What's Cooking in the NFDI4Culture Kitchen? A KG-based Research Data Integration Workflow

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

[1] FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany
firstname.lastname@fiz-karlsruhe.de
[2] Karlsruhe Institute of Technology (AIFB), Kaiserstr. 89, 76133 Karlsruhe, Germany
[3] Academy of Sciences and Literature Mainz, Geschwister-Scholl-Straße 2, 55131 Mainz, Germany firstname.lastname@adwmainz.de

Abstract. The National Research Data Infrastructure (NFDI) aims to provide a standardized and sustainable research data infrastructure across diverse domains, facilitating efficient research and scientific advancement. Despite encompassing a wide range of scientific disciplines, NFDI consortia share a foundation of common goals and concepts, emphasizing collaboration and data interoperability. Leveraging interconnected data offers new research opportunities, but requires availability in Linked Open Data (LOD) format. On example of NFDI4Culture, this paper addresses challenges of heterogeneous and isolated cultural heritage research data, and discusses efforts and results towards the creation of NFDI4Culture-KG, including the establishment of a research data index, implementing an ETL (Extract-Transform-Load) environment, and engineering lightweight semantic representations.

Keywords: research data · knowledge graphs · infrastructure · cultural heritage.

1 Introduction

Efficient research, preservation, reproducibility, and compliance of data are essential for the success of research projects and the advancement of scientific knowledge. The National Research Data Infrastructure (NFDI)[4] is a German national initiative with the aim of providing an organized, standardized, and sustainable research data infrastructure for diverse research domains, which are covered by respective NFDI consortia. For example, NFDI-MatWerk[5] is providing infrastructure for Materials Science Engineering, NFDI4DataScience[6] - in the domains of Data Science & Artificial Intelligence, MaRDI[7] – for mathematical research data, NFDI4Memory[8] – for historical research, etc. Despite covering specific scientific disciplines, all NFDI consortia share similar

[4] https://www.nfdi.de
[5] https://nfdi-matwerk.de/
[6] https://www.nfdi4datascience.de/
[7] https://www.mardi4nfdi.de/
[8] https://4memory.de/

goals and concepts such as structure, organization, people, institutions, areas of expertise, data repositories, devices, and much more [6]. When interconnected, such data may open up new research horizons, answering questions across research domains such as "Which materials contribute to the preservation and restoration of cultural heritage objects?", "What mathematical formulas were employed in the production and distribution of goods in ancient and medieval economies?", "What data science techniques are applied to the study of archaeological artifacts to identify ancient trade routes?". However, leveraging diverse research data for such insights requires the data to be available as Linked Open Data (LOD). To ensure this, a comprehensive workflow, which includes data discovery, harvesting, preprocessing, mapping to existing ontologies, integration into a Knowledge Graph (KG), etc. is required. This paper discusses the workflow created for this purpose on the case of research data in NFDI4Culture.

NFDI4Culture[9] is an NFDI consortium responsible for establishing an infrastructure for tangible and intangible CH research data following the FAIR principles[10]. The consortium's research data originates from a wide range of disciplines, including architecture and performing arts. This heterogeneity extends beyond disciplinary boundaries, encompassing variations in data standards, formats, and preparation status. While some datasets are already available as LOD, other collections are still being digitized or reside in legacy data repositories [9]. Moreover, most of the data provided to NFDI4Culture still exists in isolated silos. Despite the extensive use of authority files and controlled vocabularies like GND[11], VIAF[12], Wikidata[13], or Iconclass[14] for data organization, the data often lacks indexing and is not available for querying or and federation, making the navigation among various repositories impossible and difficult to traverse.

This paper reports on current efforts and results on building the NFDI4Culture-KG[15]. Specifically, it focuses on key aspects such as the creation of a research data index for cultural heritage research data, the establishment of a flexible data harvesting and integration pipeline, and the engineering of lightweight semantic representation for the research data. The NFDI4Culture-KG will act as a single point of access to decentralized research data resources [7], and will aggregate diverse and isolated data from the research landscape and thereby enable discoverability, interoperability and reusability of cultural heritage data.

The remainder of the paper is structured as follows. Section 2 discusses related work. In Section 3, an overview of the research data, the challenges and the standards are described. Section 4 contains the main contribution, followed by proof-of-concept in Section 5. Section 6 concludes this paper.

[9] https://nfdi4culture.de/
[10] https://www.go-fair.org/fair-principles/
[11] https://www.dnb.de/
[12] https://viaf.org/
[13] https://wikidata.org/
[14] https://iconclass.org/
[15] https://nfdi4culture.de/resources/knowledge-graph.html

2 Related Work

Throughout the last years, a number of research data infrastructure projects have been established for the humanities and related fields in Germany, all of whom NFDI4Cuture is cooperating with closely. Besides efforts within the NFDI, CLARIN[16] provides a research infrastructure for language, social and culture data, and DARIAH[17] provides data management and analysis services and tools to support research in the arts and humanities. Portal Kleine Fächer[18] maps the research landscape of rare disciplines, e.g. theatre studies or biblical archaeology, in Germany. AGATE[19] collects metadata on long-term research projects of German academies. Although all portals share the general goal of supporting digital research in the cultural heritage domain, they each offer a specialized infrastructure for researchers in their respective fields. NFDI4Culture has a broader focus on the cultural heritage sector and aims to interconnect research resources within the domain and across domains.

While NFDI4Culture is focused on providing infrastructure and services for the German cultural heritage research community, inline with these efforts are platforms that allow for storing and providing access to specific digital collections. For example, the German Digital Library (DDB)[20] is a digital platform that provides access to millions of digitized cultural heritage resources from libraries, archives, and museums across Germany. In addition to providing access to digital collections and archives, the DDB offers researchers search and discovery tools and other services. Linked (Open) Data Finland [4] is a framework that provides access to LOD from various domains, including Finnish cultural heritage. The platform provides various tools for researchers, such as SPARQL endpoints and visualizations to access and explore. The Dutch Digital Heritage Network[21] is a cooperative platform with the objective of enabling access and reuse of digital heritage collections in the Netherlands. The platform utilizes a KG-based method to establish connections among the collections and to enhance the metadata. Additionally, the Digital Public Library of America (DPLA)[22] is a digital platform that provides access to millions of heterogeneous digital resources from Galleries, Museums, Archives, and Libraries (GLAM) across the United States. Europeana[23] is a European initiative to provide access and connect digital collections across European GLAM institutions. Overall, the main contrast between NFDI4Culture and the mentioned digital libraries and platforms is in their objectives and areas of focus. While the latter aim to make digital collections accessible for exploration and discovery, NFDI4Culture's primary goal is not only to provide a centralized index of decentralized resources, but to also harvest and integrate meaningful research metadata as described in Section 4.1.

Ontologies are used to represent FAIR research data and enable their exploration across

[16] `https://www.clarin.eu/`
[17] `https://www.dariah.eu/`
[18] `https://www.kleinefaecher.de/`
[19] `https://agate.academy/`
[20] `https://www.deutsche-digitale-bibliothek.de/`
[21] `https://netwerkdigitaalerfgoed.nl/`
[22] `https://dp.la/`
[23] `https://www.europeana.eu/`

data sets, data portals and repositories. As one of the most prominent, the VIVO ontology [2] represents researchers and the full context in which they are working, including their outputs, interests, accomplishments, and associated institutions. The DCAT vocabulary[24] enables publishers to describe datasets and data services in a catalog on the Web. The goal is to ensure discoverability, enable a decentralized approach to publishing data catalogs and providing the possibility for federated search for datasets across catalogs in multiple sites. The Core Data Set for Research (KDSF)[25] represents research information for the German academic system with the goal to harmonize and standardize the research reporting system at universities and research institutions. CIDOC-CRM[26] is designed for the representation of cultural heritage information and the documentation of cultural objects, events, and the relationships between them. Although CIDOC-CRM is not specifically tailored for representing research data in a broader scientific or academic context, many researchers and institutions in the cultural heritage and digital humanities domains use this reference model to ensure the interoperability of their research data. ArCo is a knowledge graph of the Italian Cultural Heritage and consists of various ontology modules describing the data from the General Catalogue of the Italian Ministry of Cultural Heritage [1]. The NFDI4Culture ontologies (further explained in Section 4.2) are created in a modular approach and are mapped to a number of existing ontologies, including BFO[27], DCAT[28], DCTerms[29], and schema[30]. The aim is to 1) represent the research metadata of the cultural heritage communities described in section 3 and 2) ensure the interoperability with further NFDI consortia and communities including material science, mathematics, and data science.

3 Research Data Landscape

NFDI4Culture encompasses a vast landscape of heterogeneous data sources coming from different research disciplines. Despite their interconnectedness, these disciplines approach research from diverse perspectives, addressing diverse research questions. This section aims at providing an overview of the research data landscape utilized within culture-related research. It involves five academic disciplines contributing their research data to NFDI4Culture, while, given that culture data is inherently embedded within historical context, also incorporating historical research data provided by NFDI4Memory.

Architecture. The architectural research data landscape encompasses resources essential for advancing architectural practice and research. This includes graphical elements like drawings, sketches, photographs, and maps, along with multimedia recordings such as audio and video files capturing interviews, lectures, and design processes. Additionally, it comprises digital models like Building Information Modeling (BIM) files and

[24] https://www.w3.org/TR/vocab-dcat-3/
[25] https://kerndatensatz-forschung.de/
[26] https://www.cidoc-crm.org/
[27] https://basic-formal-ontology.org/
[28] https://www.w3.org/TR/vocab-dcat-3/
[29] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
[30] https://schema.org/

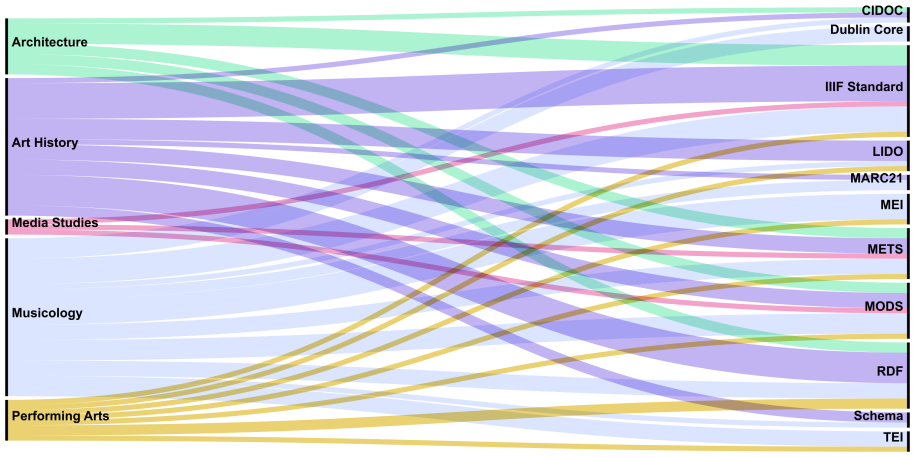

Fig. 1. Standards used in data portals in the Research Data Landscape of NFDI4Culture

CAD drawings. This diverse repository of data addresses key research questions in architecture, such as ”What tools should be used to reconstruct monuments that are no longer in existence?”, ”What construction materials and techniques were used to build an architecture object?”, ”Which phases were included into the process of conservation of an architecture object?”, ”Who worked on reconstructing an architecture object?”.

Art History. Art history involves examining visual artifacts and the visual representation within the multifaceted contexts – cultural, political, religious, economic, and artistic – that shape the visual aspects of a work of art. In this domain, research data encompasses diverse forms of artistic expression, photographs, paintings, sculptures, performances installations, audio-visual art works. Research questions may include ”What religious and spiritual symbolism is encoded in ornaments depicted in the architectural elements of an architecture object?”, ”What common motifs and themes are present in stained glass windows throughout a certain time period?”, ”What historical books by librettists include prints showing a certain iconographic subject?”.

Musicology. Musicological research aims at exploring cultural, historical, and aesthetic dimensions of music and musical practices. Musicological research data ranges from audiovisual recordings, musical instrument collections, biographies, to digitized musical scores, archival manuscripts, letters, diaries, concert programs, etc. Relevant research questions include: ”How do changes in instrument design and construction affect musical performance and composition?”, ”What events influenced a composer’s creative process and musical output?”, ”How do composers use harmony and rhythm to convey meaning?”.

Performing Arts. In the context of NFDI4Culture, the research of performing arts is related to theater and dance studies. The manifestation of the performing arts is based on performing events – intangible cultural heritage [8]. The data landscape ranges from audiovisual recordings of dance choreographies, theatrical plays, interviews; to written

materials, e.g choreographic notations, scripts, theater programs; to photographs and sketches of the stages, etc. Possible research questions include "Which theatres produced the literary works of a certain writer?", "Which actors were depicted on a stage during a certain time period and geography?", "How close is the performing arts work to its literary source?".

Media Studies is focused on examining content, historical evolution, and impact of diverse media forms and media systems on the society. Among the research data are diverse prints, e.g. magazines, newspapers, leaflets, audiovisual recordings, e.g. radio broadcasts, television. Potential research questions in this domain are "What media of certain period are classified as propaganda?", "Who was the most published author in underground newspapers during certain events, e.g. the student movement in the 1980s?", etc.

History. Historical sciences focus on analyzing past events, their causes, consequences, and correlations to understand the trajectory of human societies and cultures. Historians rely on primary and secondary sources and navigate through a diverse landscape of data, including books, letters, diaries, or transcripts, their editions, graphical or artistic sources such as maps, diagrams, or figurative artifacts. Research questions include "What influenced a particular political leader during a certain moment in history?", "How did cultural norms and societal expectations affect women's opportunities across diverse regions and eras?", "What drove the expansion of empires and civilizations across regions?" [5,3]. Political, religious, cultural history and other subdomains often have their own conventions and standards for organizing and analyzing historical data. Popular research data infrastructures include WissKI[31] utilizing CIDOC-CRM and Wikibase[32] employing the Wikidata data model, posing numerous challenges for data integration due to differences in data modeling and semantics.

Despite the apparent similarity of culture-related research data across the disciplines, the absence of the common workflow of data maintaining and storage across the disciplines and institutions results in creation of data silos. Moreover, each discipline adheres to varied formats and standards for data management, as visualized in Fig. 1. NFDI4Culture is aiming at aggregating diverse and isolated data from the cultural research landscape. However, achieving this is challenging, since data standards often require individual unique workflow solutions for data integration. Additionally, due to the great volume of data and frequent privacy sensitivity of cultural data, centralized storage solutions are beyond the scope of the project. On the other hand, federated infrastructures present their own challenges, as they require significant technological efforts from individual legacy institutions to provide reliable endpoints. In the following section, a hybrid solution pipeline for harvesting, integration, and interconnection of cultural research data is described.

4 Research Data Graph of NFDI4Culture

The research data targeted by NFDI4Culture is characterized by significant diversity, it spans across various research disciplines and is managed by various data providers,

[31] `https://wiss-ki.eu/`
[32] `https://wikiba.se/`

including universities, research institutes, art and music schools, academies, galleries, libraries, archives, museums, and individual researchers. The NFDI4Culture-KG aims at connecting research data, data sets, tools, infrastructures, and services from diverse data providers within the NFDI4Culture subject areas using Linked Open Data and facilitating FAIR principles.
In [9], specific requirements for the integration of NFDI4Culture data from the community were outlined, along with steps to address these requirements. One of the solutions was for data providers to implement a light-weight Culture Graph Interchange Format (CGIF)[33] to easily structure and index their legacy data, which would enhance the interoperability of the resources. However, despite the plan for community involvement in expanding the NFDI4Culture-KG, practical challenges have arisen. Many participants lack the necessary resources or technological expertise to contribute effectively. Recognizing this, the consortium is adapting its approach to provide additional support and solutions.
The NFDI4Culture-KG can conceptually be divided into Research Information Graph (RIG) and Research Data Graph (RDG). The RIG aims at exploration and retrieval of index and metadata for NFDI4Culture project resources, data resources, data services, persons and institutions, as e.g. "What data sets were provided to NFDI4Culture by libraries?", "What tools are used for automated architectural reconstruction?", "Who is a contact point of a certain data set?". While RDG's goal is to represent and interconnect the content of the data collections, and to answer research questions, as e.g. "Which historical books are written by librettists and include prints showing the iconographic subject "Hercules at the Crossroads?" (all user stories are available here[34]). This section describes a workflow towards the implementation of the RDG.

4.1 ETL Enviromnent – the Culture Kitchen

Integrating cultural research data poses significant challenges due to its existence in diverse formats and silos across various repositories. Additionally, the lack of LOD standards, SPARQL endpoints, and consistent metadata further complicates the integration process. Moreover, often even harvesting data from these repositories becomes a formidable task. Recognizing these challenges, an ETL (Extract, Transform, Load) environment "Culture Knowledge Graph Kitchen"[35] has been designed. This environment enables the systematic harvesting, cleaning, mapping and integration of research cultural data into the NFDI4Culture-KG. It consists of six modular workflow components, adaptable for independent use or within a comprehensive automated ingest routine, which aims at converting diverse data feeds[36] into a standardized format compatible with the NFDI4Culture-KG. Once harvested and integrated, resources are accessible

[33] `https://docs.nfdi4culture.de/ta5-cgif-specification`
[34] `https://nfdi4culture.de/resources/user-stories.html`
[35] `https://gitlab.rlp.net/adwmainz/nfdi4culture/knowledge-graph/culture-kg-kitchen`
[36] In the context of the ETL process, discussions often revolve around data feeds rather than datasets due to the emphasis on continuous streams

through a SPARQL endpoint[37] and SHMARQL[38] – a SPARQL endpoint explorer developed in NFDI4Culture, as well as a dashboard[39] for analysis and visualizations. This section will briefly outline the data integration workflow.

Step 1: Consume. First, an RDF/Turtle action file is generated, outlining schema-based definitions for the ingestion workflow, linking data feeds to metadata in the RIG. Due to the diverse nature of data sources, flexible routines to follow the steps are implemented that delegate individual tasks to specialized workflows and tools. Depending on the input data, those workflow tools cover various integration needs, including generating persistent identifiers, harvesting embedded metadata (CGIF[40]), harvesting and transforming LIDO[41] annotated data, extracting data from dedicated endpoints with unique annotation schemes, mapping harvested data to the Culture Ontology (see Section 4.2), and storing the converted data in the ETL environment/Culture Knowledge Graph Kitchen.

Step 2: Rinse. After storing the resulting RDF/Turtle files of the resources, the data is cleaned to harmonize the harvested data feed with the associated data file and prevent conflicts with information present in the consumed resources that may contradict the definitions in the action files. Based on the information in the action file, triples are added or deleted.

Step 3: Oven. The third component streamlines the Git commands. This facilitates efficient management of version control, allowing for easy status checks, branch switching, and handling of bulk operations. The system is designed to accommodate different modes of operation, including dry runs for testing, committing changes, and pushing modifications to remote branches. Changes in the metadata of data feeds are committed to the ETL environment's Git repository.

Step 4: Stash. If changes are pushed, a fourth component enables the management and generation of data directories – "stashes". Depending on whether a data directory for a specific data feed already exists, a stash is either updated or created and then made available via a SPARQL endpoint and SHMARQL.

Step 5: Deliver. To avoid downtimes, building the new endpoint is realized through a Docker-based delivery workflow. If a new endpoint needs to be built, environment variables are updated to point to the new container and port, and the new container is started. After the new SPARQL endpoint is available, the old container is stopped and removed.

Step 6: Bill. The last component provides statistics about the integrated data feeds through the Culture Knowledge Graph Dashboard for data analysis and visualizations based on the execution of provided SPARQL queries.

The Culture Kitchen serves as a crucial means in providing index data and extracting essential information based on community requirements for culture related domains. Specifically tailored workflows for every use case are developed within the Culture Kitchen to map the extracted data to the NFDIcore Ontology and the Culture Ontology,

[37] `https://nfdi4culture.de/sparql`
[38] `https://nfdi.fiz-karlsruhe.de/shmarql`
[39] `https://nfdi4culture.de/go/kg-kitchen-dashboard`
[40] `https://docs.nfdi4culture.de/ta5-cgif-specification`
[41] `https://www.lido-schema.org/schema/latest/lido.html`

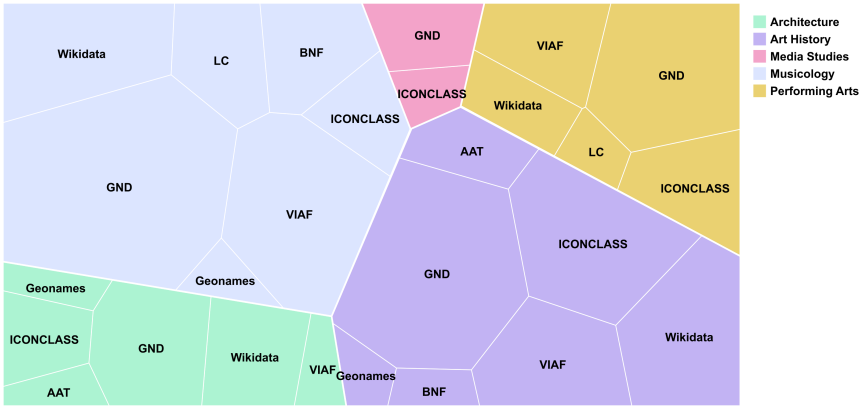

Fig. 2. Distribution of external vocabularies used in data portals across research disciplines in NFDI4Culture

aligning it with community standards and needs and ensuring interoperability. This curated data is then published in the NFDI4Culture-KG, where it is accessible through a SPARQL endpoint and SHMARQL. While not all original data may be fully integrated into the NFDI4Culture-KG, the focus remains on incorporating the most pertinent information that ensures interoperability across culture domains.

4.2 Culture Ontology within NFDI

Similarly to NFDI4Culture, each NFDI consortium established an interoperable research data infrastructure tailored to its specific domain. To facilitate this, the NFDIcore ontology[42] was developed and serves as a top level ontology for representing metadata about NFDI resources such as individuals, organizations, projects, data portals, etc. Recognizing the diverse needs of consortia, NFDIcore establishes mappings to a wide array of standards across domains, including the Basic Formal Ontology (BFO), schema.org, DCTERMS, and DCAT, which is crucial for advancing knowledge representation, data exchange, and collaboration across diverse domains. To answer domain-specific research questions, NFDIcore is extended following a modular approach, as e.g. with the **c**ul**t**ure **o**ntology module (CTO)[43], **m**at**w**erk **o**ntology module (MWO)[44].
In NFDI4Culture, the RIG facilitates exploration and retrieval of index and metadata for NFDI4Culture resources, while the primary objective of the RDG is to represent and interconnect the content of distributed data collections to address domain related research questions. Numerous cultural institutions describe their data and provide links to external vocabularies, see Fig. 2. However, these links typically remain underutilized,

[42] `https://github.com/ISE-FIZKarlsruhe/nfdicore/tree/main`
[43] `https://gitlab.rlp.net/adwmainz/nfdi4culture/knowledge-graph/culture-ontology`
[44] `https://git.rwth-aachen.de/nfdi-matwerk/ta-oms/mwo`

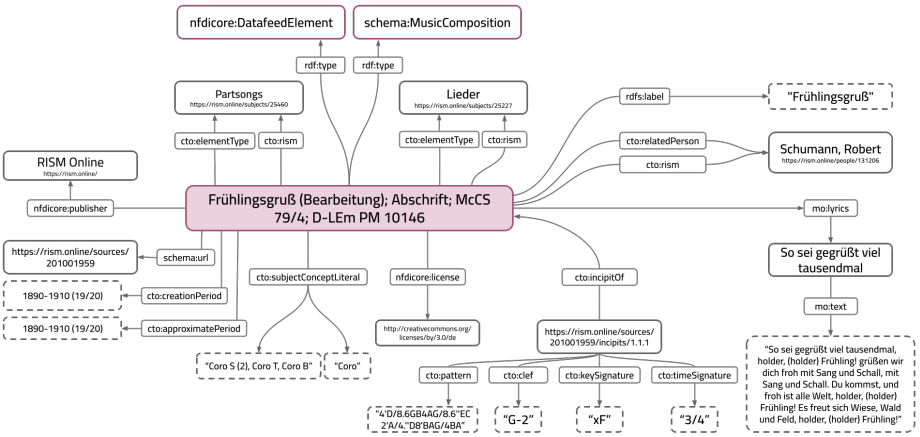

Fig. 3. An example of representing a musical composition "Frühlingsgruß" with CTO.

serving little purpose beyond manual exploration. The CTO seeks to transform this by extracting essential information from source data, including these links, and enable data interconnection and querying. Instead of representing all levels of granularity of culture data across every discipline comprehensively, the main goals of CTO are indexing the data, materializing links to external vocabularies and standardizing diverse formats.

Based on the project research questions, the scope and the research culture data, the CTO provides a light-weight vocabulary to achieve this level of research granularity in the RDG: the ontology establishes a connection between a data resource (stored and described in the RIG) and its individual component – `cto:DatafeedElement`. This component is associated with its corresponding type, e.g. `schema:MusicComposition`, `schema:ImageObject`, `nfdicore:Event`, etc. Each data feed element can then be characterized by its types (`cto:elementType`), subjects or themes (`cto:subjectConcept`), and related concepts, including creative works, persons, locations, and temporal information. This is achieved by linking the element to the corresponding concepts in the external vocabularies, e.g. `cto:gnd, cto:wikidata, cto:iconclass` (for example, see Figure 3[45]). Additionally, for more detailed research, each data feed element is linked to its source file (`cto:sourceFile`), where all the information contributed by a data provider is stored and represented using the domain-appropriate standard.

The following section demonstrates the CTO and the Culture Kitchen on the example of the RISM Online service as part of the musicology domain.

5 Proof of Concept: Musicology

RISM, the Répertoire International des Sources Musicales[46], is a service dedicated to cataloging and preserving musical sources from various historical periods. They maintain centralized databases using relational database management systems to organize metadata about musical sources, including composer, title, date, and location. RISM is part of the musicology data landscape described in Section 3 and therefore integral to capture in the NFDI4Culture-KG. The RISM Online service itself offers a multifaceted complex search within its own database. However, in order to get a simple and lightweight overview of composition types, licenses, mentioned persons or time frames the navigation is rather complex, takes time and expertise, and a federated search beyond the RISM Online database is not possible.

Corresponding to the above mentioned pipeline of the NFDI4Culture Kitchen, the RISM Online Harvester[47] was created for a smooth integration into the Research Data Graph (RDG). It was designed to fetch and process JSON-LD data from the RISM API and to create DataFeedItems for musical compositions, enriching them with detailed information such as composer, lyricist, musical keys, and other relevant metadata. The KG integration steps include: 1) Capturing RISM Online in the NFDI4Culture Portal and RIG[48]. 2) Harvesting all source records with music incipits from RISM online. 3) Cleaning the data of any invalid IRIs. 4) Converting all data from JSON-LD to Turtle for data ingest. 5) Merging harvested files to a single Turtle file. 6) Mapping the turtle file to the CTO ontology. 7) Integrating the data into the NFDI4Culture-KG for search and exploration.

Fig. 3 visualizes the representation of the musical composition (schema:MusicComposition) ”Frühlingsgruß” with CTO. Users can query connections of this nfdicore:DatafeedElement with related persons, lyrics, date and type information within and beyond the RISM Online dataset by means of federation. To further satisfy the requirements by musicologists, a service for an incipit search[49] was provided based on the aforementioned steps. The provided Culture Kitchen Dashboard reveals that currently the NFDI4Culture-KG contains (among other items) 1.9M music sheets with incipits and 93K persons as a part of RISM[50]. A similar procedure is implemented for further domains described in Section 3. The modeling examples are documented on the Web[51].

---

[45] more examples from other disciplines can be found on the Web: `https://nfdi.fiz-karlsruhe.de/4culture/rdg`

[46] `https://rism.online/`

[47] `https://gitlab.rlp.net/adwmainz/nfdi4culture/knowledge-graph/rism-online-harvester`

[48] `https://nfdi4culture.de/id/E4227`

[49] `https://nfdi4culture.de/kg-incipit-search.html`

[50] `https://nfdi4culture.de/go/kg-kitchen-dashboard`

[51] `https://nfdi.fiz-karlsruhe.de/4culture/rdg`

6 Conclusions

This paper contributes the current efforts on building the NFDI4Culture-KG with a focus on the creation of a research data index for cultural research data, the establishment of a flexible data harvesting and integration pipeline, and the engineering of lightweight semantic representation for the research data via ontologies. A proof-of-concept was described based on the musicology domain. Future work includes the provision of Culture Kitchen based harvesting and integration tools for further research data standards and formats as well as a deepened interconnection with other NFDI consortia.

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

References

1. Carriero, V.A., Gangemi, A., Mancinelli, M.L., Marinucci, L., Nuzzolese, A.G., Presutti, V., Veninata, C.: ArCo: The Italian Cultural Heritage Knowledge Graph. In: The Semantic Web–ISWC 2019: 18th Int. Semantic Web Conference, Auckland, New Zealand, Proceedings, Part II 18. pp. 36–52. Springer (2019)
2. Corson-Rikert, J., Mitchell, S., Lowe, B., Rejack, N., Ding, Y., Guo, C.: The vivo ontology. In: VIVO: A Semantic Approach to Scholarly Networking and Discovery, pp. 15–33. Springer (2012)
3. Hiltmann, T.: Vom Medienwandel zum Methodenwandel: Die fortschreitende Digitalisierung und ihre Konsequenzen für die Geschichtswissenschaften in historischer Perspektive. In: Konzepte, Methoden und Kritiken Digitaler Geschichtswissenschaft, pp. 13–44. De Gruyter Oldenbourg (2022)
4. Hyvönen, E.: Linked Open Data Infrastructure for Digital Humanities in Finland. DHN 2020 (2020)
5. Piersma, H., Ribbens, K.: Digital Historical Research: Context, Concepts and the Need for Reflection. BMGN - Low Countries Historical Review **128**(4), 78–102 (2013)
6. Sack, H., Schrade, T., Bruns, O., Posthumus, E., Tietz, T., Norouzi, E., Waitelonis, J., Fliegl, H., Söhn, L., Tolksdorf, J., Steller, J.J., Azócar Guzmán, A., Fathalla, S., Zainul Ihsan, A., Hofmann, V., Sandfeld, S., Fritzen, F., Laadhar, A., Schimmler, S., Mutschke, P.: Knowledge Graph Based RDM Solutions: NFDI4Culture - NFDI-MatWerk - NFDI4DataScience. In: 1st Conf. on Research Data Infrastructure (2023)
7. Tietz, T., Bruns, O., Fliegl, H., Posthumus, E., Schrade, T., Sack, H.: Knowledge Graph-basierte Forschungsdatenintegration in NFDI4Culture. In: Busch, A., Trilcke, P. (eds.) DHd2023: Open Humanities, Open Culture (2023)
8. Tietz, T., Bruns, O., Sack, H.: A Data Model for Linked Stage Graph and the Historical Performing Arts Domain. In: Proc. of Int. WS on Semantic Web and Ontology Design for Cultural Heritage (SWODCH) co-located with ISWC (2023)
9. Tietz, T., Bruns, O., Söhn, L., Tolksdorf, J., Posthumus, E., Steller, J.J., Fliegl, H., Norouzi, E., Waitelonis, J., Schrade, T., Sack, H.: 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)