

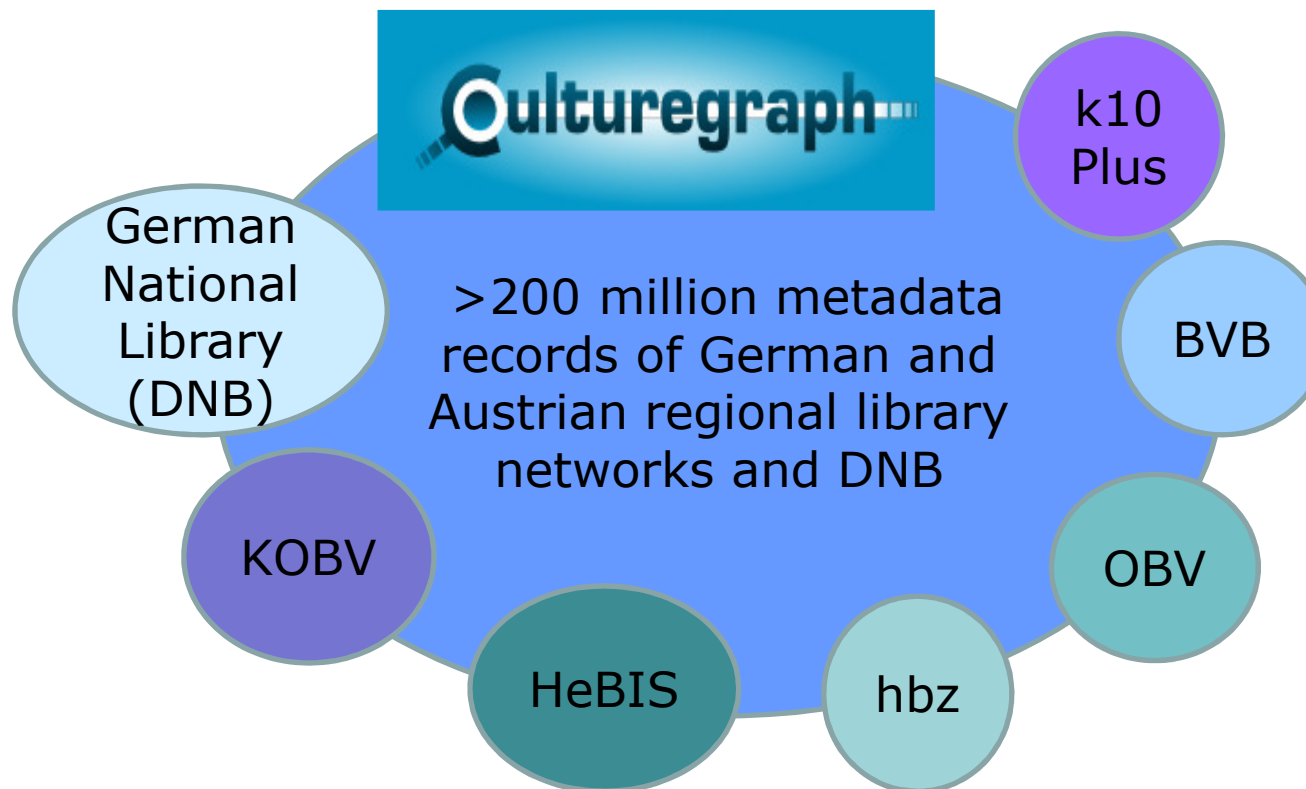
Data science & Data analysis at German National Library (DNB)

1. Data analysis

- **Work clustering in Culturegraph**
- **Focus on persistent identifiers**
- **Automatic indexing**

2. Research support

Data basis



Data science & Data analysis at German National Library (DNB)

Hauptsachtitel	Das Meer Als Versprechen
Zusatz	Bedeutung Und Funktion Von Seeherrschaft Bei Thukydides
Inhalt	Inhaltsverzeichnis Inhaltstext Inhaltstext
Person	aut Kopp, Hans 1127785095
Körperschaft	pbl Vandenhoeck & Ruprecht
Umfang	303 Seiten
Erscheinungsjahr	2017
Material	book
Erscheinungsort	Göttingen
Herausgeber	Vandenhoeck & Ruprecht
Schlagwort	655 Hochschulschrift 4113937-9 650 Seeherrschaft 4077302-4 600 Thucydides 4138063-0
Klassifikation	DDC 355 DDC 880 DDC 938 DDC 938.05
Standard-Identifizier	OCLC 961837268 EKL DNB1115738097
Verlags-Identifizier	GTIN14 9783525253243 ISBN13 9783525253243

Hans Kopp



Das Meer als Versprechen

title+remainder of title-author/contributor-
volume-publication type:
meeralsversprechenbedeutungufunktionvonse
eherrschaftbeithukydides-kopphans-X-text



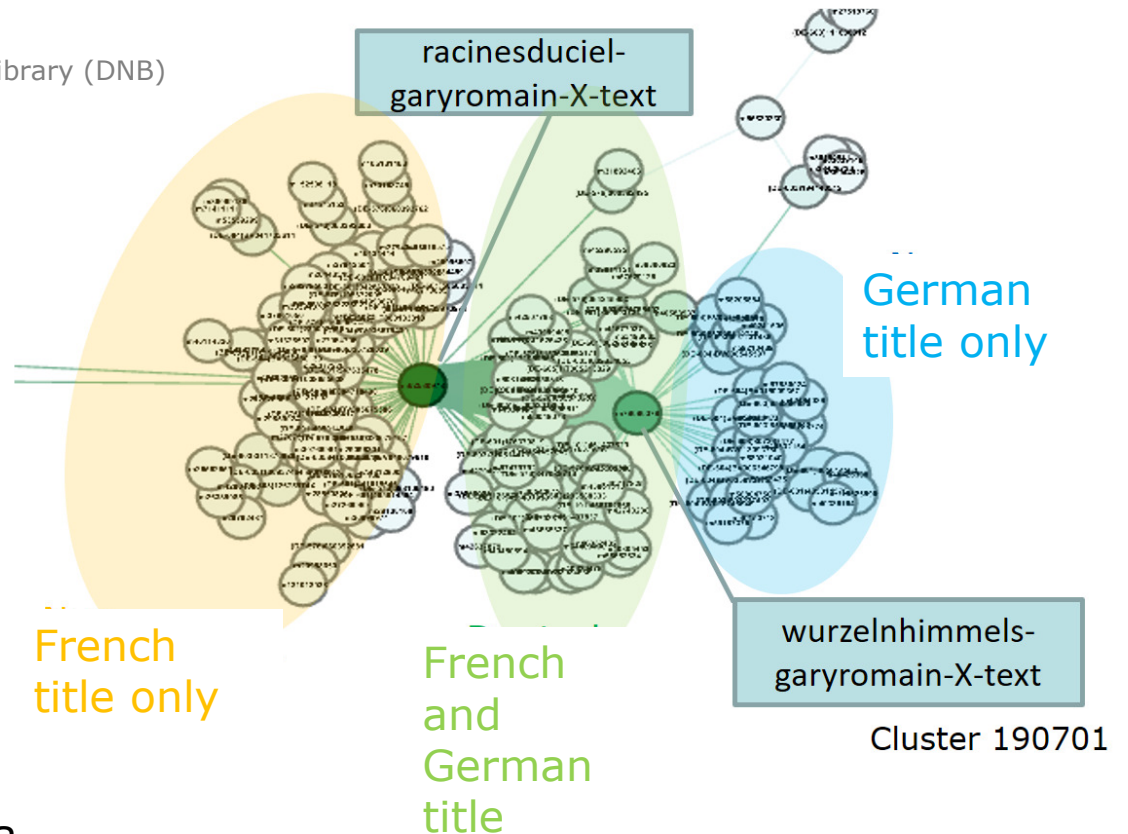
Control number-volume-title-publication
type:
oclc961837268-X-meeralsversprechen-text

Control number-volume-title-publication
type:
DNB1115738097-X-meeralsversprechen-text

ISBN-volume-title-publication type :
9783525253243-X-meeralsversprechen-text

Work cluster

- Each record possesses one or more matchkeys
- Clusters accumulate identical keys
- All records sharing at least one identical key with a member of a cluster belong to the cluster



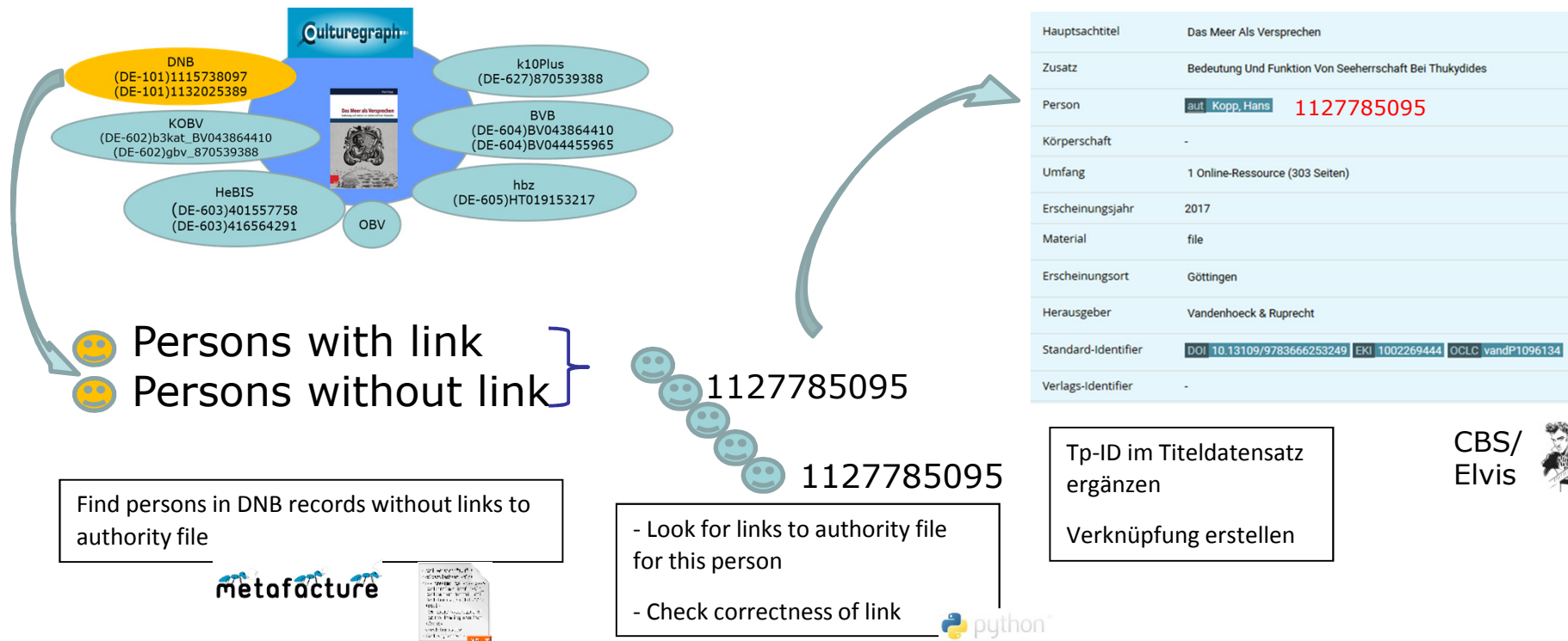
Possible transfers of metadata content

Hauptsachtitel	Das Meer Als Versprechen
Zusatz	Bedeutung Und Funktion Von Seeherrschaft Bei Thukydides
Person	aut Kopp, Hans
Körperschaft	-
Umfang	1 Online-Ressource (303 Seiten)
Erscheinungsjahr	2017
Material	file
Erscheinungsort	Göttingen
Herausgeber	Vandenhoeck & Ruprecht
Standard-Identifizier	DOI 10.13109/9783666253249 EKI 1002269444 OCLC vandP1096134
Verlags-Identifizier	-

Transfer

Hauptsachtitel	Das Meer Als Versprechen
Zusatz	Bedeutung Und Funktion Von Seeherrschaft Bei Thukydides
Inhalt	Inhaltsverzeichnis Inhaltstext Inhaltstext
Person	aut Kopp, Hans 1127785095
Körperschaft	-
Umfang	303 Seiten
Erscheinungsjahr	2017
Material	book
Erscheinungsort	Bristol, CT
Herausgeber	Vandenhoeck & Ruprecht
Schlagwort	655 Hochschulschrift 4113937-9 650 Seeherrschaft 4077302-4 600 Thucydides 4138063-0
Klassifikation	DDC 938.05 RVK FH 26175 RVK NH 5850
Sachgruppe	DNB 355 DNB 880 DNB 930
Standard-Identifizier	DNB 1115738097 OCLC 968670393 EKI DNB1115738097
Verlags-Identifizier	GTIN14 9783525253243 ISBN13 9783525253243

Transfer of links to person authority files from work clusters



Enrichment with content descriptions

```
*****
*****
3. Analyse, welche Inhalte im Bündel und/oder Titel vorhanden sind:
*****
*****
RSWK-Folgen im Titel vorhanden
FormSLW/Angaben zum Inhalt im Titel vorhanden
DDC-Notationen im Titel nicht vorhanden

RSWK-Folgen im Bündel vorhanden
Schlagwörter im Bündel vorhanden
FormSLW/Angaben zum Inhalt im Bündel vorhanden
Klassifikationen im Bündel vorhanden
DDC-Notationen im Bündel nicht vorhanden
DDC/Synthetische Komponenten im Bündel nicht vorhanden
```

DNB records with sequences of index terms

None

DNB records without sequence of index terms

Transfer of
sequences of
index terms from
other DNB-
records

If none of the preceeding

Transfer of
sequences of index
terms from title
records from other
regional library
networks

If none of the preceeding

Transfer of
authority
controlled index
terms from GND

If none of the
preceeding

Transfer of index
terms from other
thesauri

Transfer to dedicated data fields for external content with provenance information

Same proceeding for transfer of DDC notations

Statistics work clusters (2/2022)

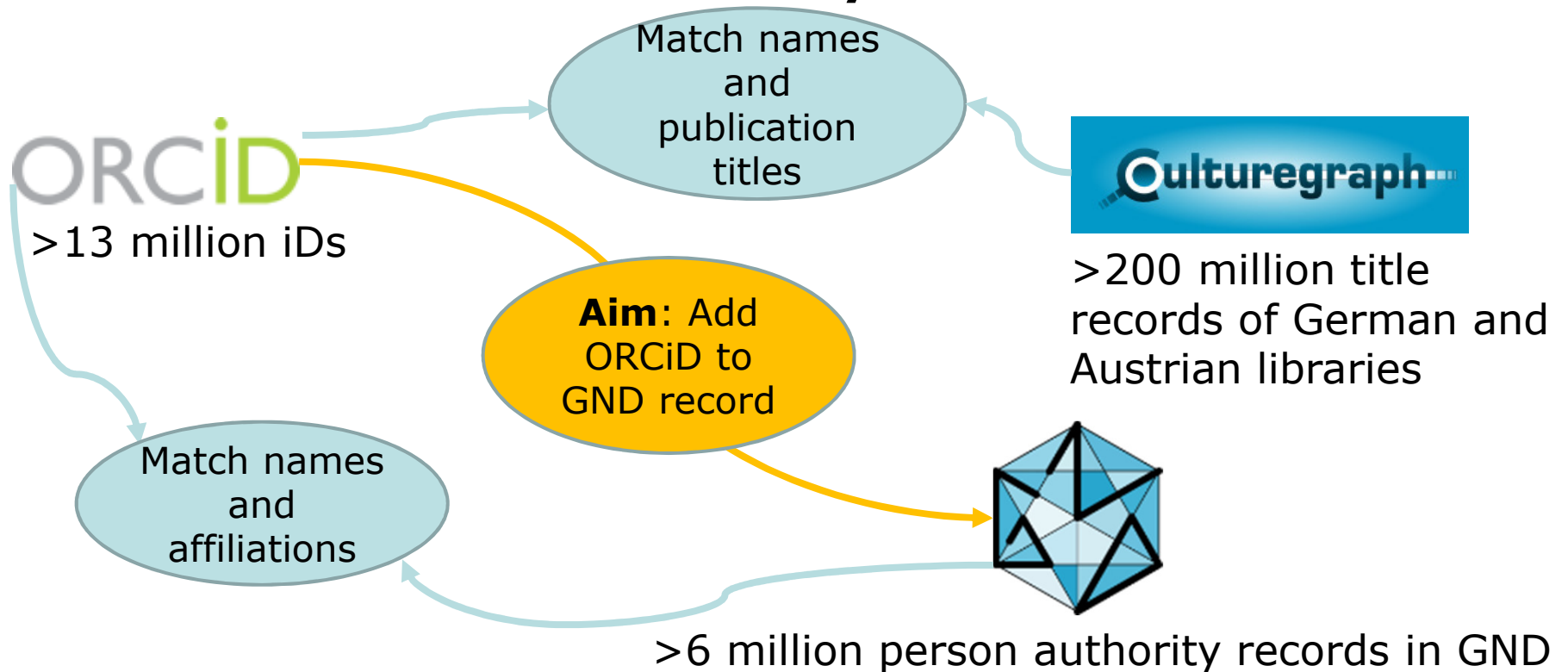
- 129,546,109 records processed (excluded: dependent parts, serials)
- 19,419,514 clusters with more than one member
- comprising 99,267,013 records
- Average cluster size (when more than one member): 4.7

Focus on persistent identifiers

- Establish and strengthen networks of persistent identifiers by indentifying links and enriching title and authority records
 - GND
 - ORCID
 - ISNI
- Automatic processes examine persistent IDs in title records and authority file records daily and create links
- User claimings serve as a basis for links and suggestion for new person authority records

Matching ORCiD-Culturegraph

Enrichment of title and authority records with ORCiDs



Automatic assignment of indexing information

- Automatic subject indexing with Annif.
- Automatic Indexing: GND descriptors
- Automatic Classification
 - DDC Subject Categories
 - DDC Short Numbers
- Tests of different associative approaches (TFIDF, fasttext, omikuji ...)
- Evaluation of assignment of GND descriptors: 68% (very) useful

Data science & Data analysis at German National Library (DNB)

1. Data analysis

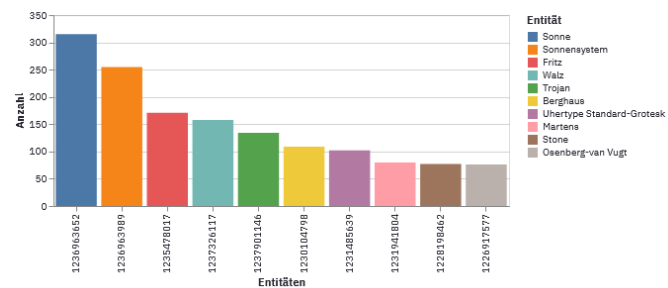
2. Research support

- Visualization of authority file
- Text+: Research data management
- Research support in Digital Humanities

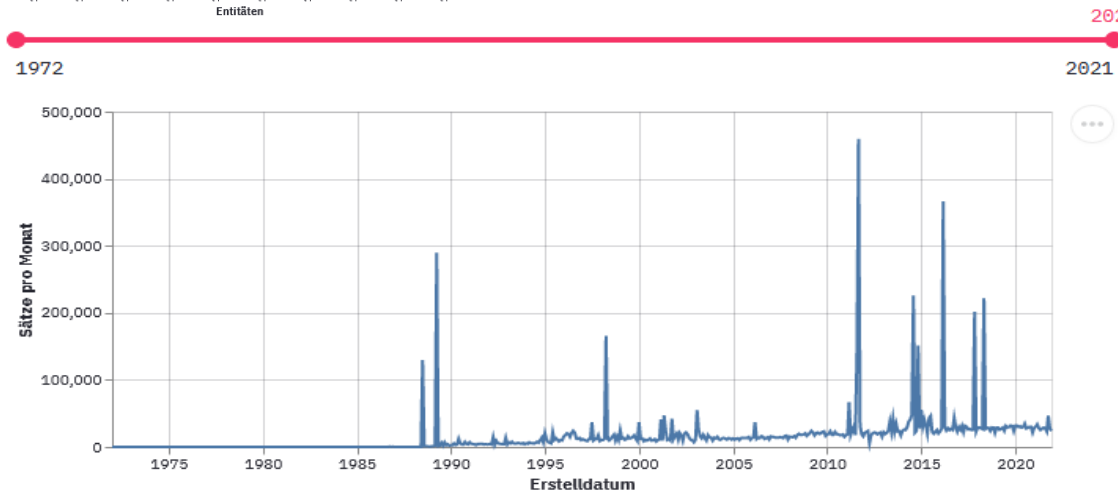
GND Dashboard

TOP 10 GND-Newcomer

TOP 10 der GND-Entitäten, die in den letzten 365 Tagen angelegt wurden.

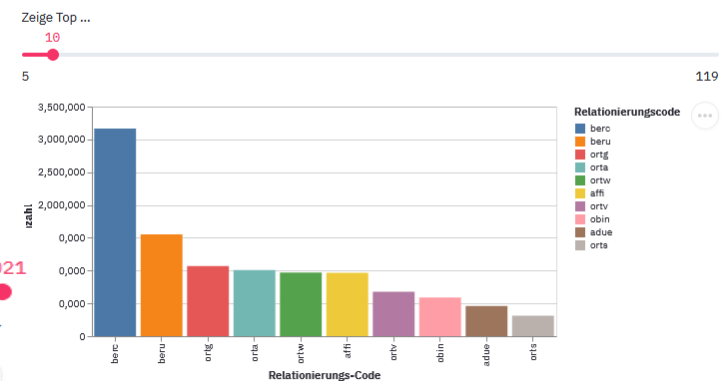


Interactive
Visualization of
statistics derived
from the Integrated
Authority File (GND)



Relationen

GND-Datensätze können mit anderen Datensätzen verlinkt (»relationiert«) werden. Die Art der Verlinkung wird über einen Relationierungscode beschrieben. Hier sind die am häufigsten verwendeten Relationierungs-codes zu sehen. Die Auflösung der wichtigsten Codes gibt es [hier](#).



Relationen zwischen Entitäten gesamt: 13.290.389

<https://share.streamlit.io/deutsche-nationalbibliothek/gnd-dashboard/main/dashboard/gnd-app.py>

Text+

- Project funded by German Research Foundation as part of National Research Data Infrastructure (NFDI)
- Preserve text- and language –based research data for long term and facilitate use in academia
- OA content (metadata, TOCs)

- ▼ DNBLab: Free collections of digital objects
- ▼ DNBLab: Interfaces
- ▼ DNBLab: Tutorials
- ▼ Events

Data sets

https://www.dnb.de/EN/Professionell/Services/WissenschaftundForschung/DNBLab/dnblab_node.html

DNBLab

- Theme-based data sets for download
 - Digitised tables of contents (>2 million)
 - Free online publications (>500,000)
 - Free online dissertations (>245,000)
 - Free digital sheet music (>2,600 items)
 - Articles from German-language Exile Journals 1933-45 (>230,000 articles)
 - And more
- Jupyter Notebooks as tutorials and examples
 - Access data via different interfaces
 - Title analysis in MARC21 metadata
- Online tutorials
- https://www.dnb.de/DE/Professionell/Services/WissenschaftundForschung/DNBLab/dnblab_node.html

Digital Humanities research support

- Collaborations with researchers in individual projects where DNB provides selected metadata content or access to full texts on the premises (e.g. Children and Youth literature, dime novels)
- DNB offers a yearly DH call. Applicants can obtain access to DNB data and infrastructures to work on Digital Humanities projects involving DNB content.
- DH Fellowship: investigate DNB's metadata and digital content for research purposes and receive a funding for a period up to 12 months
- participation in NFDI4Culture a consortium funded in the German National Research Data Infrastructure (NFDI) framework concentrating on cultural heritage

Thank you very much for your attention.



a.vorndran@dnb.de