

# From Floppy Disks to 5-Star LOD: FAIR Research Infrastructure for NFDI4Culture

Tabea Tietz<sup>1,2</sup>, Oleksandra Bruns<sup>1,2</sup>, Linnaea Söhn<sup>3</sup>, Julia Tolksdorf<sup>3</sup>, Etienne Posthumus<sup>1</sup>, Jonatan Jalle Steller<sup>3</sup>, Heike Fliegl<sup>1</sup>, Ebrahim Norouzi<sup>1,2</sup>, Jörg Waitelonis<sup>1</sup>, Torsten Schrade<sup>3</sup>, and Harald Sack<sup>1,2</sup>

<sup>1</sup> FIZ Karlsruhe – Leibniz Institute for Information Infrastructure,  
Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany  
`firstname.lastname@fiz-karlsruhe.de`

<sup>2</sup> Karlsruhe Institute of Technology (AIFB), Kaiserstr. 89, 76133 Karlsruhe, Germany

<sup>3</sup> Academy of Sciences and Literature Mainz, Geschwister-Scholl-Straße 2, 55131 Mainz, Germany  
`firstname.lastname@adwmainz.de`

**Abstract.** NFDI4Culture is establishing a knowledge graph-based infrastructure for research data on material and immaterial cultural heritage in the context of the German National Research Data Infrastructure (NFDI) in compliance with the FAIR principles. The NFDI4Culture Knowledge Graph is developed and integrated with the Culture Information Portal to aggregate heterogeneous and isolated data from the culture research landscape and thereby increase their discoverability, interoperability and reusability. This paper presents the research data management strategy in the long-term project NFDI4Culture, which combines a CMS and a knowledge graph-based infrastructure to enable an intuitive and meaningful interaction with research resources in the cultural heritage domain.

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

## 1 Introduction

The research landscape in the domain of cultural heritage in Germany is characterized by its immense diversity and heterogeneity. The research community covers a wide range of academic fields, e.g. architecture, art history, theatre, musicology, and media studies, using and creating research resources in various data types from texts and images over 3D models to music sheets and AV-files. A particular challenge in the area of cultural heritage data is furthermore the variance in data quality and the availability of research data in various formats. Some projects provide their data as 5-star Linked Open Data (LOD) and make them accessible via open interfaces, in other projects they are not available in queryable form or they are still being digitized or reside in legacy data repositories. However, the efficiency and quality of research depends on the way research resources can be found and accessed, interconnected, published and reused as well as on finding expertise in research projects.

The National Research Data Infrastructure (NFDI)<sup>4</sup> is a German national initiative with the aim of providing an organized, standardized, and sustainable research data infrastructure across all scientific domains. NFDI4Culture is one out of overall 27 NFDI

<sup>4</sup> <https://www.nfdi.de>

consortia and is responsible for establishing an infrastructure for research data on tangible and intangible cultural assets in compliance with the FAIR principles<sup>5</sup>. Within NFDI4Culture, a Knowledge Graph (KG) is developed and integrated with the Culture Information Portal<sup>6</sup> with the goal to aggregate heterogeneous and isolated data from the research landscape focused on by NFDI4Culture and thereby increase the discoverability, interoperability and reusability of cultural heritage data. The NFDI4Culture Knowledge Graph will act as a single point of access<sup>7</sup> to decentralized research resources provided by both NFDI4Culture and the research landscape via a SPARQL endpoint [7]. The NFDI4Culture-KG integrates research resource metadata provided by the culture community, which involves datasets, services, repositories, and software as well as organizations and persons involved in research and best practices to be followed. The Culture Information Portal provides a unified access to these decentralized resources and enables institutions, academies, universities and individual researchers to use and extend the data in the KG.

Goal of this paper is to present the research data management (RDM) strategy in the long-term project NFDI4Culture (running from 2020-2025), which combines a content management and knowledge graph-based infrastructure to enable an intuitive and meaningful interaction with research resources in the cultural heritage domain. This paper contributes intermediate results on the Culture Information Portal, the NFDI4Culture-KG and related ontologies. Although NFDI4Culture is an ongoing project, all contributions described in the paper are publicly available and already in use by the consortium and the research community it encompasses.

This paper is structured as follows: Section 2 gives an overview of initiatives and software solutions relevant to the NFDI4Culture approach. In Section 3, the challenges and requirements with respect to the data integration and the portal infrastructure are summarized, followed by a Culture Information Portal and NFDI4Culture-KG architecture description in section 4. First exploration and data integration use cases are provided in section 5 and section 6 concludes this paper.

## 2 Related Work

In recent years, there has been a growing emphasis on Research Data Management (RDM) and compliance with FAIR principles. However, due to the large amount of data generated and the need for collaboration among researchers, the process of storing data has become highly complex. As part of this effort, NFDI4Culture is an infrastructure project at national level in Germany with multiple areas of focus and goals. This section presents work related to the NFDI4Culture initiative, in particular the development and provision of tools and resources for the creation and management of Knowledge Graphs; development of research infrastructure for connecting and managing research data, and services for exploration and access of cultural heritage resources.

One of the key components of the NFDI4Culture infrastructure is the use of KGs, which have shown to provide useful means in RDM due to their compliance to the FAIR data

<sup>5</sup> <https://www.go-fair.org/fair-principles/>

<sup>6</sup> <https://nfdi4culture.de/>

<sup>7</sup> <https://nfdi4culture.de/resources/knowledge-graph.html>

principles and the possibility to represent data and its relationships in a flexible and scalable way. In this context, there exist several software that facilitate the creation and management of research KGs in the cultural heritage domain. For instance, Wikibase<sup>8</sup> is an open-source software that enables collaborative KG creation and management. However, the current limitations of Wikibase in integrating external ontologies and developing custom input forms may restrict semantic expressiveness and intuitive interaction with the NFDI4Culture-KG. On the other hand, WissKI<sup>9</sup> is a web-based platform for RDM that allows scholars to create, manage, and publish research data with features such as customizable taxonomies and semantic search capabilities. WissKI is designed for the implementation of CIDOC-CRM [4]. Due to its semantic expressiveness, CIDOC-CRM is a popular choice in the GLAM sector (involving Galleries, Libraries, Archives and Museums). However, its high level of complexity and abstract workflows makes it difficult to use for those who are not already familiar with the standard [3]. Further software for KG creation is described and evaluated in [6]. Overall, NFDI4Culture aims to represent relevant information in compliance with the FAIR principles, provide them as LOD, and offer an intuitive search and access point while remaining compatible with existing infrastructures. Therefore, a solution was needed that integrates seamlessly with the existing technical infrastructure (such as identity and access management or a registry for tools and services) and all research information curated via the Culture Information Portal (like persons, organizations, research products, services, events etc.). At the same time, the system needs to allow the easy integration of external ontologies and the application and publication of a dedicated NFDI4Culture ontology as well. Due to their current limitations, neither of the mentioned tools could be directly utilized within the portal infrastructure. In NFDI4Culture, TYPO3 CMS with an additional RDF extension layer is being used to meet the project goals and requirements (see section 4.1).

Similarly to the Culture Information Portal, several other research infrastructures have been established in other research domains in the humanities throughout the last years (especially in linguistics and philologies). NFDI4Culture closely cooperates with three other NFDI consortia in the humanities, namely NFDI4Memory<sup>10</sup> (historical data), NFDI4Objects<sup>11</sup> (archeological data), and Text+<sup>12</sup> (text and language based data). Besides the consortia of the NFDI, CLARIN<sup>13</sup> facilitates access to language resources and tools for humanities and social sciences researchers, while DARIAH<sup>14</sup> provides data management and analysis services and tools to support research in the arts and humanities. The Portal Kleine Fächer<sup>15</sup> maps the research landscape of small disciplines, e.g. theatre studies or biblical archaeology, in Germany. Likewise, AGATE<sup>16</sup> collects meta-

<sup>8</sup> <https://wikiba.se/>

<sup>9</sup> <https://wiss-ki.eu/>

<sup>10</sup> <https://4memory.de/>

<sup>11</sup> <https://www.nfdi4objects.net/>

<sup>12</sup> <https://www.text-plus.org/>

<sup>13</sup> <https://www.clarin.eu/>

<sup>14</sup> <https://www.dariah.eu/>

<sup>15</sup> <https://www.kleinefaecher.de/>

<sup>16</sup> <https://agate.academy/>

data on long-term research projects of German academies, including their relationships with institutions and researchers, as well as their disciplinary classification, research methods, and topics. 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 within the domain.

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)<sup>17</sup> is a digital platform that provides access to millions of digitized cultural heritage resources from libraries, archives, and museums across Germany. In addition to access to digital collections and archives, the DDB offers researchers search and discovery tools and other services. Linked (Open) Data Finland [5] is a platform 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. Similarly, the Dutch Digital Heritage Network<sup>18</sup> 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)<sup>19</sup> is a digital platform that provides access to millions of heterogeneous digital resources from GLAM across the United States. Europeana<sup>20</sup> 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 to provide a complex centralized index of decentralized resources, but to integrate fewer and meaningful metadata, and thus, enable exploration across the collections by means of federation.

### 3 Implementation Challenges and Requirements

The diversity of the research landscape in CH and heterogeneity of research resources provides a number of challenges faced within NFDI4Culture with regard to the technical infrastructure and knowledge representation, which are described in section 3.1. Furthermore, section 3.2 outlines several requirements that were jointly developed with domain experts and must be taken into consideration.

#### 3.1 Infrastructure and Data Integration Challenges

The research landscape that NFDI4Culture is focusing on is characterized by strong diversity. It includes not only a variety of research disciplines but also different types of

<sup>17</sup> <https://www.deutsche-digitale-bibliothek.de/>

<sup>18</sup> <https://netwerkdigitaalerfgoed.nl/>

<sup>19</sup> <https://dp.la/>

<sup>20</sup> <https://www.europeana.eu/>

organizations, such as university institutes, art and music schools, academies, galleries, libraries, archives, museums, and individual researchers. Accordingly, the research processes and resources that need to be made discoverable, interoperable, and reusable are also heterogeneous and are not only available in diverging standards and formats but also in different states of preparation. Collections are not always completely digitized and indexed, which is why only a few descriptive metadata are often available. Other datasets are already fully available as Linked Open Data (LOD) and can be easily linked to the NFDI4Culture-KG. The implementation of the FAIR data principles is a key goal of the consortium. Therefore, dedicated ontologies are provided, and measures are taken to help all actors transform their own data and resources into Linked Open Data in a sustainable and long-term manner. Research data management guided by academic communities requires the active participation of everyone. Therefore, the consortium offers extensive participation opportunities for users of all involved disciplines, as well as for artists and cultural professionals from different fields and representatives of civil society. It aims to represent the broad spectrum of different actors in the cultural heritage sector. Among other things, it is planned that the research community itself will expand the Knowledge Graph with its own resources and maintain its content. The condition for this is a technical infrastructure that allows resources to be curated, added, linked, and searched intuitively and without deep technical knowledge. This infrastructure will enable a semantic expressive representation of the data to ensure a comprehensive implementation of the FAIR principles.

### 3.2 Culture Community Requirements

The requirements for the NFDI4Culture data integration, which concerns the Culture Information Portal, the NFDI4Culture-KG and the ontology development have been derived from user stories collected together with the culture community at the beginning of the NFDI4Culture project. This community involves GLAM institutions, universities, academies, as well as individual researchers. The user stories are publicly available<sup>21</sup>. The extracted requirements have been divided into seven categories and are listed below.

#### **REQ1: Users**

The interaction with the resources in the NFDI4Culture-KG should fit the heterogeneous user base within the culture community.

- 1.1:** Users without a technical background have to be able to enter data and access and explore data by means of an easy to use system (User Interface).
- 1.2:** Users with a technical background have to be able to interact with the data in the NFDI4Culture-KG based on their specific needs as researchers and by means of automated methods (SPARQL Endpoint).

#### **REQ2: Search and Find Research Resources**

- 2.1:** Users need to find tangible and intangible CH research resources. This includes datasets, tools, software, standards and ontologies, and research projects.
- 2.2:** To enhance findability, users require contextual information about research resources, including research projects and connected organizations and persons.

<sup>21</sup> <https://nfdi4culture.de/resources/user-stories.html>

**2.3:** To enhance findability, users require content descriptions of the available resources in the Knowledge Graph.

**REQ3: Access to Research Resources**

**3.1:** Research resources have to be published with licensing information.

**3.2:** Users require a unified access to research resources.

**3.3:** To enhance accessibility, it should be possible to connect research resources to the NFDI4Culture-KG without necessarily being published as LOD.

**3.4:** Users require long-term access to research resources.

**REQ4: Interoperability of Resources**

**4.1:** Research resources need to be interoperable to enable an interconnection between resources and thus, allow users to explore the data across collections.

**4.2:** To enable data interoperability, ontologies and standards have to be published and documented.

**4.3:** Published research resources require version information.

**REQ5: Publishing and Reuse**

**5.1:** Research resources need to be published and made available for reuse based on the users' needs, which requires respective licenses. Thereby it must be possible to enable only partial datasets to be reused and it must be possible to allow reuse only for a specific purpose.

**5.2:** Persistent identifiers are required for each entry into the portal to enable long term referencing of resources.

**REQ6: Networking and Recruiting**

**6.1:** Individual researchers and institutions need the possibility to exchange their expertise on specific domains, topics, standards, technologies and services for networking purposes, to engage in citizen science tasks and for the recruitment of new staff.

**REQ7: Training and Education**

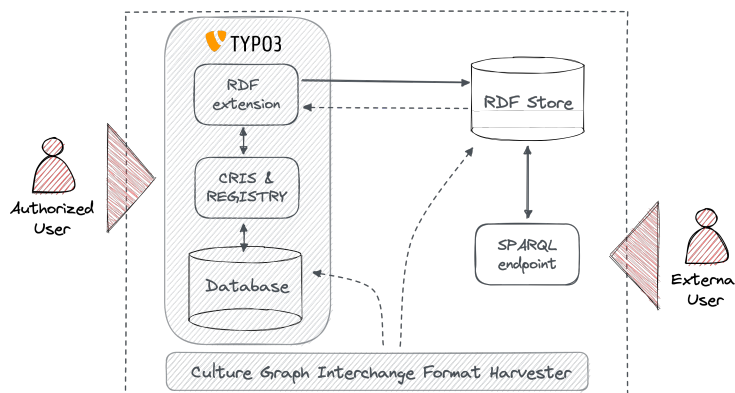
**7.1:** Users require training and education resources to teach their staff about digitizing, structuring and storing data resources.

**7.2:** Users need best practices and guidelines in order to transform their resources into structured data and LOD using open standards.

The following section 4 describes the main contribution of the paper and also explains in which way each of these requirements were met.

## **4 NFDI4Culture Research Data Integration**

This section describes the main features and achievements of NFDI4Culture initiative towards access, management and interconnection of decentralized research resources of the culture community. Firstly, the Culture Information Portal, its architecture, specifics and implementation are described. Secondly, the design methodology, development and the status of the NFDI ontologies are explored, and the NFDI4Culture-KG, its capabilities and scope are presented. Finally, the NFDI4Culture curation strategy on A-Box and T-Box level is briefly discussed.



**Fig. 1.** Overview of the Culture Information Portal Architecture

About: <<https://nfdi4culture.de/id/E3625>>

### Guideline for a FAIR Cultural Studies Research Data Management

**type**

Contribution  
Guide

**url**

<https://nfdi4culture.de/go/E3625>

**version**

1.0.1

**description**

In order to be optimally re-usable, research data should be processed according to the FAIR principles. This guideline explains what these principles mean and how they can be implemented in cultural studies and heritage collections.

**keywords**

Digital Humanities  
Guideline  
Authority Data

**author**

Angela Kaius

**contributor**

Task Area 2: Standards, data quality and curation

**Fig. 2.** The screenshot of resource E3625 in culture portal

## 4.1 Culture Information Portal

NFDI4Culture aims to provide a unified, intuitive access point to the decentralized research data of the NFDI4Culture research community and all other services of the consortium through the Culture Information Portal<sup>22</sup>. These services include an overarching helpdesk, reusable guidelines for all areas of research data management, and a consortium-wide identity and access management for collaborative tools. Additionally, the Culture Information Portal provides information on current events and news from the consortium, as well as participating institutions, individuals and projects. One important requirement for all NFDI consortia is the integration of resources and services with international infrastructures like the European Open Science Cloud (EOSC)<sup>23</sup>. To

<sup>22</sup> <https://nfdi4culture.de/>

<sup>23</sup> <https://eosc-portal.eu/>

achieve this, NFDI4Culture has taken the approach to implement the portal as a web based Current Research Information System (CRIS) that adheres to international standards for CRIS systems, namely the CERIF standard of euroCRIS<sup>24</sup>. CERIF provides a data model for all entities in the domain of NFDI4Culture (like persons, organizations, products/datasets, events etc.) and their relations towards each other. For the implementation, the TYPO3 extension "Academy Current Research Information System"<sup>25</sup> was used and developed further. By using a CMS for data curation, the consortium is able to provide a web-based user-friendly interface where non-expert users can easily contribute resources such as technical and community services, software tools, datasets and at the same time provide metadata about those resources (REQ1.1, REQ3.2).

The CRIS data of the portal is also made available as Linked Open Data in accordance with the FAIR principles. The TYPO3 LOD extension<sup>26</sup> provides a semantic layer that is implemented directly on top of the CMS's relational database with a configurable IRI generator and resolver for all data sets, as well as an RDF serializer for all database content to ensure persistent identification and long term referencing of resources (REQ5.2). All resources captured in the Culture CRIS are published using a standardized LOD API with various RDF serializations (e.g., RDFa, Turtle, JSON-LD, etc.) using the Hydra Core Vocabulary<sup>27</sup>. This allows the data to be curated and continuously expanded in a decentralized manner for the Culture Knowledge Graph. The data published via the CMS's LOD API is integrated into the actual Knowledge Graph using ingest routines. Oxigraph<sup>28</sup> serves as a native RDF store with a python-based wrapper for lightweight deployment of the public SPARQL endpoint, which is directly integrated into the Culture Information Portal through a graphical interface. The implementation of required functionalities in TYPO3 was found to be the best solution to ensure a seamless and future-proof integration into the existing infrastructure of the consortium, making direct use of the IAM component for user management, long-term sustainability beyond the current funding period through TYPO3 LTS releases, convenient and easy-to-use form templates for web based data curation and at the same time providing the freedom to integrate external ontologies as well as using dedicated NFDI ontologies such as the NFDIcore and NFDI4Culture ontologies.

## 4.2 NFDI Ontologies and the NFDI4Culture Knowledge Graph

This section presents ontologies that are used to represent research data resources, events, agents, publications, etc. in the domain of NFDI. Additionally, the NFDI4Culture-KG which serves as a unifying platform for all research-related data collections, infrastructures, software tools, and services in the field of NFDI4Culture is presented.

**A Modular Approach: NFDIcore and its Extensions.** While each consortium within the NFDI has its specific focus and requirements with respect to its domain, they are

<sup>24</sup> <https://eurocris.org/services/main-features-cerif>

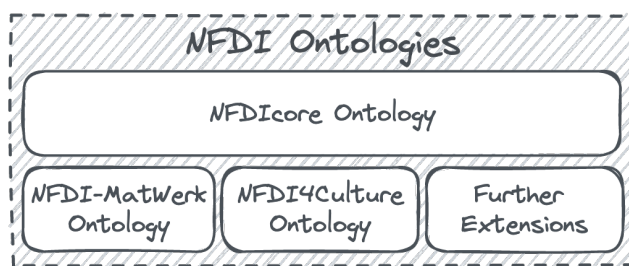
<sup>25</sup> <https://github.com/digicademy/academy>

<sup>26</sup> <https://github.com/digicademy/lod>

<sup>27</sup> <https://www.hydra-cg.com/spec/latest/core/>

<sup>28</sup> <https://github.com/oxigraph/oxigraph>





**Fig. 3.** Overview of the NFDI Ontologies

united in their commitment to providing a comprehensive, sustainable, and interoperable research data infrastructure that supports researchers across different fields and disciplines. Inline with these efforts, NFDIcore is a generic ontology which is being developed to represent research resources, e.g. data sets, providers, persons, areas of expertise across the NFDI consortia (REQ6.1, REQ7.2). It serves as the basis for further domain specific ontologies, e.g. the NFDI4Culture ontology, NFDI-MatWerk ontology (materials science) which take into account the respective domain specific needs (see figure 3). Such modular approach is beneficial to improve the consistency and clarity of the representation of research resources by an establishment of standard vocabularies and structures, it boosts the reusability and helps to uncover new relationships and insights by enabling knowledge discovery across different domains. Ontology extensions by further NFDI consortia are currently ongoing work.

The current version of the NFDIcore ontology<sup>29</sup> comprises 33 classes and 60 properties that cover various aspects of research data management, such as data access and sharing, citation, provenance, and security. Its development has been guided by established standards and best practices in data management, including the FAIR Data Principles. To ensure interoperability, NFDIcore links to around 20 external vocabularies, e.g. SCHEMA<sup>30</sup>, FRAPO<sup>31</sup>, Dublin Core<sup>32</sup>, and PROV-O<sup>33</sup>. Additionally, there are extensions that cater to the unique requirements of different research fields, e.g. NFDI4Culture<sup>34</sup> and NFDI-MatWerk<sup>35</sup> [1] ontologies.

The development of the NFDI ontologies follow a user-centered design and evaluation methodology ([2]). For example, for NFDI4Culture ontology a set of requirements for different user groups is developed incrementally and iteratively (see section 3.1). Since NFDI ontologies are intended to be dynamic, evolving and to adapt to the changing needs of the NFDI community over time, every version of the ontologies is accessible and referential online (REQ4.2). Addressing REQ2.1, the ontologies contribute to enhancing findability and interoperability of cultural heritage (CH) research resources by

<sup>29</sup> <https://github.com/ISE-FIZKarlsruhe/nfdicore/tree/v1.1.0>

<sup>30</sup> <https://schema.org/>

<sup>31</sup> <https://sparontologies.github.io/frapo/current/frapo.html>

<sup>32</sup> <https://www.dublincore.org/>

<sup>33</sup> <https://www.w3.org/TR/prov-o/>

<sup>34</sup> <https://nfdi4culture.de/ontology.html>

<sup>35</sup> <https://nfdi-matwerk.pages.rwth-aachen.de/ta-oms/mwo/doc/index.html>

enabling connections within CH domain and across different consortia. For this, various disparate materials, e.g. data sets, portals, softwares, tools, creative collections, etc. of the community are consolidated into a single coherent - the central class of the ontology *nfdicore:Resource* and its sub-classes. Every resource can be described with rich metadata (see figure 2), including information on responsible organisations, contact points, content descriptions, license information, version, media types, etc. (REQ2.3, REQ3.1, REQ4.3, REQ5.1).

**NFDI4Culture Knowledge Graph.** The NFDI4Culture project aims to improve the accessibility and reusability of cultural heritage data by integrating diverse and isolated research resources through a KG. For this, the NFDI4Culture-KG is being developed and integrated with the Culture Information Portal. It provides a single point of access to decentralized research resources. The NFDI4Culture-KG utilizes LOD and the NFDI ontologies to connect research data collections, tools, infrastructures, and services within the subject areas of NFDI4Culture. The NFDI4Culture-KG consists of the *Research Information Graph* and the *Research Data Graph*, and can be explored and retrieved through a SPARQL endpoint<sup>36</sup> (REQ2.2). The *Research Information Graph* contains an index and metadata for NFDI4Culture project resources, services, persons, and institutions, while the *Research Data Graph* enables the integration of information about specific content of the NFDI4Culture resources, e.g. metadata about objects in a museum collection. The workflow to integrate resources into the *Research Data Graph* currently involves two components: (1) Knowledge from external RDF-based resources is accessed in the NFDI4Culture-KG via native SPARQL federation. (2) For the integration of structured legacy data that is not published as LOD (REQ3.3), a lightweight interchange format that provides important data attributes has been developed for ingestion of index data feeds. The Culture Graph Interchange Format (CGIF) specification is published online<sup>37</sup> (REQ4.2).

The current version of the NFDI4Culture-KG contains information about 219 persons and 183 organisations. It describes 229 contributed resources, among which are e.g., 23 data portals and 76 news articles. The NFDI4Culture-KG is meant to be a constantly changing and evolving representation of the connections between research entities within cultural heritage domain and archaeology. The graph is intended to be flexible so that it can accommodate the addition of new data and the refinement of existing data based on new discoveries or improved understanding.

### 4.3 Community Participation and Curation Strategy

As mentioned above, NFDI4Culture lives from the participation of the entire culture community. For the Culture Information Portal and the NFDI4Culture-KG infrastructure, this means to enable users to provide research resources and to participate in ontology development. The provision of research resources and thus, the extension of the KG with instance data will be enabled for authorized and registered users directly in the portal via pre-defined input forms (A-Box). The ontologies (NFDIcore and the

<sup>36</sup> <https://nfdi4culture.de/resources/knowledge-graph.html>

<sup>37</sup> <https://nfdi4culture.de/go/E3712>

NFDI4Culture extension) are under constant development. Expert users familiar with ontology design are invited to participate in discussions directly in github to provide feedback and make modeling suggestions (T-Box).

In this section, the RDM infrastructure for NFDI4Culture was described including the developed ontologies and the NFDI4Culture-KG. In the following section 5, query and data integration use cases will be briefly discussed.

## 5 Culture Research Data Exploration and Discussion

The Culture Information Portal and the NFDI4Culture-KG are publicly available and can be used to explore research resources within the research community. As outlined in Section 4, the majority of the community's requirements have already been fulfilled, while for the remaining ones, a plan was presented for addressing them as part of this long-term project. However, the community itself is also sharing a responsibility in this task (e.g. REQ4.1). With appropriate guidelines and best-practices<sup>38</sup> provided to the culture community, all stakeholders should be able to transform their data into structured data or LOD themselves. Once this is accomplished using open standards and resources are connected to the NFDI4Culture-KG, it will be possible to find interconnections between data collections and to explore the content in the research domain of NFDI4Culture in a more meaningful way. This way, researchers can better evaluate whether a resource is valuable to their own research as well. Relevant queries include: (1) Which datasets contain a painting collection, were contributed by a museum and are encoded via CIDOC-CRM? (2) What datasets are available which contain AV-files with music of a certain composer? What is the license agreement of these AV-files? (3) What datasets contain performance documentation videos that are allowed for re-use in the context of research? (4) Which datasets contain books by librettists that feature prints depicting the motif of Hercules at the Crossroads?

The integration of full LOD resources into the *Research Data Graph* is performed via SPARQL federation. For instance, *Linked Stage Graph* [8] is a KG with a public SPARQL Endpoint<sup>39</sup> containing historical performance data and is represented as a data portal in the NFDI4Culture-KG<sup>40</sup>. By means of federated queries, useful information about these resources can be integrated into the NFDI4Culture-KG. These information may include: (1) Which performances are presented in the KG and in which year did they take place? (2) For which performances are photographs provided? (3) Which performances were originally authored by William Shakespeare?

Data portals providing structured data on the Web can also easily contribute their resources to the *Research Data Graph* by making them available in the Culture Graph Interchange Format<sup>41</sup>. For instance, the Corpus Vitrearum Medii Aevi Deutschland<sup>42</sup> (CVMA) provides data on medieval and early modern stained glass that was integrated

<sup>38</sup> <https://nfdi4culture.de/resources/knowledge-base.html>

<sup>39</sup> <https://slod.fiz-karlsruhe.de/sparql>

<sup>40</sup> <https://nfdi4culture.de/id/E3590>

<sup>41</sup> <https://nfdi4culture.de/go/E3712>

<sup>42</sup> <https://corpusvitrearum.de/>

into the NFDI4Culture-KG and made accessible for further research according to aspects such as locations, epochs, and motifs. This integration enables the exploration of research resources beyond a single dataset. For instance, it is now possible to query<sup>43</sup> for depictions with ICONCLASS<sup>44</sup> “wild man” in three different data collections, CVMA, Wikidata and the Rijksmuseum Amsterdam.

## 6 Conclusion

This paper presents the ongoing efforts of the NFDI4Culture consortium to provide an infrastructure for tangible and intangible cultural heritage research data with regard to the FAIR principles. In particular, the contributions of this paper include the presentation of the Culture Information Portal, the NFDI4Culture-KG and its ontologies to enable an intuitive and meaningful interaction with cultural heritage research resources and their metadata. All components are publicly available and used by the community. Next steps with respect to data integration involve the possibility for collaborative provision and maintenance of instance data and the registration and curation of datasets via the portal. Furthermore, the federation of the NFDI4Culture-KG with Wikidata, DDB, Europeana and other KGs will be extended.

**Acknowledgement.** This work is funded by DFG (German Research Foundation) under project number 441958017.

## References

1. Bayerlein, B., Birkholz, H., Fliegner, S., Grundmann, J., Hanke, T., von Hartrott, P., Schilling, M., Waitelonis, J.: PMDco: platform material digital ontology. Version 2.0.0-alpha3 (March 24th 2023), <https://w3id.org/pmd/co>
2. Bravo, M., Luis, H.R., Reyes, J.: Methodology for Ontology Design and Construction. *Contaduría y Administración* **64**, 134 (03 2019)
3. Brodaric, B., Neuhaus, F.: Ontological Analysis and Modularization of CIDOC-CRM. In: *Formal Ontology in Information Systems: Proceedings of the 11th International Conference (FOIS 2020)*. vol. 330, p. 107. IOS Press (2020)
4. Doerr, M.: The CIDOC CRM, an Ontological Approach to Schema Heterogeneity. In: *Dagstuhl Seminar Proceedings. Schloss Dagstuhl-Leibniz-Zentrum für Informatik* (2005)
5. Hyvönen, E.: *Linked Open Data Infrastructure for Digital Humanities in Finland*. DHN 2020 (2020)
6. Nasarek, R., Rossenova, L.: *Linked Open Data Management Services: A Comparison. (Dataset)*. In: *DHd2023 Open Humanities Open Culture (DHd2023)* (2023), <https://doi.org/10.5281/zenodo.7738424>
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., Waitelonis, J., Zhou, K., Felgentreff, P., Meyer, N., Weber, A., Sack, H.: *Linked stage graph*. In: *SEMANTICS Posters&Demos* (2019)

<sup>43</sup> <https://nfdi4culture.de/go/kg-query-cvma>

<sup>44</sup> <https://iconclass.org/>