Interlinking Research Data and Services in the Historical Sciences with MemO and the NFDI4Memory Knowledge Graph

Sarah Rebecca Ondraszek^{1,2,*}, Tabea Tietz^{1,2}, Jörg Waitelonis¹ and Harald Sack^{1,2}

Abstract

The German National Research Data Infrastructure (NFDI) aims to harmonize the flow of data from science and research. Consequently, the NFDI4Memory consortium, responsible for data from historically oriented sciences, also pursues this bigger objective. The goal is to realize a FAIR digital research environment in which users can discover content through shared semantic structures on previously unconnected materials.

Essential to this effort is the development of the NFDI4Memory Ontology (MemO) and the NFDI4Memory Knowledge Graph (MemO KG). In combination, they build the ground for the NFDI4Memory Data Space, supporting federated searches and semantic interoperability. As a modular extension of the interconsortial mid-level ontologies NFDIcore and the Culture Ontology, MemO incorporates domain-specific concepts from the historical sciences, among others, including the harmonization of metadata and the detailed representation of provenance. The MemO KG serves as a central index, harmonizing metadata for research data, institutions, researchers, and services. This infrastructure lays the groundwork for a unified point of access to research data across disciplines and consortia. Thereby, it fosters new modes of exploration and research data acquisition in historical research.

Keywords

Digital humanities, knowledge graphs, ontologies, knowledge extraction, research data management

1. Introduction

The NFDI (National Research Data Infrastructure) framework is a German initiative with the goal of systematically interconnecting and publishing research data across all academic disciplines, such as natural sciences, engineering, life sciences, social sciences, humanities, and earth sciences. The NFDI4Memory consortium focuses specifically on the historically oriented sciences and is dedicated to the creation of long-term and sustainable research data infrastructures that address the particular requirements of historical research. These are, among others, the modeling of provenance information about historical documents or the different hierarchical structures in archives [1, 2].

Historical research data are often scattered across a wide range of specialized repositories, such as the European Holocaust Research Infrastructure (EHRI Portal) [3], which aggregates metadata about dispersed Holocaust-related collections and holdings of numerous archives. The German Digital Library (DDB) aggregates various types of sources, such as museum artifacts and archival documents, in a unified access point. Another example is FactGrid, a collaborative Wikibase graph database for researchers with a historical research interest. It remains a challenge to interlink these sources and their heterogeneous research data, given the different metadata standards in use, the varied representation formats, and a multitude of access points [4, 5].

¹FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Eggenstein-Leopoldshafen, Germany

²Institute of Applied Informatics and Formal Description Methods (AIFB) of KIT, Karlsruhe, Germany

⁵th International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment, Nov 2024, Nara, Japan *Corresponding author.

Sarah-rebecca.ondraszek@fiz-karlsruhe.de (S. R. Ondraszek); tabea.tietz@fiz-karlsruhe.de (T. Tietz); joerg.waitelonis@fiz-karlsruhe.de (J. Waitelonis); harald.sack@fiz-karlsruhe.de (H. Sack)

^{© 0009-0003-7945-6704 (}S. R. Ondraszek); 0000-0002-1648-1684 (T. Tietz); 0000-0001-7192-7143 (J. Waitelonis); 0000-0001-7069-9804 (H. Sack)

^{© 2025} Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

In NFDI4Memory, the task area Data Services aims to enhance the interconnectivity of existing heterogeneous data collections by aggregating and providing a centralized access point to historical research within a single data space. Key roles in this endeavor can be attributed to the Memory Ontology (MemO) and the NFDI4Memory Knowledge Graph (MemO KG) [6], which serve as a foundational framework for (federated) search and exploration in the NFDI4Memory Data Space. The latter is NFDI4Memory's digital infrastructure for scientific access to research data and services. Additionally, it forms the basis for innovative services that can be tested in a dedicated Data Lab [7].

This paper presents ongoing work on MemO and the MemO KG. A crucial aspect of developing these two is considering the specificities of historical data and the historical research domain. Forming an index that enables cross-connections between decentralized sources, the MemO KG represents research data, institutions, researchers, their expertise, and the services they offer. The underlying ontology, MemO, is based on NFDIcore and the Culture Ontology (CTO) [8, 9], both of which are maintained by NFDI4Culture, another NFDI consortium that deals with research data from material and immaterial cultural heritage.⁴

Central to the integration of diverse resources from repositories in the historical domain is providing flexible representations for inherently heterogeneous sources; in this case, modularity enables reuse, adaptation, and extension for specific disciplinary contexts without requiring reinvention from scratch in each approach [10, 11].

MemO is a modular extension of the aforementioned ontologies. In the context of NFDI4Memory, MemO adds concepts that represent historical research information and data. Based on competency questions (CQs) developed in collaboration with domain experts, it harmonizes data and facilitates the interconnection of fragmented resources. This includes, among other aspects, the representation of archival materials, hierarchical relationships between documents, provenance descriptions, and document-content-entity-relationships (such as mentioned persons, events, ...).

The system allows for interoperability both within the NFDI4Memory context and across other NFDI consortia, fostering connections and cross-disciplinary reuse.

In this way, MemO and the MemO KG, in combination with the NFDI4Memory Data Space, do not aim to replace existing systems. The goal is to complement and connect portals such as EHRI or the DDB with a semantic layer. This aligns with the overarching NFDI goal of creating a federated landscape in which individual data providers remain autonomous while contributing to an interconnected ecosystem [1, 11].

The paper introduces related works in section 2. Subsequently, section 3 explains how NFDIcore and CTO are tied to MemO, which functions as a foundation for the following section. Section 4 describes MemO's development, beginning with the CQs and the alignment with NFDIcore/CTO, going over to concrete modeling examples to visualize the application scenarios. These aspects are harmonized in section 5, which offers a data story for a concrete use case in NFDI4Memory. Section 6 wraps this paper up with a conclusion.

2. Related Work

To create MemO and the MemO KG in a sound manner, it is crucial to build on established practices in ontology development and KG construction. This section outlines related initiatives and foundational methodologies in three key areas: ontology design in NFDI consortia, KGs as metadata indexes, and technical infrastructures for access and querying.

2.1. Knowledge Graphs as Decentralized Indices

The use of KGs as indices for distributed and heterogeneous data is increasingly common in digital humanities and cultural heritage infrastructures. These KGs do not replicate data sources; instead, they

³https://4memory.de/ueber-4memory/task-areas/data-services/

⁴https://nfdi4culture.de/index.html

describe and connect them using shared ontologies and authority vocabularies [12].

For example, ArCo, the Italian cultural heritage KG, models cultural heritage objects and corresponding metadata, offering standardized access through a comprehensive ontology network [13]. Similarly, the Odeuropa project was a European H2020-funded research project aimed at capturing and investigating smell history and heritage.⁵ In the project, information about smell heritage and information about its creation and perception [14] is captured in a KG. The underlying data model is aligned with ontologies such as CIDOC CRM [15] and PROV-O [16] for the multifaceted representation of provenance information. This tackles the challenge of representing both events revolving around a smell, as well as the original textual or image fragment to connect a smell with the metadata of the work from which it has been extracted.

At the European level, the Europeana Knowledge Base and Entity API aggregate multilingual cultural heritage metadata by linking to external authorities [17, 18]. Collectively, these projects demonstrate the potential of KGs to serve as a semantic backbone for integrating research across institutions and disciplines.

2.2. Technical Infrastructures, Interfaces, and Querying

The NFDI4Memory Data Space requires a reliable technical architecture that combines modular views of the data with standardized application programming interfaces (APIs) to represent the multitude of heterogeneous metadata collected in the MemO KG. The NFDI4Culture KG serves as a model for this endeavor and is currently accessible via a SPARQL endpoint, SHMARQL (a SPARQL endpoint explorer), and a dashboard for analysis and visualization [11]. NFDI4Culture has also implemented the Culture Information Portal, a web-based Current Research Information System (CRIS) that adheres to international standards for CRIS systems. For the web interface, TYPO3 CMS with an RDF extension was applied and extended for implementation [19]. Modular API stacks such as this have also been proposed for domain-agnostic infrastructures, such as the Europeana API suite, which provides a RESTful search interface [17]. The NFDI4Memory Data Space employs a multi-API approach to facilitate exploratory searches and structured queries throughout the distributed ecosystem [12].

3. Reusing the NFDIcore and the NFDI4Culture Ontology

To facilitate interoperability across NFDI consortia, the NFDIcore ontology has been designed as a mid-level ontology [8]. The scope of NFDIcore encompasses the representation of the organizational and research structure within NFDI, as well as data and information management, including datasets, portals, and collections, geographical and contextual information, software, specifications, standards, services, processes, and licensing. As a mid-level ontology, NFDIcore has been aligned with the Basic Formal Ontology (BFO 2020) [20, 21] and reuses ontologies, including the Information Artifact Ontology (IAO), the Software Ontology (SWO) [22], and the EDAM ontology [23]. By adhering to established data standards, the ontology ensures consistent and sustainable accessibility, sharing, and reuse of research data.

While all 26 NFDI consortia share overarching goals and concepts, each consortium also faces individual requirements and challenges, such as domain-specific standards, workflows, and methods for discovering research data. Therefore, a modular ontology design structure has been developed, which facilitates the development of ontology extensions tailored to each consortium and domain, thereby adhering to the specific domain requirements. A domain module that extends NFDIcore is the NFDI4Culture ontology (CTO) [9], which is being reused in this presented work. CTO builds on NFDIcore, with the key objective of facilitating the integration of cultural heritage research data into the NFDI4Culture Knowledge Graph, which is made available through the Culture Information Portal⁶. The primary scope of the CTO is to represent cultural heritage research data within the NFDI4Culture data

⁵https://odeuropa.eu

⁶https://nfdi4culture.de/

index, thereby providing a single point of access to decentralized cultural heritage research resources. Its primary focus is the creation of a lightweight index of cultural heritage research data provided by the culture community, including but not limited to the subject areas of musicology, performing arts, media studies, architecture, and art history [19, 11].

4. The NFDI4Memory Ontology and Knowledge Graph

MemO is being developed as a domain ontology and a modular extension of NFDIcore and CTO, following an iterative development strategy, closely aligned with the eXtreme Design (XD) methodology [24]. Its purpose is to represent research data and services in the historical sciences, based on CQs that were derived through consultations with domain experts and gathered in collaboration with the community. These CQs primarily focus on the discoverability of historical sources, the traceability of provenance, and the ability to link different documents or items through shared entities, such as people, places, or historical events.⁷⁸

4.1. Foundations of MemO

The MemO KG encompasses two subgraphs on a conceptual level: The Research Information Graph (RIG) for structured data on researchers, projects, services, institutions, and their connections. The Research Data Graph (RDG) focuses on metadata and contextual information about sources, such as archival materials and their provenance.

An extensive list of CQs can be found in GitHub. Central to it are concepts in historical research relevant to both RIG and RDG; they address aspects like: Is the data part of a collection? (CQ14), Is a document a copy, an edition, or an original? (CQ21), or Which data link people, events, and places in a specific epoch? (CQ53-CQ55).

4.2. Modularity and Alignment with NFDIcore and CTO

MemO's modular structure aligns with and extends NFDIcore and CTO. Building on these, MemO introduces modular extensions for provenance and extended document representation, adding hierarchical relations (e.g., in the context of archival items and collections), as well as subject categories.

4.2.1. Further Reused Ontologies

In correspondence to how CTO is being developed, MemO builds on standardized upper-level ontologies to provide a well-defined and standardized semantic structure. Next to NFDIcore and CTO, the following ontologies have been reused:

- Basic Formal Ontology (BFO 2020): BFO was selected as the top-level ontology due to its design, broad applicability, and ability to integrate with various ontologies [20].
- Information Artifact Ontology (IAO): The Information Artifact Ontology was partially reused to describe data feeds, creative works, and material entities. The central class reused in MemO is iao:information content entity. Since IAO does not yet fully support BFO 2020, certain relevant concepts could not be reused directly. As also introduced in the CTO, NFDIcore-specific classes, such as dataset, document, and identifier, were introduced to fill the gaps [25].
- Schema.org: In MemO, as well as in CTO, schema.org was particularly used for describing creative works.¹⁰

 $^{^7} All\ MemO\ resources\ are\ available\ on\ GitHub:\ https://github.com/ISE-FIZKarlsruhe/memo/tree/main.$

⁸The current MemO web resource: https://nfdi.fiz-karlsruhe.de/4memory/.

⁹https://github.com/ISE-FIZKarlsruhe/memo/blob/main/docs/competency-questions.md

¹⁰ https://schema.org/

- RiC-O (Records in Contexts Ontology): MemO reuses structures from RiC-O to describe hierarchies in archives. Since MemO functions only as an index and does not need the full-fledged information that can be provided with RiC-O, only the part to represent collections and individual elements (rico:RecordSet, rico:Record, rico:Instantiation, rico:includes, rico:has or had holder, rico:has or had instantiation) is reused. These parts are aligned to information content entities and material entities in the BFO/IAO [26].
- PROV Ontology (PROV-O): Although the provenance ontology PROV-O is not directly used in MemO, there is an alignment with BFO available, which makes MemO compatible with descriptions provided in PROV-O [16].

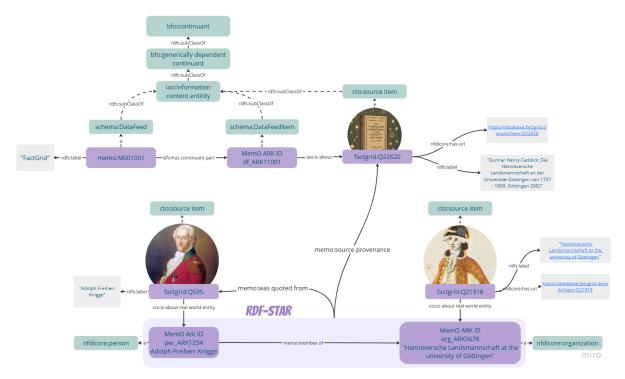


Figure 1: Representation of the modeling of provenance with MemO. Shows the ingestion of multiple cto: source items from a schema: DataFeed, and the representation of provenance information for claims made within them.

4.3. Conceptual Extensions as Provided in MemO

In the current version, MemO introduces properties to capture provenance and hierarchical document relations relevant for historical research. The following properties were defined in MemO:

- memo:resource provenance links a source (such as a document, a web page, etc.) to the schema:DataFeed resource it was ingested from (provenance information concerning the KG population).
- memo: source provenance specifies the provenance of a piece of information, e.g., the membership of a person to an organization provided in a repository, as established in a piece of secondary literature. It links a claim or information content entity to a source that is cited as evidence for its validity, mostly a reference to a publication.
- memo: was quoted from indicates that a claim is a literal quote from another cto: source item.
- memo: member of allows for the representation of membership relations (e.g., a person being a member of an organization).

In addition, as shown in Figure 1, MemO employs RDF-star to enable statements about statements, particularly to model provenance metadata at the level of individual triples.

4.4. Modeling Examples

To articulate MemO's capability to harmonize decentralized sources, concrete modeling cases from key infrastructures are integrated into the NFDI4Memory Data Space, functioning as use cases.

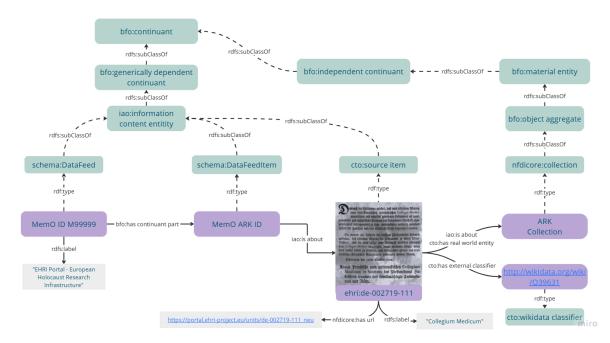


Figure 2: Representation of the modeling of resources with MemO. Shows the ingestion of schema: DataFeeds – so, entire data repos – and connected individual documents (cto:source_items).

EHRI provides metadata on Holocaust-related archival holdings, including country reports, descriptions of archival institutions, archival descriptions, and items, as well as their own or applied vocabularies and authority sets. In this use case (see Figure 2), MemO models metadata for an archival description from the portal as a schema: DataFeed, which is, by definition, an iao:information content entity that is composed of one or more schema:DataFeedElements, each of which is also an iao:information content entity. In the context of NFDI4Culture and NFDI4Memory, information on DataFeeds is collected manually.

Accordingly, a resource identified by the Archival Resource Key (ARK) identifier ehri:de-002719-111, representing the archival description of the "Collegium Medicum" collection, is a cto:source item and is linked to its corresponding physical entity. This real-world entity is associated with an nfdicore:collection and is assigned semantic identifiers, including authority links to external classifiers such as Wikidata (e.g., Q39631). The digital metadata record itself is classified as a schema:DataFeed and schema:DataFeedItem, allowing it to be treated as an information content entity (iao:information content entity) that is part of a broader data aggregation. ¹¹

The DDB aggregates digitized cultural materials. Its sister component Archivportal-D does the same for archival holdings. While the general integration of resources into the KG remains the same, via the representation of external classifiers or related entities (such as persons or documents), cross-connections to other portals, like EHRI, can be established. As shown in Figure 3, a relation between two previously unrelated archival documents could be drawn using the shared concept Q39631, the external Wikidata classifier referring to a physician.

The current version of the NFDI4Memory KG links central services of the participating institutions

¹¹IAO class to describe a generically dependent continuant that is about some thing. For more information, see the definition in the official documentation.

¹²https://www.archivportal-d.de/?lang=en

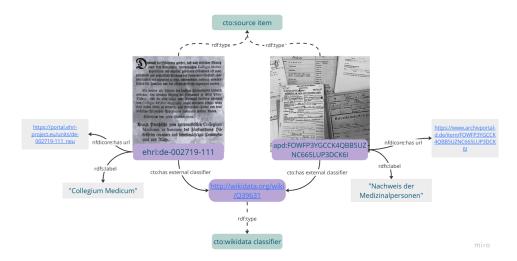


Figure 3: Representation of the interlinking of previously unrelated resources with MemO. The example displays two archival documents, one from EHRI and the other from Archivportal-D, that share the Wikidata classifier Q39631 (physician).

via common disciplines, technologies, and standards, as well as endpoints. Additional information can be gathered through a federated search to external repositories (e.g., Wikidata). ¹³

5. Tangible MemO: A Data Story About Cross-Connections

To illustrate the practical utility of the MemO KG, Figure 4 presents a concrete research scenario in which a fictional member of the NFDI4Memory community utilizes the offered approach. This data story, therefore, follows a historian's inquiry into the life and influence of a historical physician associated with early modern European universities. The central figure in this scenario is Wilhelm Bernhard Trommsdorff. Beginning his medical studies in Erfurt around 1756, he went on to a distinguished career serving as both the personal physician to Carl Theodor von Dalberg, the governor of the Electorate of Mainz, and a professor at the University of Erfurt's medical faculty. The research begins with a targeted inquiry: to identify documents referencing Trommsdorff across disparate archival collections. Traditionally, without the MemO KG, the researcher would face the challenge of conducting thorough searches through numerous unconnected archival repositories. By querying with relevant keywords (e.g., 'physician,' 'Erfurt,' 'Wilhelm Bernhard Trommsdorff'), using the interface with a facetted search, or using SPARQL, the researcher can access cross-connections enabled by MemO properties and identify relevant materials across different repositories from a single entry point. The MemO KG connects entities, including persons, organizations, and items. Shared authority data and classifiers enable the system to provide relevant materials from both FactGrid and Archivportal-D, revealing a more comprehensive set of sources.

6. Conclusion and Future Work

This position paper outlines the ongoing implementation of the integrated system, comprising MemO, the MemO KG, and the NFDI4Memory Data Space.

As a modular extension of NFDIcore and the CTO, MemO enables the structured representation of information on heterogeneous research data originating from multiple sources within the NFDI4Memory community. With queries crafted in collaboration with domain experts, the ontology addresses their needs directly and strives to provide reusable modules to describe provenance, contextual metadata, and other relevant aspects, thereby forming the schematic backbone of the Data Space. Thus far, it is possible

¹³https://www.wikidata.org/wiki/Wikidata:Main Page

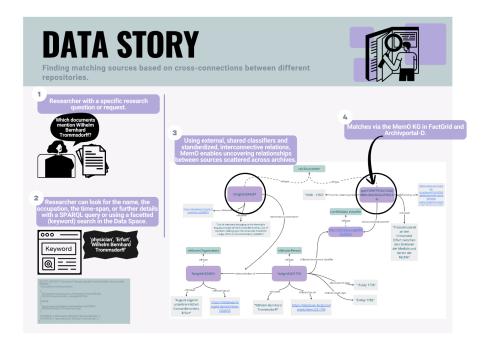


Figure 4: Data story illustrating how the MemO KG connects related sources across repositories.

to generate cross-connections between previously unconnected material from different institutions via the shared metadata, such as external classifiers, ingested into the KG using the ontological schema of MemO.

Next steps include finalizing the alignment of MemO with the latest versions of NFDIcore and CTO, expanding the ontology with additional domain modules (e.g., archival structures), and continuing to populate the KG using a structured pipeline.

Acknowledgments

This work was created as part of the NFDI consortium NFDI4Memory (www.4memory.de). We gratefully acknowledge the financial support of the German Research Foundation (DFG) – project number 501609550.

Declaration on Generative Al

During the preparation of this work, the authors used Grammarly and DeepL for grammar and spelling checks. The authors take full responsibility for the publication's content.

References

- [1] N. Hartl, et al., Nationale Forschungsdateninfrastruktur (NFDI), Informatik Spektrum 44 (2021) 370–373. URL: https://doi.org/10.1007/s00287-021-01392-6. doi:10.1007/s00287-021-01392-6.
- [2] J. Paulmann, et al., NFDI4Memory. Consortium for the historically oriented humanities. Proposal for the National Research Data Infrastructure (NFDI), 2022. doi:10.5281/zenodo.7428489.
- [3] T. Blanke, et al., The european holocaust research infrastructure portal, J. Comput. Cult. Herit. 10 (2017). doi:10.1145/3004457.
- [4] A. Meroño-Peñuela, et al., CLARIAH: Enabling Interoperability Between Humanities Disciplines with Ontologies, Applications and Practices in Ontology Design, Extraction, and Reasoning 49 (2020) 73–90. doi:10.3233/SSW200036.

- [5] E. Davis, B. Heravi, Linked data and cultural heritage: A systematic review of participation, collaboration, and motivation, J. Comput. Cult. Herit. 14 (2021). doi:10.1145/3429458.
- [6] S. Ondraszek, et al., NFDI4Memory Ontology (MemO), 2024. URL: https://nfdi.fiz-karlsruhe.de/ 4memory/ontology/.
- [7] T. Holste, M. Razum, Der NFDI4Memory Data Space als digitales Ökosystem für die historisch arbeitenden Wissenschaften, Gedächtnisinstitutionen und Informationsinfrastrukturen, ABI Technik 44 (2024) 259–267. doi:doi:10.1515/abitech-2024-0039.
- [8] J. Waitelonis, et al., NFDIcore Ontology, 2025. URL: https://nfdi.fiz-karlsruhe.de/ontology/3.0.1.
- [9] T. Tietz, et al., NFDI4Culture Ontology (CTO), 2025. URL: https://nfdi4culture.de/ontology/3.0.0.
- [10] B. O'Neill, L. Stapleton, Digital cultural heritage standards: from silo to semantic web, AI amp; SOCIETY 37 (2022) 891–903. doi:10.1007/s00146-021-01371-1.
- [11] O. Bruns, et al., What's cooking in the NFDI4Culture kitchen? A KG-based research data integration workflow, in: 4th Workshop on Metadata and Research (objects) Management for Linked Open Science (DaMaLOS), co-located with ESWC, 2024.
- [12] J. J. Steller, et al., Communities, Harvesting, and CGIF: Building the Research Data Graph at NFDI4Culture, in: Book of Abstracts DHd2024, Zenodo, Passau, Deutschland, 2024, pp. 131–135. doi:10.5281/zenodo.10698301.
- [13] V. A. Carriero, et al., ArCo: The Italian Cultural Heritage Knowledge Graph, in: The Semantic Web ISWC 2019, Springer International Publishing, Cham, 2019, pp. 36–52. doi:10.1007/978-3-030-30796-7_3.
- [14] P. Lisena, et al., Capturing the Semantics of Smell: The Odeuropa Data Model for Olfactory Heritage Information, in: The Semantic Web, Springer International Publishing, Cham, 2022, pp. 387–405. doi:10.1007/978-3-031-06981-9_23.
- [15] M. Doerr, The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata, AI magazine 24 (2003) 75–75.
- [16] T. Lebo, et al., PROV-O: The PROV ontology, Technical Report, World Wide Web Consortium, 2013.
- [17] C. Concordia, et al., Not just another portal, not just another digital library: A portrait of Europeana as an application program interface, IFLA Journal 36 (2010) 61–69. doi:10.1177/0340035209360764.
- [18] A. L. Silva, A. L. Terra, Cultural heritage on the Semantic Web: The Europeana Data Model, IFLA Journal 50 (2024) 93–107. doi:10.1177/03400352231202506.
- [19] T. Tietz, et al., From floppy disks to 5-Star LOD: FAIR research infrastructure for NFDI4Culture, in: 3rd Workshop on Metadata and Research (objects) Management for Linked Open Science (DaMaLOS), co-located with ESWC, 2023.
- [20] B. Smith, et al., Basic Formal Ontology (BFO) Version 2020-08-26, 2020. URL: http://purl.obolibrary.org/obo/bfo.owl.
- [21] R. Arp, et al., Building Ontologies with Basic Formal Ontology, The MIT Press, 2015. doi:10.7551/mitpress/9780262527811.001.0001.
- [22] J. Malone, et al., The Software Ontology (SWO): a Resource for Reproducibility in Biomedical Data Analysis, Curation and Digital Preservation, Journal of Biomedical Semantics 5 (2014) 25. URL: https://doi.org/10.1186/2041-1480-5-25. doi:10.1186/2041-1480-5-25.
- [23] M. Black, et al., EDAM: the Bioscientific Data Analysis Ontology, Poster presented at F1000Research, 11 (ISCB Comm J): 1, 2022. doi:10.7490/f1000research.1118900.1, version 1; not peer-reviewed. Open access.
- [24] E. Blomqvist, et al., Engineering Ontologies with Patterns The eXtreme Design Methodology., in: Ontology Engineering with Ontology Design Patterns Foundations and Applications, 2016, pp. 23–50. doi:10.3233/978-1-61499-676-7-23.
- [25] S. Arabandi, et al., Information artefact ontology (iao), http://purl.obolibrary.org/obo/iao.owl, 2024. Accessed: 2025-04-27.
- [26] International Council on Archives, Expert Group on Archival Description (EGAD), Records in contexts ontology (ric-o), https://ica-egad.github.io/RiC-O/, 2023. Version 1.0. International Council

on Archives.