The Development Model and Implications of Germany's Research Data Infrastructure (Postprint)

Authors: Zhou Lei, Yang Ping, Yuan Rubing

Date: 2023-04-01T16:16:03+00:00

Abstract

[Purpose/Significance] Research data infrastructure is an indispensable supporting infrastructure for the efficient management and utilization of research data and the efficient operation of digital research. [Method/Process] This study takes German research data infrastructure projects as the research object and analyzes the construction models of research data infrastructure from the perspectives of differences in task content between national and regional levels, institutional organization for construction, and relevant management mechanisms. [Results/Conclusion] The study finds that German research data infrastructure has different focuses at different levels: at the national level, it focuses primarily on data standards and tools; at the local level, it focuses primarily on services and training. In its construction, it adopts a strategy of extensive cooperation and complementary capabilities for institutional organization; in management, it features an emphasis on unified management and oversight. Finally, recommendations are offered for the construction of research data infrastructure in China.

Full Text

Preamble

Title: Construction Mode and Implications of German Research Data Infrastructure

Authors: Zhou Lei[1],[2], Yang Ping[1],[2], Yuan Rubing[1],[2]

Affiliations: [1]Beijing Institute of Science and Technology Information, Beijing 100044 [2]Beijing Science and Technology Strategic Decision Consulting Center, Beijing 100044

Abstract: *[Purpose/Significance]* Research data infrastructure constitutes an indispensable supporting foundation for the efficient management and

utilization of research data and the effective operation of digital research. *[Method/Process]* This study examines German research data infrastructure projects, analyzing their construction mode from the perspectives of national and regional task differentiation, organizational structures, and related management mechanisms. *[Result/Conclusion]* The findings reveal that German research data infrastructure exhibits distinct emphases across different levels: at the national level, the focus lies on data standards and tools, while at the regional level, services and training take priority. Construction adopts a strategy of extensive collaboration and complementary capabilities among institutions, while management emphasizes unified oversight and supervision. The paper concludes with recommendations for China's research data infrastructure development.

Keywords: research data; infrastructure; research data infrastructure; construction mode

Classification Numbers: G259.1; G311

DOI: 10.13266/j.issn.0252-3116.2020.21.017

Research data comprises data generated during scientific research, experimental development, and related activities, as well as original and derived data obtained through observation, monitoring, investigation, and other means for scientific research purposes [1]. With the advent of the "fourth paradigm" of scientific research, research data has become a crucial strategic resource driving technological development and social progress in the information age [2], making research data infrastructure construction a significant research proposition [3]. The scope of data infrastructure is broad; according to Wikipedia, any electronic facility that facilitates data sharing and consumption qualifies as data infrastructure. Consequently, any information service or tool required for effective research operation can be considered research data infrastructure. Currently, research data infrastructure construction exhibits characteristics of diversity, fragmentation, and constant change, with widely varying organizational and management approaches. The two primary construction modes are individual construction and alliance construction [9]. Individual construction, the most common approach, involves research institutions independently organizing and archiving data from their research projects. This is the predominant method adopted by most institutions in China for sharing research data. However, this mode typically involves limited data volume, struggles to ensure continuity and completeness, and offers limited potential for value enhancement. Alliance construction primarily involves data centers collaborating with universities/governments or several institutions forming sharing alliances to collect and share research data. Examples include the Netherlands' 4TU.Centre for Research Data, jointly established by four university libraries [10], and the Inter-university Consortium for Political and Social Research (ICPSR) in the United States, created by the University of Michigan with over 700 academic institution members and more than 500,000 data archives [11]. In China, similar models include the Tsinghua University China Economic and Social Data Research Center [12] and the Ren-

min University of China Survey and Data Center [13]. While this mode offers better operational continuity and data reliability than individual construction, it remains relatively fragmented and isolated compared to the vast volume of research data. Moreover, although data is partially aggregated, no integrated value-added service process has been formed that encompasses data usage, provision, intermediate technologies, and legal frameworks.

The European Union pioneered foundational research on research data infrastructure, with projects such as the 2009 High-Level Expert Group on Scientific Data (HLEG) [4], the 2010 "Mapping the European Research Infrastructure Landscape" project [5], GRDI2020 [6], and the 2011 "Pan-European Collaborative Data Infrastructure" project [7], primarily addressing the benefits, challenges, and action pathways of research data infrastructure construction. Subsequently, countries including the United States and Australia began conducting related research and proposing numerous research data infrastructure framework models [8]. However, practice reveals that current research data infrastructure exhibits diverse, fragmented, and constantly changing characteristics, with organizational management approaches varying significantly.

Germany, as a major global research force and a key initiator and participant in open science, began implementing research data management services in numerous universities starting in 2013, establishing institutional research data repositories. Alliance-based research data centers also gradually emerged, such as the German Social Science Data Center GESIS [14] and the LMU-ifo Economics and Business Data Center [15]. Against this backdrop, Germany's federal and state governments initiated unified research data infrastructure development projects. During the project process, they progressively defined their roles, forming an integrated research data infrastructure. This study examines federal/state-level research data infrastructure projects to analyze their construction mechanisms and provide references for China's research data management and national research data infrastructure establishment.

2 Overview of German Research Data Infrastructure Development

Germany's early research data infrastructure primarily consisted of institutional and thematic data repositories, such as Heidelberg University's heiDATA [16], Ludwig Maximilian University of Munich's OpenDataLMU [17], and Humboldt University of Berlin's MediaRepository [18]. Subsequently, alliance-based research data repositories emerged, including GESIS and the LMU-ifo Economics and Business Data Center. In response to relatively dispersed information and data, Germany's federal and state governments launched unified research data infrastructure development projects, with a timeline shown in Figure 1 [Figure 1: see original paper].

In January 2014, Baden-Württemberg's Ministry of Science, Research and the Arts launched Germany's first state-level research data infrastructure project,

bwFDM [19]. Executed jointly by nine universities in the state, Phase I comprised two sub-projects: bwFDM-Communities and bwFDM-Info. The first phase focused on disciplinary needs investigation, while the second aimed to establish a complete research data management information platform (forschungsdaten.info) for researchers. Phase II, bw2FDM, proposed establishing a state research data center. In May 2016, the Council for Information Infrastructures (Rat für Informationsinfrastrukturen, RfII) proposed establishing a National Research Data Infrastructure (Nationale Forschungsdateninfrastruktur, NFDI) [20], publishing the first discussion paper in April 2017. This formally proposed that NFDI should be built on extensive collaboration among research data stakeholders and achieve interconnection and supplementation on existing foundations [21]. In March 2018, the committee released a second discussion paper, "Cooperation as an Opportunity," clarifying NFDI's design objectives and characteristics, and explicitly stating that "NFDI is a distributed, networked infrastructure for data producers and users, meeting Germany's generic and domain-specific research data management needs through reliable and sustainable services" [22]. In August 2018, the Alliance of German Science Organizations released a discussion paper on developing a federal-state agreement for establishing NFDI [23], which was approved in November 2018 by the Joint Science Conference (Gemeinsame Wissenschaftskonferenz, GWK), comprising federal and state government authorities. In December 2018, RfII published a third discussion paper [24], proposing to construct NFDI using a consortium model, which received approval from GWK in December and is being implemented by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) [25]. During the NFDI discussion process, various states also launched their own research data infrastructure projects (referred to as LNFDI following German conventions), including Hesse's HeFDI project (Hessische Forschungsdateninfrastruktur, involving 11 state universities) and North Rhine-Westphalia's NFDI Landesinitiative project (implemented across state universities/research institutions through DH.NRW).

Currently, over 400 data repositories are registered on re3data in Germany, covering natural sciences, humanities and social sciences, engineering, and life sciences, essentially establishing a nationwide repository system. In the service domain, since 2013/2014, Bielefeld University, Humboldt University of Berlin, University of Göttingen, and Heidelberg University were among the first German universities to initiate research data management services, accumulating rich practical experience. Presently, more than half of comprehensive and technical universities provide research data management services. At the technical level, national-level projects involving tool software (such as the data management plan software RDMo) and domain-specific data management (such as the engineering research data management solution DIPL-ING) have concluded, establishing the foundation for implementing a national research data infrastructure. This analysis demonstrates that German research data infrastructure construction can be categorized by dimension: national-level and state (regional) level construction; by approach: primarily through alliances/consortiums; and

by method: primarily through interconnection and supplementation.

3 Analysis Perspectives and Data Sources

3.1 Analysis Perspective

The ultimate purpose of research data infrastructure construction is implementation. Therefore, this study's analysis of construction modes is grounded in an implementation perspective. Currently, more systematic research data infrastructure frameworks include implementation plans developed by institutions such as the European Commission's High-Level Expert Group (HLEG) and Australia's Research Data Infrastructure Committee (RDIC). The EU HLEG views research data infrastructure as a flowing circulatory system, emphasizing how different participants and data/services should interconnect and operate within the infrastructure [26]. Australia's RDIC emphasizes efficiency, arguing that research data infrastructure should coordinate cooperation across national, domain, and institutional levels to avoid duplication and reduce waste [3]. Specifically, data providers supply raw data and utilize IT institutions' technologies for data management, organization, and aggregation. Data users provide clear research needs while using these data for research and generating results, potentially becoming data providers themselves. IT institutions must supplement relevant technologies, software, and services according to research needs, facilitating current research and serving as powerful tools for other institutions to participate in the research data ecosystem. Furthermore, service agencies are crucial in research data infrastructure implementation, acting as intermediaries between data providers, users, IT institutions, and research data. Moreover, research data infrastructure is an open system; only with service agencies can more researchers and institutions be attracted to participate in data sharing and infrastructure development. Finally, existing data repositories require further integration and networking through research institutions and IT technologies to serve as continuously utilized data assets within the ecosystem, as shown in Figure 2 [Figure 2: see original paper]. The entire system must operate efficiently to avoid resource waste. Therefore, the core of research data infrastructure construction is to clarify the tasks of each component, the organizations undertaking these tasks, and their interrelationships, while ensuring efficient system operation and minimizing unnecessary consumption. This analysis will focus on tasks, implementing organizations, and management mechanisms that guarantee efficient operation.

3.2 Data Sources

German research data infrastructure projects have been comprehensively implemented, making project execution the most appropriate basis for analyzing construction modes. Currently, Germany's federal and state-level research data infrastructure construction projects are listed in Table 1 . By the end of 2019, the first round of NFDI project applications concluded, with 22 projects submitted [27]. At the state level, seven states including Baden-Württemberg and

Hesse have launched research data management infrastructure projects, with North Rhine-Westphalia and Thuringia still focusing on pilot projects, while earlier participants like Baden-Württemberg and Hesse have begun developing data center projects or related infrastructure.

4 Tasks of National and State Research Data Infrastructure

The August 2018 discussion paper issued by the Alliance of German Science Organizations clarified that German research data infrastructure construction would be implemented at both national and state (local) levels. Therefore, analysis should begin with the positioning and tasks at the macro-level national and state tiers, examining project execution guidelines to understand their construction purposes and functions.

4.1 NFDI: Technical Specifications

According to project documents, NFDI aims to systematically and standardize research data management, transforming Germany's current situation of dispersed, project-based data storage across the scientific system. It provides long-term data storage, backup, and accessibility while networking data domestically and internationally to ensure third parties can easily and systematically access and combine these data for analysis [28], with specific construction requirements shown in Table 2 . These requirements mandate that NFDI be executed by multi-institutional consortia working closely with disciplinary communities, external research data management service organizations, and other consortia. Under these conditions, NFDI develops unified, generic products and services including data collection, processing, standards, quality, management services, and training (see Figure 3 [Figure 3: see original paper]). The construction core revolves around data and the development of standards and necessary technical elements surrounding these data. First, research data infrastructure tasks ensure data accessibility in distributed storage repositories and enable data flow among different stakeholders. Therefore, establishing data collection, processing, and exchange standards forms the foundation for interconnectivity and data sharing. Simultaneously, in research data infrastructure, researchers are the source of data generation, and unlike literature, data processing remains primarily researcher-dependent due to professional constraints. Consequently, unified and standardized services and training for data generation are necessary to ensure fundamental data quality. Establishing a series of standards around specific disciplines or fields represents crucial foundational work at the national level.

4.2 LNFDI: Service-Oriented Approach

The Hesse Research Data Infrastructure Project (HeFDI), initiated in 2017, has produced relatively complete research outcomes and serves as the primary reference. HeFDI's construction content is shown in Table 3 , executed by a con-

sortium of 11 universities in the state. The "hard" components involving data repositories and management plan software tools are jointly completed by the 11 universities, while the "soft" components concerning research data policies and consulting services are uniformly developed through inter-university collaboration. HeFDI focuses on developing unified research data policy strategies and consulting services across the state to popularize research data management and provide related services. At the 11 universities with better foundations, unified data policies and models are used to establish necessary hardware including data management plan software and data repositories, forming a service network as shown in the "LNFDI" portion of Figure 3 [Figure 3: see original paper]. Thus, LNFDI's construction approach mirrors the federal level through inter-institutional cooperation to develop unified research data management service products such as tools, policies, training, and consulting. However, LNFDI's core purpose is service, manifested in service uniformity and standardization. Currently, research data management services are generally provided by university libraries or related institutions. LNFDI establishment can directly address potential issues of non-uniform service standards or varying service depth due to institutional differences in research data engagement, fundamentally ensuring research data quality.

4.3 Relationship and Characteristics of NFDI and LNFDI

4.3.1 Mutual Assistance As analyzed above, NFDI's core is data, primarily addressing data sharing and usage at the macro level, while LNFDI's core is service, focusing on establishing unified service models and products for researchers at the meso/micro level. These two aspects are mutually supportive and reinforcing. First, NFDI provides data and technical guarantees, offering extensive foundational research data, metadata standards, and a series of standardized technical service elements. Collaboration with external research data management technology organizations opens channels for identifiers and copyrights domestically. Second, LNFDI provides the foundation for NFDI. The essence of research data infrastructure is data, and beyond large facilities and major projects, vast amounts of research project data constitute the main body of research data and an important source for value enhancement. This foundation rests on data sharing and standardization, with LNFDI establishing numerous institutional-level research data repositories, primarily in universities, thus providing original support for NFDI and adding data source channels so NFDI does not start from "zero." Third, LNFDI provides experimental prototypes for NFDI. Data aggregation and interdisciplinary applications originate from research project practices, and grassroots research institutions like universities make needs more easily identifiable while enabling technical development at manageable data scales. Finally, LNFDI enhances researchers' data culture and literacy for NFDI. A sound research data infrastructure requires researchers to continuously share their data, and LNFDI services both improve researchers' data literacy and provide convenient conditions for participation in data sharing, offering sustained momentum for NFDI construction.

4.3.2 Formation of a National-Regional Two-Level Ecosystem This analysis demonstrates that Germany has established a national-regional two-level research data infrastructure ecosystem. At the hardware level, NFDI provides national infrastructure meeting high-volume, foundational research data storage needs, while LNFDI provides regional infrastructure closer to grassroots researchers for daily data sharing, with both interconnected through linking solutions. At the software level, NFDI provides domain standards, research data products, and services including metadata standards, processing workflows, training, and consulting, ensuring effective data generation and exchange throughout the ecosystem. LNFDI can use these outcomes to provide related services, serving researchers and institutions at the micro level while attracting increasing participation in data sharing to continuously expand and improve the entire ecosystem.

5 Organizational Structure

Organizations are systems that combine according to certain structural forms and activity patterns to achieve common goals, serving as the implementation entities of research data infrastructure. After clarifying tasks, organizational approaches form the basis for dividing, grouping, and coordinating work. Therefore, analyzing specific institutional organization methods and structures reveals internal relationships and operational modes within the system. This analysis uses the NFDI4Culture project [30] (National Research Data Infrastructure for Cultural Heritage) as an example to clarify organizations and related management mechanisms.

5.1 NFDI Organizational Structure: Broad Cooperation

Project documents indicate that NFDI's execution-level 主体 is the project consortium, the responsible entity for infrastructure construction. Broad cooperation is both a project requirement and the foundation for future infrastructure application. NFDI execution consortia can be divided into three parts: lead executors, joint executors, and participants. The first two participate as project execution institutions in infrastructure development, while participants serve as supplementary support, providing specialized disciplinary knowledge and product testing during development. Using the NFDI4Culture consortium as an example (see Figure 4 [Figure 4: see original paper]), this consortium constructs research data infrastructure in architecture, art history, musicology, theater, film, and media science. The main applicant is the German Academy of Sciences and Literature Mainz. Among joint applicants are four universities, three IT research institutions, and one comprehensive cultural institution, with 14 relevant academic societies, 26 specialized research institutes, and 25 international data management organizations such as UNESCO, Getty, PHAROS, ORCID, DataCite, and Wikimedia participating. The consortium assigns different tasks according to each institution's strengths: for instance, the Berlin Prussian Cultural Heritage Foundation, Germany's largest cultural institution,

provides rich foundational data resources for the ecosystem, while information institutions like the German National Library of Science and Technology (Technische Informationsbibliothek, TIB) primarily undertake foundational research data digitization tasks.

Cooperation with academic research groups, other data management organizations, and project consortia represents a construction requirement. Key characteristics include: (1) Comprehensive cooperation with disciplinary research groups. Beyond the four participating universities among joint applicants, participants include 14 academic societies and 26 disciplinary research institutes covering Germany's entire disciplinary system. Since research groups are the ultimate users and producers of research data, project outcomes have broader applicability and facilitate product promotion and popularization. (2) Essential international research data management organizations. Among participants are 25 international data management organizations, as many domains have established foundations and regulations for metadata schemes, data publishing, and identifiers, such as the commonly used Digital Object Identifier (DOI), Open Researcher and Contributor ID (ORCID), and discipline-specific identifier systems. Additionally, one-stop discovery systems are necessary components of any resource facility, requiring cooperation with foreign data repositories. Therefore, international partners with professional data management knowledge can provide assistance and experience technically, while enhancing research data influence and researcher participation in application. (3) Emphasis on cooperation with other consortia for generic information technology. In the NFDI4Culture Letter of Intent [31], NFDI4Culture explicitly identifies that its cultural heritage data shares common technologies with MaRDI (Mathematical Research Data Initiative), NFDINeuroscience (National Neuroscience Research Data Infrastructure), and Text+ (Research Data Infrastructure for Language and Textual Data) in mathematical modeling, imaging, text encoding, and metadata. Cooperation on generic technologies is a prerequisite for developing and establishing unified services and standards. In research data infrastructure, common technologies exist not only in encoding, metadata, identifiers, interoperability, and specialized terminology, but also in hardware systems (e.g., unified identity access management platforms, cloud infrastructure, computing power), legal and copyright issues (e.g., sensitive data, commercial data), and data governance (e.g., data culture, incentive mechanisms, publishing support).

5.2 LNFDI Organizational Structure: Regional Libraries/Computing Centers

Since LNFDI's core is service with the goal of coordination and unification, it differs from NFDI. In LNFDI, university libraries and computing centers are the main project implementers, with organizational design focusing on ensuring collaborative coordination among relevant departments across different universities. Specifically, the HeFDI project working group comprises libraries, computing centers, and disciplinary representatives from 11 universities in the

state, coordinating research data management services across state universities. Its primary role is ensuring unified technical standards and standardized service strategies. Additionally, the state library information technology service alliance HeBIS provides technical support to the working group, while participating universities' vice presidents for research form a steering committee to coordinate top-level research data policies and project progress liaison, ensuring consistent organizational frameworks.

6 Efficient Guarantee Mechanisms

The overall research data infrastructure system requires efficient operational guarantee mechanisms. First, since NFDI adopts a project fund application approach, the project management system ensures effective operation. Second, the ecosystem requires numerous data repositories whose construction consumes substantial resources without national-level unified planning, making repository usage mechanisms effective for reducing resource waste.

6.1 Management Structure: Unified Management, Enhanced Coordination and Supervision

NFDI governance adopts a strategic-executive-regulatory triangle structure, specifically including the Consortium Assembly, Domain Council, Board of Directors, and Expert Committee [32] (see Figure 5 [Figure 5: see original paper]). Key characteristics include: (1) At the strategic level, the Domain Council ensures project advancement and practicality. The Domain Council is the strategic decision-making body comprising the Consortium Assembly, German Alliance of Scientific Academies, GWK, and Board of Directors. The German Alliance of Scientific Academies, composed of 10 national research institutions and academic organizations, represents virtually all German research institutions and scientific needs/interests. Since research data sharing involves academic recognition and incentive mechanisms, NFDI also requires scientific community support. GWK, as a government organization and funding agency representing financial stakeholders, oversees fiscal funds. Both are supported operationally by the Board of Directors, which comprises professional information institutions with specialized resource services and IT capabilities, enabling strategic planning that reflects both research needs and national strategic directions. (2) At the execution level, emphasis is placed on project construction coordination to reduce duplication and ensure unified governance. The Consortium Assembly comprises representatives from each consortium, primarily coordinating and unifying content and technology across disciplinary research data infrastructure development. As research data infrastructure shares numerous common framework issues beyond technology—including governance and legal aspects—the Consortium Assembly enhances service standardization, avoids duplication, and ensures unified data governance. (3) At the regulatory level, an independent supervisory body is established. The DFG appoints an Expert Committee to evaluate funding applications, while

the DFG itself oversees project execution without direct affiliation with specific management institutions, ensuring relatively independent and objectively reliable supervision. As Germany's largest research funding organization, the DFG possesses extensive project funding experience.

6.2 Data Repository Construction: Shared Utilization Model

Establishing regional foundational data repositories is both a central task requirement and an important node for NFDI interconnection and interoperability. This analysis examines HeFDI's repository usage mechanism for reference. HeFDI primarily adopts a cooperative shared-use model for institutional data repositories, where several qualified universities in the state construct repositories that other partner universities use and pay for according to usage volume. This model's benefits and necessity include: (1) Lowering barriers to using specialized data repositories. Universities with smaller research data generation volumes need not build institutional repositories independently. (2) Reducing repository interconnection difficulty. Avoiding numerous small repositories ensures each repository's quality. (3) Leveraging existing repositories. Some universities already have repositories with professional service teams and models that can be fully utilized.

7 Lessons and Implications for China

Currently, China has established a data policy system comprising government, industry, and domain data centers, with sharing practices encompassing government agencies, research institutions, enterprises, and social forces. However, this concurrent development has resulted in insufficient overall planning for research data infrastructure. Most scientific data remains with project teams or individuals, and even institutions with research data centers face issues such as duplicated construction, limited data volume, and emphasis on construction over utilization. Some centrally managed scientific data remains isolated due to divergent technologies and standards, making compatibility difficult and limiting value enhancement. Therefore, Germany's research data infrastructure construction model offers strong reference value for China's development.

7.1 Urgent Need for Systematic Research Data Infrastructure Construction

China's research data infrastructure primarily originates from individual universities, the Chinese Academy of Sciences system, and specialized research institutes. The most important components include 20 national scientific data centers such as the "National High Energy Physics Scientific Data Center" and 30 resource repositories [33]. Additionally, universities like Wuhan University and Fudan University have gradually established data management platforms and collaboratively advanced research data management [34]. However, in terms of disciplinary composition, these centers focus mainly on basic disciplines, with

insufficient engineering and social science data. Regarding data source types, they emphasize foundational data over cutting-edge scientific data. In contrast, Germany's research data infrastructure includes not only foundational data (physics, biology, ecology) but also technical data (materials, engineering, medical health) and socio-cultural data (social sciences, cultural heritage), forming a more complete system. Vertically, it features both national and local infrastructure, networked through data repositories to avoid duplication while enabling effective data interoperability. In March 2018, the General Office of the State Council issued the "Measures for Scientific Data Management" (hereinafter "Measures") [35], clarifying research data's strategic resource status and providing governance frameworks. In April 2020, the "Opinions of the Central Committee of the Communist Party of China and the State Council on Building a More Perfect Market-oriented Allocation System and Mechanism for Factors" [36] affirmed research data's economic and social resource value as an important data element component. Drawing from Germany's model, China should establish a national research data resource catalog, standardize data repository disciplinary systems, and adopt a national data hub-regional/institutional data node model within specific disciplines to accelerate research data infrastructure construction.

7.2 Strengthening Research Data Infrastructure Ecosystem Construction

Extensive data is the fundamental guarantee for research data infrastructure. As open science influence deepens, more scholars and institutions will participate in research data sharing. Therefore, research data infrastructure must maintain openness, requiring processes, standards, software, and tools to form open interfaces. Since people are the source of research data, unified, standardized consulting services and research data management literacy education are essential components for forming a research data culture. Thus, constructing research data infrastructure essentially involves establishing an entire product suite integrating data, services, standards, consulting, and training to form an open research data ecosystem. As Chinese institutions establish research data platforms, a batch of related standards and specifications have been released, such as disciplinary core metadata standards, data schema description rules, and methods from national scientific data centers and platforms, as well as the Chinese Academy of Sciences' research data standard system [37], providing a certain technical foundation. However, technology promotion and optimization depend on frequent usage. Drawing from Germany's model, consulting, training, and services should first be incorporated into research data center construction frameworks, then developed in terms of content, methods, and qualifications. At the application level, funding and encouragement should support institutional libraries/computing centers in trial promotion and popularization among researchers and students.

7.3 Differentiating National and Local Construction Tasks and Methods

The core of research data infrastructure is data, which has diverse sources: large-scale, relatively static foundational research data, and smaller-scale, relatively dynamic cutting-edge research data from projects. These data generate a series of tools, services, and training products targeting different audiences, necessitating differentiated construction tasks and methods at national and local, specialized and generic levels. In Germany's model, the national level focuses on foundational data construction, developing relevant metadata standards, and establishing integrated service models, while the local level emphasizes project-oriented data construction. Through coordinated and unified research data management plans and consulting services, it improves micro-level research data management services and popularization, providing data-level supplementation for the national infrastructure and ensuring data sources and quality. The "Measures" indicate that China's research data management implements a national coordination system with division of responsibilities among departments and regions. Therefore, at the national level, different disciplinary or related-direction national research data centers should be established, focusing on data standards, tools, and software. At the local level, priority should be given to funding alliances of institutional libraries/computing centers to strengthen unified services. For universities or research institutions with existing research data infrastructure platforms or capabilities, encouragement should be provided to undertake or establish local-level (e.g., provincial) research data centers, while other units purchase services to cover resource consumption costs. This effectively concentrates data while avoiding duplication and facilitating interoperability with national platforms.

7.4 Strengthening Integrated Development of Research Data Infrastructure

Currently, China's research data infrastructure construction primarily follows institutional self-building or national coordination models. Using national scientific data centers as an example, their construction relies on national science and technology resource sharing service platforms, with the Chinese Academy of Sciences and national ministries as supervisory departments and affiliated institutes as responsible entities. This model can rapidly and effectively establish foundational research data. However, scientific data's application value lies in interconnectivity, while China's construction entities (national, local, or university-level) are relatively dispersed with non-uniform standards and specifications, limiting data value enhancement. Coordinated development of national research data infrastructure is therefore crucial. Drawing from Germany's model, a strategic management institution should be established comprising disciplinary, information service, and resource representatives to coordinate data centers nationally, avoid duplication, enhance cooperation, and provide unified solutions for generic technologies. At the construction level, project funding

should be adopted with special funds to enable qualified and willing research units to apply jointly and facilitate participation by resourceful and technologically capable social institutions, forming national/local research data infrastructure construction entities. Following fund management methods, science and technology management departments at all levels should serve as funding responsible entities, with fund committees or management institutions as supervisory bodies, strengthening pre-construction evaluation, in-construction supervision, and post-construction assessment to reduce potential construction-without-utilization, numerous "static" platforms, and infrequent data updates, while postponing construction for immature conditions.

7.5 Expanding Cooperation Scope in Research Data Infrastructure Construction

Research data infrastructure construction requires extensive cooperation, representing not only technical capability requirements but also effective means for enhancing application and popularization of data centers and products. Currently, China's research data centers are primarily built along departmental lines with data as the core, resulting in insufficient application breadth and applicability, isolated data, and inadequate discovery systems. Internationally, authoritative domain-specific repositories such as Dryad and GenBank have formed extensive cooperation with journals, scholars, and research institutions. Drawing from Germany's model, first, construction should enhance broad cooperation with research institutions and groups to both promote products and accelerate transformation from users to contributors. Second, cooperation on generic information technology should be strengthened; Germany requires identifying potential generic technologies among different consortia to reduce development difficulty and facilitate data interconnection and retrieval. Additionally, national platform construction should enhance cooperation with foreign research data infrastructure management organizations to improve management and operational levels and increase international authority. Foreign institutions have accumulated rich experience in data center construction and data standardization, while authoritative systems exist internationally for repository certification and services in data publishing. Finally, strengthening data linking and one-stop retrieval requires cooperation among different institutions across countries. At the local level, cooperation among service agencies should be enhanced, encouraging existing alliances such as university library alliances to develop unified research data management services.

References

[1] Wei Haiyan, Fan Xinyi. Research on Data Librarian Competency Requirements and Service Content in Research Libraries[J]. Library, 2019(4): 85-91. [2] Liu Ziheng, Zeng Liying. Investigation and Comparative Analysis of University Research Data Management and Sharing Platforms in China[J]. Information and Documentation Services, 2017(6): 90-95. [3] Wang

Jing, Wang Yanbing. Investigation and Analysis of Foreign Research Data Infrastructure Research and Practice[J]. Information and Documentation Services, 2016(6): 99-104. [4] GRDI2020. Unlocking the full value of scientific data[EB/OL]. [2020-01-20]. http://www.grdi2020.eu/Pages/Unlock.aspx. [5] European Science Foundation. MERIL Portal[EB/OL]. [2020-01-23]. http://portal.meril.eu/conversesf/static/about.grdi202.eu/Default.aspx. [6] FAASH. GRDI2020[EB/OL]. [2020-01-05]. http://www.grdi202.eu/. [7] EUDAT Ltd. EUDAT Collaborative Data Infrastructure[EB/OL]. [2020-01-20]. http://eudat.eu/. [8] BMBF. Nationale Forschungsdateninfrastruktur[EB/OL]. [2020-01-04]. https://www.bildung-forschung.digital/de/nationale-forschungsdateninfrastruktur-2542.html. [9] Chu Wenjing, Li Shuning. Research on the Construction of Scientific Data Alliance Management Model in China[J]. Library and Information Science Research, 2019(14): 51-57. [10] TU Delft Library. 4TU.Centre for Research Data[EB/OL]. [2020-01-05]. https://researchdata.4tu.nl/en/. [11] University of Michigan. ICPSR[EB/OL]. [2020-03-31]. https://www.icpsr.umich.edu/icpsrweb/. [12] Tsinghua University. Tsinghua University China Economic and Social Data Research Center[EB/OL]. [2020-01-05]. http://www.tcdc.sem.tsinghua.edu.cn/. [13] Renmin University of China. Renmin University of China Survey and Data Center[EB/OL]. [2020-01-05]. http://nsrc.ruc.edu.cn/. [14] Leibniz-Institut für Sozialwissenschaften. GESIS[EB/OL]. [2020-01-20]. https://www.gesis.org/home. [15] ifo Institute. LMU-ifo Economics & Business Data Center[EB/OL]. [2020-01-20]. https://www.ifo.de/en/EBDC. [16] Heidelberg University. heiDATA[EB/OL]. [2020-01-20]. https://heidata.uni-heidelberg.de/. [17] Ludwig-Maximilians-Universität München. OpenDataLMU[EB/OL]. [2020-01-20]. https://data.ub.uni-muenchen.de/. [18] Humboldt-Universität zu Berlin. Media Repository Humboldt-Universität zu Berlin[EB/OL]. [2020-01-20]. https://media.hu-berlin.de/. [19] Karlsruher Institut für Technologie. bwFDM[EB/OL]. [2020-01-20]. https://bwfdm.scc.kit.edu/. [20] Rat für Informationsinfrastrukturen. Leistung aus Vielfalt[EB/OL]. [2020-01-20]. http://www.rfii.de/?wpdmdl=1998. [21] Rat für Informationsinfrastrukturen. RfII-DISKUSSIONSPAPIER “SCHRITT FÜR SCHRITT-ODER: WAS BRINGT WER MIT?”[EB/OL]. [2020-01-20]. http://www.rfii.de/download/rfii-diskussionspapier-2017/. [22] Rat für Informationsinfrastrukturen. RfII-DISKUSSIONSPAPIER “ZUSAMMENARBEIT ALS CHANCE”[EB/OL]. [2020-01-20]. http://www.rfii.de/download/rfii-diskussionspapier-maerz-2018/. [23] Allianz der Wissenschaftsorganisationen. Diskussionspapier zur Ausarbeitung einer bund-Länder-Vereinbarung zum Aufbau einer nationalen Forschungsdateninfrastruktur[EB/OL]. [2020-01-20]. https://www.fraunhofer.de/content/dam/zv/de/ueber-fraunhofer/wissenschaftspolitik/10/20180808%20Disku
[24] Rat für Informationsinfrastrukturen. Wide impact for research: NFDI Consortia as Stakeholders[EB/OL]. [2020-01-20]. http://rfii.de/?wpdmdl=3818. [25] DFG. Agreement between the federal government and the Länder concerning the establishment and funding of a National Research Data Infrastructure (NFDI)[EB/OL]. [2020-01-20]. https://www.dfg.de/download/pdf/foerderung/programme/nfdi/blv_{en}.pdf.
[26] Research data e-infrastructures: framework for action in H2020[EB/OL].

[2020-03-24]. https://indico.cern.ch/event/233119/contributions/492496/attachments/388387/540129/H2020-Framework-4-Action.pdf. [27] DFG. NFDI statistics on applications[EB/OL]. [2020-01-20]. https://www.dfg.de/download/pdf/foerderung/programme/nfdi/191212_{{{nfdi}}{{statistik}}
[28] DFG. Guidelines for Consortia National Research Data Infrastructure (NFDI)[EB/OL]. [2020-10-12]. http://www.dfg.de/formulare/nfdi100/nfdi100_{en}.pdf.
[29] Philipps-Universität Marburg. HeFDI[EB/OL]. [2020-01-20]. https://www.uni-marburg.de/de/forschung/kontakt/eresearch/projekte-und-netzwerke/hefdi.
[30] NFDI4Culture. NFDI4Culture[EB/OL]. [2020-01-15]. https://nfdi4culture.de/.
[31] DFG. NFDI4Culture-Consortium for research data on material and immaterial cultural heritage[EB/OL]. [2020-01-20]. https://www.dfg.de/download/pdf/foerderung/programme/nfdi/a
[32] DFG. Nationale Forschungsdateninfrastruktur[EB/OL]. [2020-01-07]. https://www.dfg.de/foerderung/programme/nfdi/index.html. [33] Ministry of Science and Technology. Notice on the optimized adjustment list of national science and technology resource sharing service platforms[EB/OL]. [2020-01-05]. http://www.most.gov.cn/mostinfo/xinxifenlei/fgzc/gfxwj/gfxwj2019/201906/t20190610{147031}.htm.
[34] Qu Baoqiang. Research on the Construction and Development Countermeasures of China's Scientific Data Infrastructure[J]. Information Engineering, 2020, 6(1): 11-21. [35] General Office of the State Council. Measures for Scientific Data Management[EB/OL]. [2020-04-02]. http://www.gov.cn/zhengce/content/2018-04/02/content_{5279272}.htm.
[36] Central Committee of the Communist Party of China and the State Council. Opinions on building a more perfect market-oriented allocation system and mechanism for factors[EB/OL]. [2020-04-09]. http://www.gov.cn/zhengce/2020-04/09/content_{5500622}.htm. [37] Hu Lianglin, Zheng Xiaohuan, Zhu Yanhua, et al. Research and Practice on the Chinese Academy of Sciences Scientific Data Standard System[J]. Library, 2019(11): 6-10.

Author Contributions: Zhou Lei: Topic selection, data collection and analysis, manuscript writing; Yang Ping: Research framework guidance, manuscript revision and finalization; Yuan Rubing: Participated in manuscript revision.

Enlightenment and Construction Mode of German Research Data Infrastructure

Zhou Lei[1,2], Yang Ping[1,2], Yuan Rubing[1,2]

[1]Beijing Institute of Science and Technology Information, Beijing 100044
[2]Beijing Science and Technology Strategic Decision Consulting Center, Beijing 100044

Abstract: *[Purpose/Significance]* The research data infrastructure is an indispensable supporting infrastructure for the efficient management and utilization of research data and the efficient operation of digital scientific research. *[Method/Process]* This paper takes German research data infrastructure projects as the research object and analyzes the construction mode of research data infrastructure from the perspectives of national and regional task content differences, construction organization, and related management mechanisms. *[Result/Conclusion]* The study found that German research data infrastructure has

different emphases at different levels. At the national level, the focus is on data standards and tools; at the local level, the focus is on services and training. In construction, it adopts a strategy of extensive cooperation and complementary capabilities; in management, it has the characteristic of focusing on unified management and supervision. Finally, suggestions are provided for China's research data infrastructure construction.

Keywords: research data; infrastructure; research data infrastructure; construction mode

Note: Figure translations are in progress. See original paper for figures.