engineering RDM. It adopts a method-oriented, user-centred approach to ensure engineering research data becomes FAIR. It identifies typical methods and workflows in engineering research through seven archetypes, addressing various needs such as bespoke experiments, engineering software, provenance tracking, and more. NFDI4Ing plans to coordinate its efforts based on the DFG classification of engineering research areas, supporting five community clusters and offering seven base services including quality assurance, research software development, metadata, repositories, training, and data discovery.
Second Selection Round 2021	**BERD@NFDI (Humanities and Social Sciences)** The consortium NFDI for Business Administration, Economics and Related Data aims to create a comprehensive platform for collecting, (pre)processing, analysing, and preserving subject-specific data. It will support integrated management of algorithms and data throughout the research cycle, with a particular emphasis on unstructured (big) data like video, image, audio, text, or mobile data. The initiative will address the challenges of expanded empirical research by fostering community building, providing publicly available and online accessible datasets, enhancing data documentation and preservation according to the FAIR principles. It strives for offering an algorithm repository and benchmarks, supplying computing and storage power for data analysis, and providing a wide range of APIs for interaction with external systems.
	DAPHNE4NFDI (Natural Sciences) The consortium Data from Photon and Neutron Experiments collaborates with key scientific users and research facilities in photon and neutron science to enhance data management practices. Through DAPHNE, the photon and neutron community aims to: (1) enhance metadata capture via user-driven online logbooks, improving experiment documentation; (2) establish a community repository for processed data, reference databases, and analysis code, enhancing data access and enabling reuse; (3) deploy user-developed analysis software on facility infrastructure, enabling wider access to advanced analysis tools.
	FAIRmat (Natural Sciences) The FAIR Data Infrastructure for Condensed Matter Physics and Chemical Physics of Solids aims to establish a federated FAIR Data Infrastructure for materials data through the FAIRmat Portal. It will enhance metadata schemes and ontologies and facilitate efficient data exchange to advance condensed-matter and materials physics, benefiting engineering as well. Additionally, FAIRmat seeks to encourage researchers to share even seemingly irrelevant data and provide support and training within and beyond its community.

	Consortium (Prior Scientific Disciplines Covered) and Description
Second Selection Round 2021	**MaRDI (Natural Sciences)** The Mathematical Research Data Initiative focuses on establishing standards for certified Mathematical Research Data, designing confirmable workflows, and providing community services. Its main goal is to implement FAIR principles across mathematics and its applications, ensuring interoperability of data handling and driving new research. The initiative revolves around four pillars: computer algebra, scientific computing, statistics and machine learning, and interdisciplinary mathematical research, targeting certified data, software developments, and service provision.
	NFDI4DataScience (Engineering Sciences) Considering the growing significance of Data Science and AI across various scientific domains, the consortium will advance FAIR and open data infrastructures, encompassing code, models, data, and publications through an integrated approach. The initiative aims to enhance transparency, reproducibility, and fairness in subject-specific projects by making digital artefacts accessible, interlinking them, and offering innovative tools and services.
	NFDI4Earth (Natural Sciences) The NFDI Consortium for Earth System Research meets the digital requirements of Earth System Sciences, which aim to comprehend Earth's functioning and interactions, addressing global change challenges. It is a community-driven initiative providing researchers with FAIR, coherent, and open access to relevant Earth System data, along with innovative RDM and data science methods.
	NFDI4Microbiota (Life Sciences) The aim of NFDI for Microbiota Research is to enhance the availability and quality of microbial research results by providing training and fostering interdisciplinary collaboration. This will be achieved through the establishment of a German microbial research network and the development of a cloud-based system for storing, integrating, and analysing microbial data, particularly omics data, across all life science fields. By focusing on education and services, NFDI4Microbiota aims to improve the efficiency and excellence of microbial research in Germany. Additionally, it will serve as a central resource and connecting hub for NFDI consortia working with microbiological data, including GHGA, NFDI4Biodiversity, NFDI4Agri, and others.
	NFDI-MatWerk (Engineering Sciences) The National Research Data Infrastructure for Materials Science and Engineering focusses on integrating decentralised data and metadata, experimental and numerical workflows, and a materials ontology to enhance interoperability and reproducibility of research data processing in its field. Data use profiles of Participant Projects are analysed to identify key scientific scenarios, informing the development and review of Infrastructure Use Cases for the continuous improvement of the infrastructure.
	PUNCH4NFDI (Natural Sciences) The consortium Particles, Universe, Nuclei and Hadrons for NFDI aims to develop solutions for the challenges of increasing data volumes and complexity. It embodies FAIR principles and will provide a science data platform comprising a data lake, cloud computing systems, data transformation layer, and user-friendly interface for the PUNCH community and the broader NFDI.
	Text+ (Humanities and Social Sciences) The Language and Text-Based Research Data Infrastructure seeks to establish a data infrastructure for Humanities disciplines and beyond, particularly focusing on language and text research. It will be designed to be flexible, scalable, and adaptable to various discipline-specific needs. By providing easy access to high-quality research data, Text+ aims to facilitate methodological diversity, which is essential for fostering innovative and transdisciplinary research. Text+ primarily emphasises Collections, Lexical Resources, and Editions, areas with well-established research traditions and mature methodological paradigms. These domains require distinct yet cross-disciplinary practices for data generation, curation, and management.

	Consortium (Prior Scientific Disciplines Covered) and Description
Third Selection Round 2022	**NFDI4Memory (Humanities and Social Sciences)** The Consortium for Historically Working Humanities aims to advance historical research by providing RDM guidelines and digital services for a diverse research community undergoing digital transformation. Key objectives include identifying innovations, evaluating their value, fostering development, and promoting wider adoption. Additionally, the consortium seeks to bridge methodological gaps by integrating digital and analogue methodologies. Ultimately, NFDI4Memory will establish a portfolio of key services, creating a data infrastructure that systematically links scholarship with digital resources.
	NFDI4Objects (Humanities and Social Sciences) To empower the Research Data Infrastructure for the Material Legacies of Human History Community and enhance its contributions to the entire NFDI initiative, nine key objectives have been identified: (1) providing reliable and interoperable data services; (2) raising awareness of data quality; (3) improving research data quality; (4) implementing, aligning, and expanding standards; (5) promoting professionalism and qualification; (6) strengthening cross-cutting topics across NFDI; (7) integrating NFDI4Objects into the national and international landscape; (8) ensuring collaborative governance and sustainability; (9) increasing diversity among consortium participants.
	NFDI4BIOIMAGE (Life Sciences) The National Research Data Infrastructure for Microscopy and Image Analysis aims to facilitate the sharing and re-use of bioimaging data across disciplines, fostering new discoveries. A key aspect is the creation of a standardized digital object bundling images with metadata and promoting interoperability. With expertise spanning plant biology to neuroscience, NFDI4BIOIMAGE advocates for standardization while ensuring compatibility with various data types and RDM systems. To unlock the full potential of bioimage data, the consortium will provide access to cutting-edge analysis tools, including AI-based methods in scalable cloud environments. NFDI4BIOIMAGE collaborates with GHGA and DataPLANT, contributing to the definition of FAIR data objects.
	NFDI4Energy (Engineering Sciences) The National Research Data Infrastructure for Interdisciplinary Energy Systems Research focuses on addressing research challenges in transitioning energy systems to net zero greenhouse gas emissions. It emphasises the digitalisation of energy systems, particularly cyber-physical energy systems, which impact technical, social, and societal aspects. The consortium aims to provide services for FAIR data, models, and processes, enabling traceability, reproducibility, and transparency of results. It promotes societal involvement, collaboration between research institutes and businesses, and simplifies model integration. Key services include competency guidance, best practices, data and software registry, simulation coupling, and transparency. The goal is to create an open and FAIR research ecosystem for energy systems research, accommodating various research stages and components.
	NFDI4Immuno (Life Sciences) The National Research Data Infrastructure for Immunology aims to enhance understanding of the immune system by integrating data from diverse experimental technologies. It focuses on harmonising data representations and metadata standards, enabling consistent annotation of datasets, promoting FAIR practices, operating a network of repositories, and developing tools for standardised data analysis. Collaboration with national and international partners is central to its objectives, including close ties with immunological societies and other NFDI consortia.

	Consortium (Prior Scientific Disciplines Covered) and Description
Third Selection Round 2022	**FAIRagro (Life Sciences)** The FAIR Data Infrastructure for Agrosystems Research addresses challenges in agriculture by focusing on integrated research and data approaches. It aims to establish a well-organised domain for agrosystem research, enabling FAIR and quality-assured RDM and providing innovative RDM services. Six use cases target key research areas such as crop breeding and digital farming, piloting FAIRification of agrosystem data. The initiative will establish a central portal, connect repositories, facilitate data analysis, and provide support services. It also addresses quality and legal security challenges, collaborating closely with other NFDI consortia and the agrosystem community.
	NFDIxCS (Engineering Sciences) The National Research Data Infrastructure for and with Computer Science (CS) consortium strives to establish an infrastructure adhering to FAIR principles for storing and managing complex CS data and software. The consortium seeks to modernise publication processes, promote interdisciplinary collaboration, and share CS knowledge. Through collaboration with various partners, it aims to create an open environment supporting CS researchers' data needs while fostering cultural change and responsible data handling practices.
Basic Services	**Base4NFDI (Union of Consortia)** Base4NFDI is a joint initiative of all NFDI consortia and develops basic services for NFDI. It will create the basis for better FAIRness of research data in a cross-disciplinary perspective. For this purpose, common, technical services will be developed together with experts for the data in the different research disciplines. Base4NFDI supports common solutions to avoid parallel developments. Already existing services are thereby adapted or extended to be usable for researchers from other disciplines.

TABLE 2: CURRENT BASE4NFDI PROJECTS

Table 2 refers to section 2.1.5 in this publication (pp. 11 – 12).

The descriptions were taken from https://base4nfdi.de/projects.

	Basic Service and Description
Initialisation Phase	**PID4NFDI** Persistent Identifier Services for the German National Research Data Infrastructure will design a work programme to build an NFDI foundation service on established PID infrastructures.
	TS4NFDI Terminology Services for NFDI is a cross-domain service for the provision, curation, development, harmonisation and mapping of terminologies.
	DMP4NFDI NFDI Basic Service for Data Management Plans is a centralised Basic Service for managing data management plans and software management plans across NFDI.
	Jupyter4NFDI Jupyter4NFDI addresses the fragmented deployment of Jupyter Notebooks across NFDI consortia by offering a centralised service. This service intends to simplify access, improve user experience, and extend Jupyter's reach to a wider audience within and beyond NFDI.
	KGI4NFDI Knowledge Graph Infrastructure for the German National Research Data Infrastructure advocates for a central and reusable Knowledge Graph (KG) Infrastructure to enhance interoperability within the research domain and aims to provide essential components including a KG registry and a service for accessing KGs across NFDI projects.
Integration Phase	**IAM4NFDI** Identity and Access Management for the German Research Data Infrastructure is concerned with connecting and expanding existing and emerging Identity and Access Management Systems in a way that researchers from different domains and institutions are able to access digital resources within NFDI as easily as possible, including access to and exchange with external infrastructures and resources like EOSC.

TABLE 3: GAIA-X LIGHTHOUSE PROJECTS

Table 3 refers to section 2.3.6 in this publication (p. 32).

The project descriptions are taken from: https://gaia-x.eu/who-we-are/lighthouse-projects/

Remark: Exceeding Gaia-X Lighthouse Projects, the Lighthouse Data Spaces represent an exemplary level of alignment with Gaia-X principles and standards, signifying a distinguished commitment to advancing digital sovereignty and European value-creation. See https://gaia-x.eu/who-we-are/lighthouse-data-spaces/. The Lighthouse Data Spaces appointed by Gaia-X must not be confused with and are not directly related to the European Data Spaces. The latter are currently being developed in 14 domains in which different projects are involved. For an overview, see section 2.4 and https://digital-strategy.ec.europa.eu/en/policies/data-spaces.

	Project, Domains and Description
Lighthouse Data Spaces	**AGDATAHUB (Agriculture)** AGDATAHUB supports the agricultural sector by providing tools to ensure data sovereignty for farmers and promotes sustainable, innovative agriculture based on the pillars of innovation, community, European sovereignty, and fair governance. The goal is to strengthen the European agricultural sector through innovative data systems and advisory tools, providing farmers with easy, secure access to their data.
	Catena-X (Mobility) Catena-X Automotive Network is developing a collaborative, decentral organised data and services ecosystem along the entire automotive value chain that enables companies – from manufacturers and SME suppliers to recyclers – to benefit from the advantages of data-based business while being protected under European law with regard to data sovereignty and data protection. The aim is to increase the efficiency of both industry-specific collaboration and internal company processes, while at the same time meeting legal requirements.
Lighthouse Projects	**ACCURATE (Manufacturing)** The ACCURATE project seeks to enhance the competitiveness of European manufacturing companies and value chains by improving sustainability, performance stability, resilience, and management of unexpected events. It will develop innovative solutions and validate them through three key use cases in important European industries, including aerospace, automotive, biotech, energy, medical, and mobility sectors.
	Boot-X (Infrastructure & Technology) Boot-X is a component of Huawei EDS (Exchange Data Space) focused on enabling cross-border data exchange and ensuring compliance with Gaia-X and International Data Spaces Association standards, particularly for data exchange between Chinese and European industries. The Boot-X Connector is compatible with the Eclipse Data Space connector and includes additional features such as local data usage policy management, self-sovereign identity federation, and compliance monitoring.
	COOPERANTS (Aeronautics & Space) The project Collaborative Processes and Services for Aeronautics and Space advances the digitisation process in the aerospace industry, aiming to implement more efficient working methods and production processes across the entire lifecycle of aerospace vehicles in the future, while also strengthening the competitiveness of the industry in Germany and Europe. The coordinating organisation is the German Aerospace Center – Institute of Space Systems. The project is funded by the German Ministry for Economic Affairs and Climate Action.

	Project, Domains and Description
Lighthouse Projects	**Data4Industry-X (Manufacturing, Industry 4.0)** Data4Industry-X is a secure and compliant data exchange platform to enable using distributed data from various factories, organisations, or countries. It addresses the challenge of data exchange in decentralised manufacturing, aiming to improve efficiency, productivity, and competitiveness while reducing carbon footprints. As part of the international Manufacturing-X initiative, Data4Industry-X follows Gaia-X standards and complies with European data regulations like the Data Act. It supports Europe's digital transformation and industrial sovereignty by enhancing cross-border and cross-industry data exchanges to build a sustainable industrial ecosystem.
	ELINOR-X (Smart Cities & Urban Data) ELINOR-X aims to pioneer a novel approach to data collaboration, facilitating seamless and targeted sharing between the public and private sectors. Through individually tailored and legally binding data sharing agreements, data will be routed directly from providers to consumers without centralised storage. ELINOR-X lays the groundwork for future digital twins, empowering government and public service providers to address critical challenges through data-driven decision-making.
	Energy Data-X (Energy) Energy Data-X aims to establish a common data space for the German energy industry within Gaia-X. Data will stay decentralised with the owners but can be shared and exchanged securely according to compliance rules. This initiative is expected to enhance energy industry processes and foster innovative, data-driven applications and business models, significantly contributing to German and European energy and climate goals.
	EONA-X (Mobility, Transport & Tourism) EONA-X endeavours to establish a trusted environment for accessing data sets and promoting applications in mobility, transportation, and tourism. Key emphasis lies in enhancing multimodal trip optimisation to support the EU's zero emission goals outlined in the Mobility Strategy. The initiative has developed a robust demonstrator utilising real-time data sourced from operational systems in the airline and rail sectors.
	EuProGigant (Manufacturing, Industry 4.0) European Production Giganet, is a project focused on calamity-avoiding self-orchestration of value chain and learning ecosystems within manufacturing. The project aims to address core issues regarding the intelligent and sovereign utilisation of data in manufacturing. Its objectives include implementing machine connection and machine-oriented data processing through Gaia-X-compliant edge architecture to bolster resilience in the value-creation ecosystem.
	EuroDaT (Finance) The goal of the European Data Trustee (EuroDaT) is to establish a neutral data trustee in accordance with the EU's Data Governance Act and to promptly implement Gaia-X-compatible exemplary use cases. EuroDaT is designed to enable companies, academia and public authorities to merge data quickly, securely, legally compliant and largely automatically and make them accessible for joint analysis. In this way, data that cannot be shared for reasons of data protection or antitrust law, for example, can be made accessible for consolidated analyses by the data providers or third parties. The project is funded by the German Ministry for Economic Affairs and Climate Action.
	GAIA-X4Future Mobility (Mobility, Transport) The GAIA-X4 Future Mobility project is situated within the mobility sector of the German Gaia-X Hub. Comprising six projects, its primary objective is the Gaia-X-based integration of future mobility applications. Given the product-centric nature of these applications, effective data-driven collaboration with manufacturers, suppliers, service providers, and users is a top priority.

	Project, Domains and Description
Lighthouse Projects	**Health-X dataLOFT (Health)** Health-X dataLOFT places citizens at the centre of providing, accessing, and controlling their own health data. The goal is to develop transparent cloud-based applications in accordance with Gaia-X standards in highly relevant areas of healthcare such as health prevention, healthy ageing, and clinical care. Concepts from the Medical Informatics Initiative, as well as legally binding Gematik standards and solutions from the telematics infrastructure, are integrated for networking healthcare sectors and enabling integrative data utilisation. The coordinating organisation is the Charité Universitätsmedizin Berlin. The project is funded by the German Ministry for Economic Affairs and Climate Action.
	TEAM-X (Health) The Trusted Ecosystem of Applied Medical Data eXchange project aims to create a secure digital data ecosystem based on the Gaia-X infrastructure, facilitating data-driven healthcare innovations. It seeks to overcome sectoral boundaries in healthcare, enhancing prevention, prediction, personalisation, and patient participation. The project will make currently inaccessible health data available while ensuring patients have full control over who can access their encrypted data. TEAM-X works closely together with Health-X dataLOFT to use synergies. The project is funded by the German Ministry for Economic Affairs and Climate Action.
	Mobility Data Space (Mobility) The Mobility Data Space is a project aimed at shaping the future of the mobility sector. It encompasses a wide range of stakeholders, including vehicle manufacturers, ride-share services, public transport operators, navigation software companies, research institutes, bike-sharing companies, and others. One of the project's primary objectives is to promote Gaia-X-compliant data exchange, fostering competition in innovative, environmentally sustainable, and user-friendly mobility solutions. The project is funded by the Federal Ministry of Digital Affairs and Transport (BMDV).
	Omega-X (Energy) Omega-X is an Energy Data Space designed to enhance the European data economy. It encompasses a federated infrastructure, a data marketplace, and a service marketplace, facilitating data sharing among diverse stakeholders and showcasing its benefits through specific energy use cases. Importantly, it ensures scalability and interoperability with other data space initiatives.
	Prometheus-X (Education & Skills) Prometheus-X aims to create a cloud-to-edge infrastructure for education and skills, connecting services and data in an accessible data space ecosystem. The project is working on six use cases and developing 21 building blocks, including consent management, skills interoperability, decentralised AI training, and a crowd-tagging system.
	SCSN (Electronics Supply Chain) The Smart Connected Supplier Network (SCSN) fosters data sharing within manufacturing environments. This facilitates rapid, secure, and efficient data exchange within the high-tech supply chain. As part of the project, the team will support the development of new application clusters within SCSN and expand its scope to encompass various industries.
	Structura-X (Cloud Services) Structura-X is a project focused on advancing European cloud infrastructure. Its aim is to empower current Cloud Service and Infrastructure Providers (CSPs) to achieve Gaia-X certification for their data and infrastructure services. The project seeks to establish an ecosystem of independent CSPs, coordinated by a shared layer of federation certification and labelling services, leveraging Distributed Ledger Technology. Success will be evaluated through a series of Minimal Viable Products, which will be fully portable and interoperable across various CSPs, certified by either Gaia-X or a designated certification authority.

TABLE 4: COMPARISON OF THE FEDERATED DATA INFRASTRUCTURES AND THE EUROPEAN DATA SPACES FRAMEWORK

Table 4 refers to section 3 in this publication (p. 45).

	History	Objectives
NFDI	RfII recommends the founding of NFDI in 2016. Federal Government-Länder agreement by the Joint Science Conference (GWK) of 26 November 2018. Foundation of NFDI e. V. in 2020/21. Accompanied by the DFG and its expert committee on the NFDI.	Systematically unlocking data resources for science and research. Establishing standards in RDM. Developing services for RDM. Facilitating sustainable storage and continued utilisation of research data. Alignment with the FAIR principles.
EOSC	First mentioned in an EU Commission communication in 2015. Start of the 1st phase on 23 November 2018, pilot phase 2017-12/2020. Foundation of the EOSC Association (AISBL) in summer 2020. Memorandum of Understanding with the European Commission in early 2021, formalising the co-programmed partnership (project funding via Horizon Europe).	Developing a web of FAIR data and services for science: - Data infrastructure network with access to data, tools and services for science (and industry), not consistently 'open'. - As open as possible and as restricted as necessary. Alignement with the FAIR principles. Council of the European Union defined EOSC as one of 20 pilot measures to deepen and strengthen the new ERA.
Gaia-X	Founding starts with an initiative of the German and French Governments in October 2019. Foundation of the Gaia-X AISBL in January 2021. Brings together a broad range of organisations (large companies and SMEs, developers and users of technology, industrial players, and members of academia).	Enabling the creation of common data spaces based on European values. Full alignment with the objectives of the EU's data strategy. Digital sovereignty, cloud and data standards to guarantee safe and trustworthy data exchange. Network of independent actors in decentralised and interoperable systems, based on transparency and security.
European Data Spaces	Key part of the EU data strategy of 2020. Associated legislative acts on the EU level (e.g. Data Governance Act in 2022, Data Act in 2024). Most advanced is currently the development of the European Health Data Space (Political agreement of EU Council and Parliament on the EC's proposal for the EHDS in spring 2024).	Implementing a 'single market for data' through sector-specific data spaces, realised through a federated architecture. Full alignment with European values of privacy, data protection and security.

	Governance	Financing	Stakeholders / Participants
NFDI	**NFDI e. V.:** Association bodies: - Scientific Senate - NFDI Consortia Assembly - NFDI-Directorate - General Assembly - Board of Trustees - (Sections) Directorate and office co-ordinate the Association. High involvement of the scientific community via the different boards.	2019-2028 by the Federal (90 %) and Länder (10 %) governments, up to EUR 90 million annually. No membership fees for stakeholders of the NFDI e. V.	26 NFDI consortia and one network of consortia (Base4NFDI). 200 (co)applicant organisations.
EOSC	**EOSC Association:** President and Board of Directors. Involved in tripartite governance with EOSC Steering Board (EU-MS) and EU Commission. Association members. Advisory Groups with Task Forces that deal with key issues of the implementation, participation of the scientific community.	2017-2020: EUR 350 million (Horizon 2020). 2021-2027: Up to EUR 1 billion, of which half from the European Commission (Horizon Europe) and half from Member States (via in-kind contributions). Membership fees for participating in the Association.	250 members of the EOSC Association (85 of which are observers, 28 mandated organisations). Advisory Groups with Task Forces that deal with key issues of the implementation.
Gaia-X	**Gaia-X Association (AISBL):** Board of Directors, Management Board, Advisory Boards and Committees. National Gaia-X Hubs. National and central contact points. Involvement of the open-source community. Lighthouse projects as Use Cases.	Funding from Germany (BMWK) until 2024: EUR 117.4 million and from France until 2025: EUR 150 million. Annual contributions of the members: commercial organisations EUR 5-75 000, non-commercial organisations EUR 2 500.	Umbrella organisations like: - IDS(A) - BDVA/DAIRO - eco – Association of the Internet Industry, DE-CIX Research organisations like: - DFKI, DLR, Charité, Fraunhofer
European Data Spaces	European Data Innovation Board to guide and advise sector-specific governance arrangements. Support through DSSC and SIMPL.	Blend of EU funding programmes, most notably Digital Europe, Horizon Europe, Connecting Europe Facility and partially the Recovery and Resilience Facility (no exact figures available).	European institutions and EU Member States, Public authorities on the local, regional and national level, universities and research organisations, Civil society organisations, technology providers, standardisation bodies, other specific bodies.

	Technical Structure and Planned Services	Cooperation
NFDI	Scientific-methodical- and generic services: Technical and organisational solutions for research communities. NFDI-wide basic services for ensuring interoperability, e.g. Identity & Access Management, PID Services, Terminology Services.	NFDI Association is the mandated organisation of Germany in the EOSC Association. Via consortia, thematic sections and the consortium network for basic services (Base4NFDI). With Gaia-X, currently project-based: 'FAIR Data Spaces'.
EOSC	EOSC EU Node (Core, Exchange, FAIR Data Federation). EOSC-CORE (until 2025: central services and basic functions such as AAI, PID, metadata standards, and catalogues). Exchange infrastructure services and exchange application services. EOSC Portal Catalogue and Marketplace: reference system for repositories and research infrastructures.	With members of the EOSC Association via funding of the EU Framework Programmes. With networks and associations such as CESAER, GÉANT, OpenAIRE, PRACE and RDA. With ESFRI projects and other EU-funded projects and research infrastructures.
Gaia-X	Federation Services (AAI, catalogues, smart contracts). Data Spaces with customised standards for user communities (private and public). Federated Cloud Services Catalogue containing over 500 cloud services meeting Gaia-X requirements.	With large enterprises and SMEs in domains and national hubs, e.g. Atos, SAP, Software AG, etc. With research institutions in several projects.
European Data Spaces	- Federated Architecture - Interoperability Frameworks - Data Governance Mechanisms - Secure Data Sharing infrastructure - Cloud and Edge Computing	Envisaged: cross-sector collaboration, involvement of public-private partnerships, engagement with academia, engagement with standardisation efforts.

Conference Report

SHAPING DATA SPACES IN GERMANY AND EUROPE – THE IMPULSE OF SCIENCE

Herrenhausen Conference of the Council for Scientific Information Infrastructures (RfII) and the Volkswagen Foundation on 24th-25th April 2023, in Hanover, Germany.

In April 2023, the RfII, supported by the Volkswagen Foundation, hosted a conference at Herrenhausen Palace in Hanover titled 'Shaping Data Spaces in Germany and Europe – The Impulse of Science'.

More than 220 participants from various scientific disciplines and institutions, as well as representatives from science policy, research funding, and infrastructures, came together to discuss current challenges in the development of research data infrastructures. The conference successfully continued a format that had garnered significant attention with the first RfII-conference on 'The Data Quality Challenge' in February 2020 at the same venue – the Xplanatorium in Herrenhausen Palace, ideally suited for such purposes.

The aim of 'Shaping Data Spaces in Germany and Europe – The Impulse of Science' was to stimulate discourse on the present and future of existing and emerging research data infrastructures in Germany and Europe and to enhance the exchange among various involved and enabling actors in the scientific system. To prepare for the conference, the RfII had published a working paper in March 2023. It outlined the genesis and current status of the three most significant initiatives – the German National Research Data Infrastructure (NFDI), the European Open Science Cloud (EOSC), and Gaia-X – and provided initial suggestions for further steps towards roll-out, integration and goal archievment of the respective infrastructures to the various stakeholders and user communities in the scientific system.

In addition to science policy keynotes by Martina Brockmeier, President of the Leibniz Association and former Chair of the German Science and Humanities Council, and Karel Luyben, President of the EOSC Association and Rector Magnificus Emeritus of Delft University of Technology, the programme included panel discussions, keynote presentations, and workshops on the goals and challenges of data infrastructures for research, both in general and from various disciplinary and institutional perspectives. The panel sessions gave opportunities for questions and discussions while in the workshops proporsitions for the conditions of success for the future design of data infrastructures were debated. The exchange was further facilitated through information and poster stands in the corridors of the conference venue. The individual programme segments were moderated by Lars Bernard, Georg Schütte, Stefanie Speidel, and Jan-Martin Wiarda.

Welcome Addresses

The conference was opened with welcome addresses from representatives of science policy by **Jens Brandenburg**, Parliamentary State Secretary to the Federal Minister of Education and Research (BMBF), **Joachim Schachtner**, State Secretary to the Lower Saxony Minister of Science and Culture, **Georg Schütte**, Secretary General of the Volkswagen Foundation, and **Petra Gehring**, Chair of the German Council for Scientific Information Infrastructures (RfII). All welcome addresses highlighted challenges associated with digital transformation, but also questions of societal and ecological developments for which data based empirical evidence is expected from higher education and research institutions. A qualitatively outstanding database was emphasised as crucial for addressing these tasks. Access to and use of data, especially – but not exclusively – in scientific research, must be facilitated, necessitating the long-term availability of infrastructures for generating, archiving, quality assessment, and secure transfer of data. The entire scientific system is called for actively accompanying the development of such infrastructures in the spirit of a transformation process, addressing new questions for the conception and organisation of research in general.

Petra Gehring stressed that the RfII sees active use and participation in the development of data infrastructures as an important dimension of scientific practice. In this context, the conference aimed to provide another impulse for the necessary discourse to accompany the required 'cultural change' in science, ultimately contributing to the methodological and theoretical advancement in numerous scientific disciplines and research fields.

Keynotes

The keynotes by **Martina Brockmeier** and Karel Luyben highlighted the significance of a 'cultural change' driven by comprehensive societal digitisation – and the role of higher education and research within this transformation process.

According to Brockmeier, a cultural change in science and humanities is urgently needed, particularly when it comes to criteria and procedures used to evaluate scientific quality and achievements. Research data infrastructures and data management play a central role in this complex – they enable data driven research performance and quality assurance and therefore must be considered in quality assessments. However, they should not be viewed in isolation from but rather in a strong relationship to genuine research processes and other research-enabling infrastructures. Brockmeier highlighted the necessity of long-term financial support, beyond project funding, for the permanent safeguarding of research data, especially for long-term archiving. Legal certainty is also crucial as an enabling condition for sharing and accessing research data. As another aspect of cultural change, Martina Brockmeier mentioned the training and recruitment of personnel needed to develop and implement high-quality data infrastructures. She sees adequate remuneration of highly qualified scientific and science-supporting personnel as indispensable to make higher education and research institutions more attractive as employers in competition with the private sector. This would help to retain at least a considerable quantity of young talents in the scientific system.

In his keynote, **Karel Luyben** also emphasised the necessity of a cultural change in higher education and research for his vision of establishing a 'Web of Scientific Insights', a vision significantly advanced by the construction of the EOSC. To explain this kind of web, Luyben outlined the ideal image of a digital network of scientific insights based on the FAIR principles, enabling the integration of existing and future data sources, aligning with European data requirements, and providing an open offering of scientific information and services for non-scientific actors as well.

Luyben particularly highlighted the opportunities that a well-understood EOSC could offer to European universities to evolve into international 'champions'. EOSC could also provide a stable infrastructure for the public and private sectors and contribute to addressing societal challenges in the future. Success criteria on the path to these goals – especially for universities – include networking in the scientific system (e.g. through European university networks like the European University Association), establishing a new pool of scientific and science-related employees such as 'Data Stewards', and the pan-European establishment of digital or data competence centres. He concluded his keynote with another vision: by 2040, at least 50 per cent of globally relevant research data (data, publications, software) should be available according to FAIR principles, as a visible achievement of a successfully implemented and utilized EOSC.

Introductory Panel Discussion

The first panel discussion moderated by Georg Schütte – featuring **Heide Ahrens**, Secretary General of the German Research Foundation (DFG), **Ursula Staudinger**, Rector of TU Dresden, **York Sure-Vetter**, Director of the NFDI e. V., and **Klaus Tochtermann**, Director of the EOSC Association and Director of the Leibniz Information Centre for Economics – shed initial light on the goals and challenges of different data infrastructures for research. The participants examined the current state from their respective institutional perspectives. It was collectively acknowledged at the beginning that Germany had laid a good foundation for future data-intensive research with the establishment of the NFDI. Now, the task is to further develop and expand NFDI according to research demand, especially to encompass the entirety of the scientific system and its disciplines.




The participants in the discussion, particularly Heide Ahrens, emphasised the importance of a strong and widely visible commitment from research policy actors indicating long-term funding availability for NFDI. This would be crucial for fostering trust within the scientific community. The significance of a long-term perspective is further driven by the growing pressure on scientific organisations to utilize commercial providers for specific data infrastructures and services, leading to the potential risk of new dependencies and silos. Consequently, there is a pressing need for prompt action from both the scientific community and policymakers.

Ursula Staudinger underscored that research funders also have a responsibility to strengthen incentive structures for supporting digital transformation. Positive changes can often be effectively advanced through initiatives from relevant institutions. The European and, prospectively, international integration of different data infrastructures and their linkage across various levels present a significant challenge.

Klaus Tochtermann clarified that the involvement of German actors and individuals in the EOSC process is positively regarded. He advocated for continued close coordination with European and international partners. However, he pointed out that EOSC is under pressure to implement scientifically adequate services and infrastructures for the broader scientific community promptly, especially due to the competition from commercial entities. An early sign from research policy regarding the long-term funding of NFDI is also deemed necessary by York Sure-Vetter to retain professionals in the 'Data Literacy' domain of research – a domain currently being developed in the NFDI process.

According to Sure-Vetter, delaying decisions on the sustainability of NFDI would be equivalent to a showstopper and send a detrimental sign to the entire scientific system, especially concerning the international perception of Germany as a hot spot for science and innovation. All panel participants unanimously agreed that recruiting, training, and

retaining professionals pose a demanding task for all scientific institutions, given the overall societal situation of a shortage of IT skills and the often not attractive compensation and career structures posed by collective labour agreements in the public sector.

The scientific system has demonstrated high flexibility with the establishment of NFDI and certain initiatives within: e.g. Base4NFDI, a joint effort of all consortia to build basic services. While this flexibility can be further leveraged, encouraging signs of sustained support from research policy are urgently needed to maintain the momentum.

In concluding the discussion, Georg Schütte emphasised the dynamic that has emerged on individual and institutional levels in this field over the past years. Despite foreseeable constraints in state and societal resources, the financial support for long-term tasks in building and expanding scientific data infrastructures should not diminish. He called for a joint strategy for digital transformation at the interfaces of scientific organisations and universities, federal and state governments, European institutions, and economic and civil society actors. This strategy should secure future investments in NFDI and EOSC. Schütte highlighted the invaluable contribution of EOSC to achieving common standards in the provision and use of data within the framework of shared values across Europe, extending far beyond the scientific system.

The introductory panel discussion illustrated the current status of building various data infrastructures for research in Germany and Europe, revealing the imminent tasks. A key challenge lies in the necessity of an early commitment from research policymakers regarding further funding for research data infrastructures, especially considering the reliable development of personnel under highly competitive conditions. The urgent need for further intensification of international collaboration, with suitable points of connection between infrastructure initiatives, was emphasised by all participants in the discussion, including considerations related to security issues, and the expansion of technical interoperability toward convergent standards for basic services.

Evening Lecture

In the Evening Lecture, the Egyptologist **Antonio Loprieno**, President of All European Academies (ALLEA) and former President of the Austrian Science Council took conference participants on a journey through the history of academic ‘drift’ in the wake of media upheavals in the past.

Loprieno's lecture focused on the role that universities and scientific academies as institutions have played in such processes of change in the past and what can be learned for the future. Contrary to the common notion that universities are at the forefront of digital transformation, Loprieno argued that neither today nor in the past have the established scientific institutions acted as drivers of media upheavals and the associated social changes. Instead, they have been subject to a certain ‘innovative inertia’, in which the academic elite defends established disciplinary bastions against external pressures for modernisation. For his own field, archaeology and the historical sciences, Loprieno illustrated this tendency by pointing out how technology-driven methodological innovations are initially segregated into

the realm of so-called 'historical auxiliary sciences'. Similarly, he critically examined the establishment of 'Digital Humanities' alongside the disciplinarily established humanities and strongly advocated for the seamless integration of digital techniques and methods into the basic canon of each discipline. Accordingly, the demonstration of a natural handling of digital tools, methods, and infrastructures and the ability to convey such skills in research and teaching should be necessary requirements for the appointment of professors today. Towards the end of his lecture, Loprieno surprised the audience with the suggestion to organise forgetting along with the archiving of more and more data in the course of digital transformation: the FAIR principles should be supplemented with a D for 'Disposable' if research wants to stay capable of acting in the sea of digital data.

Session 1: Perspectives on Research Practices

The contributions of the first session, moderated by **Stefanie Speidel**, Director at the National Center for Tumor Diseases / UCC Dresden, focused on the digital challenges of research practices and whether data infrastructures offer suitable solutions for their management and development. **Cordula Artelt**, Director of the Leibniz Institute for Educational Trajectories, reflected on the current state and potential requirements for improvement, especially in research data-related policies, from the perspective of empirical educational and social research. According to her, surveys, such as those conducted within the National Educational Panel Study (NEPS), face various challenges, especially regarding costs, the willingness of households and institutions to participate, and difficulties in collecting contextual data. Increased availability and linkage of data of various types from different public sources, as well as harmonisation of legal frameworks for data access, would be desirable. She highlighted that the differing access regimes of the 16 Länder pose a significant hurdle for comparative educational research, particularly in the area of researching school education and its conditions. Complex negotiation processes on data access issues slow down the research process and have consequences for short-term delivery of evidence-based policy advice.




For **Ursula Klingmüller**, a systems biologist at the German Cancer Research Centre and a member of the German Ethics Council, adapting data management systems to the requirements and working conditions of research institutions is a fundamentally important step, crucial for the quality of research data. Using a life sciences research project as an example, she emphasised that significant time resources are consumed before negotiation processes on data access and modalities of data linkage are completed, allowing the actual research to begin – in the specific case example, two years. In her field, interdisciplinary cooperation with medical informatics and data scientists is particularly necessary – an aspect that should be given more consideration in corresponding funding opportunities. Similar to Martina Brockmeier, Klingmüller advocated

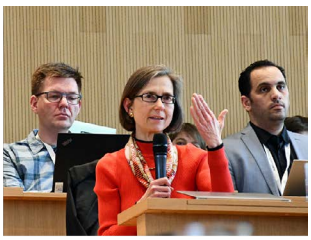
for revising evaluation systems within research and strengthening scientific reputation opportunities, for instance, for creating outstanding datasets or developing analysis tools. Ursula Staudinger added that the establishment of data infrastructures has a significant epistemic dimension: these new infrastructures contribute to enabling new forms of trans-disciplinarity and, consequently, mutual entanglement between scientific disciplines and research fields.

Eva Barlösius addressed the autonomy of research in the context of data infrastructures. She explored what happens to research data from the perspective of researchers when they are integrated into a publicly accessible data infrastructure. Based on interviews, she demonstrated that the transfer of research data into outward data infrastructures resembles a process of decontextualisation, where data is detached from the immediate research process and the specific research questions of the individuals conducting the research. This has an essential, identity-forming dimension for researchers, as they perceive the collected data, not least, as a core element of 'their' research. The autonomy of research could be negatively affected if the idea of later transferring data into infrastructures already influences the collecting and processing of research data. The same holds true, if research data could only be archived and shared through large data infrastructures. According to Barlösius, the established form of 'Peer to Peer Data Sharing' within the scientific community, which is less formalised and more prevalent in institute contexts and working groups, should continue to be possible complementarily to data sharing through large infrastructures.

The presenters in Session 1 were unanimous in their perception that there is still insufficient cooperation in the academic system when it comes to setting up services. Many institutions are still developing their own solutions or using repositories they have set up themselves, which are often not found by researchers from outside the institution. The development of common goods is a better approach, as long as it is made transparent who can access the data archived there, under what conditions and for what purposes.

Session 2: Perspectives of Research Institutions

The contributions of the second session focused on the organizational perspective of universities and non-university research institutions regarding the challenges of digital transformation and the institutional prerequisites necessary for successful data infrastructures. The keynote speeches, moderated by **Lars Bernard**, featured presentations by **Simone Fulda**, at this time President of Kiel University and Chair of the Commission for Research Information in Germany (KFiD); **Torsten Schrade**, Academy Professor for Digital Humanities at Mainz University and the Academy of Sciences and Literature; and **Chris Eberl**, Deputy Director at the Fraunhofer Institute for Mechanics of Materials. They highlighted that the digital transformation process and the associated development of data infrastructures are subjects of ongoing negotiation within the academic com-

munity. Success in this endeavour requires a collective effort from all scientific institutions and strategic collaboration, which may involve painful adjustments for some institutions when they have to open previously 'owned' resources and structures, making them interoperable and accessible for external and common use.

Simone Fulda advocated for establishing research information, i. e., data about research, as a complementary pillar alongside research data and research data management. Research information provides in-depth insights into the scientific system, its organisations, infrastructures, personnel, and the overall scientific process. It can enable new research questions, contribute to the development of institutional data strategies, and enhance the steering and strategic capabilities of universities. Fulda emphasised the responsibility of university leadership for data sovereignty and suggested guidelines and incentive systems, such as supporting FAIR principles, to improve research conditions, adapt to digital transformation, and enable researchers to make the best use of infrastructures. Early competency development and the introduction of new job profiles within universities were also highlighted.

Torsten Schrade focused on small and medium-sized research institutions and their role in building federated data infrastructures. He outlined the expectations of academies regarding infrastructures, emphasising long-term provision of research data, software, and results, as well as low-threshold use, trust-building, and certification. Small and medium-sized research institutions increasingly take on responsibilities, supporting various data infrastructures and coordinating transdisciplinary negotiation spaces. They act as 'brokers' between infrastructures, scientific communities, and the broader public, promoting institutional permeability of data infrastructures.

Chris Eberl expanded the previous perspective on scientific institutions to include the view of industry and the prevailing interests there, but also the parallel needs of industry and public science. He stated that science has a duty to clearly communicate its concerns and to demonstrate the added value of cooperative data sharing to industry in particular. Scientific institutions must ensure that they are not left behind in certain areas of digital change. In order not to fall behind the innovators in the economic system, more agile working methods should be taken into account – as well as the cost efficiency of one's own actions. These should include necessarily measures to avoid isolated solutions or the creation of duplicate structures. The interconnection of different data spaces should be a collective task involving public research, industry, society, and politics. While politicians need to ensure financial support, researchers should proactively design and standardise the connection between the data spaces, using principles such as FAIR but also considering quality requirements and sustainability.

In summary, Session 2 revealed a high degree of consensus among presenters on the diagnosis that institutional development strategies of universities and non-university research institutions in digital transformation extend beyond local or regional dimensions and also have to overcome the spirit of domain and location-oriented competition. The construction and expansion of data infrastructures are collective tasks aiming at a common good, transcending systemic boundaries and involving economic actors and other societal institutions in order to develop data-driven business models that generate innovation and prosperity in the long term.

Science Policy Panel Discussion

The second day of the conference began with a panel discussion moderated by Jan-Martin Wiarda, titled 'Sustainably Positioned Infrastructures as Success Conditions for Data-Intensive Science.' The panel participants, **Heide Ahrens**, **Rüdiger Eichel** (Head of Department 'Research, Innovation, Europe' at the Lower Saxony Ministry of Science and Culture), **Wolfgang Marquardt** (at this time Chair of the Board of Directors of Forschungszentrum Jülich), and **Simone Rehm** (Vice-Rector for Information Technology at the University of Stuttgart), agreed that the German scientific system needs better coordination of activities and more coherent funding to avoid fragmentation risks. Fragmentation was already observed at the level of federal and Länder ministries, where responsibilities for various aspects of digital transformation fall into different departments and sections, often lacking clear rationale.




Simone Rehm highlighted the success conditions for basic services. These services must meet scientific demands, be well-known, enable low-threshold use, ensure sustainability, and consider costs significantly. She emphasised the responsibility of university leadership for data sovereignty and suggested guidelines and incentive systems.

Heide Ahrens emphasised that while the DFG could initiate developments through project funding, it could not provide long-term funding. The institutions of the scientific system and research policy in federal and Länder governments are asked to design transitions between funding formats, allowing early exploration of long-term solutions for sustaining data infrastructures. She stressed the importance of persistent development of existing solutions and structures instead of introducing new services and structures continuously. Project funding alone cannot offer permanent solutions for incremental development.

Wolfgang Marquardt addressed research policymakers directly, arguing for a swift commitment on ensuring a durable construction and maintenance of NFDI. He emphasised the need for suitable business models for federated data infrastructures that have low entry barriers, i.e. no or only moderate access costs – but at the same time make clear that the service-provided requires a cost input that must be counter-financed. He pointed out that future costs, not yet considered in financial projections, would arise, especially in updating hardware (at least every 7 to 10 years) at universities and non-university research institutions.

Rüdiger Eichel advocated for reminding Federal Government and Länder about the commitment, to decide on further funding of NFDI by 2026 – as already outlined in the agreement between the federal state and the Länder from 2019. He also called for activating Länder-level research policies to engage more universities in NFDI and support already active organisations. The goal is to broaden the entire process. Also, effective evaluation mechanisms, such as assessing usage intensity and evidence of innovations resulting from public investments, need to be considered. Eichel emphasised that research policy, particularly when dealing with permanent, sustained investments, com-

petes with other well-justified claims from various departments for budget allocations. It should not be forgotten that political and parliamentary majorities must always be organised for the required long-term support services. He considered aligning the entire scientific system to the positive impulses of NFDI in terms of archiving, accessibility, and data quality as a way to overcome fragmentation and preserve the strengths of German scientific institutions, while at the same time maintaining its differentiated portfolio of organisations and fields of scientific knowledge production.

Participants in the conference plenum expressed additional regulatory ideas to boost the digital transformation of the scientific system. A tariff agreement specific for research and higher education personnel, redistributing financial resources within universities to cope with the high-cost pressures of managing local data infrastructures, was proposed. The effectiveness of the entire system and the currently (in most cases) spatially dispersed data infrastructures should be critically evaluated, considering actual impact and usage both within and outside the scientific community. Additionally, evaluation processes for publications, research projects, and structures, as well as appointment procedures, should be more aligned with the challenges of digital transformation and continuously integrate data issues. Wolfgang Marquardt also highlighted the entrepreneurial nature of operating scientific infrastructures, calling for political action to create solid frameworks.

Workshops

Considering diverse disciplinary and institutional challenges, conference participants split into four moderated workshops. The workshop moderators – Lars Bernard, Friederike Fless, (President of the German Archaeological Institute – DAI), Petra Gehring, and York Sure-Vetter – aimed to collaboratively develop theses on the further development of data infrastructures in research. The discussions focused on the following questions:

- Scientific Practice and FAIR Research Data Management: Who Drives the Change?
- How Do the Leading Institutions of the Scientific System Integrate Infrastructures for Research Data into their Performance Processes?
- How Can Supply and Demand for Research Data Management Services Be Reliably Balanced?
- Research and Teaching Plus X? Shaping Digital Transformation as a Permanent Task for Science Policy and Research Funding

The workshop moderators presented condensed theses on these questions during the closing podium session.

Closing Panel, Summary and Outlook

The conference highlighted the need for extensive coordination among different institutional, political, and scientific levels, as well as between various national, European, and international initiatives for data infrastructures. The concluding panel discussion, moderated by **Jan-Martin Wiarda**, brought together the workshop moderators for a final discussion on the developed theses and an overall conclusion.

Lars Bernard presented the workshop discussion on 'Scientific Practice and FAIR Research Data Management: Who Drives the Change?' He emphasised that the activation and recognition of individual commitment to using and improving data infrastructure offerings are still not sufficient. These aspects need more attention from professional societies, and NFDI has the task of mobilising scientific communities more effectively.

Petra Gehring highlighted the significance of federated data infrastructures for research in her workshop, 'How Do the Leading Institutions of the Scientific System Integrate Infrastructures for Research Data into their Performance Processes?' The connectivity of local structures to 'big' data infrastructures is crucial for the success of building a well-integrated scientific data ecosystem – and also for the attractiveness of the universities or research institutions embedded in this ecosystem. She also underscored the importance of speed in the digital transformation of research, considering the rapid pace of overall digital change.

York Sure-Vetter, focusing on 'How Can Supply and Demand for Research Data Management Services Be Reliably Balanced?' reiterated the characteristics that good services must fulfil from the perspective of researchers. These services should be well-known, meet existing needs better than commercial alternatives, be reliable and quality-assured, and be modularly expandable.

Friederike Fless addressed the research policy level with the results from the workshop 'Research and Teaching Plus X? Shaping Digital Transformation as a Permanent Task for Science Policy and Research Funding.' She emphasised that it is crucial for scientific institutions to spread information about the added value of data infrastructures to justify and secure the necessary investments in the eyes of politicians and the wider public. It is essential to clarify what science understands by the often-mentioned 'cultural change' and how it contributes to societal innovation and prosperity.

In the discussion with conference participants, additional aspects were discussed. The acceptance of federated data infrastructures under construction depends significantly on motivating individual researchers to contribute their data and use data collected by others, especially in humanities and social sciences. Specific incentives, coupled with expanded training opportunities and general competency development across disciplines, could be effective. The envisaged Research Data Act by the German government

was mentioned, aiming to overcome certain obstacles related to data access and linking data from different (especially public) sources. Bringing more transparency into the complex legal situation with regard to data sharing, including the development and implementation of data trustee models, as well as harmonising state (Länder), federal, and European law in the upcoming data spaces, were also topics on the wishlist for political action.

As a concluding remark, Petra Gehring stressed that NFDI has initiated a structural change in the entire scientific system within a very short time. This has been achieved in an unusual, cooperative way, with the GWK acting innovatively and boldly in favour of higher education and research. The establishment of NFDI is a fast process on the timescale, not only in comparison to the parallel development of EOSC. Despite the progress made, the fact that infrastructure issues are immediately relevant to research should be more deeply ingrained in the entire scientific system's consciousness. Active participation of public research in industry initiatives, especially Gaia-X and the different European Data Spaces, is a crucial future task. Publicly funded research can and must play a role in these fields, showcasing its innovation potential.

CONTRIBUTORS

MEMBERS OF THE COUNCIL

June 2024

Representatives of Scientific Users

Prof Dr Marion Albers
University of Hamburg – Faculty of Law

Prof Dr Stefan Decker
FIT – Fraunhofer-Institute for Applied Information Technology

Prof Dr Petra Gehring (Council Chair)
TU Darmstadt – Department of Philosophy

Prof Dr Kurt Kremer
MPI-P – Max-Planck-Institute for Polymer Research

Prof Dr Anne Lauber-Rönsberg
TU-Dresden – Chair of Civil Law, Intellectual Property,
Media and Data Protection Law

Prof Dr Wolfgang Marquardt
formerly Forschungszentrum Jülich

Prof Dr Stefanie Speidel
NCT/UCC – National Center for Tumor Diseases Dresden

Prof Dr Joachim Wambsganß
ZAH – Centre for Astronomy of Heidelberg University

Representatives of Information Infrastructure Facilities

Prof Dr Sören Auer
TIB – Leibniz Information Centre for Science and Technology and University Library

Prof Dr Lars Bernard (Deputy Chair)
TU Dresden

Prof Dr Barbara Helwing (Deputy Chair)
Vorderasiatisches Museum – Staatliche Museen zu Berlin / Preußischer Kulturbesitz

Prof Dr Beatrice Rammstedt
GESIS – Leibniz Institute for the Social Sciences

Prof Dr Gerhard Sagerer
Bielefeld University

Katrin Stump
SLUB – Saxon State and University Library Dresden

Prof Dr Ramin Yahyapour
GWDG – IT in science

N. N.

Representatives from the Federal and State (Länder) Governments

Rüdiger Eichel
Ministry of Science and Culture of Lower Saxony

Dr Christopher Karmann
Federal Ministry of Education and Research

Prof Dr Andrea Kienle
Ministry of Culture and Science of North Rhine-Westphalia

Marion Steinberger
Federal Ministry of Education and Research

Representatives of the Public

Dr Anke Beck
Frontiers

Dr h. c. Marit Hansen
State Data Protection Commissioner of Schleswig-Holstein

Christine Regitz
SAP SE

Dr Harald Schöning
Software AG

Developments in the European Context – Committee

Prof Dr Lars Bernard (head), Dr Anke Beck, Prof Dr Stefan Decker, Prof Dr Wolfgang Marquardt, Prof Dr Klaus Tochtermann (guest), Prof Dr Joachim Wambsganß, Prof Dr Ramin Yahyapour

Head Office

The committee received substantive and organisational support from the RfII Head Office from Daniel Zdun.

Editing of the Report

Dr Stefan Lange, Daniel Zdun

The RfII would like to thank all participants and speakers at the Herrenhausen Conference *Shaping Data Spaces in Germany and Europe – The Impulse of Science* (Datenräume in Deutschland und Europa gestalten – Impulse der Wissenschaft) for their valuable input.