

There are a million ways to describe the world

How to share your authority without losing control

NFDI Satellite Event at the JCDL Cologne 2022
Slidedeck: <https://doi.org/10.5281/zenodo.6683365>



Introduction: Barbara Fischer, DNB.

Agenda



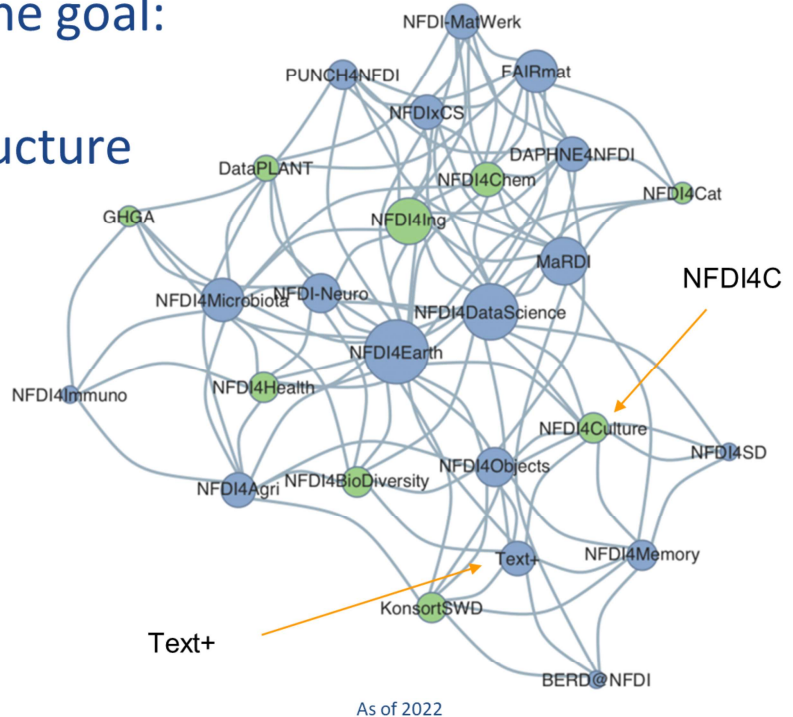
1. A backbone to the National Research Data Infrastructure (NFDI):
The Gemeinsame Normdatei (GND), *Jürgen Kett, head of the Office of Library Standards at the German National Library*
2. The perspective of NFDI4Culture, *Desiree Mayer, project manager at SLUB Dresden,*
3. The perspective of Text+, *Stefan Buddenbohm, project manager at SUB Göttingen*
4. Flipping the perspective - How is it in your country? - *Discussion with all participants*



Different disciplines, one goal: The National Research Data Infrastructure



Try out: [NFDI@Wikidata](https://www.wikidata.org/wiki/NFDI)



- National Research Data Initiative forms with federal funding a framework of research data infrastructure consortia.
- 3 funding rounds 2020/2021/2022 with 19 consortia as of 6/2022.
- Consortia usually with disciplinary scopes and linkages to already existing building RI blocks (e.g. ERICs or national projects).
- Up to 30 consortia in the end, with currently also a basic services consortium in the application process: Base4NFDI.
- NFDI as member of the [EOSC association](https://www.eosc-portal.eu/).
- A cluster of humanities and cultural science affiliated consortia/applications: the Memorandum of Understanding group consisting of NFDI4Culture, Text+ and NFDI4Memory, NFDI4Objects.

Meet your hosts:

1. **Stefan Buddenbohm** is contributing to research infrastructure projects since in 2009 at the Göttingen State and University Library. He is a long-time DARIAH-afficionado and currently working in the NFDI consortium [Text+](#) in the infrastructure/operations task area. He is particularly interested in community engagement.
2. **Jürgen Kett** is the head of the German National Library's Agency for Standardisation (AfS, DNB) and accountable to open, enrich and modernize the Integrated Authority File (GND)
3. **Desiree Mayer** studied musicology and philosophy at the Eberhard Karls University in Tübingen and received her doctorate in musicology from the Ludwig Maximilian University in Munich. As a scientist, she has been working e.g. at the Austrian Academy of Sciences in Vienna for the Digital Catalogue of the Works of Anton Bruckner. Since February 2021 she is working for [NFDI4Culture](#) at Saxon State and University Library Dresden, SLUB.
4. **Regine Stein** is the co-head of the Research & Development Department of the Göttingen State and University Library and Infrastructure/Operations speaker of [Text+](#). She is involved in standardisation work for documenting cultural heritage, focusing on data quality aspects of integrating and linking the information.
5. **Thorsten Trippel** is Part of the [Text+](#) coordination team. Studied English and Mathematics at Bielefeld University, Germany. For his Ph.D. thesis he re-used lexical resources provided by other researchers. Rich metadata and semantic interoperability are part of his major interests.
6. **Barbara Fischer** is the [GND](#) Liaison Counsel at the German National Library's Agency for Standardisation (AfS, DNB). She strives to increase the quality of metadata through Authority Control to foster retrieval and linked open data.



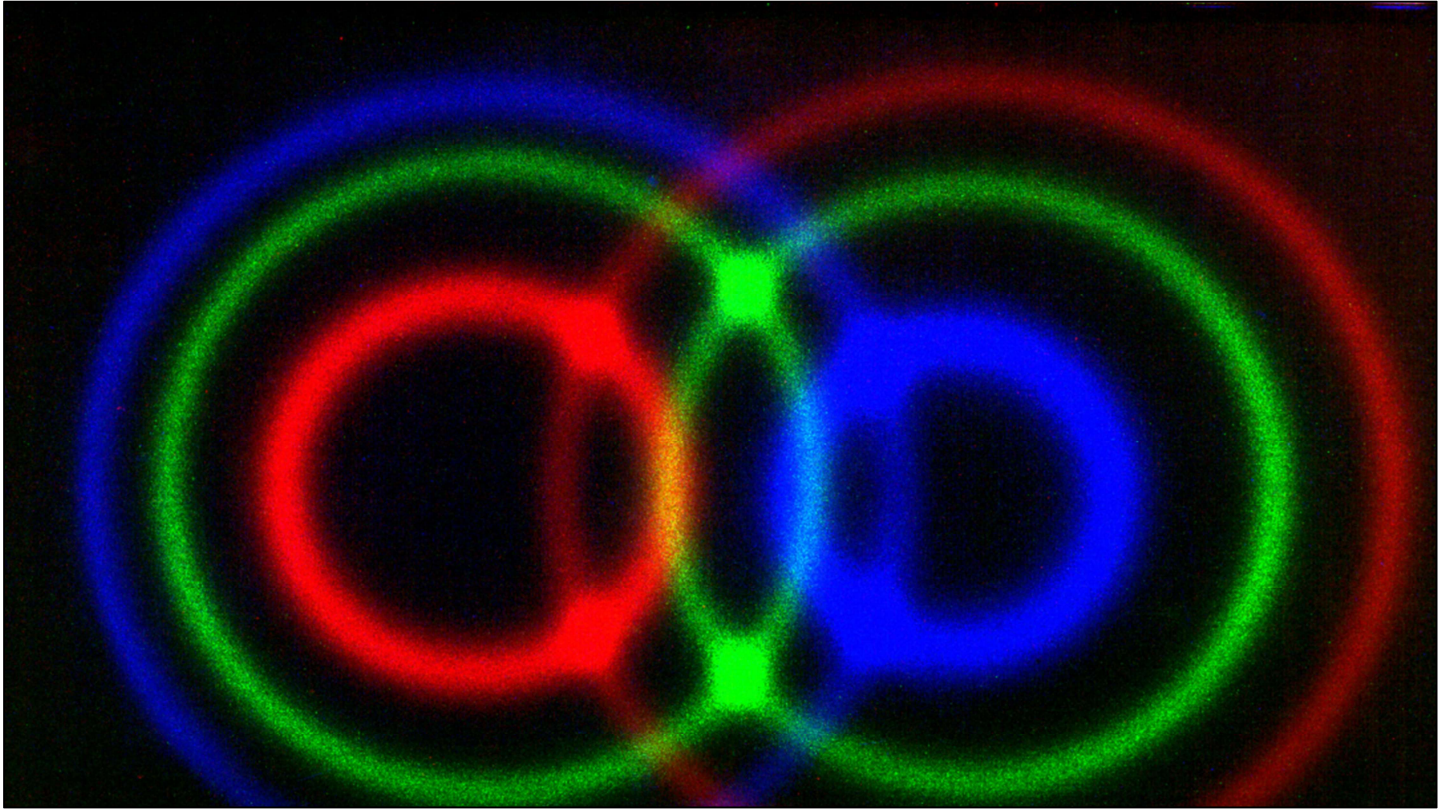
Gemeinsame Normdatei

A backbone to the
National Research Data Infrastructure

<https://gnd.network>

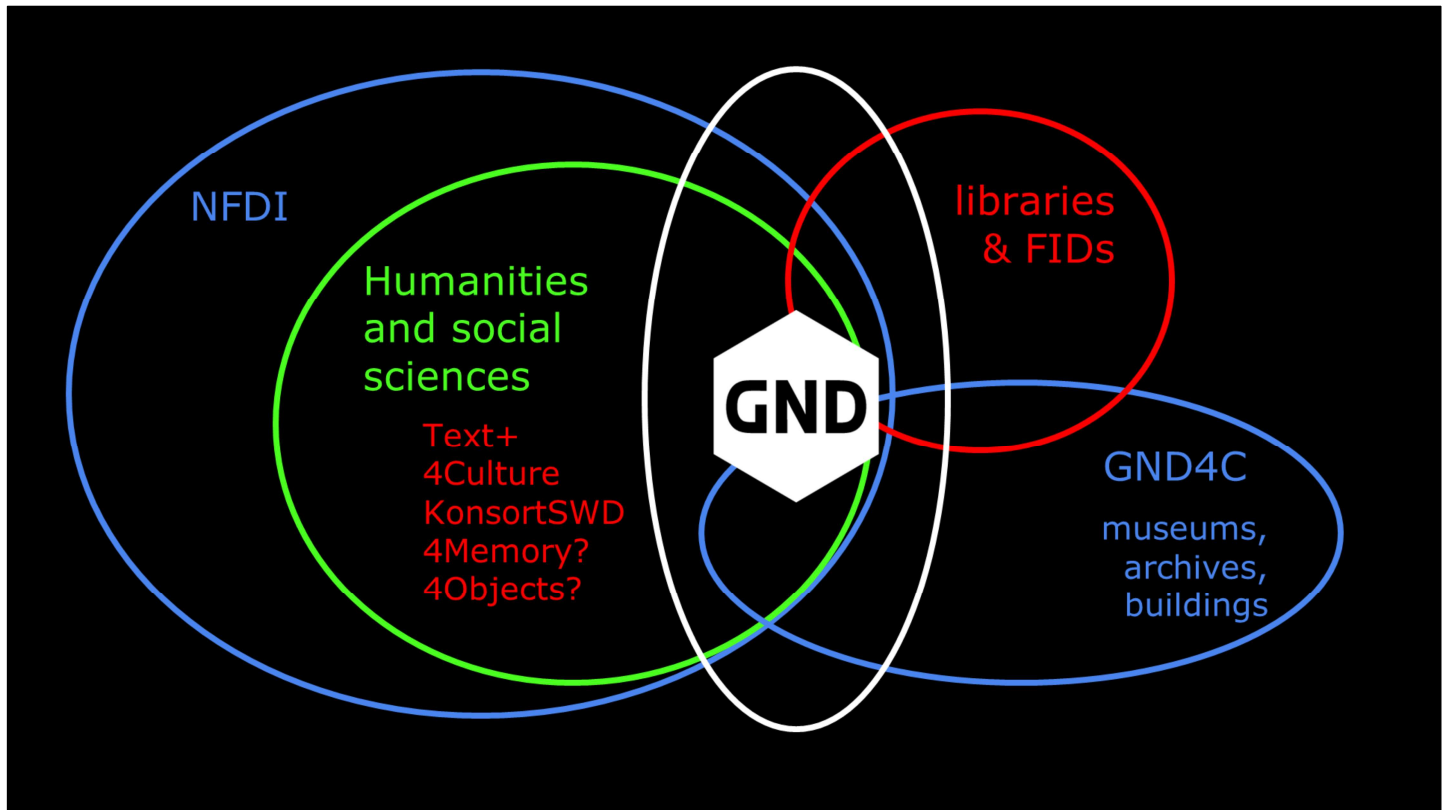
5

Presentation: Jürgen Kett, DNB.



This picture shows a three-photon interference. Thus, three photons have been successfully entangled with each other. After entanglement, these particles are in a common state that persists even over long distances.

This phenomenon reminded me of our work on the GND and NFDI: We are now trying to connect different networks and communities - or better, intertwine them.



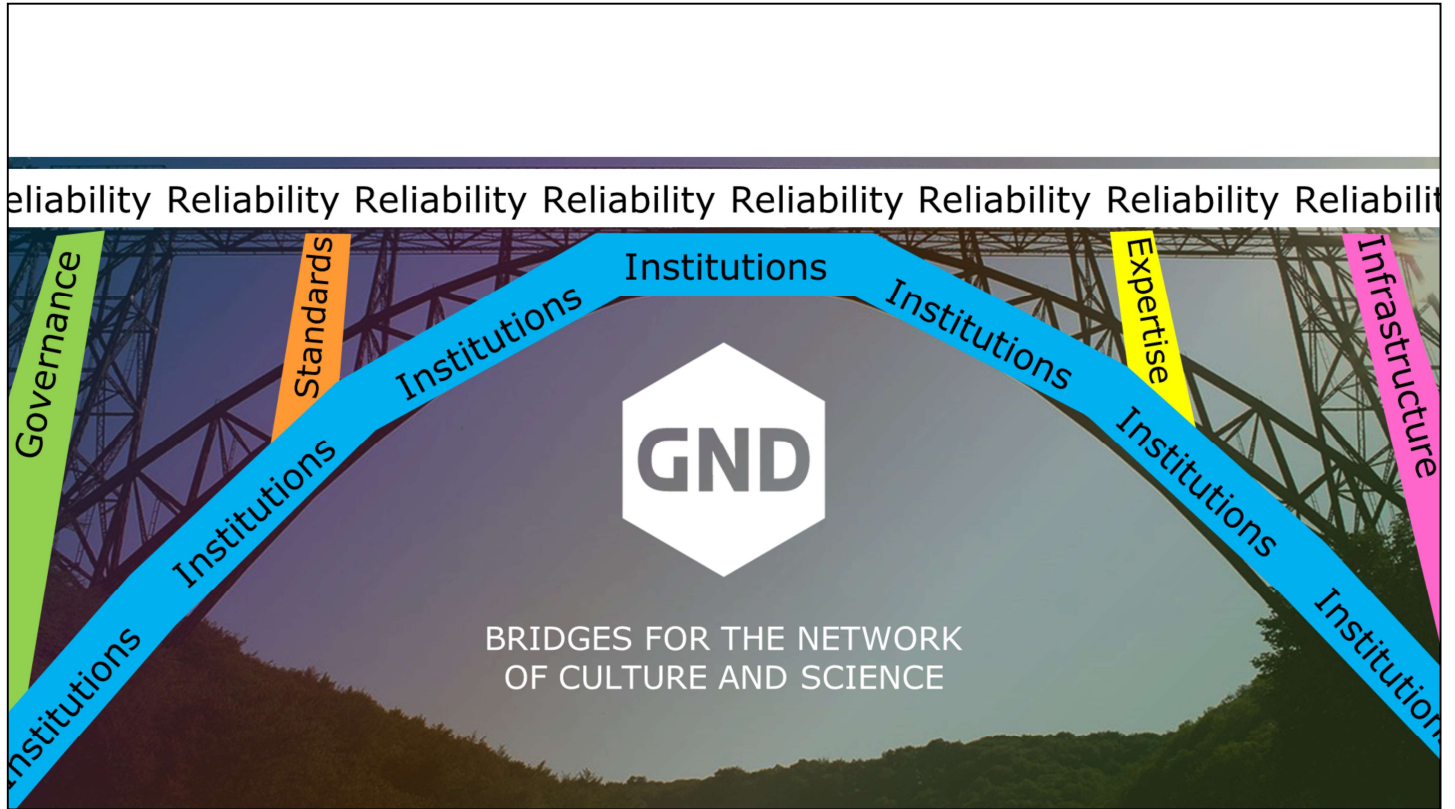
In NFDI GND is intended to act as a mediator or common interface. To remain in the language of physics: like a kind of boson mediating between the communities and their data.

In the NFDI, the humanities and social science play a particularly important role here, as the GND is particularly widespread in these areas and already fulfills many requirements in terms of content.

However, the GND is also used in some natural sciences and also has the potential there to generally play an important function as a mediator. It also brings with it the comprehensive knowledge graph of the existing GND community.



The function of the GND is to provide reliable bridges. These are bridges between data, the systems in which that data resides, and the associated communities. The GND can therefore also be seen as a kind of common language. A kind of lingua franca of the world of descriptive data. Malicious tongues claim that it is a kind of Esperanto of the librarians, which they now want to push on everyone.



The most important feature of the GND is its institutionally supported reliability. The GND is supported by over a thousand institutions. The reliability results from the permanent responsibility of these institutions and the standards, expertise and infrastructure that they share.

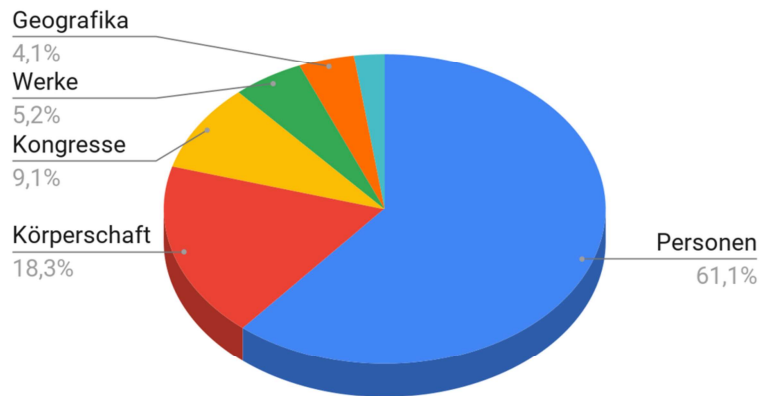


GND is a dataset

10 Mio. entities (05/2022)

- persons: 6.000.000
- corporate bodies: 1.800.000
- congresses: 890.000
- works: 510.000
- geo locations: 400.000
- subject headings: 220.000

Annual growth varies depending on projects



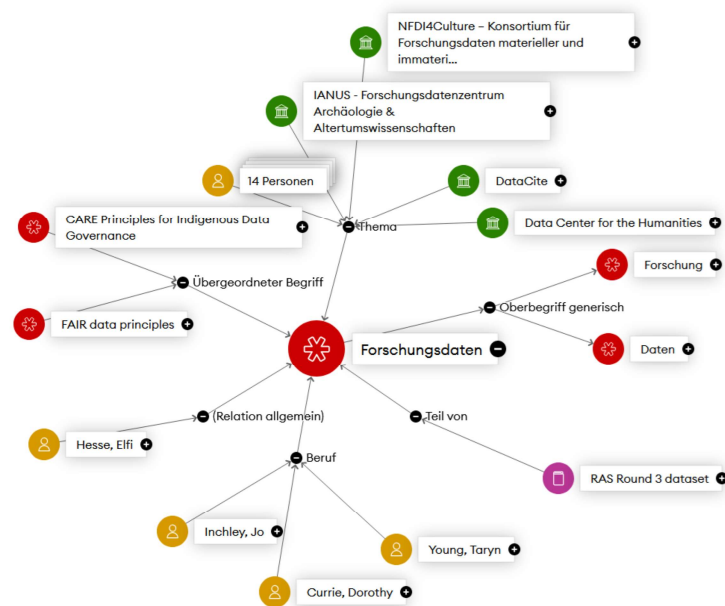
The GND is essentially a data set. It contains entries on persons, corporations, congresses, works, geographical locations and subject headings. This data set is supplemented according to the needs of the GND community.

Currently, there are about 10 million entities in the GND.

Each record uniquely represents an entity that can be machine-readably referenced by a persistent identifier. The data is recorded in a standardized format and according to fixed rules and standards.

The information in a GND record is therefore primarily used to uniquely identify an entity and to disambiguate it. It is not a data aggregation platform.

<https://explore.gnd.network/gnd/1098579690/relation>



As you can see in this screenshot from the GND Explorer, the individual GND entities are semantically linked to each other. The GND itself therefore already forms a knowledge graph of concentrated normalized records.

GND is a PID system for a distributed knowledge graph

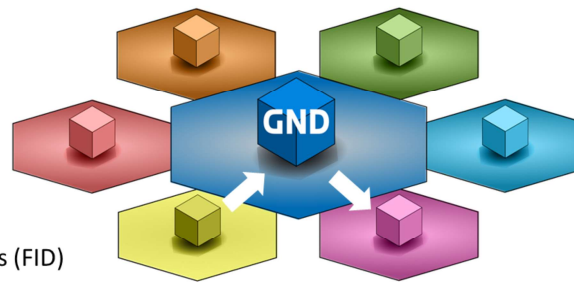
collections

libraries
archives
museums
portals

research data

research databases
specialised information services (FID)
research projects

mappings to other PI systems, ontologies and thesauri



incoming and outgoing connections

ORCID

Wikidata /Wikipedia

VIAF, LCSH, DDC, RAMEAU,
EMBNE, Nsogg,

STW, TheSoz, AGROVOC,
MeSH, T-PRO

...

12

The second part of the data are the links to the outside. They form the GND data network. The GND entities act as anchor points to which the connected data sets can refer. The second part of the data are the links to the outside. They form the GND data network. Here, the GND entities act like anchor points to which the connected datasets can refer.

The currently connected data sources essentially consist of:

- Metadata collections from libraries, archives, museums from German-speaking countries and beyond, and portals such as the Deutsche Digitale Bibliothek, Europeana, the newspaper portal, the Archivportal-D.
- Mappings to other PI systems and thesauri worldwide. These links allow GND users to move between different language spaces and subject spaces.
- Furthermore, connections to research data already exist. These can be from very different fields, but the majority of projects so far come from the field of humanities and social science. For example, research databases are enriched and linked with the GND, or entities in text documents are tagged via the GND.

The GND is community driven

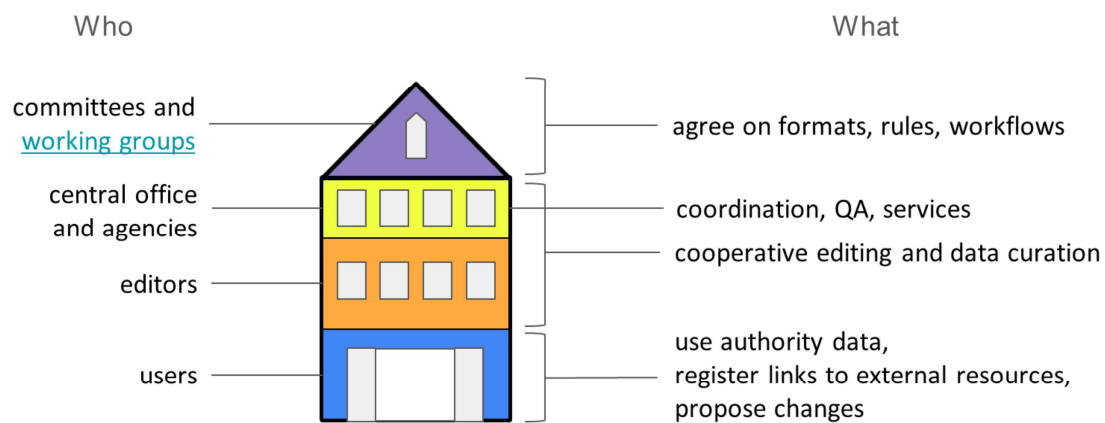
- The GND is community driven and strives to integrate a broad range of communities
- The right to contribute includes the task to care
- It regulates the amount of data the community shares



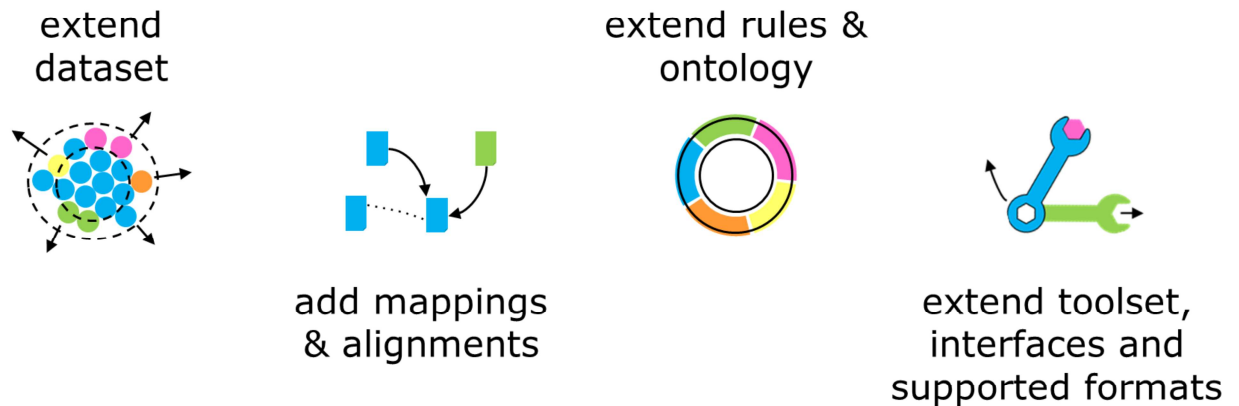
We are often asked which data and entities actually belong in the GND.

https://www.saemereien.ch/bilder/kk_dropper_uploads/kommunaler-stadtgarten.jpg

Governance: GND is a shared house



Integration of new communities requires specific adaptations and extensions



Using the GND is basically easy. The data is available under free license and can be searched and analyzed for matching via numerous tools and APIs.

In many areas of application, one can already get quite far with the existing possibilities.

However, in order to really integrate a new community fully, further development is usually required in various areas:

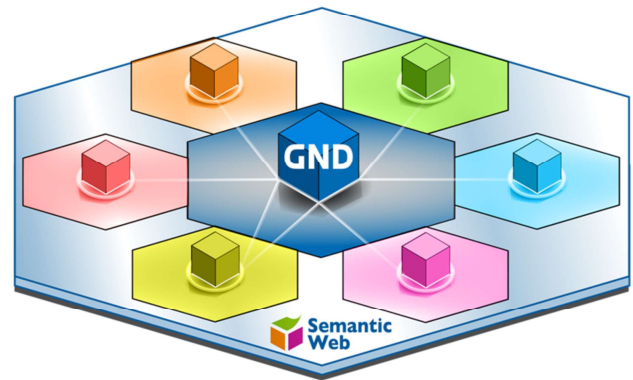
- Firstly: In most cases, it makes sense to systematically expand the data set so that frequently needed entities that were previously missing from the GND are already available to the new community.
- Secondly: It may also be necessary to map the GND to other knowledge systems in order to achieve compatibility with the standards of the respective community.
- Thirdly: Occasionally it is also necessary to slightly extend existing GND ontology to meet partner requirements.
- Fourthly: it is often necessary to enable the connection to existing systems and workflows by supporting additional APIs and formats. Also, sometimes new tools are needed for mappings and data management.

Overall goal in NFDI

Provide researchers with ideal opportunities to use the GND as an entry point and backbone for the NFDI knowledge graph.

For this purpose, researchers should have simple possibilities:

- to integrate GND into their systems and workflows
- to link their research data to GND
- to make their research data visible to the GND and NFDI community
- and to propose and supplement GND records themselves.





23th June 2022 | ACM/IEEE JOINT CONFERENCE ON DIGITAL LIBRARIES | NFDI Satellite Event

NFDI4Culture

Consortium for research data on material and immaterial cultural heritage

Speaker

Desiree Mayer | SLUB Dresden



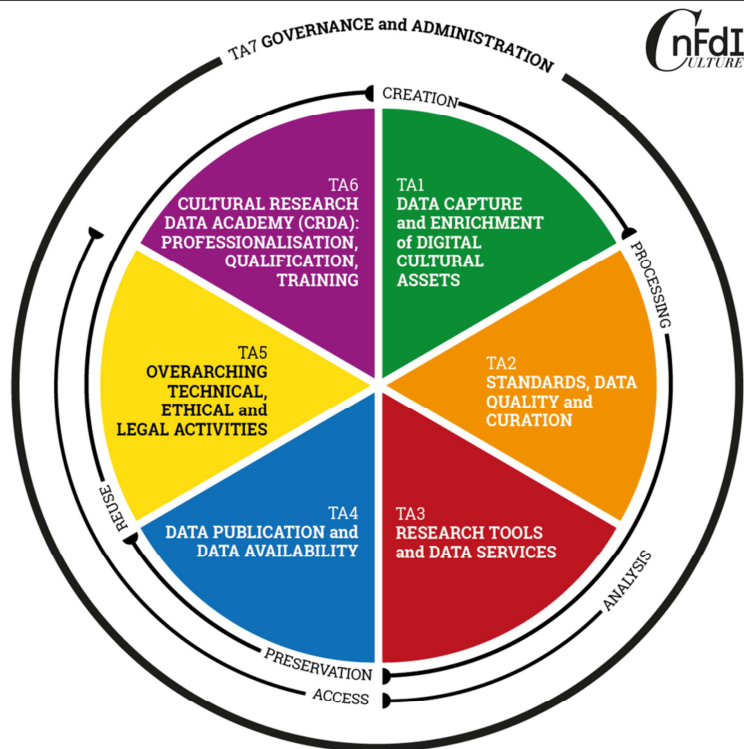
Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) - 441958017

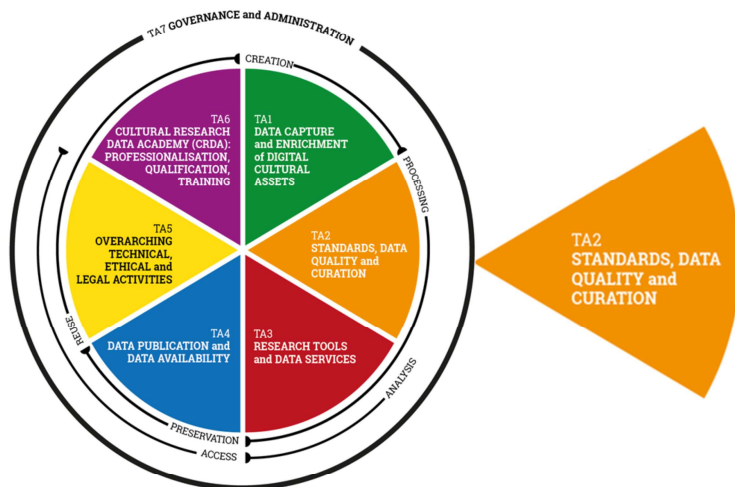
17

Presentation: Desiree Mayer, SLUB Dresden

- within the [Nationale Forschungsdateninfrastruktur \(NFDI\)](#) (National Research Data Infrastructure Germany)
- addresses research data on tangible and intangible cultural assets.
- needs-based infrastructure for research data
- community of interest: architecture, art history, musicology, theatre, dance, film and media studies
- **document, evaluate, operate and improve the services provided by members of our community.**

- 6 Task Areas aligned to the Research Data Life-Cycle
- From Digitization to Data Literacy and Training
- Cultural Coordination Office and Technical Coordination Office
- Crosscutting Teams: e.g. Team Helpdesk, Team Guidelines

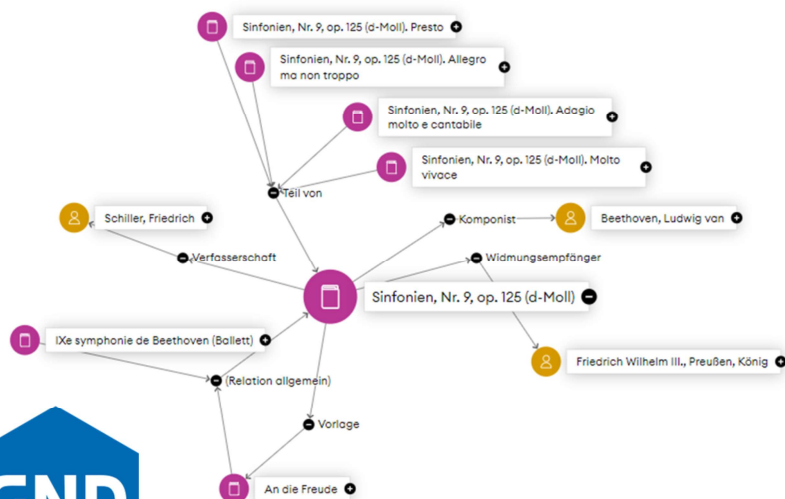




Aims

- Definition, development and application of data standards & criteria for quality assurance
- Guidelines for RDM in the field of non-textual cultural heritage
- Authority data & reference vocabulary development

Ludwig van Beethovens Symphony No. 9



Authority Data

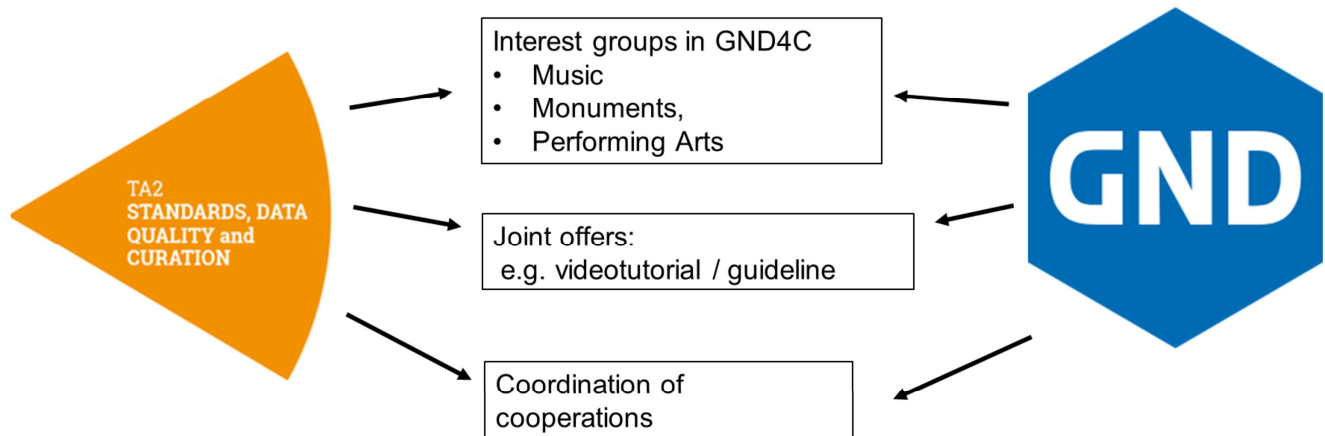
- Growing importance for humanities and cultural studies
- allow unique references and identification
- digital connection between dispersed cultural assets



Source: GND Explorer
<https://explore.gnd.network/>

The goal: Dataquality

FAIRness



Task Area for Datastandards, Quality and Curation:



- Acts according to needs in the different disciplines



- **Work Authority-Data**



- Buildings: GND-Agency
- Musicology: Cooperation RISM and GND
- Performing Arts: Working group within GND



TA2
STANDARDS, DATA
QUALITY and
CURATION

Buildings: GND-Agency

- improve the coverage of objects
- requirements of a non-librarian user group to the GND
- service point for national heritage agencies, building research and history of art and culture



TA2
STANDARDS, DATA
QUALITY and
CURATION

Music(ology): GND-RISM-cooperation

- Enabling the creation of authority records for musical works in cooperation with GND via Muscat
- Huge advantage for the community to work within one system and being able to contribute to GND
- Enlarge quantity and quality of Authority Data





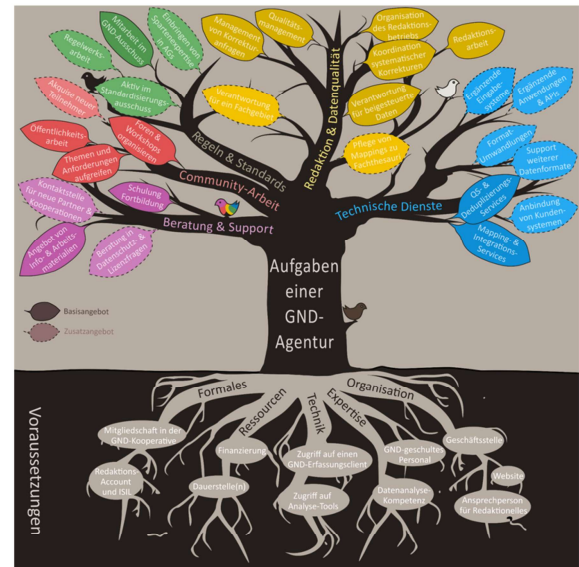
Performing Arts: Interest group

- Find a way to model authority data together with the community
- structural connection of production and reception processes is not adequately covered by the usual categories of work



Goals/Achievements:

- **Participation** through agency and collaborations
- **Representation:** Adding new Data
- **FAIRification** through using authority data
- **Networking**
- **Collaboration** together with Text+



Source: [J. Kett \(DNB\) CC BY SA](#)



nationale
Forschungsdaten
Infrastruktur
for CULTURE



NFDI4Culture

Consortium for research data on material and immaterial cultural heritage

Thank you very much for your attention!

DFG Deutsche
Forschungsgemeinschaft

Gefördert durch die Deutsche Forschungsgemeinschaft (DFG) - 441958017

Lizenzangabe: CC BY-NC-SA Desiree.Mayer@slub-dresden.de

THEODOR FONTANE: NOTIZBÜCHER
HRSG. VON GABRIELE RADECKE

Notizen

NOTIZBUCH-EDITION • (META-)EDITION • PROJEKT • ZITATION • NUTZUNG

67BETA, 28R

FAKSIMILE

TRANSKRIPTION

XML

The perspective of

Faksimile

Transkription

Code

Inhaltsübersicht

„Befehl dem Herrn“

„Denn der Tod ist“

Der Hausarzt – von Gustav von Mose

Des Meeres und der Liebe Wellen – von Grillparzer

Die Geier-Wally – von Wilhelmine von

Die Geier-Wally – von Wilhelmine von

(Notizen: Theaterkritik) (8.10.1881)

27r

28r

Sammlungen - Lexikalische Ressourcen - Editionen

Text+

Infrastruktur / Betrieb

Stefan Buddenbohm, Göttingen State and University Library

NFDI Satellite Event at the JCDL Cologne 2022

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

NIEDERSÄCHSISCHE STAATS- UND
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

SUB

DFG Deutsche
Forschungsgemeinschaft


Das NFDI-Konsortium Text+ wird gefördert durch die
Deutsche Forschungsgemeinschaft (DFG) - Projektnummer 460033370

Background: Theodor Fontane: Notizbuch 67, Hr. sg. von Gabriele Radecke. In: Theodor Fontane: Notizbücher. Digitale genetsch-kritische und kommentierte Edition. Hr. sg. von Gabriele Radecke. Blatt Blatt 28r. <http://fontane-nb.dariah.eu/edition.html?id=/xml/data/22/jtc.xml&page=28r>. Version 0.1 vom 20.09.2018, abgerufen am 23.6.2022

Presentation: Stefan Buddenbohm, SUB Göttingen

Agenda (max. 15 minutes, 7 slides)

Perspective of TEXT+ on the GND:

- practical example (text corpora)
- short introduction into Text+ with particular focus on the data domains and cluster
- authority files/ GND in Text+ and current state of affairs/challenges (infrastructural perspective regarding search and interoperability, GND Agentur Text+)
- authority files/GND as community effort: GND-Forum Text+
- Slidedeck: <https://doi.org/10.5281/zenodo.6683365>
-  unless indicated otherwise

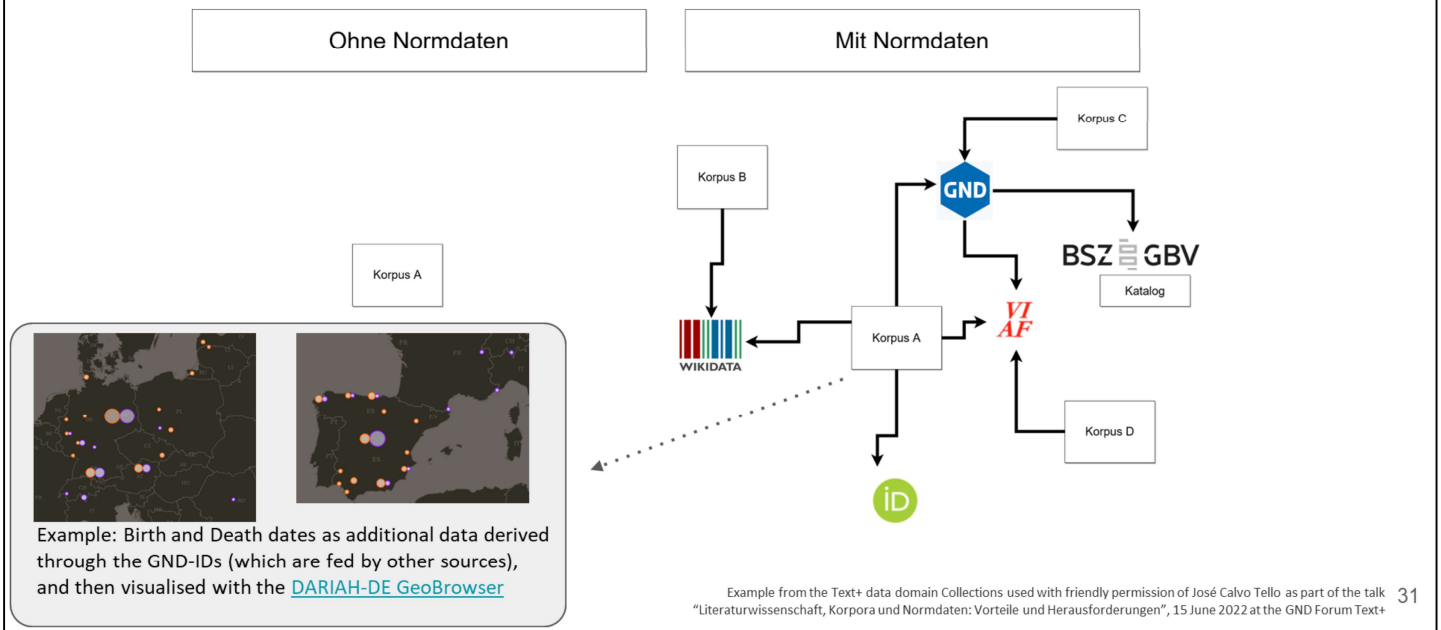


The Text+ applicant institutions. All contributing institutions may be looked up at: <https://www.text-plus.org/ueber-uns/beteiligte-institutionen/>

30

- Text+ is an NFDI consortium focussing on text- and language based research data. We address humanities and cultural sciences communities.
- In the following 15 minutes I will give you an introduction into the consortium's approach towards authority files and GND and highlight a few aspects crucial to our work. The main aspect I'd like to point you to is the community or communicative dimension of the topic.

Potential of the GND as hub for complementary information



- Let me start my talk with a very small practical example to illustrate the use and potential of the GND within our Text+ collections data domain.
- On the left side you see the case of a research undertaking which uses stand-alone corpora, in this case Korpus A.
- On the right side you can easily recognise the benefits resulting by the linking of various corpora through authority files, not only the GND. Striking the eye are the relations connecting Korpus A now with the GND, which is fed by other corpora. In our example we only highlight the birth and death dates of authors not included in Korpus A but can now be used applying the GND-IDs and visualised with a service like the DARIAH-DE GeoBrowser potentially allowing for an additional level of insight. In this case a variability between the German authors and the Spanish authors could be addressed and contribute to the initial research project.
- This additional level of enriched information may be found in the metadata OR - as is often the case with text- and language-based research data - directly in the data.
- So the task for us in Text+ is at least twofold:
 - investigate possible linkages of sources to the GND (or other authority files) AND
 - enriching our research data, e.g. tagging entities in a way that is interoperable with the GND

Text+

Aim: Text- and language based research data are at the centre of the Text+ research data infrastructure. Text+ focuses on three data domains: collections, lexical resources, and editions.

Strength: Cooperation and well networked with specialised information services (FIDs), academic societies of the SSH community (e.g., through the Text+ Scientific Coordination Committees). Large amounts of existing data sets for reuse by scholars of many disciplines in the humanities.

Experience: Long-standing reputation for providing distributed infrastructure components and services (e.g., AAI, PID, repositories, storage, computing resources, etc.).

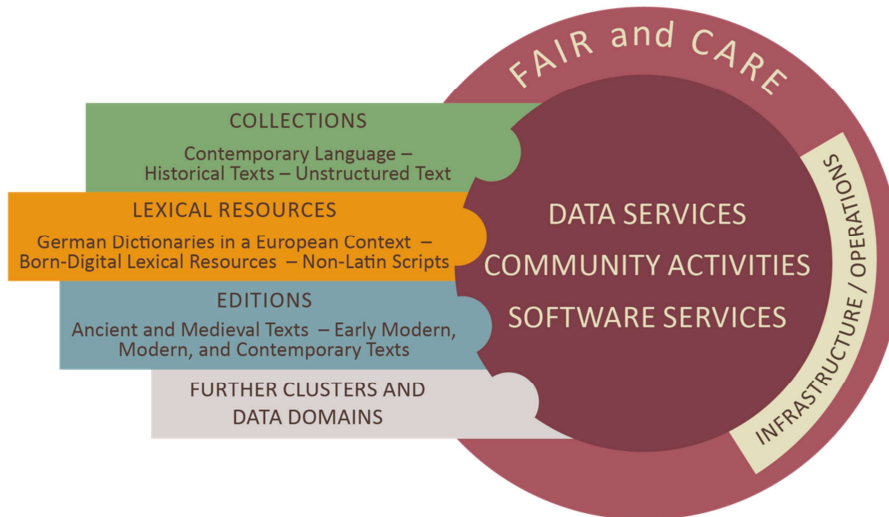


The Text+ applicant institutions. All contributing institutions may be looked up at: <https://www.text-plus.org/ueber-uns/beteiligte-institutionen/>

32

- Text- and language based research data are at the centre of our research data infrastructure. Text+ focuses on three data domains:
 - collections,
 - lexical resources, and
 - editions.
 - Perspectively additional data domains are possible.
- We are funded since autumn 2021. Of course Text+ is not a completely new activity from the scratch but integrates building blocks from its wide array of partners. Among these partners are universities, academic libraries, data centres of the Digital Humanities, members of the Union of German Academies of Sciences and Humanities, non-university research institutions of the Max Planck Society and the Leibniz Association. Text+ also includes leading data centres that provide robust and persistent services for a distributed research data infrastructure.
- We are well networked with the specialised information services (FIDs) and the academic societies of the SSH community (e.g., through the Text+ Scientific Coordination Committees).
- The partners come with a long-standing reputation for providing distributed infrastructure components and services (e.g., AAI, PID, repositories, storage, computing resources, etc.). and the portfolio of Text+ includes building blocks from CLARIAH-DE, CLARIN and DARIAH.

Data Domains and Thematic Clusters

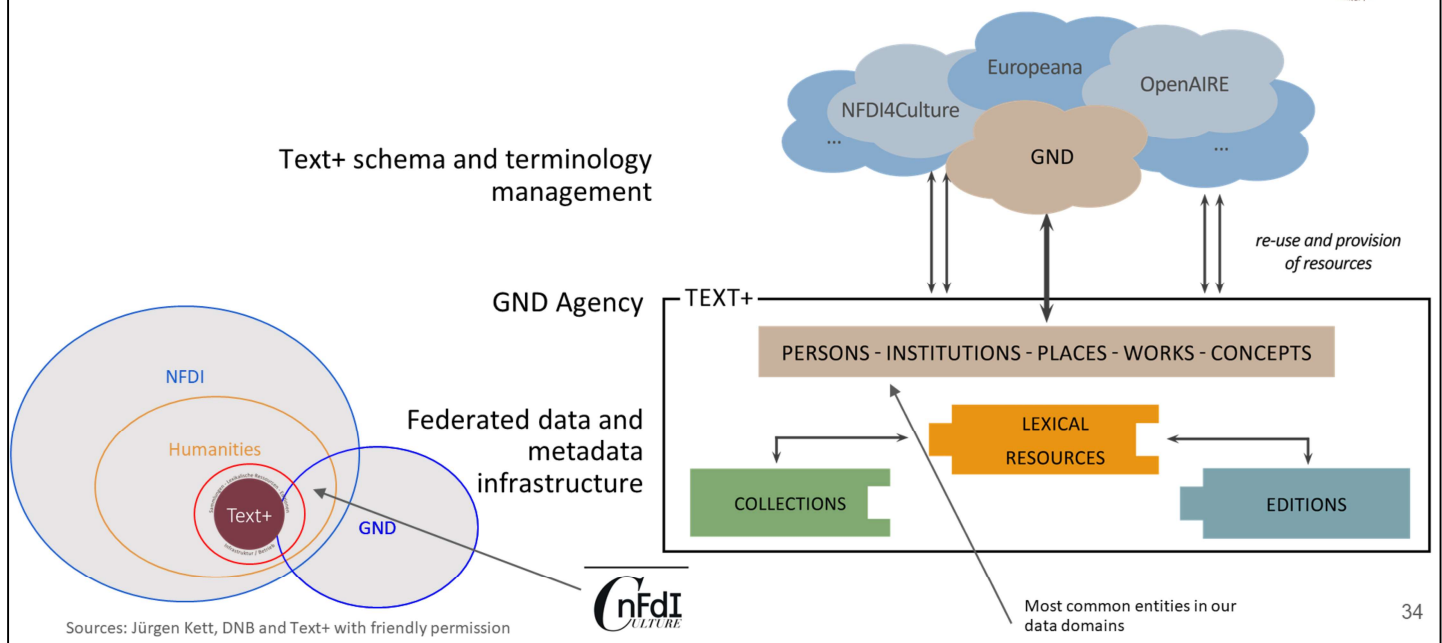


Source: Text+ with friendly permission

- Clusters bundling activities related to specific subtypes of data and research methods in a data domain.
- A cluster usually consists of various specialised data and competence centres.
- Trade-off between generic and domain-specific services to be addressed by Infrastructure/Operations.
- Modular structure open for additional domains and clusters.
- Clusters come with specific requirements regarding authority files and ownership of resources (common challenge for other consortia as well).

- The thematic clusters - as seen in the left boxes below the domain titles - are at the core of the Text+ Research Data Management strategy. The clusters bundle activities related to specific subtypes of data and research methods in one data domain. Each cluster consists of specialized data and competence centres.
- The clusters offer services, which build on generic services provided by the task area Infrastructure/Operations and are then tailored to cluster-specific requirements. At this point you have to deal with a trade-off between generic and domain-specific services. This trade-off has to be addressed by Infrastructure/Operations, which not only comes with a technical dimension but also with a communication-related one.
- The modular structure of Text+ is open for additional domains and clusters.
- The conclusion with regard to authority files and the GND at this point is: clusters come with specific requirements regarding authority files and ownership of resources. This is very likely a common challenge for other consortia as well. Particularly the ownership aspect of resources will be of interest for us later on.

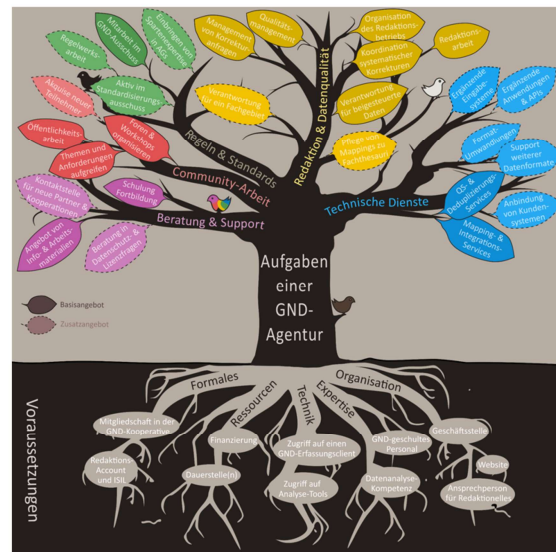
Search and Interoperability in Text+ (TA IO)



- To put a spotlight on one of the Text+ offers, I will briefly present our search and interoperability approach, which is visualised on the right side of this slide.
- We will implement a federated data and metadata infrastructure based on existing components. It ensures findability and unified presentation of all data and metadata on different levels of granularity.
- We will implement a GND agency for text- and language-based research data as a new service, which extends the scope of the GND, the Integrated Authority File of the German National Library, for research data in alignment with the relevant standardization committees. Schema management and export services allow for integration of Text+ data with Knowledge Graphs, namely those especially relevant for the NFDI and for the humanities.
- In the left bottom corner you see a visualisation of the overlaps with regard to the GND. The humanities circle includes other NFDI4Culture, which was presented before.

GND Agentur Text+

- agency as contact point for the Text+ community and perspective for the data domains/clusters in the scope of Text+
- good example for the joint/network character of authority files/GND
- currently focussing on establishing reliable working/communication relations with the data domains AND pursuing implementation scenario for linking of resources to the GND (= e.g. 3 reference editions in IO M3)
- [Kett, Jürgen/Kudella, Christoph/Rapp, Andrea/Stein, Regine/Trippel, Thorsten \(2022\): Text+ und die GND – Community-Hub und Wissensgraph. In: Zeitschrift für Bibliothekswesen und Bibliographie \(ZfBB\) 69 1-2/2022. S. 37-47.](https://doi.org/http://dx.doi.org/10.3196/1864295020691262)
<https://doi.org/http://dx.doi.org/10.3196/1864295020691262>



Source: [J. Kett \(DNB\) CC BY SA](https://www.dnb.de)

- Beyond the important infrastructural view on the topic as seen on the slide before, I would like to shift our focus here to the community or communication aspect of the topic.
- Why is this aspect so important? The title of our workshop sessions gives it away: There are a million ways to describe the world - How to share your authority without losing control.
- I wouldn't describe our outreach- and networking approach as a means of control but as coordination or alignment process. The cooperation with the community ensures that both sides keep track of each other and find opportunities to align their activities to make clear which benefits come with the use of authority files and standards .
- The GND Agentur Text+ will pursue this goal for its data domains of language- and text-based research data. The agency shall serve as contact point for the Text+ community and perspective for the data domains/clusters in the scope of Text+.
- You have seen the agency tree already before and I display it again here because this is a good example for the joint character of the GND. Other agencies for other data domains are part of the larger concept and Text+ aligns its work with this very useful schema of potential agency tasks.
- Currently in Text+ we are focussing on establishing reliable working/communication relations with the data domains AND pursuing implementation scenarios for linking of resources to the GND, with most progress in the data domain Editions.
- If you are interested in learning more on our approach towards the GND please consult the linked journal article "Text+ and the GND: Community-Hub and Knowledge Graph".

Text+

Sammlungen / Lexikalische Ressourcen - Editoren

Infrastruktur / Betrieb

NFZ DIREKTORAT

NFZ Konsortium

NFZ Konsortium

NFZ Konsortium

NFZ Konsortium

GND-Forum Text+

Hauke Bartels, Sorbisches Institut
Stefan Buddenbohm, Niedersächsische Staats- und Universitätsbibliothek Göttingen
José Calvo Tello, Niedersächsische Staats- und Universitätsbibliothek Göttingen
Barbara Fischer, Deutsche Nationalbibliothek
Jürgen Kett, Deutsche Nationalbibliothek
Harald Lordick, Salomon-Ludwig-Steinheim-Institut für Deutsch-Jüdische Geschichte
Uwe Sikora, Niedersächsische Staats- und Universitätsbibliothek Göttingen
Marcin Szczepański, Sorbisches Institut

**DEUTSCHE
NATIONALBIBLIOTHEK**

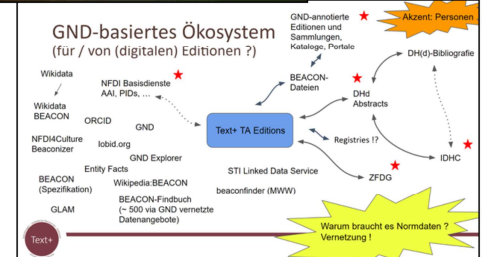
**GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN**

**NIEDERSÄCHSISCHE STAATS- UND
UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN**

SUB

Das NFZ-Konsortium Text+ wird gefördert durch die
Deutsche Forschungsgemeinschaft (DFG) - Projektnummer 460033370

Teil der
GND-Forum
15.06



Source: Harald Lordick, Salomon Ludwig Steinheim Institut

[illegible]

Source: Hauke Bartels, Sorbisches Institut

- aim: awareness raising for the benefits of contributing to/ using of authority files within the Text+ data domains
- more of a social/communicative effort than an infrastructural/technical one
- series of meetings during the coming years

- The most recent practical implementation of our community-aimed approach is the GND Forum Text, which took off a week ago and brought together 60 colleagues, mostly from the Text+ data domains but also beyond. Over the course of a day we had presentations coming from our data domains and breakout-sessions who worked on the domain-specific requirements. The results are important for the work on the GND Agentur Text+ which will be set up by the task area infrastructure operations.
- The format GND Forum is not an invention by Text+ but a proven format of the DNB within other contexts such as the 4C Forum on Performing Arts from January or a NFDI4Culture Forum a few weeks ago. For us the forum functions as a means of awareness raising, requirements engineering and the promotion of the practical implementation of authority files and the GND within Text+.

Flipping the perspective



What role do authority data and metadata standards play in your research environment? What are Your remarks on setting up a national research data infrastructure?



- We want to learn more about your practices regarding controlled vocabularies/authority files.
- Which communities of practice are relevant for you?
- What can you tell us about your relations to library specific resources and standards?
- What about the challenges/obstacles regarding internalization of?

Stay in touch

Deutsche Nationalbibliothek Jürgen Kett & Barbara Fischer

j.kett [a] dnb.de; b.k.fischer [a] dnb.de

IDS Mannheim Thorsten Trippel

trippel [a] ids-mannheim.de

SLUB Dresden Desiree Mayer

Desiree.Mayer [a] slub-dresden.de

SUB Göttingen Regine Stein & Stefan Buddenbohm

regine.stein [a] sub.uni-goettingen.de; Buddenbohm [a] sub.uni-goettingen.de

