Enabling digital research

- The German National Library as part of a (national) research infrastructure -

Dr. Peter Leinen
Head of IT
German National Library
German National Library

Frankfurt am Main

Leipzig
German National Library

- Mission
  - Collects, documents and archives all works in writing and sound published in and about Germany or in the German language since 1913

- Works
  - Books, periodicals, CDs, records, maps
  - Online publications, websites

- Basis
  - German National Library Act (DNBG)
  - Regulation on the compulsory delivery of books (PflAV)
Strategy

- Action areas
  - Expanding digital collections
  - Expanding digital indexing procedures
  - Presenting collections digitally
  - Connecting culture and science
  - Developing the learning organization
Expanding digital collections

- The digital collections have been expanded, especially in the areas of music, individual digital publications - such as e-books and e-journals - and Digitized Works
- ...

On the long run

- Collections are also available completely and permanently in digital form
Strategy

- Connecting culture and science
  - We moderate dialogues between the public, culture and science
  - We [...] are part of cooperation on strategic topics
  - We increase the visibility of our institution as a strong node in networks
Cornerstones

- 2006: Expansion of the DNB's collection mandate
- 2018: Copyright law reform in Germany
- 2021: Implementation of DSM Directive into national law
- 2016: RFII position paper: Performance through Diversity
- 2018: Decision of the GWK to establish the NFDI

**DSM:** Directive on Copyright in the Digital Single Market  
**RFII:** German Council for Scientific Information Infrastructures  
**GWK:** Joint Science Conference of the Federal Government and the Federal states  
**NFDI:** National Research Data Infrastructure
... some numbers

- Acquisition (per year)
  ~ 2,300,000 media works, of which
  ~ 1,500,000 digital objects

- Holdings
  ~ 45,000,000 media works, of which
  ~ 12,000,000 digital objects
... some more numbers

- Digital collection
  - e-books (~ 1,600,000)
  - sheet music (~ 12,000)
  - university publications (~ 330,000)
  - digital audio books (~ 75,000)
  - e-journals (~ 5,000,000)
  - e-paper (~ 3,500,000)
**TODOs**

- Backlist academic publishers
  - ≥ 6 million articles
- OA publications
  - ≥ 10 million publications
- Digital music
  - Legal and technical issues
- Web archive
  - Selection strategy
- Digitization of physical collections
  - systematically
  - topic-related
  - project-related
Cornerstones

- 2006: Expansion of the DNB's collection mandate
- 2018: Copyright law reform in Germany
- 2021: Implementation of DSM Directive into national law
- 2016: RFII position paper: Performance through Diversity
- 2018: Decision of the GWK to establish the NFDI

DSM: Directive on Copyright in the Digital Single Market
RFII: German Council for Scientific Information Infrastructures
GWK: Joint Science Conference of the Federal Government and the Federal states
NFDI: National Research Data Infrastructure
Text and data mining

- **Section 60d: Text and data mining for scientific research purposes**
  
  (1) It is permitted to make reproductions to carry out text and data mining for scientific research purposes in accordance with the following provisions.
  
  (2) ...
  
  (3) The following are, further, authorized to make reproductions:

  1. libraries and museums, insofar as they are accessible to the public
  2. individual researchers, insofar as they pursue non-commercial purposes.

- **Law regarding the German National Library**
  
  - Use of objects protected by copyright only on the premises of the National Library
Dime novels

- Pilot with University of Würzburg
- Objectives
  - Evaluate and develop methods of distance reading
- Aims DNB
  - Dare to start
  - Discover potential
  - Gather experience
Dime novels

- "Novels of the lower classes"

- Characteristics
  - Publication format (64 pages)
  - Marketing via the magazine market
  - Strong guidelines by publishers
  - Serialized novels

- Collection
  - 40,000 dime novels
  - 400 monographs of high literature
Dime novels

- Research questions
  - Genre detection
  - Complexity of the language
  - Characters - Gender
  - Action patterns
  - Forms of generating tension
  - Seriality - forms of the relationship between episode and series
  - Effects of the publishers' guidelines
  - ...
Pilot

- Lessons learned
  - Building up the specific infrastructure
  - Dealing with growing corpus
  - Native formats are not suitable for TDM
  - Gaps in the Collection
  - Listed as an example of data-intensive research in the humanities
  - Successful start, continue in a structured way
DH-Call

- **Objective**
  - Support of projects from research, science and culture
  - Up to 2 new projects/year with a focus on digital objects

- **Service**
  - Provision of metadata, digital objects and infrastructure
  - Digitization on a limited scale
  - Infrastructure
DH Projects

- Dime novels
- Otherness in Cookery Books
- Comparison of reporting on the Network Enforcement Act
- Fur market in Leipzig between the two world wars
- Secret bestsellers - the German dime novel
- Construction of a diversity corpus as a basis for algorithmic text analysis
- Foresight Extraction from Science Fiction Literature
- Directory on Children's and Youth Literature
DH Activities

- DNBLab
  - Access to free data (metadata, GND, digital objects, tables of content)
  - Interfaces to data
  - Tutorials
  - Webinars

- DH Fellowship
  - Duration 3, 6 or 12 months
  - Support by project mentors
  - Subject area 2022: “Catalogue data”
DH Projects

- Lessons learned
  - Individual requirements per project
  - Need for structured full texts
  - Scalable infrastructure
  - Management of growing corpora
  - DNB is part of the research process
Next steps

- Structured approach beyond DH projects
- Become part of a "text"-based community
  - Develop forms of legally compliant access
  - Provide services and know-how
- Build up know-how
  - Data preparation and appropriate provision
  - Basic algorithms of the DH community
Cornerstones

- 2006: Expansion of the DNB's collection mandate
- 2018: Copyright law reform in Germany
- 2021: Implementation of DSM Directive into national law
- 2016: RFII position paper: Performance through Diversity
- 2018: Decision of the GWK to establish the NFDI

RFII: German Council for Scientific Information Infrastructures
GWK: Joint Science Conference of the Federal Government and the Federal states
NFDI: National Research Data Infrastructure
National Research Data Infrastructure

- **Goal**
  - Development and long-term preservation of data holdings
    - Along the FAIR principles
    - Without disciplinary and national boundaries
    - Up to 30 consortia for broad coverage of scientific disciplines

- **Assembly**
  - In a science-driven process as a cross-linked structure of self-initiated consortia
National Research Data Infrastructure

- Interlinking NFDI-Consortia
  - MoU by consortia from the humanities and cultural studies\(^1\)
  - Leipzig-Berlin Statement on cross-cutting issues in infrastructure development\(^2\)
  - Networking in the NFDI Association

\(^1\) [https://zenodo.org/record/3265763#.Ym92pt9CSU1](https://zenodo.org/record/3265763#.Ym92pt9CSU1)

\(^2\) [https://zenodo.org/record/3895209#.Ym92Rt9CSUk](https://zenodo.org/record/3895209#.Ym92Rt9CSUk)
National Research Data Infrastructure

- Status
  - 2020: 9 consortia, including \textit{NFDI4Culture}
  - 2021: 10 consortia, including \textit{Text+}
  - 2022: application 15 consortia, including \textit{NFDI4Memory, NFDI4Objects}
    additionally Base@NFDI
Consortia from the humanities

- **Text+**
  - Research data infrastructure focused on language and text data

- **NFDI4Culture**
  - Research data on material and immaterial cultural heritage

- **NFDI4Memory**
  - *Methods that rely on data requiring historical contextualization*

- **NFDI4Objects**
  - Research data infrastructure for the material remains of human history
Consortia from the humanities

- Specialisation
- Common issues
  - Metadata
  - Authority data
  - Terminologies
  - Provenance
  - Rights and Ethnic
  - Data Literacy

NFDI

Targeted Communities:
- NFDI4Objects
- for memory
- Text+
- Communities
  - Humanities
  - NFDI
Text+

- Data domains (with specific clusters and responsibilities)
  - Collection,
  - Lexical resources
  - Editions
- Infrastructure/Operation
  - Cross-cutting issues
Text+

Communities (including, but not limited to):
- Classical Philology, Linguistics, Literary Studies,
  Social and Cultural Anthropology, Non-European Cultures,
  Jewish Studies and Religious Studies, Philosophy,
  language- and text-based research in the Social and Political Sciences
Role of a (national) library

- Role
  - Co-applicant, task area lead (collections),
  - Cluster unstructured texts with the domain collections

- Topics
  - Digital collection (born digital, digitized material)
  - Standards, Metadata standards
  - Authority File (GND), LOD
  - Legal and ethical aspects
  - Long term preservation
Cornerstones

- 2016: RFII position paper: Performance through Diversity
- 2018: Overview of research data in the DNB
- 2018: Copyright law reform in Germany
- 2018: First DH Project
- 2018: Decision to establish an NFDI
- 2019: Application phase for Text+, NFDI4Culture
- 2020: start of the annual DH Call
- 2020: Start of NFDI4Culture
- 2021: Implementation of DSM Directive into national law
- 2021: Start of Text+
Enabling digital research

... an long journey for memory organizations

- Digitization of the catalog

- Setup
  - Authority Data (e.g. GND), LOD,
  - Newspaper Database (ZDB)
  - ...

- Collect, document and archive digital/digitized works

- Make those data available to the communities in an suitable way
Enabling digital research

... an exciting journey for memory organizations
- a lot of questions concerning the collection
- technical, legal issues

- cultural issues
  - Enabling, creating free space
  - Taking risks,
  - Allowing pilots - not immediately the 120% solution