From Culture to Core: Integrating Cultural Heritage Data into Cross-Domain Research Infrastructures

Tabea Tietz[1,2], Linnaea Söhn[3], Oleksandra Bruns[1,2], Jörg Waitelonis[1], Etienne Posthumus[1], Jonatan Jalle Steller[3], Torsten Schrade[3] and Harald Sack[1,2]

[1]FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Eggenstein-Leopoldshafen, Germany
[2]Institute of Applied Informatics and Formal Description Methods (AIFB) of KIT, Karlsruhe, Germany
[3]Academy of Sciences and Literature Mainz, Geschwister-Scholl-Straße 2, 55131 Mainz, Germany

Abstract

Within the National Research Data Infrastructure (NFDI) in Germany, the NFDI4Culture consortium addresses a critical challenge: unifying access to fragmented and semantically heterogeneous cultural heritage (CH) research data scattered across institutions and disciplines. This work presents the NFDI4Culture Ontology (CTO), a strategically designed lightweight ontology that successfully bridges the gap to represent CH research resources between specialized domain requirements and interoperability demands across diverse cultural heritage fields including musicology, performing arts, and architecture. CTO is aligned with the Basic Formal Ontology (BFO) on a foundational level and extends the established mid-level ontology NFDIcore, thereby maintaining the flexibility essential for capturing domain-specific nuances and is fully integrated into operational research data infrastructures. This contribution demonstrates how domain-specific ontologies can support both highly specialized research needs and broader cross-domain interoperability through modular architecture, and provides insights into proven modeling strategies, integration workflows, and lessons learned from a productive system in the CH domain.

Keywords

Ontologies, Research Data Management, Cross-domain federation, Cultural Heritage

1. Introduction

Infrastructures for research data management are increasingly shaped by the need to represent, integrate, and make scholarly resources discoverable across disciplines and institutional boundaries. One approach that has gained prominence in this context is the use of research knowledge graphs (RKGs), which enable structured, semantically rich, and machine-understandable representations of datasets, collections, software, and their interrelations[1, 2]. RKGs rely on well-designed ontologies that define a common vocabulary and formal structure to ensure long-term interoperability and cross-domain integration.

The cultural heritage (CH) research domain presents a complex and promising application area for such approaches. It encompasses a wide range of disciplines and institutions, each with its own standards, formats, and conceptual perspectives. Within the German National Research Data Infrastructure (NFDI)[1], the consortium NFDI4Culture[2] addresses this diversity by building a centralized semantic index for CH research data, aiming to make distributed and heterogeneous resources findable and interoperable. Central for this effort is the NFDI4Culture Knowledge Graph (NFDI4Culture-KG) that integrates metadata from a wide variety of sources across disciplines such as musicology, performing arts, media studies, architecture, and art history. The NFDI4Culture-KG is made accessible through the NFDI4Culture Information Portal[3], which provides a unified point of access to the collected research data. The portal functions as an interface for discovery and reuse, while the underlying KG infrastructure supports semantic integration and querying across datasets.

ISWC 2025 Companion Volume, November 2–6, 2025, Nara, Japan
0000-0002-1648-1684 (T. Tietz); 0000-0001-8341-1187 (L. Söhn); 0000-0002-8501-6700 (O. Bruns); 0000-0001-7192-7143 (J. Waitelonis); 0000-0002-0006-7542 (E. Posthumus); 0000-0002-5101-5275 (J. J. Steller); 0000-0002-0953-2818 (T. Schrade); 0000-0001-7069-9804 (H. Sack)

[1]https://www.nfdi.de/
[2]https://nfdi4culture.de/
[3]https://nfdi4culture.de/

To support this goal, the NFDI4Culture Ontology (CTO) has been developed as a domain-specific ontology tailored to represent material and immaterial CH research data in the NFDI4Culture-KG. CTO extends the mid-level ontology NFDIcore [3] and aligns with the Basic Formal Ontology (BFO) [4] and further widely used standards. The comprehensive development of the NFDI4Culture-KG and Portal has been presented in previous work [5, 6, 7, 8]. The following section presents CTOv3.0, which fundamentally redefines the ontology based on lessons learned from its integration into an operational system. This contribution discusses the underlying requirements, design principles, and cross-domain applicability of CTO, and reflects on its role in aligning community needs with broader interoperability objectives.

CTO has been released on github[4]. A documentation is provided[5] along with a generated list of resources[6]. The NFDI4Culture-KG, built on CTO and NFDIcore, is accessible via a SPARQL endpoint[7] and currently contains more than 70m triples. Its current statistics are available through a dashboard[8]

2. Related Work

The Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [9] supports data representation in the CH and digital humanities (DH) domains but is generally less suited for interoperability with data-intensive sciences, like materials science, where BFO has become the prevailing standard. VIVO [10] integrates BFO and other standards for scholarly information systems. It was developed for the VIVO software and primarily as a research information system for universities and research institutions. Although CIDOC-CRM enables interoperability within humanities research data, its event-centered design paradigm often prevents its applicability in natural sciences, life sciences, or engineering [11]. Standards such as DCAT [12], and Schema.org [13] support harmonization, cataloging, and web integration, yet lack the semantic depth or flexibility needed for advanced research data modeling. The Scientific Knowledge Graphs Interoperability Framework (SKG-IF)[9] addresses interoperability by providing a high-level reference model for aligning heterogeneous scholarly graphs. In contrast, NFDIcore and CTO offer concrete, BFO-aligned ontologies embedded in operational infrastructures.

3. The NFDI4Culture Ontology

The NFDI4Culture Ontology (CTO) was developed with a primary focus on enabling efficient semantic indexing of research data within the German cultural heritage community. The following section outlines its design principles and requirements, describes its modular architecture and alignment with mid-level and foundational ontologies, and highlights its domain-specific applications and implementation.

3.1. Design Considerations

CTO was initially developed in close collaboration with the CH research community through user stories, competency questions, workshops, and iterative testing within the live NFDI4Culture infrastructure [14]. From this process, key design requirements emerged: The ontology had to balance simplicity with the expressivity required for effective representation of CH research data better. Particular attention was needed for domains like musicology, which requires a higher degree of granularity and complexity than others. It was also essential to improve structured referencing of authority data and external vocabularies to enhance interoperability, and to reveal connections between disparate datasets. The representation of licenses and rights had to accommodate standardized license URIs and natural language statements, depending on the format and specificity of the source metadata. Also, the ontology needed to align with

[4] https://github.com/ISE-FIZKarlsruhe/nfdi4culture
[5] https://nfdi.fiz-karlsruhe.de/4culture/
[6] https://nfdi.fiz-karlsruhe.de/4culture/ontology/
[7] https://nfdi4culture.de/resources/knowledge-graph.html
[8] https://superset.nfdi4culture.de/superset/dashboard/culture-kg-kitchen/
[9] https://skg-if.github.io/

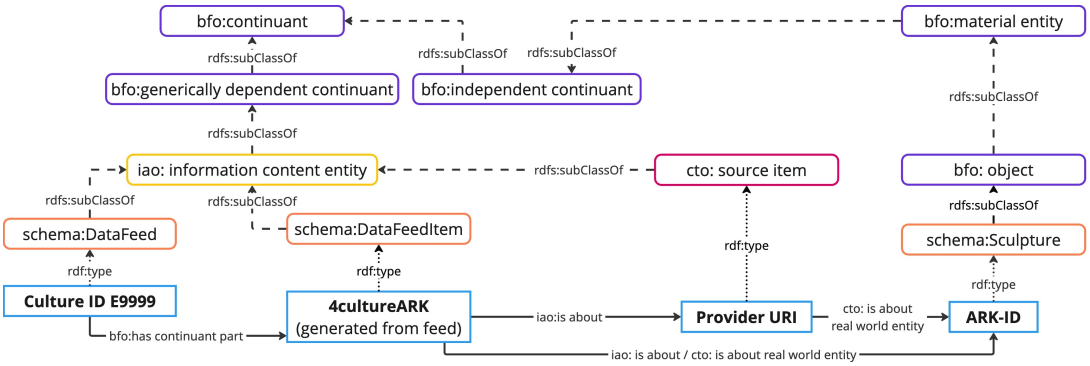

Figure 1: The core structure of the NFDI4Culture Ontology (CTO)

BFO 2020 to remain consistent with shared NFDIcore modeling practices. Ensuring the persistence and stability of identifiers, while keeping them independent from human-readable labels, was critical for maintaining referential integrity and reliable querying. Finally, the development process had to become more transparent, reproducible, and more open to collaborative contributions, supported by automated quality control mechanisms. These combined requirements informed the transition from an initially ultra-lightweight design to the current version 3.0, which provides increased semantic expressivity while preserving its suitability for integration into real-world systems.

3.2. Modular Architecture

CTO is built as an extension of NFDIcore, a BFO-aligned mid-level ontology developed collaboratively across several NFDI consortia. NFDIcore provides shared concepts for research infrastructure, datasets, persons, organizations, services, and identifiers, and is designed for reuse and extension. CTO builds on this foundation to capture domain-specific metadata for CH research resources, including cultural objects, events, performers, classifications, and media. CTO is one of four currently released domain-specific extensions of NFDIcore. The other extensions include the NFDI-MatWerk Ontology (MWO) [15] in the domain of materials science, the NFDI4Memory Ontology (MemO) [16] in the domain of history, and the NFDI4DataScience Ontology (NFDI4DSO) [17]. Next to the existing domain extensions, application oriented functional extensions are currently under development, including a provenance model and a machine learning component. This design supports interoperability across the diverse domains represented within the NFDI without sacrificing domain-specific expressivity and facilitates federated querying across a broad range of disciplines. The NFDIcore ontology has been released on github[10]. A documentation is provided[11] along with a generated list of resources[12]. NFDIcore has been evolving through regular stakeholder meetings, is maintained by an engaged community, supported defined release schedules, and clearly documented milestones.

3.3. Domain-Specific Modeling

CTO's implementation focuses on the representation of research data that is continuously harvested from the CH community via the NFDI4Culture Information Portal and integrated into the KG. As illustrated in 1, each source provided by the community is represented as a `schema:DataFeed`, with individual records modeled as `schema:DataFeedItem`, entities that receive persistent ARK identifiers[13], contain provenance metadata (e.g., license and timestamps) and function as a durable anchors in the KG. Content-related metadata are associated with a `cto:CTO_0001005` (source item), representing the

[10] https://github.com/ISE-FIZKarlsruhe/nfdicore
[11] https://ise-fizkarlsruhe.github.io/nfdicore/docs/
[12] https://ise-fizkarlsruhe.github.io/nfdicore/
[13] ARK-IDs (Archival Resource Key) are unique and persistent identifiers for information objects.

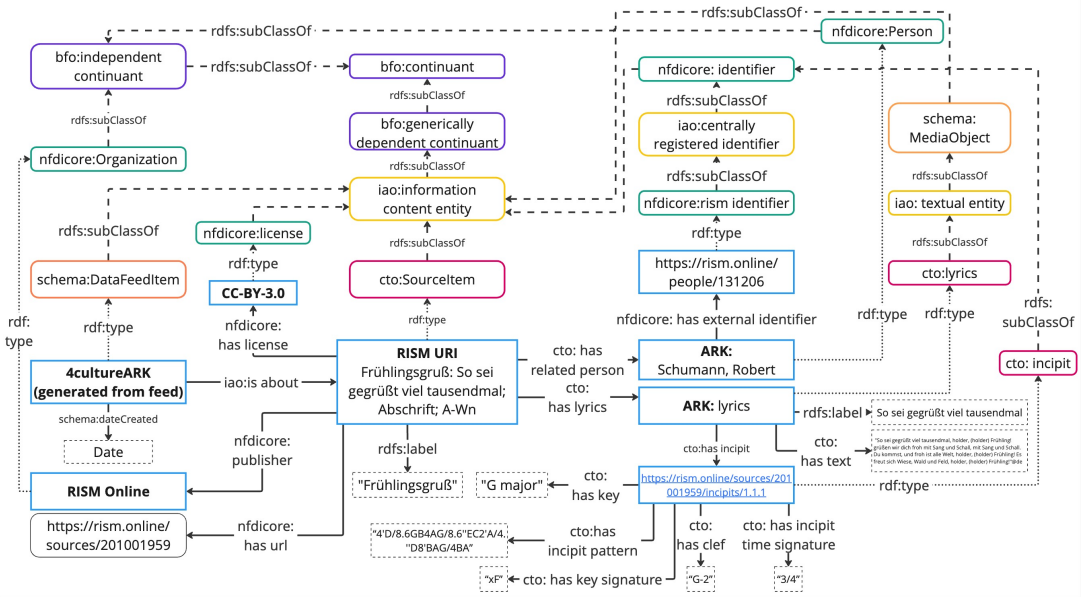

Figure 2: A CTO modeling example in the musicology domain

provided research resource. This includes media references, external identifiers, temporal metadata, and subject-specific details such as musical incipits (usually the first few bars of a musical piece, presented in standard music notation). This modeling approach supports lightweight, queryable linking across CH datasets. Figure 2 illustrates the use of CTO in the musicology domain. The source item *Frühlingsgruß* is published by the organization RISM Online[14], which aggregates musical records from international collections. The associated person *Robert Schumann* is referenced in the original provider data. To preserve the lightweight design, the relationship type only represented using the `cto:CTO_0001009` (has related person) property, because in the index it is merely relevant to know "*which persons are related to this data feed*". The same modeling approach applies to related events, organizations, and locations. An Ark-ID is assigned to the entity *Robert Schumann* in the KG due to its relation to *Frühlingsgruß* in the provider metadata. Since the RISM identifier is provided in the source metadata, and such identifiers must be queryable (e.g., "Select all entities with a RISM identifier"), the class `nfdicore:NFDI_00001016` (nfdicore: rism identifier) is introduced as a subclass of `iao:IAO_0000578` (iao: centrally registered identifier). Additionally, the representation of lyrics and incipits addresses a subject-specific requirement in the musicology domain.

3.4. Implementation

A dedicated ETL pipeline [7] transforms input metadata from partner repositories into RDF using CTO and NFDIcore and integrates the research metadata into the NFDI4Culture-KG and hence, the Culture Information Portal. The ontology development process is managed with the Ontology Development Kit (ODK), which integrates ROBOT tooling, GitHub Actions, and structured release workflows to ensure consistency, transparency, and reusability, while also supporting automated quality control through continuous validation checks [18].

3.5. Use Case

The applicability of CTO can be illustrated on the example of the digital letter edition *Ferdinand Gregorovius: Poesie und Wissenschaft. Gesammelte deutsche und italienische Briefe*[15] and their connections to musicological data from RISM Online. The Gregorovius edition contains 1,093 annotated pieces of

[14] https://rism.online/
[15] https://gregorovius-edition.dhi-roma.it/

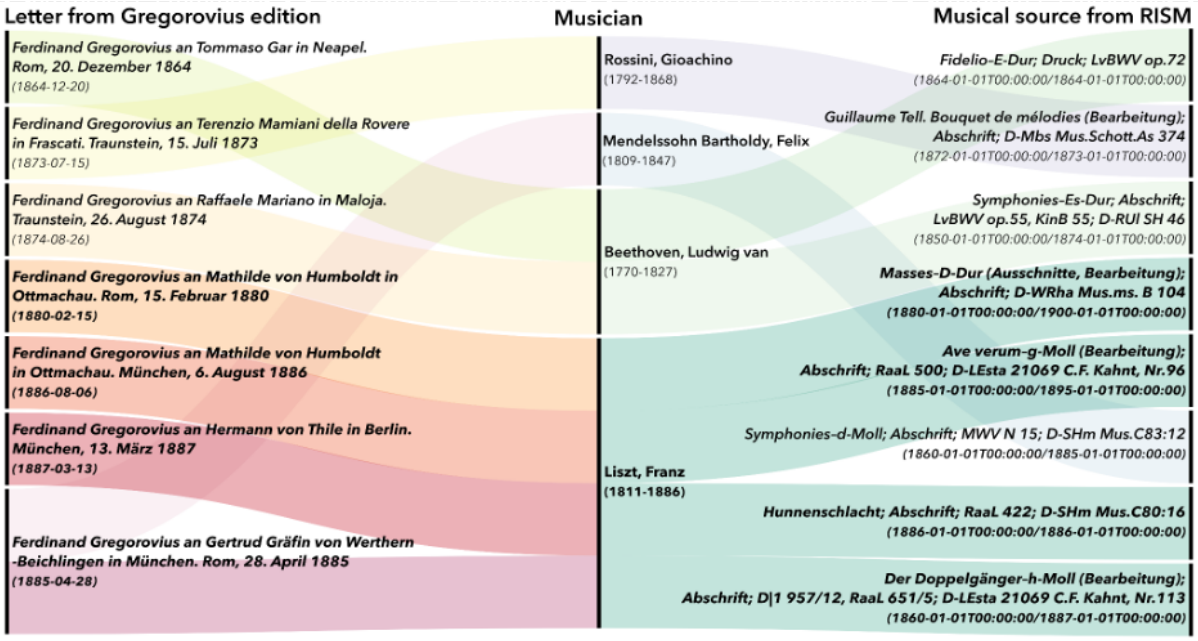

Figure 3: Example interconnections of persons in the Gregorovius letter edition and RISM Online

correspondence from the historian Ferdinand Gregorovius, whose heritage testifies a rich engagement with intellectual-historical movements and musicians of his time. Although the Gregorovius edition and RISM Online represent distinct data sources, they share overlaps in content and authority data. Prior to their integration into the KG, however, no common access point existed to query them jointly or to reveal potential interconnections for research and reuse. Figure 3 showcases connections on the level of persons between both data sets, which were revealed by means of a SPARQL query[16] [19].

4. Conclusion

CTO v3.0 offers a modular and semantically rich ontology for representing CH research data within a cross-domain infrastructure. Developed in close collaboration with domain communities and integrated into the NFDI4Culture-KG, it balances domain-specific granularity with broad interoperability via alignment with NFDIcore and BFO. Its application in real-world scenarios, demonstrates its potential for reuse, discovery, and integration across research domains. The generalizability of CTO v3.0 is further demonstrated by its direct reuse within the ontology of the NFDI4Memory consortium [16]. Future work will explore further means to support the CH community in discovering connections between distinct and previously unlinked cultural heritage datasets. A planned workshop will invite participants to query the NFDI4Culture-KG and develop data stories[17], highlighting how ontology-driven integration reveals relationships that remain hidden in fragmented collections. In addition, future work will showcase how the modular architecture facilitates cross-domain queries and enables the creation of meaningful connections across domains, including materials science, data science, and history.

Acknowledgements: This work is funded by Deutsche Forschungsgemeinschaft (DFG), project number 441958017.

Declaration on Generative AI: The authors used GPT-4 for grammar and spelling check. The authors reviewed and edited the content as needed and take full responsibility for the publication's content.

[16] https://nfdi4culture.de/go/kg-gregorovius-rism-example-musical-sources-letters
[17] https://datastories.nfdi4culture.de/

References

[1] M. Y. Jaradeh, A. Oelen, K. E. Farfar, M. Prinz, J. D'Souza, G. Kismihók, M. Stocker, S. Auer, Open Research Knowledge Graph: Next Generation Infrastructure for Semantic Scholarly Knowledge, in: Proceedings of the 10th International Conference on Knowledge Capture, 2019, pp. 243–246.

[2] O. Karras, J. Göpfert, P. Kuckertz, T. Pelser, S. Auer, Organizing Scientific Knowledge From Energy System Research Using the Open Research Knowledge Graph, arXiv preprint arXiv:2401.13365 (2024).

[3] J. Waitelonis, O. Bruns, T. Tietz, E. Posthumus, H. Beygi Nasrabadi, H. Sack, NFDIcore Ontology, 2025. URL: https://nfdi.fiz-karlsruhe.de/ontology/3.0.1.

[4] R. Arp, B. Smith, A. D. Spear, Building ontologies with Basic Formal Ontology, MIT Press, 2015.

[5] T. Tietz, J. Waitelonis, O. Bruns, E. Posthumus, H. Sack, NFDI4Culture Ontology (CTO), 2025. URL: https://nfdi4culture.de/ontology/3.0.0.

[6] O. Bruns, T. Tietz, L. Söhn, J. J. Steller, S. R. Ondraszek, E. Posthumus, T. Schrade, H. Sack, 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.

[7] O. Bruns, T. Tietz, L. Söhn, J. J. Steller, E. Poshumus, T. Schrade, H. Sack, Gotta Catch'em All: From Data Silos to a Knowledge Graph, in: Proceedings of 21st Extended Semantic Web Conference, ESWC 2024, Poster & Demos, 2024.

[8] J. J. Steller, L. C. Söhn, J. Tolksdorf, O. Bruns, T. Tietz, E. Posthumus, H. Fliegl, S. Pittroff, H. Sack, T. Schrade, Communities, Harvesting, and CGIF: Building the Research Data Graph at NFDI4Culture, in: DHd 2024 Quo Vadis DH (DHd2024), Passau, Deutschland, 2024.

[9] S. Borgo, R. Ferrario, A. Gangemi, N. Guarino, C. Masolo, D. Porello, E. M. Sanfilippo, L. Vieu, Dolce: A descriptive ontology for linguistic and cognitive engineering, Applied ontology 17 (2022) 45–69.

[10] J. Corson-Rikert, S. Mitchell, B. Lowe, N. Rejack, Y. Ding, C. Guo, The VIVO ontology, in: VIVO: A Semantic Approach to Scholarly Networking and Discovery, Springer, 2012, pp. 15–33.

[11] N. Beer, K. Herold, W. Kolbmann, T. Kollatz, M. Romanello, S. Rose, N.-O. Walkowski, Interdisciplinary interoperability, 2014.

[12] W. W. W. Consortium, et al., Data catalog vocabulary (dcat), 2014.

[13] R. V. Guha, D. Brickley, S. Macbeth, Schema. org: evolution of structured data on the web, Communications of the ACM 59 (2016) 44–51.

[14] T. Tietz, O. Bruns, L. Söhn, J. Tolksdorf, E. Posthumus, J. J. Steller, H. Fliegl, E. Norouzi, J. Waitelonis, T. Schrade, H. Sack, 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, Publisso, 2023.

[15] H. B. Nasrabadi, J. Waitelonis, E. Norouzi, K. Hubaiev, H. Sack, Nfdi matwerk ontology (mwo), https://github.com/ISE-FIZKarlsruhe/mwo, 2024. Revision: v3.0.0, accessed 2025-07-30.

[16] S. Ondraszek, et al., NFDI4Memory Ontology (MemO), 2024. URL: https://nfdi.fiz-karlsruhe.de/4memory/ontology/.

[17] G. A. Gesese, J. Waitelonis, Z. Chen, S. Schimmler, H. Sack, NFDI4DSO: Towards a BFO Compliant Ontology for Data Science, in: Proceedings of the 20th International Conference on Semantic Systems (SEMANTICS 2024), 2024.

[18] N. Matentzoglu, D. Goutte-Gattat, S. Z. K. Tan, J. P. Balhoff, S. Carbon, A. R. Caron, W. D. Duncan, J. E. Flack, M. Haendel, N. L. Harris, et al., Ontology Development Kit: a Toolkit for Building, Maintaining and Standardizing Biomedical Ontologies, Database 2022 (2022) baac087.

[19] L. C. Söhn, T. Tietz, J. J. Steller, P. Kehrein, A. Büttner, O. Bruns, E. Posthumus, J.-P. Grünewälder, J. Hörnschemeyer, C. Sander, V. Grund, H. Fliegl, H. Sack, T. Schrade, Nfdi4culture integration stories: Bridging gaps between isolated research resources, in: Digital Humanities 2025: Accessibility & Citizenship (DH2025), Zenodo, Lisbon, Portugal, 2025. URL: https://doi.org/10.5281/zenodo.16570076. doi:10.5281/zenodo.16570076.