Text+ - a next generation platform for digital libraries and language resources
- Topics, structure, bodies and connections -

Peter Leinen, German National Library
National Research Data Infrastructure

» 2011: Kommission Zukunft Informationsinfrastruktur (KII)
   “Gesamtkonzept für die Informationsinfrastruktur in Deutschland”

» 2016: German Council for Scientific Information Infrastructures (RFII)
   “Performance through Diversity”

» 2018: Joint Science Conference
   Decision to establish a national research data infrastructure (NFDI)
National Research Data Infrastructure (NFDI)

» 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

The NFDI is being built up in a science-driven process as a cross-linked structure of self-initiated consortia.
Interlinking NFDI-Consortia

 » Memorandum of Understanding by NFDI initiatives from the humanities and cultural studies\(^1\)

 » Leipzig-Berlin Statement on NFDI cross-cutting issues in infrastructure development\(^2\)

 » Networking in the NFDI Association

\(^1\) https://zenodo.org/record/3265763#.Ym92pt9CSUl

\(^2\) https://zenodo.org/record/3895209#.Ym92Rt9CSUk
Aim:

- build a research data infrastructure focused on language and text data

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
Text+ Community

» Needs and interests of the community

» Support in the creation of standards-compliant research data

» Long-term provision of research data (after project completion)

» Guidance on the use of relevant data and tools

» Support in questions regarding legal and ethical aspects of data use

» Consulting in the planning and implementation of research projects
Bodies of Text+

» Applicant institution
  » Leibniz Institute for the German Language

» Co-applicant institutions
  » Berlin-Brandenburg Academy of Sciences and Humanities
  » German National Library
  » Göttingen State and University Library
  » North Rhine-Westphalian Academy of Sciences, Humanities and the Arts

» Participants
  » 7 academies, 14 universities, 3 libraries and archives, 5 foundations and centers

Strong connections to
» Textgrid
» CLARIN-D
» DARIAH-DE
» CLARIAH-DE
Community Involvement

» 20 professional associations, 9 Specialized Information Services
  » Part of the Governance: Scientific Board, Scientific Coordination Committees
  » Advice
    » Prioritization in the curation of resources and in the development of software tools.
    » Further development of standards and best practices for research data

» User Stories (120)

» Workshops

www.text-plus.org
Text+

» Data domains (with specific clusters and responsibilities)
  » Collections
  » Lexical Resources
  » Editions

» Infrastructure/Operations
  » Cross-cutting issues
Data domains

» Measures
  » Reference Implementation
  » Portfolio Development
  » Standardization Activities
  » Community Activities
  » Software Services
Collections

» Clusters:
  » Contemporary Language
  » Historical Texts
  » Unstructured Texts

» 19 participating institutions
Collections

» Text collections
  » literary texts, non-fiction texts, newspapers, magazine texts, interviews, inscriptions, manuscripts, prints, ...

» mono- and multimodal recordings
  » speeches, dialogues, news, interviews, interaction in everyday life, ...

» sensor data
  » brain activities (EEG), eye tracking, articulography, ...

» interviews, reaction times
Collections

» first steps:

» data ingest (distributed infrastructure)

» standard-conformant preparation of text and metadata

» integration of new software services, e.g. for the connection of collections with authority files

» workshops on data curation, legal issues, etc.

» ...

www.text-plus.org
Collections (Added value)

» Text+ will

» act as a comprehensive support infrastructure for all issues regarding collections

» provide easy and transparent processes to find and obtain data from collections freely available for research

» integrate collections currently beyond the reach of researchers in a community-driven way and at a large scale

» be a hub for legal and ethical expertise related to collections

» increase standardization and interoperability for collections by Linked Open Data, controlled vocabularies and authority files
Lexical Resources

» Clusters:

» German Dictionaries in a European Context

» Born-Digital Lexical Resources

» Non-Latin Scripts
Lexical Resources

» first steps:

» common search interface to lexical resources...

» integration of all resources in the LLOD cloud; semantically linked to common data hubs

» user-oriented guidelines for data ingest

» ...
Editions

» Clusters:

» Ancient and Medieval Texts

» Early Modern, Modern, and Contemporary Texts
Editions

» Examples

» documentary or diplomatic editions
» editions on the history of documents
» historical-critical editions
Editions

» first steps:

» convey digital editorial models, offer consulting services

» foster re-use, best practices and transdisciplinary strategies

» integration of resources following the onboarding policy

» ...

www.text-plus.org
Infrastructure/Operations

» Cross-cutting
  » Interface to the data domains
  » Provider of crucial infrastructure and services for collaborative work
  » Aspires for interoperability and compatibility of services, data, standards, and roadmaps in terms of infrastructure and operation

» Constant exchange with the NFDI at large, the MoU group, and the NFDI sections to aspire interoperability, compatibility, and synergies
Infrastructure/Operations

M2 ACCESSIBILITY
- Basic Identification Services
- Long-term Archiving
- Quality Assurance of Services and Data

M1 FINDABILITY
- Registries
- Search and Retrieval

M3 INTEROPERABILITY AND RE-USABILITY
- Metadata, Authority Data, Terminologies
- Metadata Infrastructure and LOD

M4 COMMUNITY ACTIVITIES
- Web Portal
- Helpdesk and RDM Planning

M5: SOFTWARE SERVICES
- Data Processing Pipeline
- Software Development Support

www.text-plus.org
MoU by initiatives from the humanities and cultural studies

» Common topics
  » Metadata
  » Authority data
  » Terminologies
  » Provenance
  » Rights and Ethic
  » Data Literacy
Cross-cutting topics within the NFDI

» Offers from text+
  » Standards for text and language-based data
  » Authority data
  » Expertise in curating and analyzing text- and language-based resources
  » Solutions for search and interoperability
  » Know how in longtime preservation
  » Technical infrastructure services (PID, AAI)
National and international networks

» Cooperation with alliances and standardization bodies like

  » ALLEA - CLARIN ERIC - DARIAH ERIC - DataCite -
    DDB - DINI - EOSC - ESFRI - Europeana -
    HathiTrust Research Center - LIBER -
    nestor - OpenAIRE - OPERAS - RDA - ...

» guarantee the exchange of information on overarching topics and developments
Role of a (national) library

» Role
   » Co-applicant, task area lead (collections),
   » Cluster unstructured texts

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