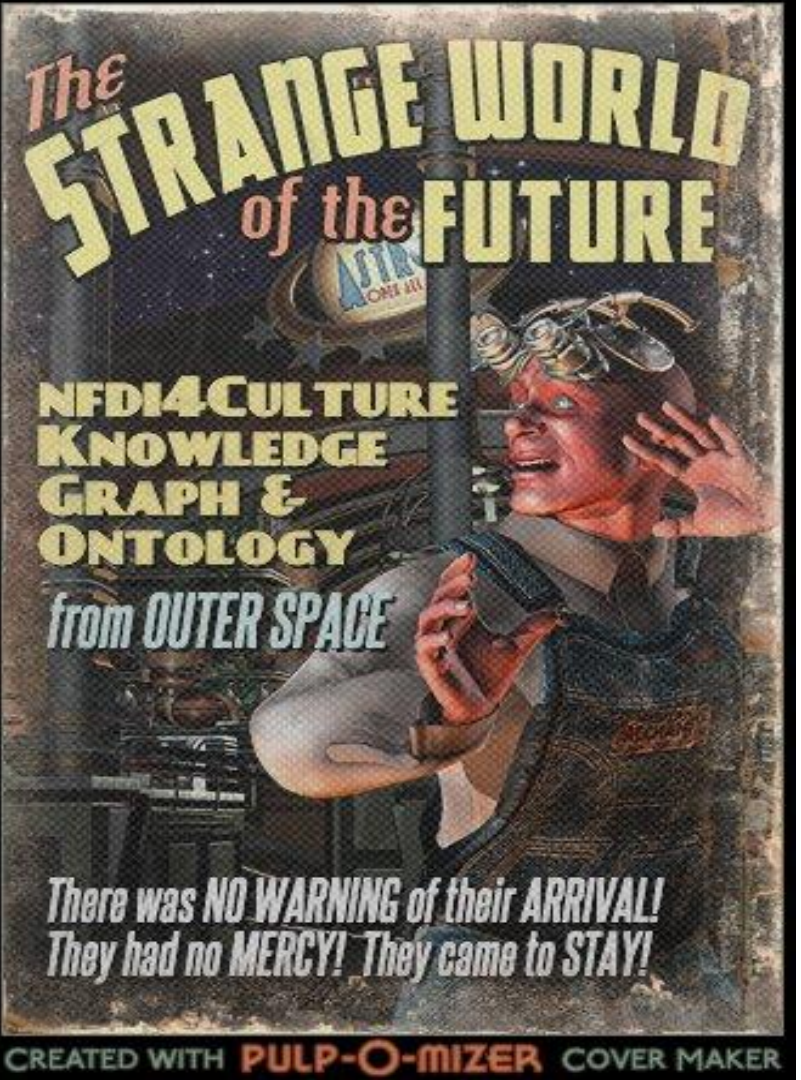


nationale Forschungsdaten Infrastruktur *for CULTURE*

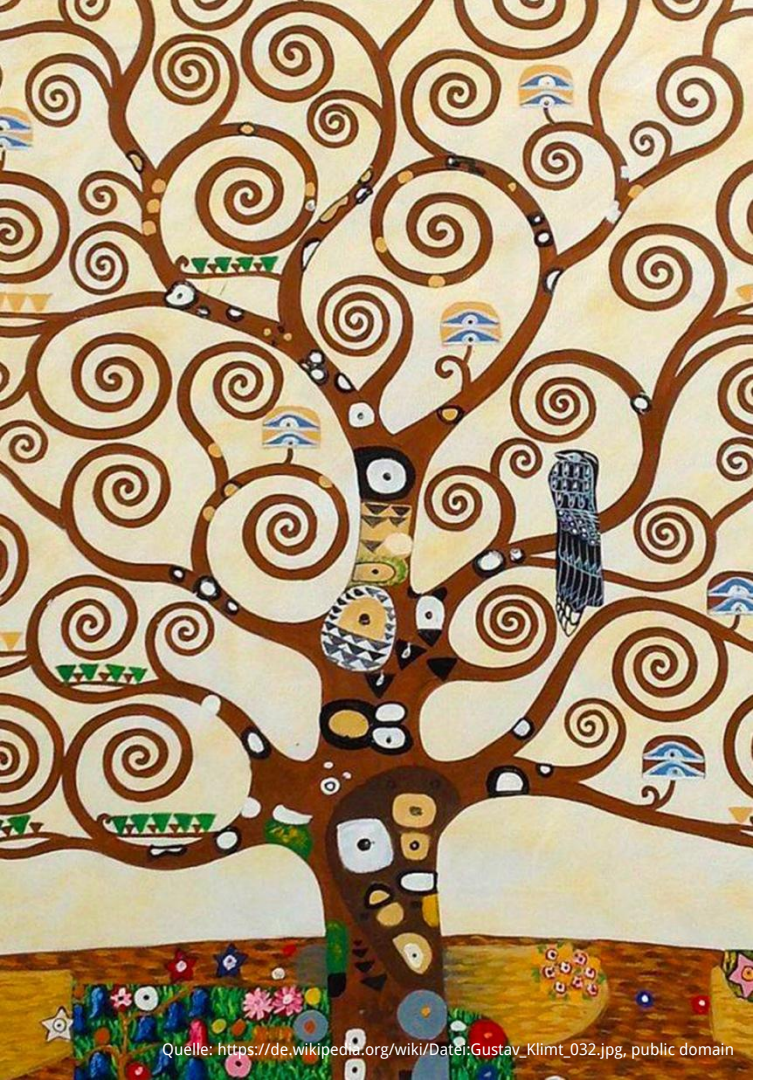
17.03.2023 | DHd2023 | Trier

Knowledge Graph-basierte Forschungsdatenintegration in NFDI4Culture

Tabea Tietz, Oleksandra Bruns, Heike Fliegl, Etienne Posthumus, Torsten Schrade, Harald Sack | CC BY 4.0

Gefördert durch die
Deutsche Forschungsgemeinschaft (DFG) - 441958017

DFG Deutsche Forschungsgemeinschaft



Outline

1. NFDI, NFDI4Culture …
   and Knowledge Graphs
2. Challenges for Data Integration
   in Cultural Heritage
3. Culture Portal and Culture
   Research Information System
4. Culture Ontology, Culture Research
   Information and Research Data Graph
5. Data Integration:
   Use Cases & Next Steps

1

NFDI, NFDI4Culture
… and Knowledge Graphs

In der Nationalen Forschungsdateninfrastruktur (NFDI) sollen die wertvollen Datenbestände von Wissenschaft und Forschung für das gesamte deutsche Wissenschaftssystem systematisch erschlossen, vernetzt und nachhaltig sowie qualitativ nutzbar gemacht werden.

nfdi Nationale Forschungsdaten Infrastruktur

DATA INTEGRATION ACROSS RESEARCH COMMUNITIES

COMMUNITIES

Architecture
Art History
Musicology
Theatre, Film and Media Studies

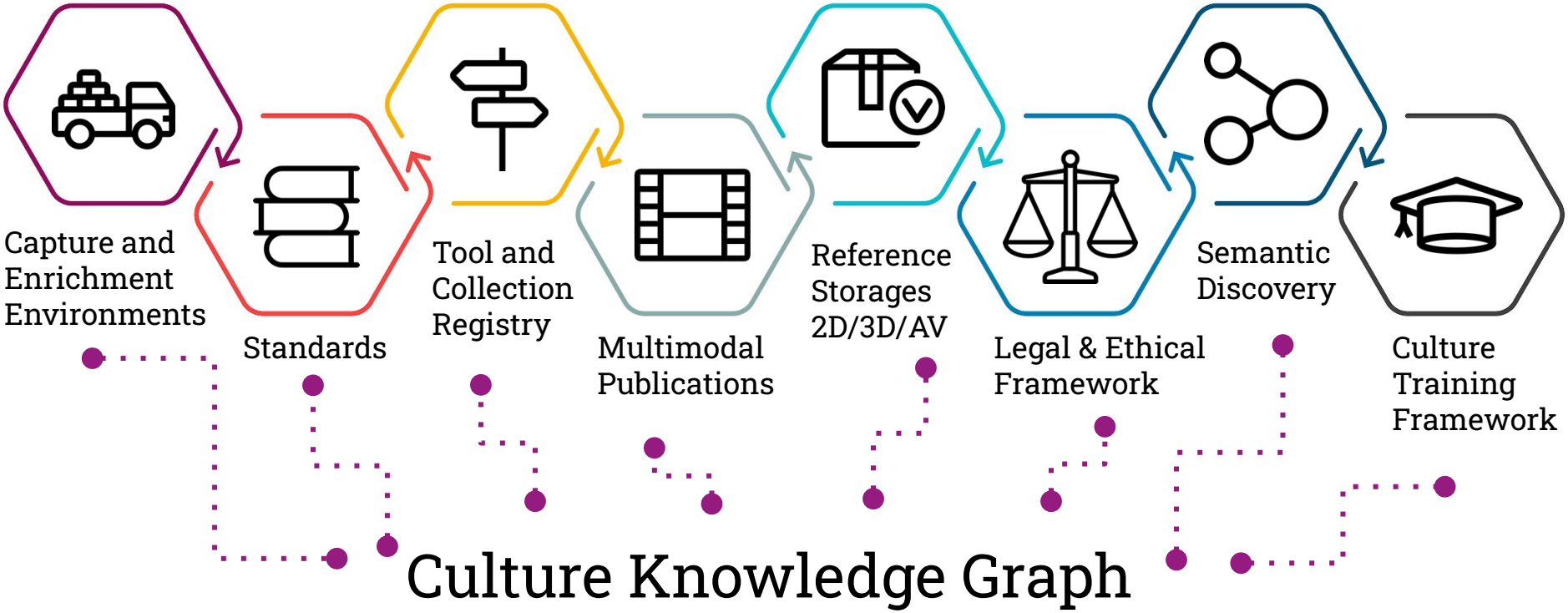
PARTICIPANTS

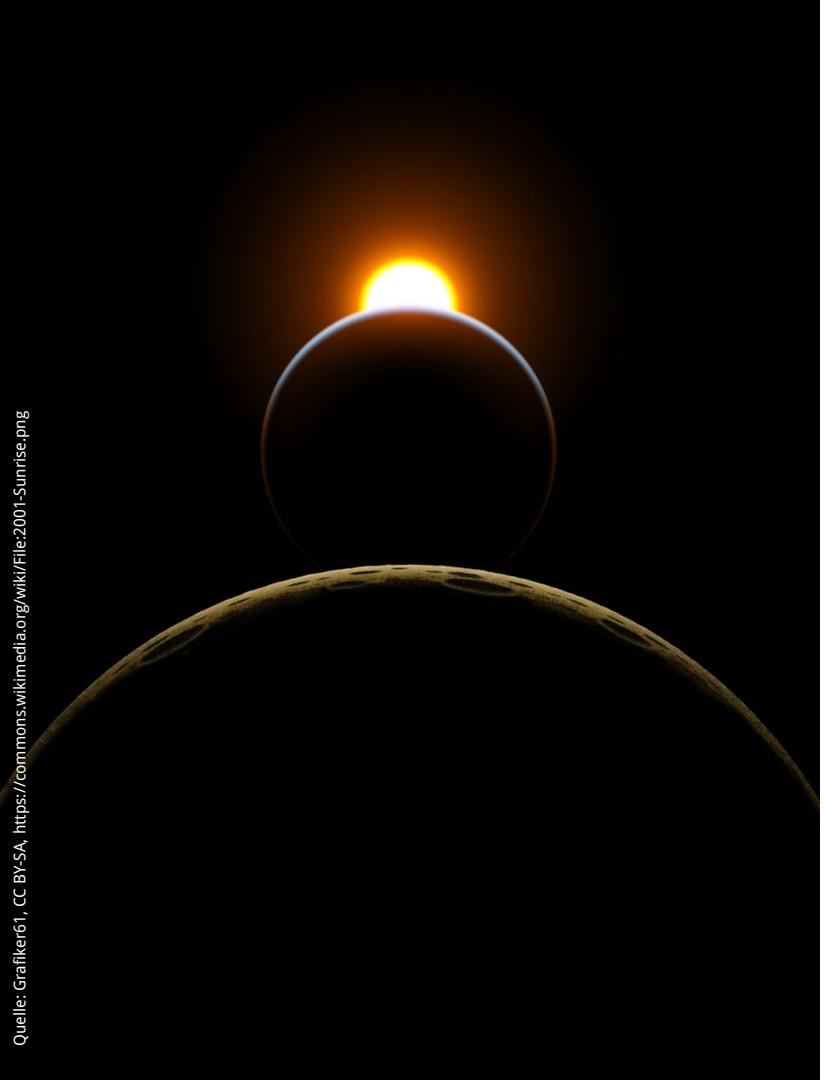
Academies
Archives
Art Colleges
Cultural Heritage Institutions
Individual Researchers
Galleries
Learned Societies
Libraries
Museums
Universities

CO-APPLICANTS

Academy of Sciences Mainz
Cologne University
FIZ Karlsruhe
Heidelberg University
Marburg University
Paderborn University
SLUB Dresden
SPK Berlin
TIB Hannover

Culture Information Portal





Quelle: Grafiker61, CC BY-SA, https://commons.wikimedia.org/wiki/File:2001-Sunrise.png

Exemplary questions

- **Which partners provide what datasets across which academic disciplines?**
- **In which 2D/3D/AV repositories can users store what kind of data?**
- **What research software is out there for which types of data?**
- **Which guidelines and/or training offers suit best for which users?**
- **What resources are in which datasets and how are they connected?**

2

Challenges for Data Integration in Cultural Heritage

Challenges

- Diversity of the culture research landscape
  - Research disciplines
  - Actors: Individual researchers, GLAM institutions, universities, academies, ...
- Heterogeneous research resources
  - Diversity in standards and formats
  - Status: from floppy discs to 5-star LOD

Magnus D from London, United Kingdom, CC BY 2.0, via Wikimedia Commons

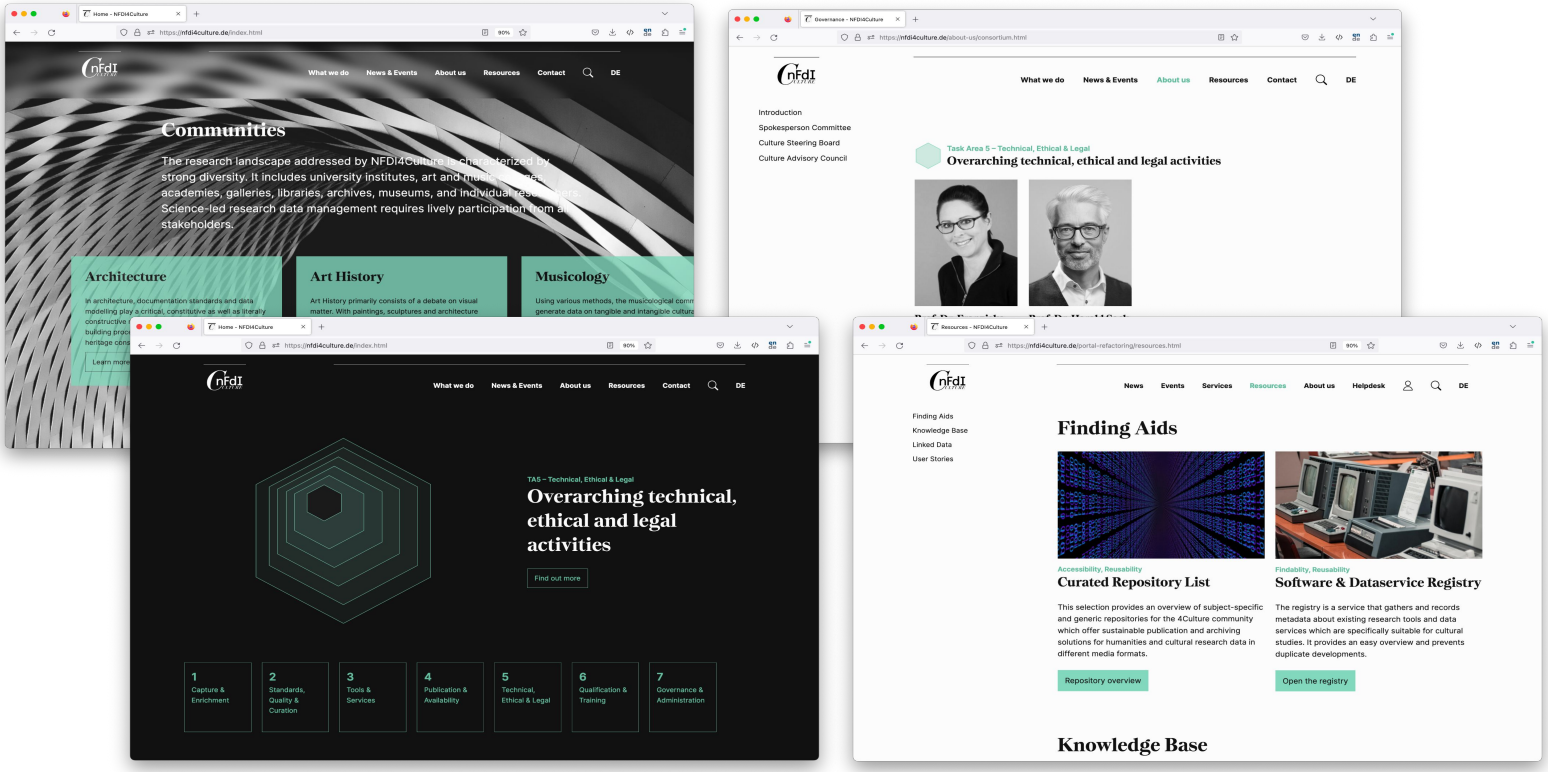
Challenges

- Enable means of participation for the culture community
- Requires a technical infrastructure that
  - allows to curate, extend, connect and access decentralized research resources
  - intuitively
  - without technical knowledge
  - by means of a single unified access

3

Culture Portal and Culture Research Information System

Culture Portal: “Things not Strings”



<https://nfdi4culture.de/>

Culture Portal: “Things not Strings”

<https://nfdi4culture.de/id/E1925>

<https://nfdi4culture.de/id/E1825>

What we do | News & Events | About us | Resources | Contact | DE

About: <https://nfdi4culture.de/id/E1925>

Harald Sack

type
Person

image

memberOf
Culture Spokespersons Committee
Role: Member

What we do | News & Events | About us | Resources | Contact | DE

About: <https://nfdi4culture.de/id/E1825>

Task Area 5: Overarching technical, ethical and legal activities

type
Organization

name
Overarching technical, ethical and legal activities

hasAcronym
Technical, Ethical & Legal

What we do | News & Events | About us | Resources | Contact | DE

About: <https://nfdi4culture.de/id/E2391>

Culture Knowledge Graph

type
Service

name
Culture Knowledge Graph

identifier
TA5M3

description
The Culture Knowledge Graph aims to be a connector for all research data produced within the NFDI4Culture subject areas. This will enable overarching search functionality and access to the distributed research data collections of the consortium and beyond. We will build this by using existing data in the Linked Open Data format as well as working with data providers to convert and integrate their digital assets. We will make use of W3C standard technologies and standard metadata formats that are already recognised by the Culture community. This development of the knowledge graph will be discussed within the 4Culture Linked Open Data Working Group. To stay up to date on their meetings, events and general news about the knowledge graph, sign up for our mailinglist: https://lists.nfdi.de/postorius/lists/nfdi4culture-lod.lists.nfdi.de.

keywords
Linked Data & Semantics

What we do | News & Events | About us | Resources | Contact | DE

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

Guideline for a FAIR Cultural Studies Research Data Management

type
Contribution
Guide

name
Guideline for a FAIR Cultural Studies Research Data Management

url
https://nfdi4culture.de/go/E3625

version
1.0.1

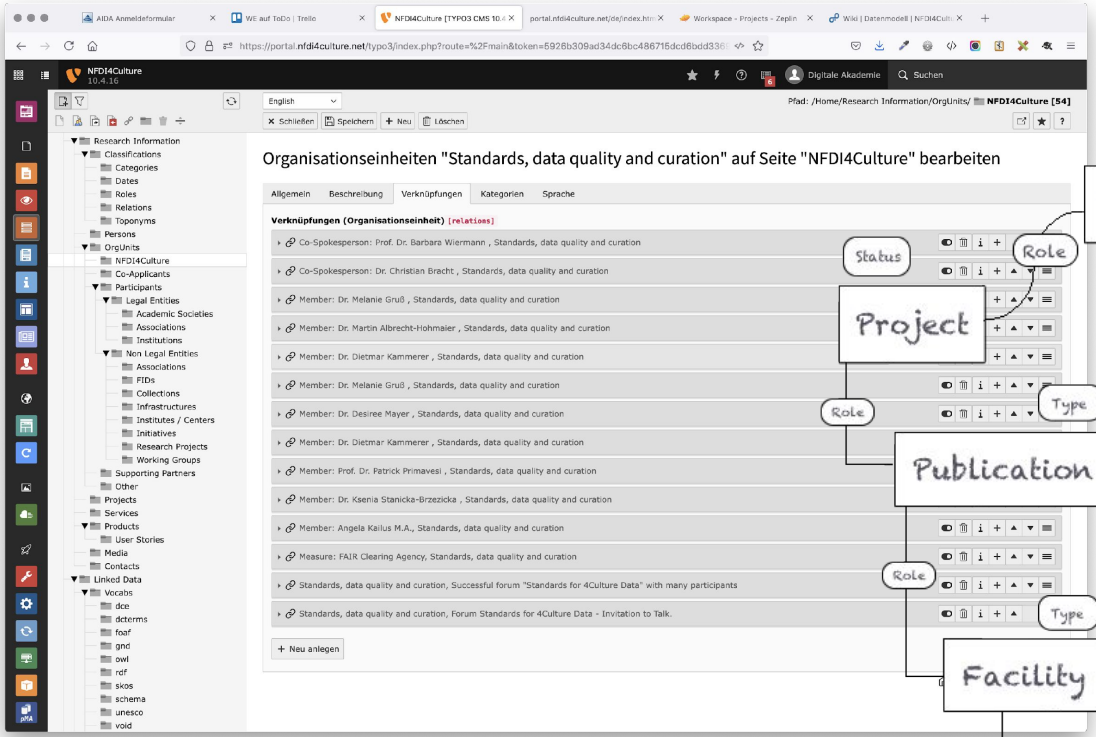
description
In order to be optimally re-usable, research data should be processed according to the FAIR principles. This guideline explains what these principles mean and how they can be implemented in cultural studies and heritage collections.

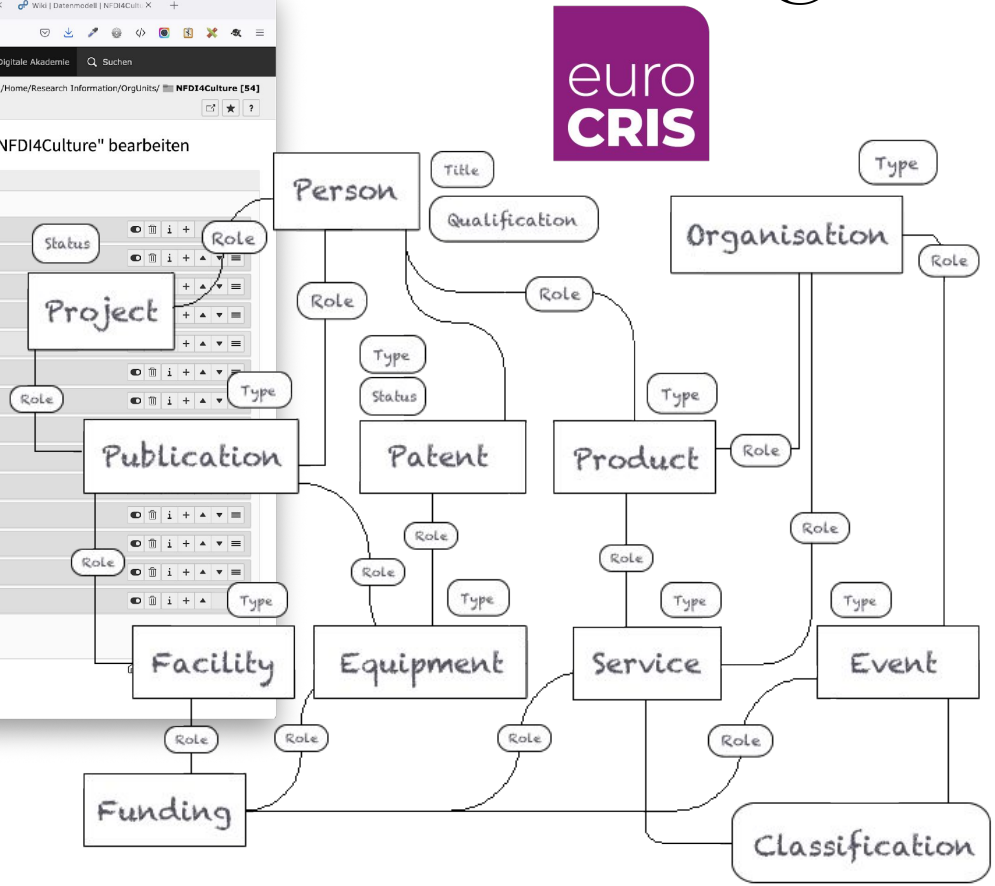
keywords
Digital Humanities

<https://nfdi4culture.de/id/E2391>

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

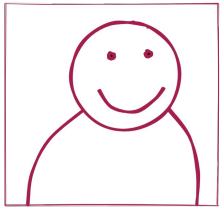
Culture Research Information System



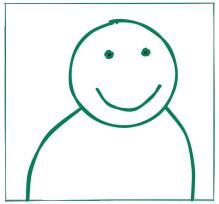


TYPO3 CRIS extension on GitHub:
https://github.com/digicademy/academy

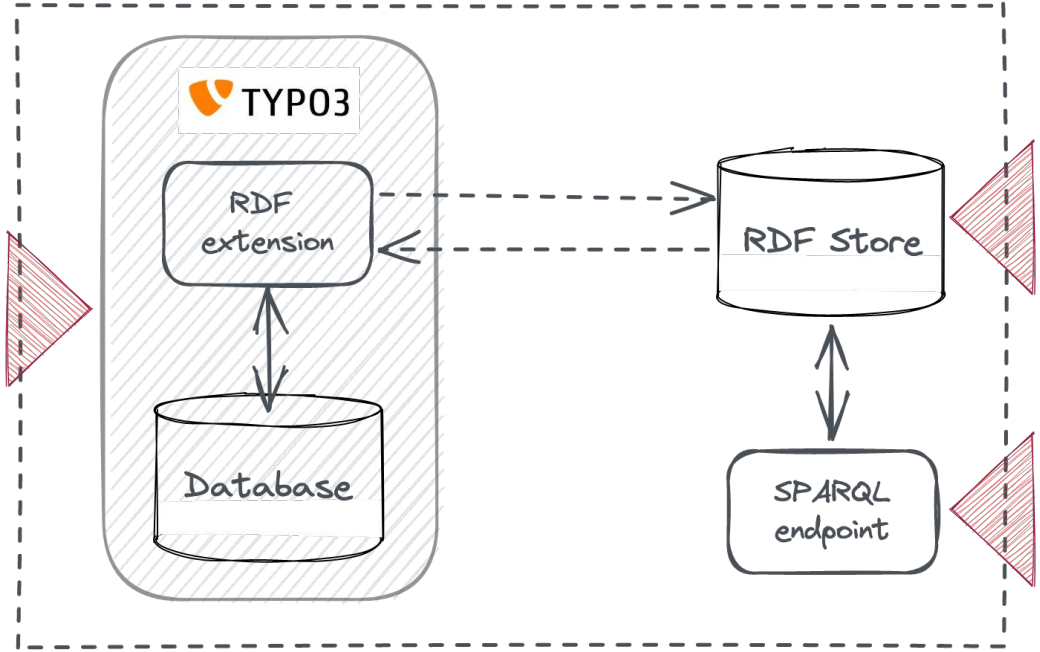
Culture Portal: Linked Data Architecture



NFDI4Culture Authorized User

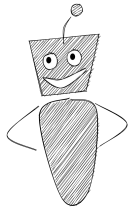


Arbitrary User



NFDI4Culture Expert User

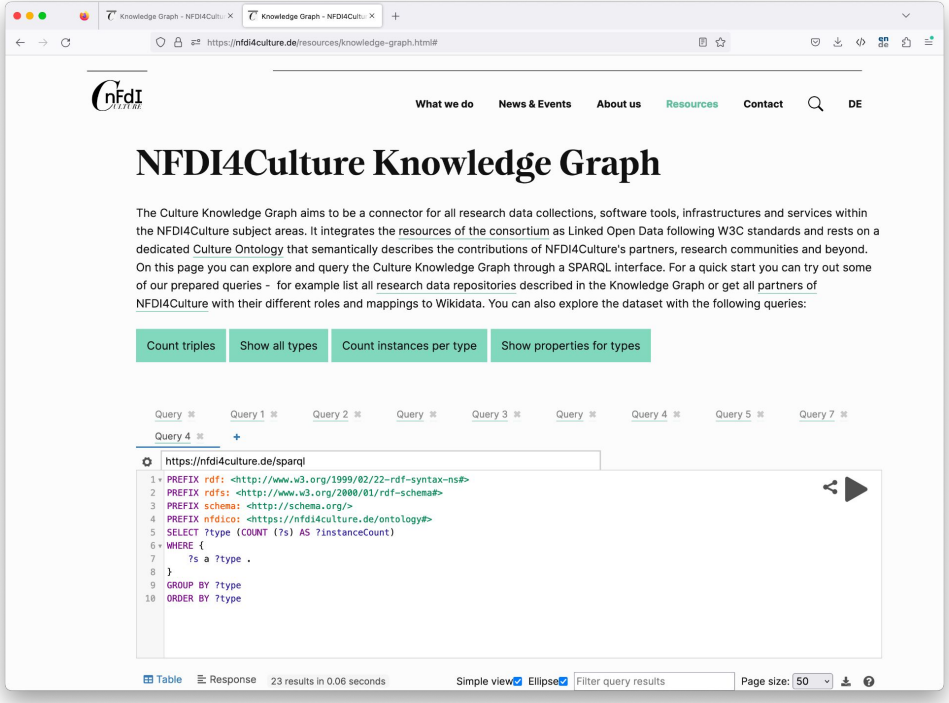
Arbitrary User

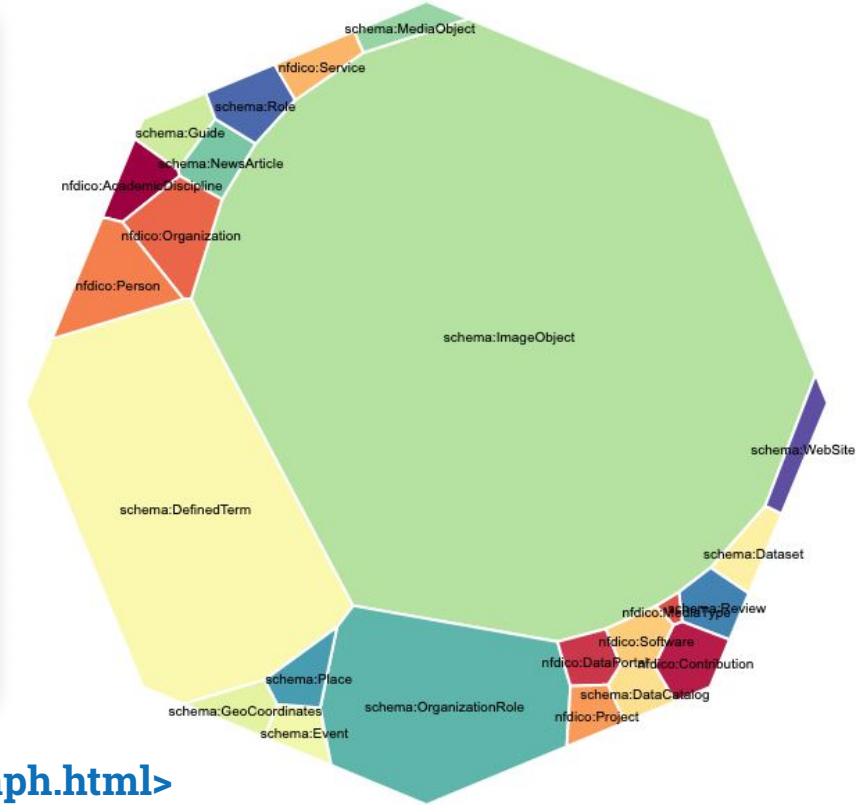


Automated Services

TYPO3 LOD extension on GitHub:
https://github.com/digicademy/lod

SPARQL Endpoint Data Exploration





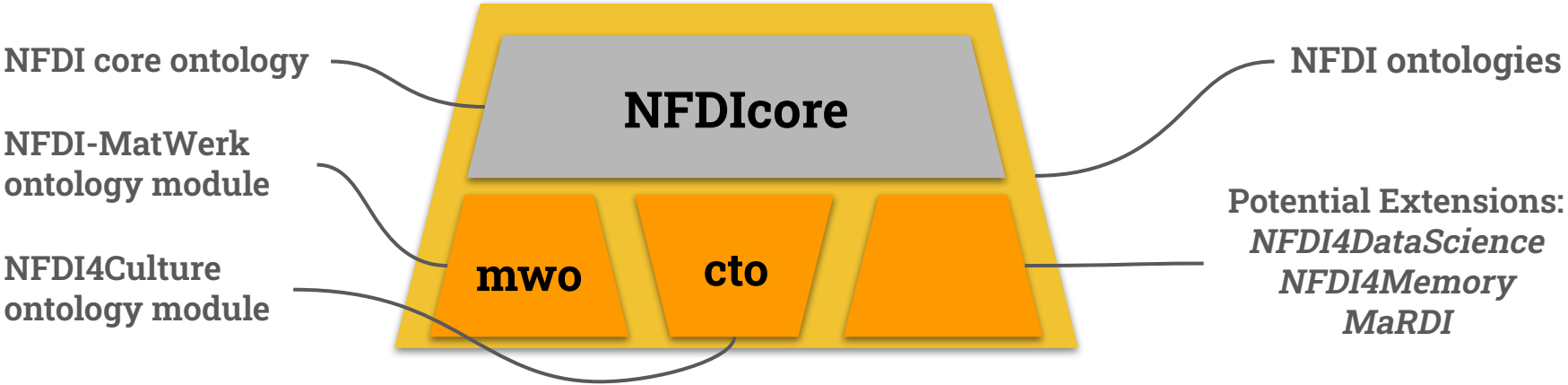
<https://nfdi4culture.de/resources/knowledge-graph.html>

4

NFDI Core Ontology, Research Information and Research Data Graph

Overview

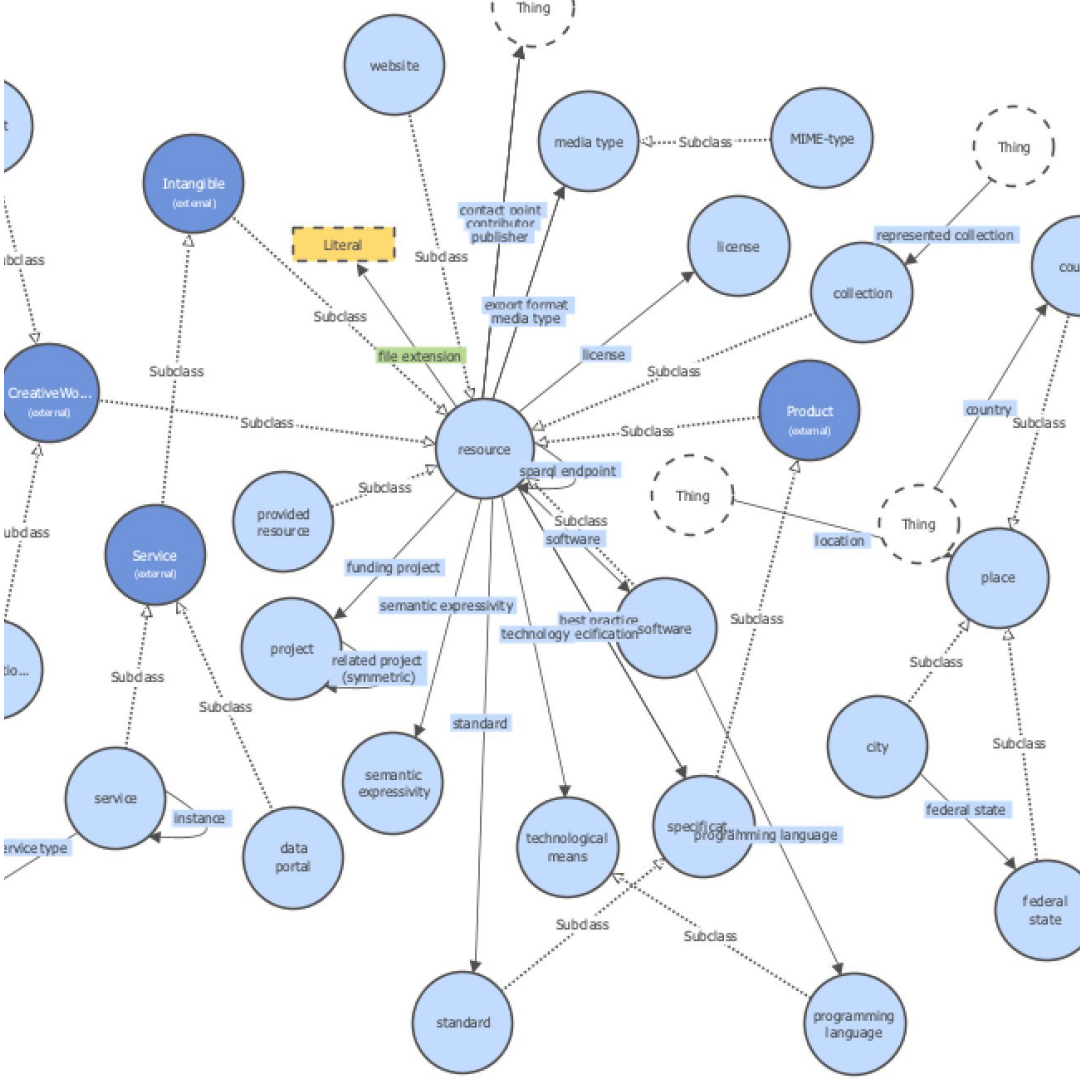
- NFDI ontologies follow modular approach
- NFDIcore v1.1 represents nfdi consortia
- NFDI modules represent specific consortia, as e.g. NFDI4Culture (cto) and NFDI-MatWerk (mwo)
- Specific requirements of consortia lead to ontology extension



NFDIcore Ontology

temporary namespace:
https://nfdi.fiz-karlsruhe.de/ontology

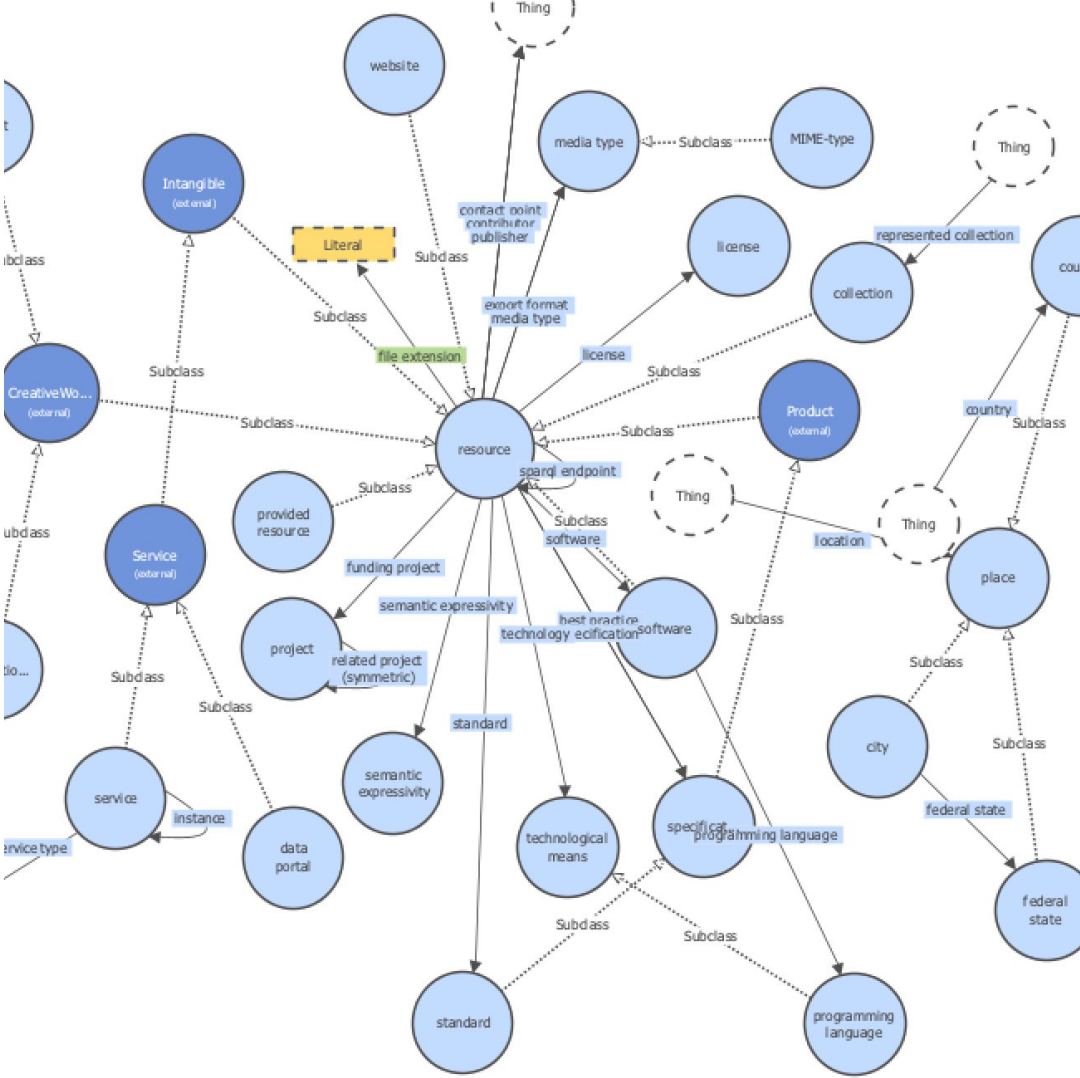
- **Represents resources of NFDI consortia, e.g. services, data sets, guidelines, etc.**
- **Bottom-up approach and in constant feedback with NFDI4Culture and NFDI-MatWerk consortia → in constant development**
- **Contains 33 classes and 60 properties**



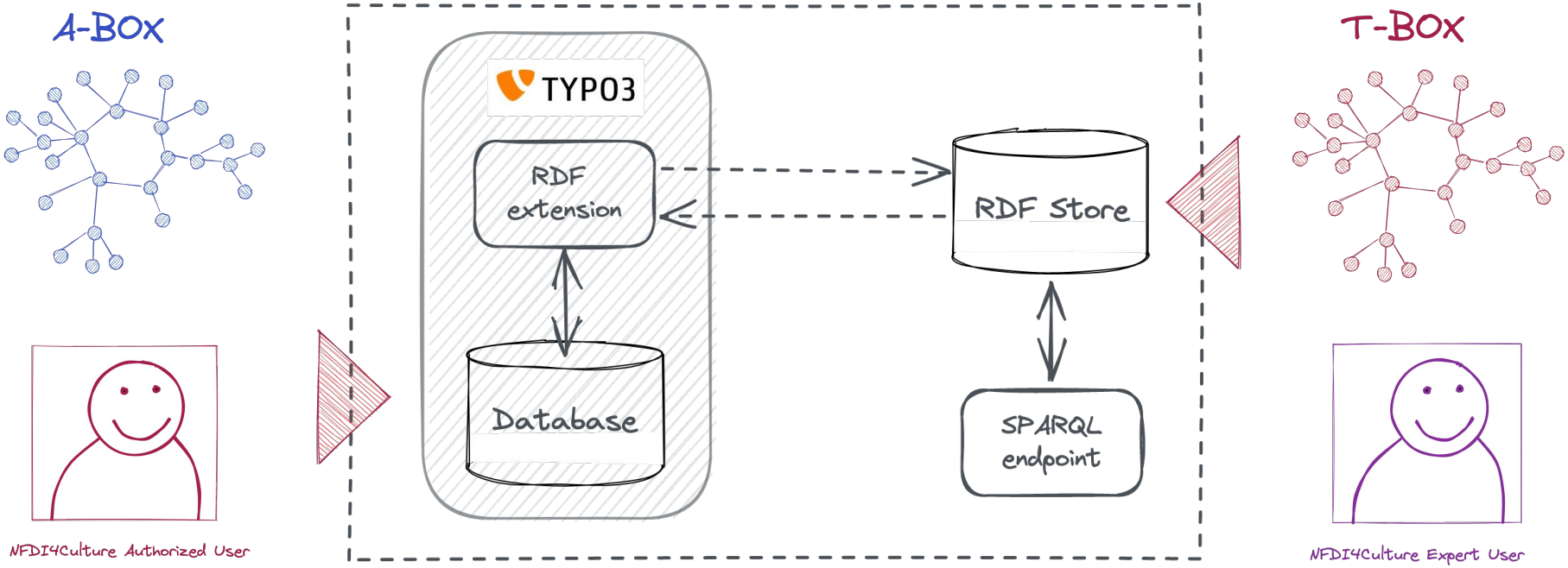
NFDIcore Ontology

temporary namespace:
https://nfdi.fiz-karlsruhe.de/ontology

- **Links to over 20 external vocabularies including schema, fabio, frapo, void, dcterms, foaf**
- **Available online**
  https://github.com/ISE-FIZKarlsruhe/nfdicore
  Most recent release: version 1.1
- **Documentation of the development version:**
  https://nfdi.fiz-karlsruhe.de/ontology_v1.1.0-alpha



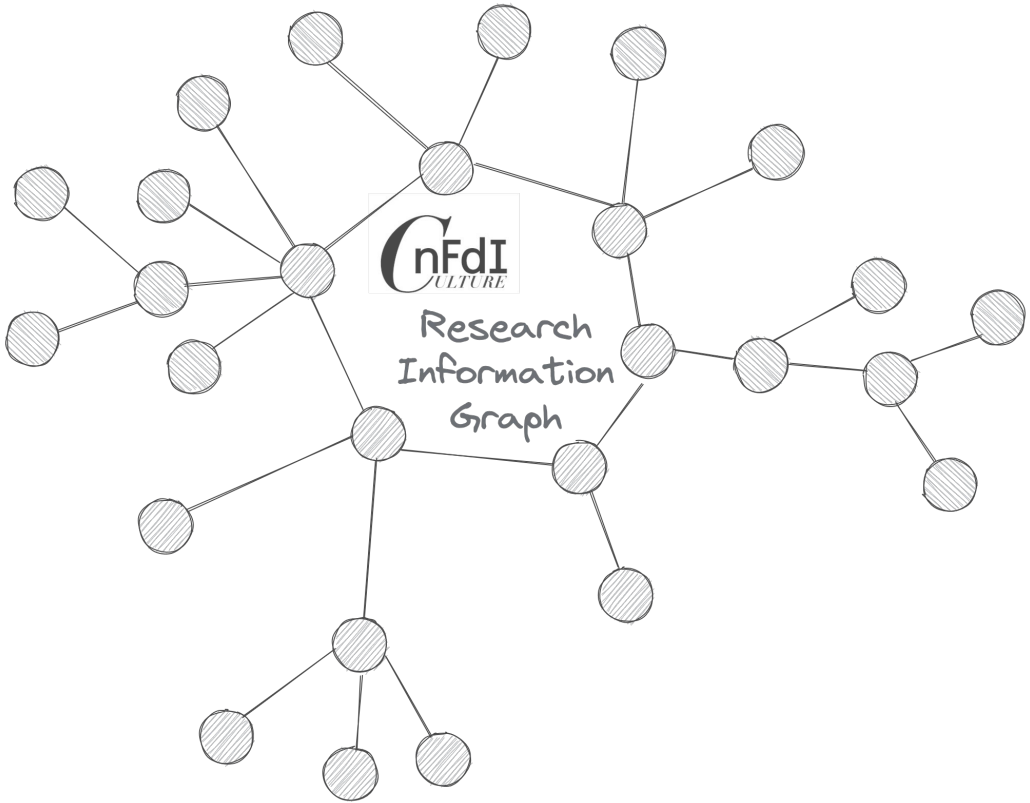
NFDI4Culture Curation Strategy



NFDI4Culture-KG v1.0 - Research Information Graph

CONTENT:

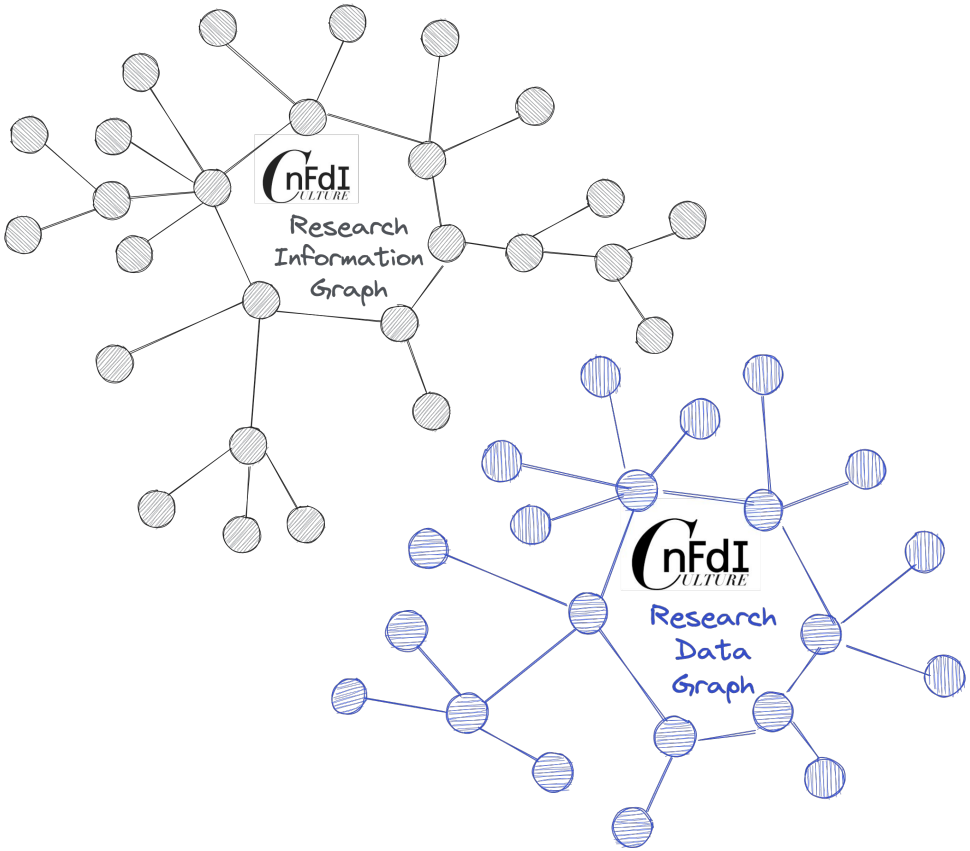
- Index and Metadata for NFDI4Culture Project Resources
  - NFDI4Culture Data Resources
  - NFDI4Culture Data Services
  - Persons and Institutions
  - Expertises



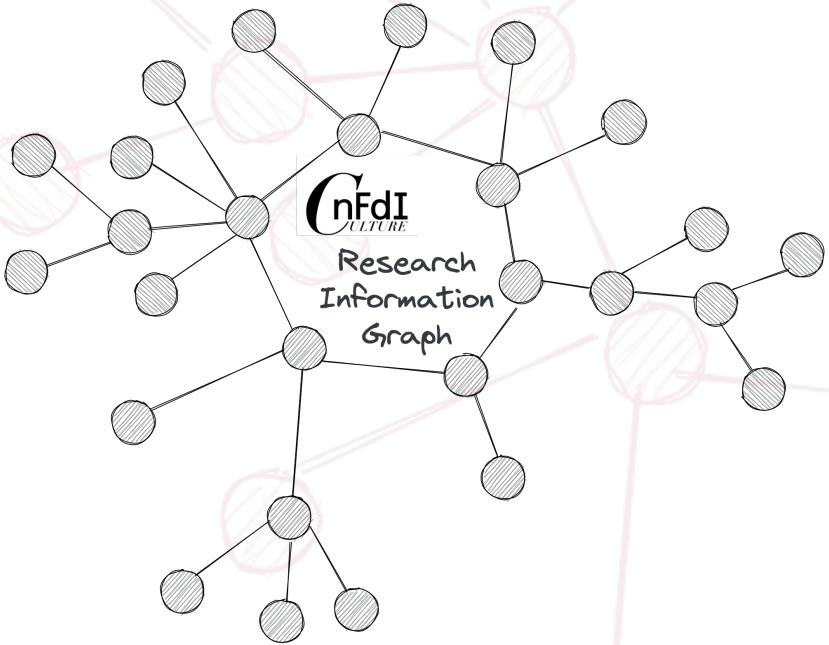
NFDI4Culture Knowledge Graph v2.0

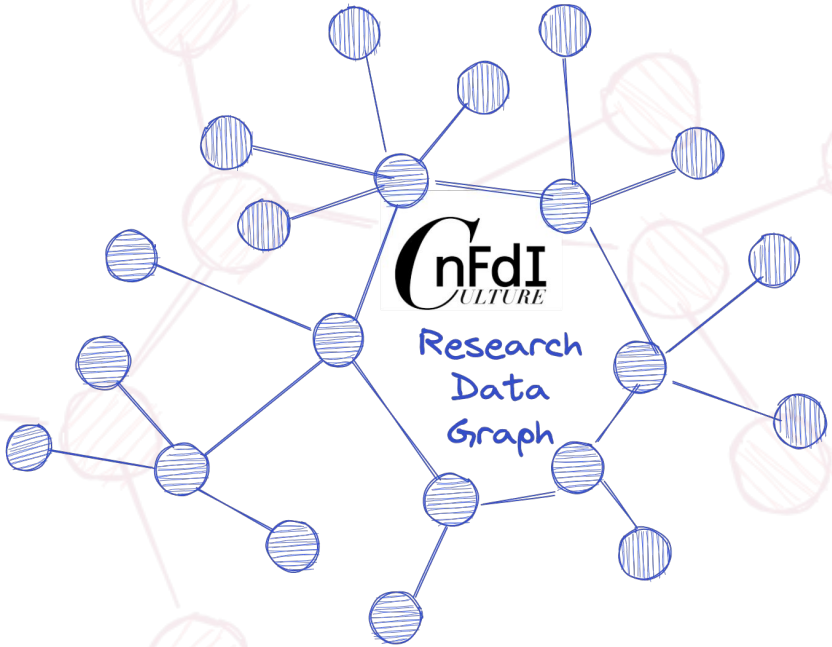
CONTENT:

- **All of NFDI4Culture-KG v1.0**
- **Plus Metadata for NFDI4Culture content as e.g., from**
  - **Fully fledged LOD resources**
  - **Structured Data resources (Legacy Data resources)**
  - **Wikibase-based resources**



NFDI4Culture Knowledge Graph v2.0



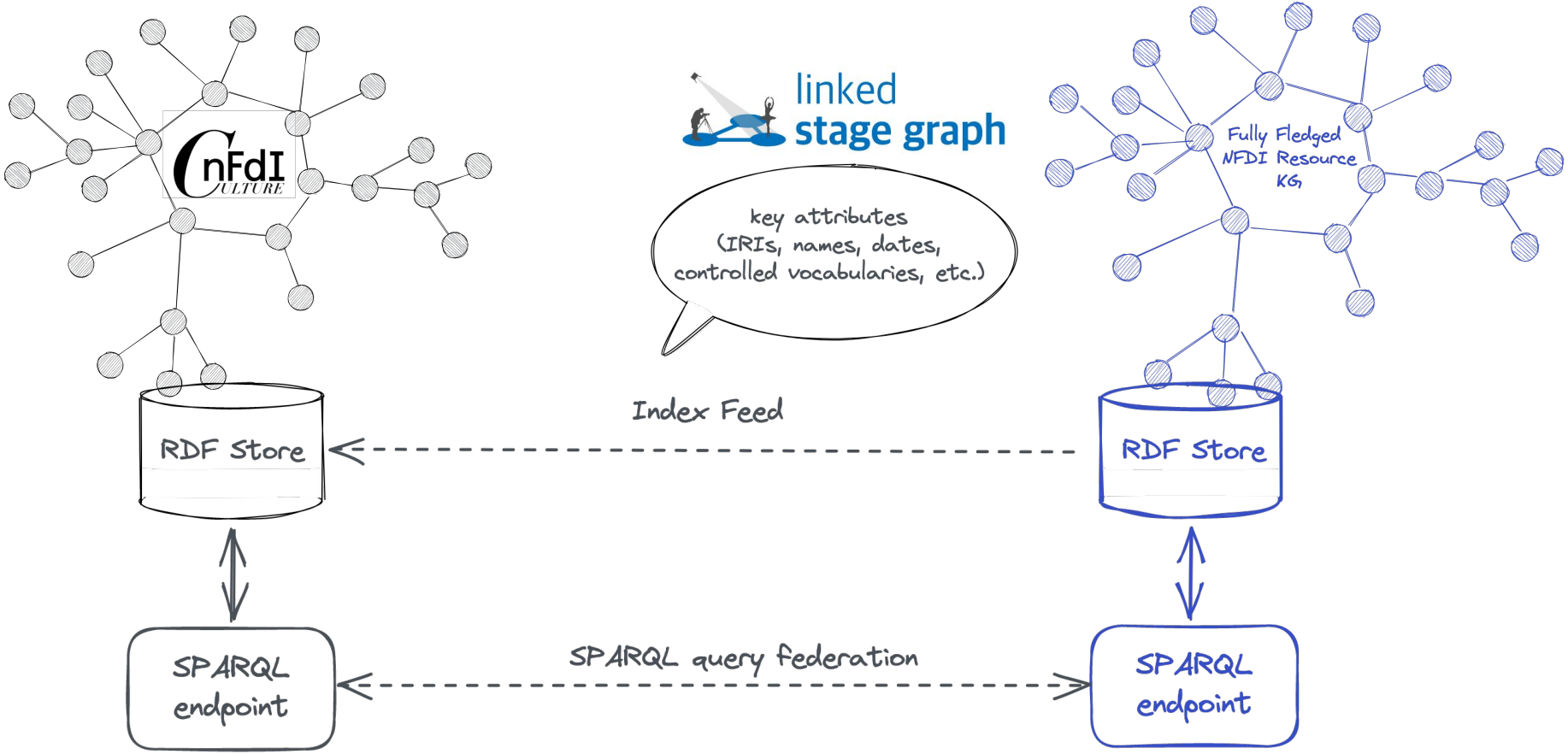


Research Information Graph + Research Data Graph = Culture KG v2.0

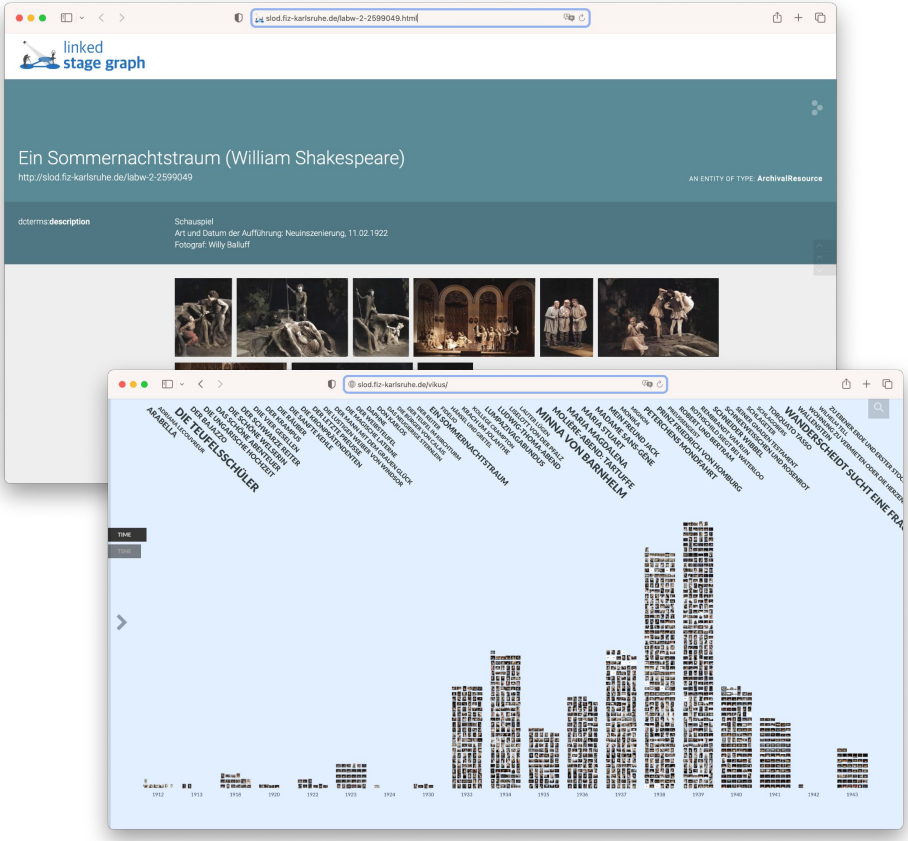
5

Data Integration: Use Cases & Next Steps

Use Case 1: Full LOD via SPARQL Endpoints



Linked Stage Graph



Knowledge Graph with photographs and metadata about the Stuttgart State Theatre from the 1890s to the 1940s

Gives insights into on-stage events like theater plays, operas and ballet performances as well as off-stage moments and theater buildings

Connections to Wikidata and GND

Means to explore the data:

- **SPARQL Endpoint**
  http://slod.fiz-karlsruhe.de/sparql
- **Vikus Viewer**
  http://slod.fiz-karlsruhe.de/vikus

Linked Stage Graph

NFDI4Culture Linked Data Resources

RDF/XML | Turtle | N-Triples | JSON-LD

About: <https://nfdi4culture.de/id/E3590>

Linked Stage Graph

type
DataPortal

name
Linked Stage Graph

mediaType
JPEG

subjectArea
Dance Studies

SPARQL Query Federation:

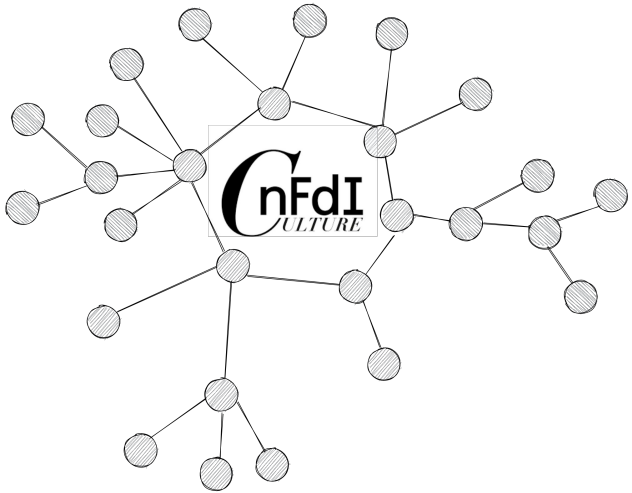
Show me all performances in Linked Stage Graph data portal, the original work they are based on and show me how many photographs are available for each performance.

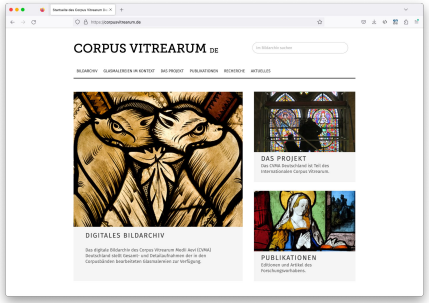
```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX schema: <http://schema.org/>
PREFIX slod: <http://slod.fiz-karlsruhe.de/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>


SELECT DISTINCT ?resource ?basedOn (COUNT(?image) AS ?imagecount)  WHERE {
  SERVICE <http://slod.fiz-karlsruhe.de/sparql> {
    ?resource schema:isBasedOn ?basedOn .
    ?resource  foaf:depiction ?image .
    }
} GROUP BY ?resource ?basedOn
```

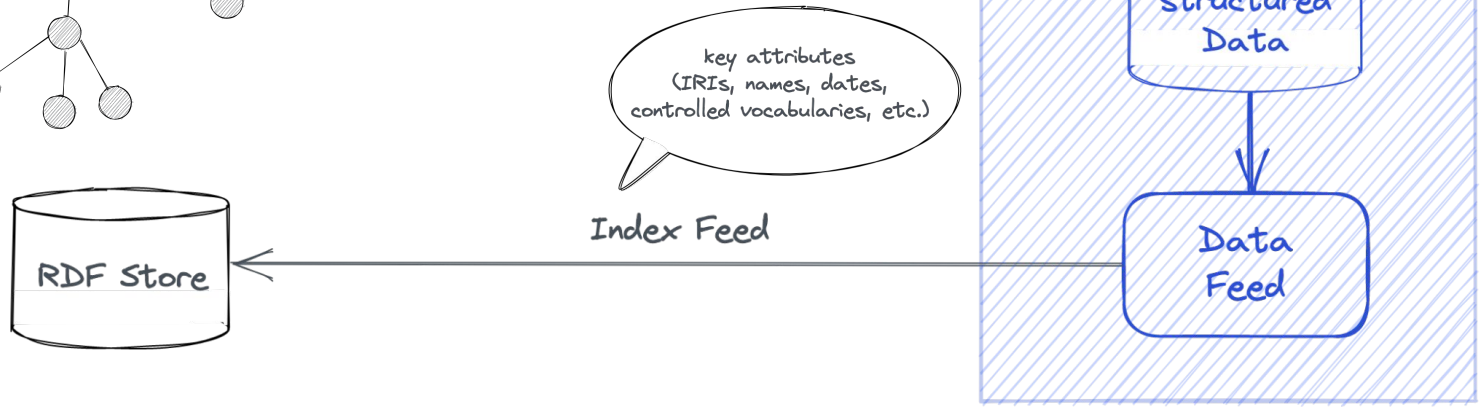
	resource	basedOn	imagecount
1	slod:labw-2-2601070	<http://www.wikidata.org/entity/Q1774386>	"72"^^<http://www.w3.org/2001/XMLSchema#integer>
2	slod:labw-2-2601341	<http://d-nb.info/gnd/4099338-3>	"56"^^<http://www.w3.org/2001/XMLSchema#integer>
3	slod:labw-2-2601341	<http://www.wikidata.org/entity/Q425837>	"56"^^<http://www.w3.org/2001/XMLSchema#integer>
4	slod:labw-2-2602552	<http://d-nb.info/gnd/300156405>	"56"^^<http://www.w3.org/2001/XMLSchema#integer>
5	slod:labw-2-2602552	<http://www.wikidata.org/entity/Q1165496>	"56"^^<http://www.w3.org/2001/XMLSchema#integer>
6	slod:labw-2-2601340	<http://d-nb.info/gnd/4099338-3>	"52"^^<http://www.w3.org/2001/XMLSchema#integer>
7	slod:labw-2-2601340	<http://www.wikidata.org/entity/Q425837>	"52"^^<http://www.w3.org/2001/XMLSchema#integer>
8	slod:labw-2-2601353	<http://d-nb.info/gnd/300162677>	"52"^^<http://www.w3.org/2001/XMLSchema#integer>
9	slod:labw-2-2601353	<http://www.wikidata.org/entity/Q657595>	"52"^^<http://www.w3.org/2001/XMLSchema#integer>
10	slod:labw-2-2599395	<http://d-nb.info/gnd/300132514>	"48"^^<http://www.w3.org/2001/XMLSchema#integer>

Use Case 2: Structured Data (via Websites)

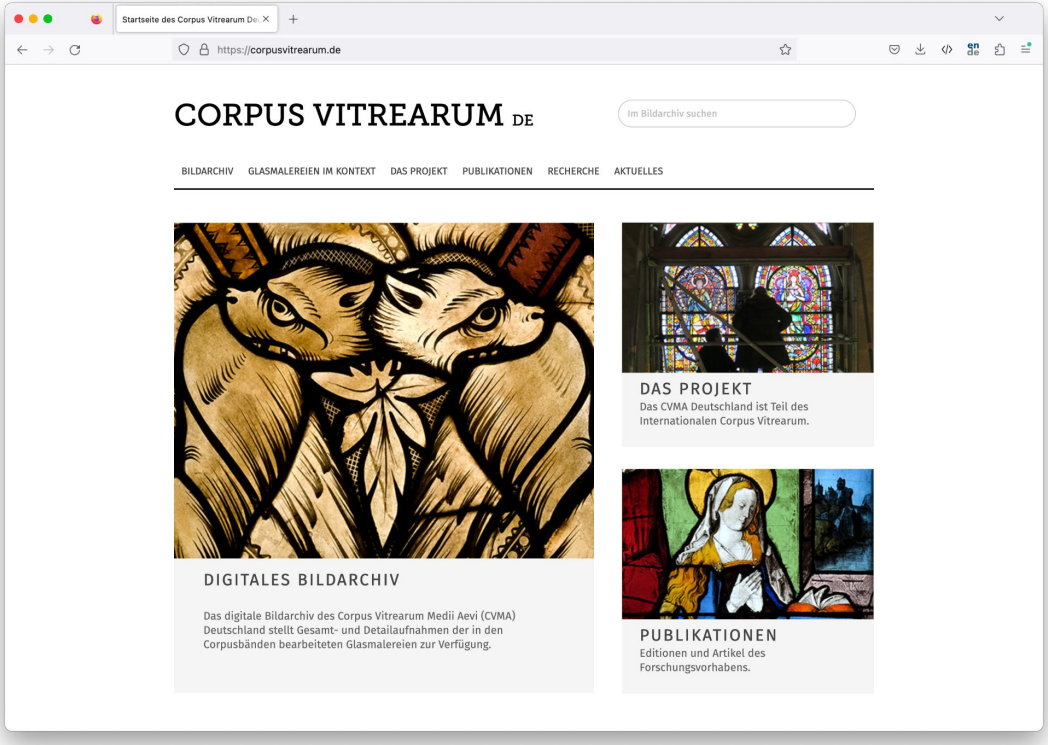
Partner: Corpus Vitrearum Deutschland







Corpus Vitrearum Deutschland (CVMA)



<https://corpusvitrearum.de>

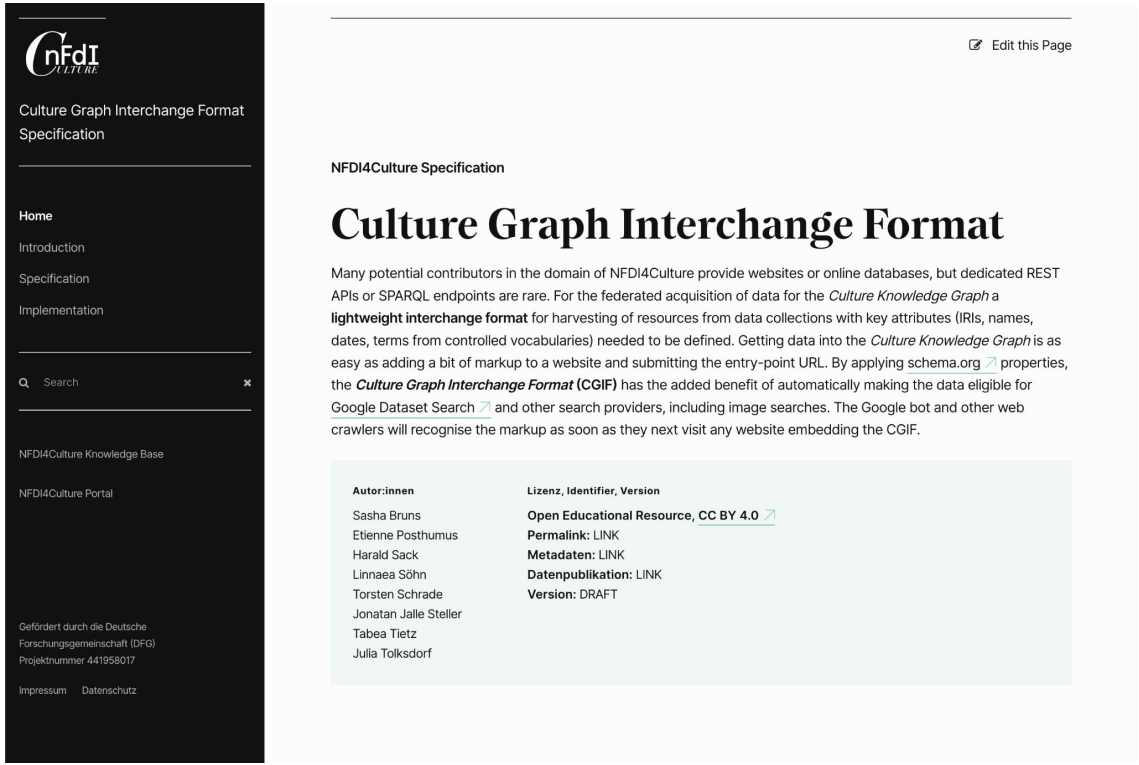
More than 7.000 high resolution digital images of medieval stained glass

XMP metadata and usage of controlled vocabularies and authority files (ICONCLASS, GND, Geonames)

Hydra Linked Data API

Structured data embedded in in web frontend of digital image archive

CVMA: Data Integration with CGIF



Culture Graph Interchange Format Specification

Home
Introduction
Specification
Implementation

Search

NFDI4Culture Knowledge Base

NFDI4Culture Portal

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) Projektnummer 441958017

Impressum Datenschutz

Edit this Page

NFDI4Culture Specification

Culture Graph Interchange Format

Many potential contributors in the domain of NFDI4Culture provide websites or online databases, but dedicated REST APIs or SPARQL endpoints are rare. For the federated acquisition of data for the *Culture Knowledge Graph* a **lightweight interchange format** for harvesting of resources from data collections with key attributes (IRIs, names, dates, terms from controlled vocabularies) needed to be defined. Getting data into the *Culture Knowledge Graph* is as easy as adding a bit of markup to a website and submitting the entry-point URL. By applying schema.org properties, the ***Culture Graph Interchange Format*** **(CGIF)** has the added benefit of automatically making the data eligible for Google Dataset Search and other search providers, including image searches. The Google bot and other web crawlers will recognise the markup as soon as they next visit any website embedding the CGIF.

Autor:innen	Lizenz, Identifier, Version
Sasha Bruns	**Open Educational Resource, CC BY 4.0**
Etienne Posthumus	**Permalink:** LINK
Harald Sack	**Metadaten:** LINK
Linnaea Söhn	**Datenpublikation:** LINK
Torsten Schrade	**Version:** DRAFT
Jonatan Jalle Steller	
Tabea Tietz	
Julia Tolksdorf	

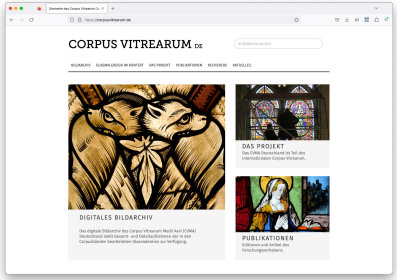
Version 1.0: https://nfdi4culture.de/go/E3712

Lightweight format based on Schema.org for harvesting resources with key attributes (IRIs, names, dates, terms)

No API required - integration into the Culture KG just by adding a bit of markup to a website

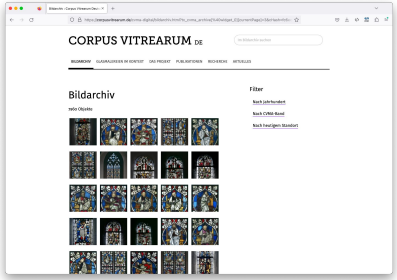
Added benefit of automatically making the website / data eligible for Google Dataset Search

CVMA: Application of CGIF Model



<https://corpusvitrearum.de>

a schema:DataCatalogue ;
schema:publisher <https://nfdi4culture.de/id/E1834> .

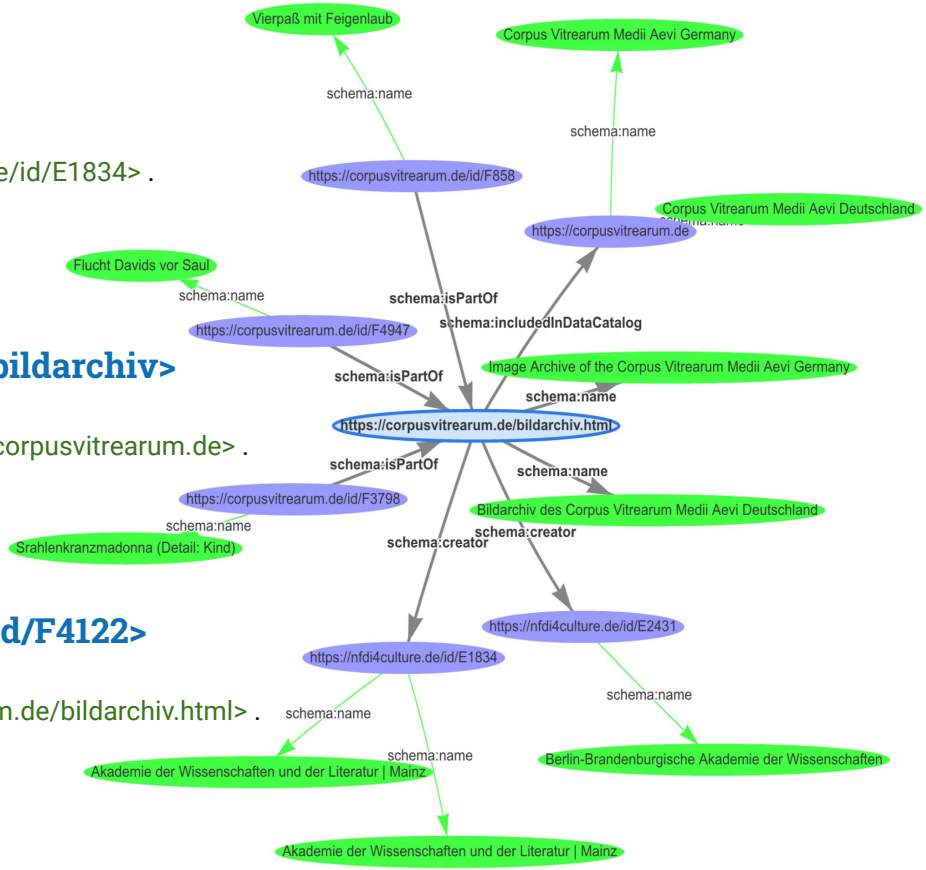


<https://corpusvitrearum.de/bildarchiv>

a schema:Dataset, schema:Datafeed ;
schema:includedInDataCatalog <https://corpusvitrearum.de> .

<https://corpusvitrearum.de/id/F4122>

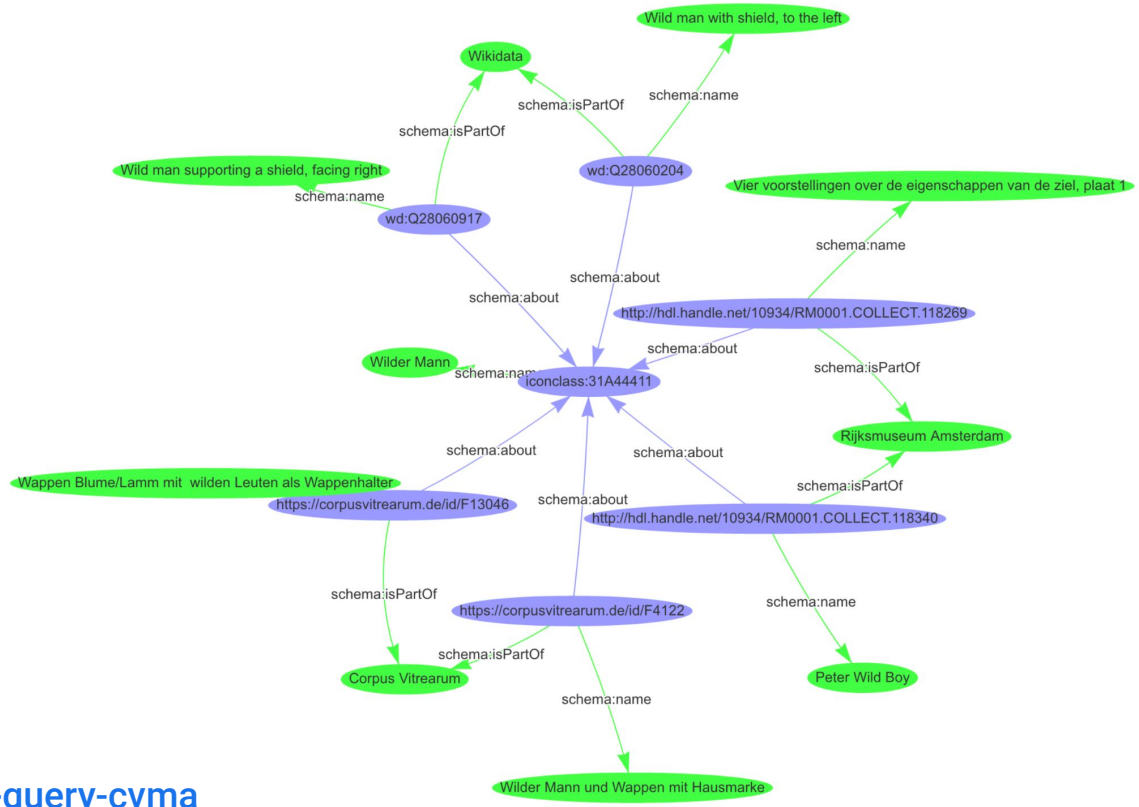
a schema:ImageObject ;
schema:isPartOf <https://corpusvitrearum.de/bildarchiv.html> .



CVMA: Data Federation via Culture KG

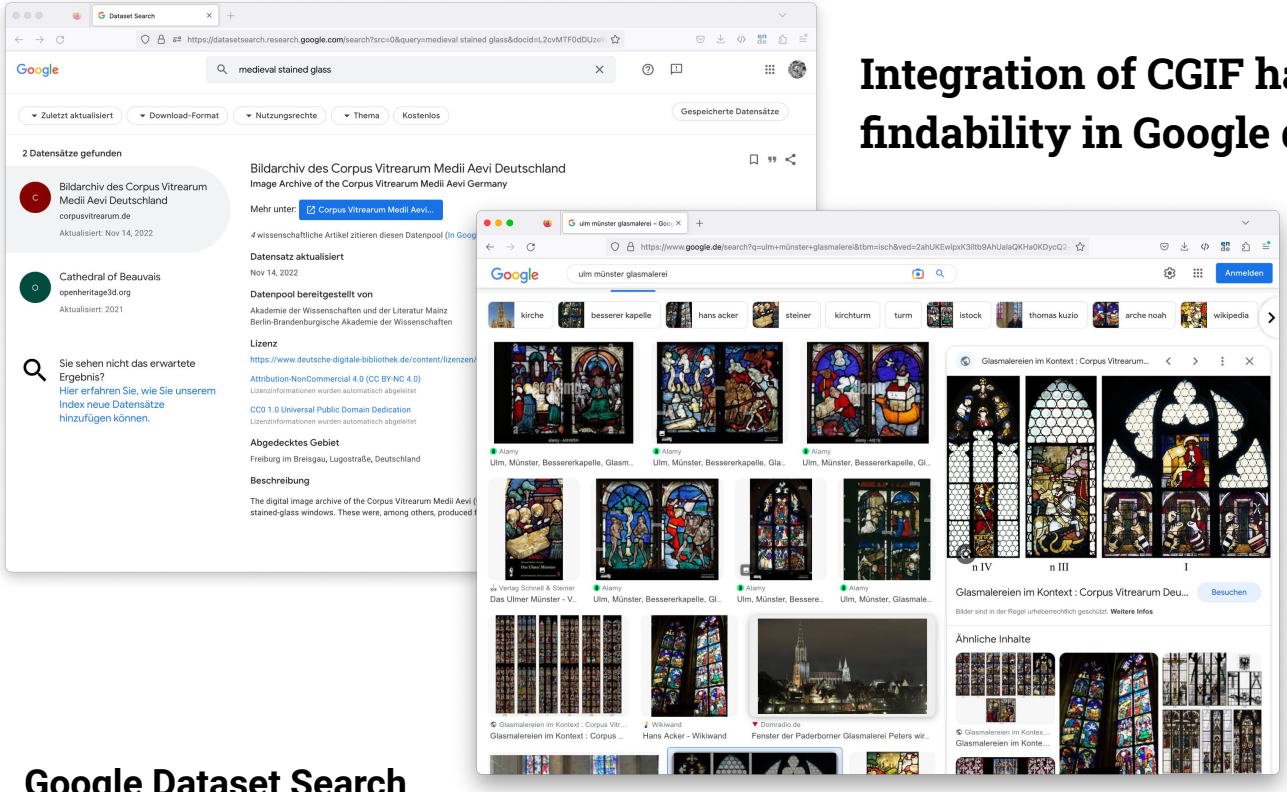
Find depictions with ICONCLASS “wild man” in Culture KG, Wikidata and Rijksmuseum

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX schema: <http://schema.org/>
PREFIX wd: <http://www.wikidata.org/entity/>
PREFIX wdt: <http://www.wikidata.org/prop/direct/>
PREFIX wikibase: <http://wikiba.se/ontology#>
PREFIX bd: <http://www.bigdata.com/rdf#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX nfdico: <https://nfdi4culture.de/ontology#>
PREFIX iconclass: <https://iconclass.org/>

SELECT ?resource ?resourceLabel ?dataCatalogue
WHERE {
   {
      SELECT ?resource ?resourceLabel ?dataCatalogue WHERE {
        ?resource schema:keywords iconclass:31A44411 .
        ?resource schema:name ?resourceLabel .
        BIND("Corpus Vitrearum / NFDI4Culture" AS ?dataCatalogue) .
      } LIMIT 2
   } UNION {
      SERVICE <https://query.wikidata.org/sparql> {
         SELECT ?resource ?resourceLabel ?dataCatalogue WHERE {
            ?resource wdt:P1257 "31A44411" .
            BIND("Wikidata" AS ?dataCatalogue) .
            SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
         } LIMIT 2
      }
   } UNION {
      SERVICE <https://api.data.netwerkdigitaalerfgoed.nl/datasets/Rijksmuseum/
         SELECT ?resource ?resourceLabel ?dataCatalogue WHERE {
            ?resource dc:subject <http://iconclass.org/31A44411> .
            ?resource dc:title ?resourceLabel .
            BIND("Rijksmuseum Amsterdam" AS ?dataCatalogue) .
         } LIMIT 2
      }
   }
}
```



Try it out: https://nfdi4culture.de/go/kg-query-cvma

CVMA: Enhanced SEO / Google Dataset Search



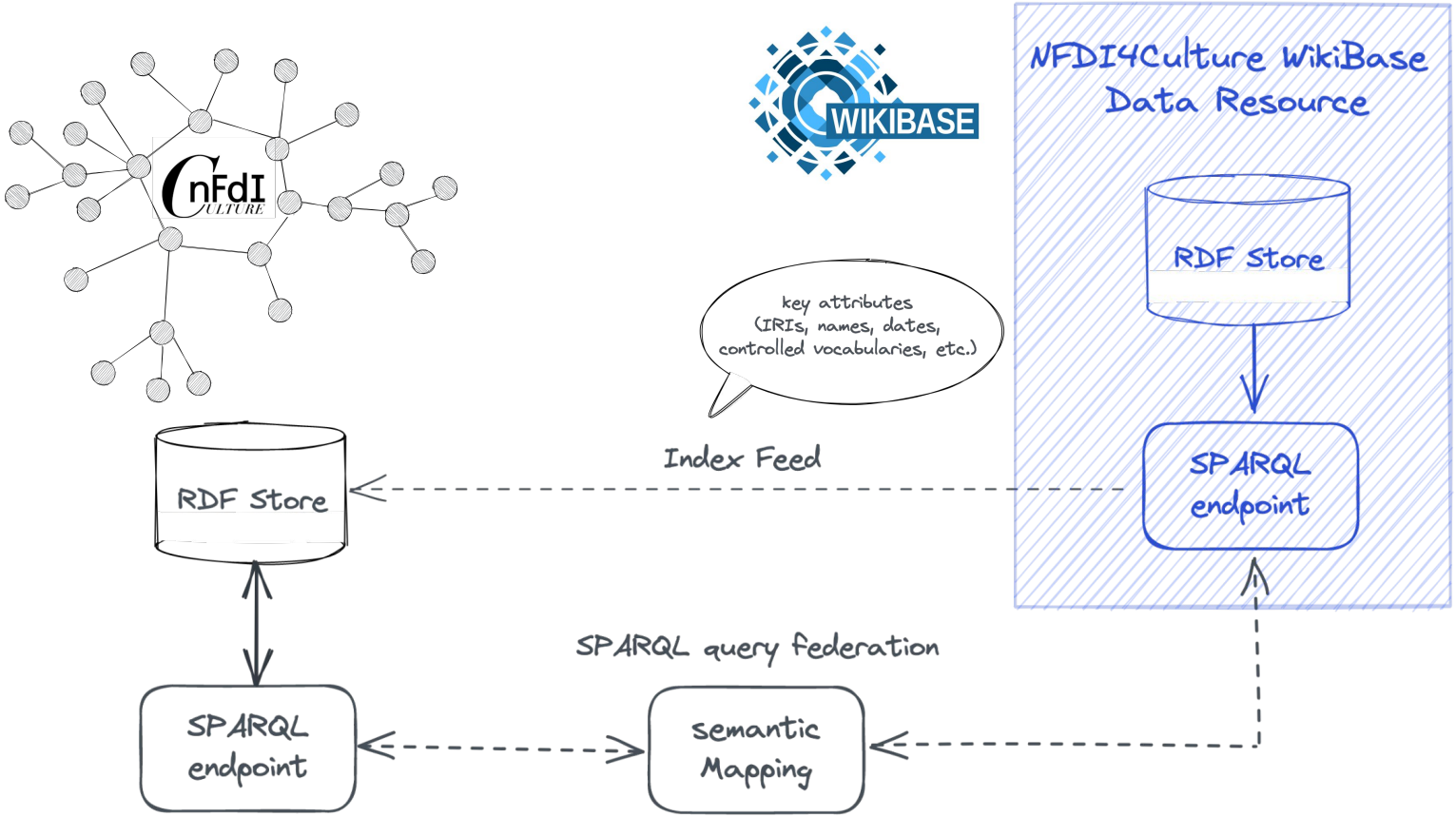
Integration of CGIF has significantly improved findability in Google dataset & image search

Read the use case report and CGIF integration tutorial:

https://nfdi4culture.de/go/cgif-cvma

Google Dataset Search
https://datasetsearch.research.google.com/help

Integration of Wikibase Knowledge Graphs



Next Steps & Outlook

- Enable interactive A-Box extension & registration/curation of datasets via portal
- Modular extension of NFDIcore
  - NFDI-MatWerk- and NFDI4Culture-Ontology
  - Planned adoption by NFDI4DataScience, MaRDI & NFDI4Memory
- Further extension of NFDI4Culture Research Information and Research Data Graph
- General adoption / implementation of CGIF by 4C partners
- Federation of NFDI4Culture KG with Wikidata, DDB, Europeana and other KGs

Quelle: Generated with stable diffusion 2.0 / Dreamstudio AI

Let’s talk

Integrate all the things !!

contact@nfdi4culture.de

Gefördert durch die
Deutsche Forschungsgemeinschaft (DFG) - 441958017