

# The NFDI4Memory Data Space App: A Federated Infrastructure and Service for Historical Research Data

Erdal Ayan<sup>1,\*</sup> and Sandra Göller<sup>1</sup>

<sup>1</sup>FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Eggenstein-Leopoldshafen, Germany

\*Corresponding author: erdal.ayan@fiz-karlsruhe.de.

## Introduction

The historical sciences face particular challenges regarding their research data: heterogeneity of sources and formats, complex provenance information, long-term interpretability, and the need to preserve context [1]. Humanities data often blur the line between research outputs and publications, highlighting the need for improved research data management and FAIR data practices [2] [3]. In fields like history, spreadsheets and relational databases are still the main tools for data analysis and management, allowing researchers to work directly with their collected data as mentioned by [4]. However, this practice has notable drawbacks: datasets are often tied to specific research questions, lack detailed provenance information, and are difficult to verify or reuse. The researchers therefore emphasize the importance of improving semantic interoperability and developing collaborative, well-structured workflows to overcome these challenges.

The NFDI4Memory Data Space [5] is a central component of NFDI4Memory. It aims to provide an integrated, trusted, and FAIR-aligned environment for comprehensive searches and discovery of historical research data. It will enable sharing and reusing data across institutions and data sources [6]. The NFDI4Memory Data Space will address the challenges by making heterogeneous and distributed data collections and services, such as the Newspaper Portal [7], the European Holocaust Research Infrastructure [8] or also datasets published on research data repositories such as RADAR4Memory [9] [10]. FAIR and easily usable by enabling innovative methods and processes in the digital humanities. We propose that this will also lead to knowledge and technology transfer, network-

ing among the partner institutions as well as the scholars working in different fields.

## Objectives: Concept and Architecture

The NFDI4Memory Data Space forms a central infrastructure as a digital ecosystem of data sources and services. It connects a variety of repository systems, archives and institutional services and enables researchers to find sources and related data across disciplinary and repository boundaries. Uniformly defined interfaces offer direct access to resources and their collection-specific metadata, as well as to other tools and services for processing and reusing data, such as text extraction using optical character recognition (OCR) technologies and conversion of schemas or formats [5].

Technically, the Data Space is built on a framework of application programming interfaces (APIs) and the MemO Knowledge Graph, based on the NFDI4Memory Ontology [11] [12]. Through these APIs searching, downloading and converting or transforming data is possible. The Data Space integrates existing infrastructures via interoperable interfaces and shared policies. This approach respects institutional autonomy while ensuring a coherent user experience and consistent metadata flow. The Data Space design follows three guiding principles:

- Interoperability through Standards – Built upon established protocols (OAI-PMH, REST) and common metadata schemas (Dublin Core, DataCite), the system connects data sources via Query, Access, and Update APIs. The shared NFDI4Memory Ontology MemO [11] enables semantic alignment across domains

and machine-actionable linking between datasets.

- **Trust and Sustainability** - Institutions are only included in the Knowledge Graph once their Update API implementation has been validated, guaranteeing reliability, transparency, and long-term availability of referenced data and services. The MemO Knowledge Graph forms a stable information backbone, enabling persistent discovery and governance.
- **Extensibility and Reusability** – The modular architecture supports the integration of domain-specific tools and services. In this way, the Dataspace can accommodate evolving research workflows and technologies. Standardized service APIs (e.g., OCR or schema conversion) support modular integration into research workflows and allow services to be swapped or expanded without breaking compatibility. Planned future APIs (such as services for standardised data or document conversion) will show that the system will enable broad reuse of both data and technical components.

## Developing a Prototype: Data Space App

One of the principal aims of the Data Space project is to facilitate the semantic findability of historical data across heterogeneous repositories and disciplinary boundaries by means of a web environment, called Data Space App. We have been developing the first prototype of the app benefiting from the modern Python framework, FastApi [13] and SPARQL-server, Fuseki-Jena [14] as the back-end solutions and Vue.js [15] as front-end implementation. To this end, comprehensive ontological mappings and crosswalks that mediate between diverse metadata schemas are being developed and integrated.

The Data Space App embodies a technological and semantic interoperability framework by marrying a standard-based API backbone with the modular ontology MemO [16] and its associated Information and Research Data Knowledge Graph. As described in the Blue Paper [17], the architecture enables diverse institutions to connect their data sources and services seamlessly under an accessible web environment, while enabling scholars to per-

form unified, cross-repository search and access via Query, Access and Update APIs. In this way, the platform not only supports technological interoperability but also advances semantic interoperability across the historical-humanities domain, fostering extensibility, reuse and sustainable data infrastructure.

NFDI4Memory Data Space will grow into a fully interoperable research infrastructure that not only connects data and services but also enables richer semantic linking across institutions, disciplines, and research workflows. Upcoming extensions – such as additional service APIs for OCR services – will further strengthen semantic interoperability and improve machine-actionable reuse of historical data. In this way, the Data Space is positioned to become a lasting backbone for digitally supported historical research in and beyond the NFDI4Memory community [18].

## Conclusion

The Data Space App, as a central user-facing component, enables unified search, discovery, and reuse of heterogeneous datasets, while ensuring institutional autonomy and verifiable trust in data and services. Through its combination of technical interoperability, semantic harmonization, and extensible modular design, the NFDI4Memory Data Space fosters collaboration and innovation across disciplines, institutions, and research workflows. It provides historians and related scholars with the means to link, contextualize, and analyze sources in novel ways, thereby enhancing both the transparency and sustainability of digital research practices. As the infrastructure continues to evolve—expanding APIs, integrating new repositories, and deepening semantic interconnections—it lays the foundation for a dynamic, community-driven ecosystem that will support historically oriented research well into the future.

## References

- [1] P. Fafalios, K. Konsolaki, L. Charami, K. Petrakis, M. Paterakis, D. Angelakis, Y. Tzitzikas, C. Bekiari, and M. Dorr, “Towards Semantic Interoperability in Historical Research: Documenting Research Data and Knowledge with Synthesis,” 2021, vol. 12922, pp. 682–

698. doi: 10.1007/978-3-030-88361-4\_40 ., arXiv:2107.13957 [cs].
- [2] L. Poljak Bilić and K. Posavec, “FAIRness of Research Data in the European Humanities Landscape,” *Publications*, vol. 12, no. 1, p. 6, March 2024. doi: 10.3390/publications12010006 ., publisher: Multidisciplinary Digital Publishing Institute.
  - [3] R. Grant, “Reusable, FAIR Humanities Data: Creating Practical Guidance for Authors at Routledge Open Research,” *International Journal of Digital Curation*, vol. 17, no. 1, pp. 15–15, 2023. doi: 10.2218/ijdc.v17i1.820 .
  - [4] P. Fafalios, K. Petrakis, G. Samaritakis, K. Doerr, A. Kritsotaki, Y. Tzitzikas, and M. Doerr, “FAST CAT: Collaborative Data Entry and Curation for Semantic Interoperability in Digital Humanities,” *Journal on Computing and Cultural Heritage*, vol. 14, no. 4, pp. 1–20, December 2021. doi: 10.1145/3461460 ., arXiv:2105.13733 [cs].
  - [5] T. Holste and M. Razum, “Der NFDI4Memory Data Space als digitales Ökosystem für die historisch arbeitenden Wissenschaften, Gedächtnisinstitutionen und Informationsinfrastrukturen,” *ABI Technik*, vol. 44, no. 4, pp. 259–267, November 2024. doi: 10.1515/abitech-2024-0039 ., publisher: De Gruyter.
  - [6] J. Paulmann, J. C. Wood, K. Ceynowa, F. Cremer, S. Daniel, D. Fähle, B. Fichtl, P. Haslinger, T. Hiltmann, R. Hohls, U. Lehmkuhl, M. Lemaire, G. Maier, O. Meiners, G. Minn, K. Moeller, A. Neuburger, C. Prinz, M. Razum, H. Sack, J. Sauter, H. Schäffler, E. Schlotheuber, S. Schmunk, A. Stanzel, H. Trischler, and T. Wübbena, “NFDI4Memory. Consortium for the historically oriented humanities. Proposal for the National Research Data Infrastructure (NFDI),” December 2022. doi: 10.5281/zenodo.7428489 ., publisher: Zenodo.
  - [7] “Deutsches Zeitungsportal - Deutsche Digitale Bibliothek,” 2025. [Online]. Available: <https://www.deutsche-digitale-bibliothek.de/newspaper>
  - [8] “The EHRI Portal,” 2025. [Online]. Available: <https://portal.ehri-project.eu/>
  - [9] “RADAR Forschungsdaten,” 2025. [Online]. Available: <https://radar4memory.radar-service.eu/radar/de/home>
  - [10] Re3data.Org, “RADAR4Memory,” 2024. doi: 10.17616/R31NJNOX .
  - [11] ISE-FIZKarlsruhe, “NFDI4Memory Knowledge Graph,” 2025. [Online]. Available: <https://github.com/ISE-FIZKarlsruhe/nfdi4memory>
  - [12] S. R. Ondraszek, T. Tietz, J. Waitelonis, and H. Sack, “Interlinking Research Data and Services in the Historical Sciences with MemO and the NFDI4Memory Knowledge Graph,” October 2025. doi: 10.5281/zenodo.17376359 ., conference Name: Scientific Knowledge: Representation, Discovery, and Assessment 2025 ((Sci-K 2025)) Publisher: Zenodo.
  - [13] “FastAPI,” 2025. [Online]. Available: <https://fastapi.tiangolo.com/>
  - [14] “Apache Jena Fuseki,” 2025. [Online]. Available: <https://jena.apache.org/documentation/fuseki2/>
  - [15] “Vue.js,” 2025. [Online]. Available: <https://vuejs.org/>
  - [16] “NFDI4Memory Ontology (MemO),” October 2025, original-date: 2025-04-29T11:41:38Z. [Online]. Available: <https://github.com/ISE-FIZKarlsruhe/memo>
  - [17] F. Bach, D. Fähle, S. Göller, T. Holste, J. Schweikert, and S. R. Ondraszek, “Blue Paper: Technische Spezifikationen für den NFDI4Memory Data Space,” Zenodo, Tech. Rep., August 2025. doi: 10.5281/zenodo.16614629 .
  - [18] D. Fähle and H. Sack, *Ein „digitaler Werkzeugkasten“ für historische Forschung mit Archivgut: Status quo und Perspektiven*. regiopen.books, June 2023. doi: 10.53458/books.166 ., iSSN: 2941-5578 Publication Title: regiopen.books.