

Template

for Submission of Letters of Intent

The successful implementation of a German national research data infrastructure is dependent on collaboration among all stakeholders, particularly among consortia. In order to support cooperation and networking among prospective consortia, the DFG requests letters of intent to submit a proposal from all potential applicants. The letters of intent will subsequently be published online at

www.dfg.de/nfdi

In addition to facilitating cross-consortia networking, the letters of intent will be used to plan the review.

Please note the following

- If you are planning to apply in September 2020, you **must** submit a **binding letter of intent** that will serve both as advance notification of proposal submission in 2020 and as **binding information on the (co-)applicants and participants** of your consortium. Proposals that have not been preceded by a letter of intent according to this template will not be considered in the second round.
- If you are planning to apply in the third round of NFDI consortia calls, which is scheduled for autumn 2021, you are encouraged to submit a **non-binding letter of intent** now.
- For 2020, consortia focusing primarily on cross-cutting topics will not be considered. The binding and non-binding letters of intent will instead be used to work towards a clearer

understanding of user and community needs and potential contributions of consortia in the area of cross-cutting topics (see number 4 below).

- Full proposals for the second round must be submitted via the elan portal. Note, however, that the letter of intent must be sent by e-mail and cannot be submitted via the elan portal.
- To enable networking, the contact details of all (co-)applicant institutions and all (co-)spokespersons will be published online.
- Binding and non-binding letters of intent must be written in English.
- Your document must be structured according to the template provided below (this rtf file, Arial 11).
- Please send your binding letter of intent (for 2020) or your non-binding letter of intent (for 2021) to the following e-mail address: nfdi@dfg.de.
- **DEADLINE: 15 August 2020**

Please note that the letters of intent you submit, including personal data submitted in accordance with the applicable procedural rules and in line with data protection information on DFG research funding, will be published on the DFG website for informational and networking purposes and to facilitate the preparation of research proposals.

Please address the following aspects in your letter of intent

1 Binding letter of intent as advance notification or non-binding letter of intent

<input checked="" type="checkbox"/>	Binding letter of intent (required as advance notification for proposals in 2020)
<input type="checkbox"/>	Non-binding letter of intent (anticipated submission in 2021)

2 Formal details

- Planned name of the consortium

Text+: Language- and Text-Based Research Data Infrastructure

- Acronym of the planned consortium

Text+

- Applicant institution

Leibniz Institute for the German Language, Mannheim
R5, 6-13, 68161 Mannheim
Prof. Dr. Henning Lobin

- Spokesperson

Prof. Dr. Erhard Hinrichs, hinrichs@ids-mannheim.de, Leibniz Institute for the German Language, Mannheim

- Co-applicant institution

Berlin-Brandenburg Academy of Sciences and Humanities, Jägerstraße 22/23,
10117 Berlin
Prof. Dr. Martin Grötschel

- Co-spokesperson

PD Dr. Alexander Geyken, geyken@bbaw.de, Berlin Brandenburg Academy of

Sciences and Humanities

▪ Co-applicant institution

German National Library, Deutscher Platz 1, 04103 Leipzig and Adickesallee
1, 60322 Frankfurt am Main
Frank Scholze

▪ Co-spokesperson

Dr. Peter Leinen, P.Leinen@dnb.de, German National Library

▪ Co-applicant institution

North Rhine-Westphalian Academy of Sciences, Humanities and the Arts, Karl-
Arnold-Haus der Wissenschaften, Palmenstraße 16, 40217 Düsseldorf
Prof. Dr. Wolfgang Löwer

▪ Co-spokesperson

Prof. Dr. Andreas Speer, andreas.speer@uni-koeln.de, North Rhine-
Westphalian Academy of Sciences, Humanities and the Arts

▪ Co-applicant institution

Göttingen State and University Library, Platz der Göttinger Sieben 1, 37073
Göttingen
Prof. Dr. Wolfram Horstmann

▪ Co-spokesperson

Regine Stein, regine.stein@sub.uni-goettingen.de, Göttingen State and
University Library

▪ Participant

Institutional Participants

Name	Institution, location
Claudia Raap	Academy of Sciences and Humanities in Hamburg
Anna Neovesky Prof. Torsten Schrade	Academy of Sciences and Literature, Mainz
Prof. Dr. Thomas O. Höllmann Eckhart Arnold	Bavarian Academy of Sciences and Humanities
Prof. Dr. Tonio Sebastian Richter	Berlin-Brandenburg Academy of Sciences and Humanities/Freie Universität Berlin
Prof. Dr. Thomas Gloning	Berlin-Brandenburg Academy of Sciences and Humanities/Giessen University
Prof. Dr. Sandra Richter	German Literature Archive Marbach
Prof. Dr. Dr. h. c. mult. Jörg Hacker	German National Academy of Sciences Leopoldina, Halle (Saale)
Prof. Dr. Ramin Yahyapour Prof. Dr. Philipp Wieder	Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen
Prof. Dr. Andreas Gardt PD Dr. Volker Harm	Göttingen Academy of Sciences and Humanities
Prof. Dr. Thomas Holstein Dr. Dieta Svoboda-Baas	Heidelberg Academy of Sciences and Humanities
Prof. Dr. Peter Burschel Dr. Johannes Mangei Dr. Hartmut Beyer	Herzog August Library Wolfenbüttel
Prof. Dr. Dr. Thomas Lippert Daniel Mallmann	Jülich Supercomputing Centre
Dr. Ulrike Lorenz Dr. Gerrit Brüning	Klassik Stiftung Weimar
Prof. Dr. Andreas Witt	Leibniz Institute for the German Language/University of Mannheim/University of Cologne
Prof. Dr. Jonathan Harrington PD Dr. Christoph Draxler	Ludwig-Maximilians-University Munich
Dr. Harald Rosenbach Dr. Michael Kaiser	Max Weber Foundation, Bonn
Prof. Dr. Elke Teich	Saarland University
Dr. Christian Winter Prof. Dr. Gerhard Heyer	Saxon Academy of Sciences and Humanities, Leipzig
Prof. Dr. Andrea Rapp	Technical University of Darmstadt/AdWL Mainz
Prof. Dr. Thomas Stäcker	Technical University of Darmstadt, University and State Library Darmstadt
Prof. Dr. Wolfgang Nagel Dr. Ralph Müller-Pfefferkorn	Technical University Dresden, Centre for Information Services and High Performance Computing
Prof. Dr. Michael Brocke Harald Lordick	The Salomon Ludwig Steinheim Institute of German-Jewish Studies, Essen
Prof. Dr. Stefan Schmunk	University of Applied Sciences Darmstadt
Prof. Dr. Andreas Henrich	University of Bamberg
Prof. Dr. Nikolaus Himmelmann Prof. Dr. Karl Ubl	University of Cologne
Prof. Dr. Andreas Blätte	University of Duisburg-Essen
Prof. Dr. Christian Mair	University of Freiburg
Prof. Dr. Kristin Bührig Prof. Dr. Michael Friedrich	University of Hamburg

Prof. Dr. Joachim Veit	University of Paderborn
Prof. Dr. Claudine Moulin Prof. Dr. Christof Schöch	University of Trier
Prof. Dr. Peter Grathwohl	University of Tübingen
Prof. Dr. Fotis Jannidis	University of Würzburg

Participating individuals

Name	Institution, location	Contribution of participant
Prof. Dr. Andreas Henrich	University of Bamberg	Operations Vice Speaker
Prof. Dr. Verena Klemm	University of Leipzig	SCC Chair Editions
Prof. Vivien Petras, PhD	Humboldt University of Berlin	OCC Chair Operations
Prof. Dr. Andrea Rapp	Technical University of Darmstadt	Scientific Vice Speaker
Prof. Dr. Sandra Richter	German Literature Archive Marbach	SCC Chair Collections
Prof. Dr. Ingrid Schröder	University of Hamburg	SCC Chair Lexical Resources
Prof. Dr. Elke Teich	Saarland University	Scientific Vice Speaker
Dr. Thorsten Trippel	University of Tübingen	Scientific Office Coordinator

3 Objectives, work programme and research environment

- Research area of the proposed consortium (according to the DFG classification system):
 - 101-02 Classical Philology
 - 104 Linguistics
 - 105 Literary Studies
 - 106 Social and Cultural Anthropology, Non-European Cultures, Jewish Studies and Religious Studies (partly)
 - 108 Philosophy
- Concise summary of the planned consortium's main objectives and task areas

The main objective of Text+ is to develop a research data infrastructure for Humanities disciplines and beyond whose primary research focus is on language and text. The Text+ research data infrastructure will be open, flexible and scalable so as to be able to adapt to emerging areas of research and to new research paradigms. Text+ is committed to foster and support innovative research for its communities of interest and to respond to diverse community needs across different academic disciplines.

Language and text are the basis of human communication and, as such, underlie all aspects of society: ethics, economy, education and research, and even the internet. In addition, languages as well as numerous texts respectively documents are internationally protected cultural heritage, whose preservation and accompanying research is given a new dimension in the digital age. "Text as data" is a fairly new phenomenon but it is instrumental to address it specifically in a national research data infrastructure (hereafter NFDI). Text+ will tackle this challenge primarily from a Humanities research perspective and will engage with other NFDI consortia that require text- and language-based resources and services as part of their research portfolios.

The primary research domains addressed by Text+ are those Humanities disciplines that create, access, collect, annotate, analyse, visualise, publish, share, and archive text- and language-based research data in machine-readable form. Text+ will offer community-built solutions to manage and reuse research data for German and for a wide range of different languages, including minority and endangered languages, multilingual resources, resources for different text genres, including scholarly publications, literary texts, letters, newspapers, inscriptions, public records, and non-fictional texts of various kinds, as well as resources for different modalities, including written texts, speech, sign language, audio and video data, and derived text sources such as dictionaries and grammars.

The materials span several thousand years of cultural heritage, often collected over many years in academy-funded, long-term projects and with an ever-increasing amount of research data resulting from thousands of ongoing research projects in the Humanities. More text- and language-based research data are made available in very large quantities by research libraries and archives. Substantial parts of these materials are available in digital form. At the same time, the majority of the sources are not yet machine-readable or not yet digitised at all. The availability gap affects practically all areas of language and text data, but especially the small and endangered languages and literatures. Thus, high-quality digitisation, the curation of such data and standardisation of formats for digitised data remain a big challenge for any NFDI initiative in the Humanities.

Initially, the Text+ consortium will focus on three types of research data, the Text+ data domains: language- and text-based Collections, Lexical Resources and Editions. These three types of data have a long tradition in the Humanities and have given rise to mature methodological

paradigms that require distinctive, yet cross-cutting practices of research data creation, curation and management. In addition, they also provide occasions for subject-specific and interdisciplinary methodological discussions and innovations. Thus, when we speak of “the three data domains”, we do not refer to data resources or formats alone, but we refer to a common encoding of knowledge and information with a common set of research methodologies that cut across different Humanities disciplines and have significant relevance for a broad range of academic disciplines outside the Humanities. In essence, by shaping Text+ around established data domains rather than by reference to individual academic disciplines or to specific research methodologies, Text+ seeks to connect the diversity of research communities and facilitates answers to both disciplinary and interdisciplinary research questions.

The development of the research infrastructure will be driven by the research needs and research priorities of its communities of interest. Continuous development of the data and service portfolios of Text+ is determined by three Scientific Coordination Committees (SCCs), for Collections, Dictionaries and Editions, respectively. The SCCs will develop and apply research-driven data and service selection criteria that will jointly ensure strict adherence to the FAIR principles.

The SCCs will collaborate closely with the Operations Coordination Committee (OCC) which manages the development of the underlying technical infrastructure and its associated services. Membership of all committees is determined by a large number of academic societies (Fachverbände) which have pledged their support and their active involvement in the governance of Text+. Four distinguished domain experts have kindly agreed to act as inaugural chairs of the SCCs and of the OCC. They are involved in the Text+ consortium as Individual Participants (see Section 2 above).

The work programme of Text+ is organized in a total of five task areas: one each for the three data domains and additionally the two task areas infrastructure/operations and administration. The Leibniz Institute for the German Language, internationally recognized for language- and text-based research and related research data provisioning, acts as applicant institution of Text+ and is responsible for the task area Administration. The work programme for the other four task areas, which are headed by the four co-applicant institutions of Text+, comprises tightly coupled measures for data services, community services, and community-driven software services. For the task area Collections, the German National Library (DNB) will serve as co-applicant institution. The co-applicant institution Berlin-Brandenburg Academy of Sciences and Humanities will coordinate the task area Lexical Resources. The task area Editions will be the responsibility of the co-applicant institution North Rhine-Westphalian Academy of Sciences, Humanities and the Arts. The co-applicant institution Göttingen State and University Library (SUB) will coordinate the task area Infrastructure/Operations.

- Brief description of the proposed use of existing infrastructures, tools and services that are essential in order to fulfil the planned consortium's objectives

Due to the distributed nature of knowledge, resources, software tools, and services that Text+ will provide for the NFDI, the consortium will put in place a geographically distributed network of data and competence centres. These will provide and further develop data and service offers based on a common framework of quality criteria while each of them will specialize in complementary thematic and methodological clusters. This network involves experts and scientific institutions from the various disciplines, a growing number of CTS-certified data centres as well as several libraries' and computing centres' infrastructures, all of them long-standing partners in Humanities' research. A common technical backbone for the federation of metadata, data and services will facilitate the seamless integration of research data and services from different data and service providers and consists of a layered set of infrastructure protocols and services. Major components of this backbone have been under active development in the research infrastructure project CLARIAH-DE and in the European Research Infrastructure Consortia CLARIN ERIC and DARIAH ERIC and will be re-used.

A set of basic infrastructure services, which include persistent identifier services, uptime monitoring services for data repositories and software tools, authentication and authorization services (AAI), will provide a solid foundation. A set of extended services will include long-term archiving of research data and support of different stages of the research data lifecycle, comprising workflows for accessing and generating digital text, platforms for the annotation, analysis, visualization, and publication of research data. A federated metadata and data infrastructure is provided together with a fine-tuned set of hosting, migration, integration and synchronization services. This will include conversion from and to RDF and linking with arising knowledge graphs within the NFDI as well as other resources in the Linked Open Data cloud. Special emphasis is given to metadata interoperability and the linkage with widely used authorities such as the Gemeinsame Normdatei hosted by Text+ co-applicant DNB. Text+ will also support the development of cutting-edge digital research methods and technologies. All partners have long experience in the development of study programmes, curricula, summer schools, workshops and consider such activities as central in developing an innovative research infrastructure.

International collaboration of Text+ includes active participation in the Research Data Alliance, in Standardization organizations such as relevant technical sub-committees of ISO, TEI and the W3C consortium, in order to be able to contribute to and promote implementation of internationally accepted standards for language- and text-based research data and research data management. At the European level, Text+ will continue to coordinate its activities with CLARIN ERIC and

DARIAH ERIC. Members of the Text+ consortium will also continue to actively contribute to the construction of a European Open Science Cloud for the Humanities and Social Sciences via HORIZON2020 projects, such as EOSC-Hub and SSHOC. Furthermore, there is close cooperation with ALLEA, the European Federation of Academies of Sciences and Humanities.

- Interfaces to other proposed NFDI consortia: brief description of existing agreements for collaboration and/or plans for future collaboration

In the area of the Humanities, extensive consultations between infrastructure providers and researchers from the communities of interest have taken place in preparation of the NFDI. During a workshop series, in the context of the subsequently founded Geisteswissenschaftliches Forum NFDI, several more open meetings, and an open call for contributions we identified the actual needs of Humanities scholars across a wide range of disciplines. In addition, interfaces to three other NFDI consortia from the humanities and cultural studies (NFDI4Culture, NFDI4Memory, and NFDI4Objects) have been established in the course of several coordination meetings. These interfaces and the planned collaboration are described in a Memorandum of Understanding (MoU), which is open for integrating further initiatives. This MoU has been published electronically at <https://doi.org/10.5281/zenodo.3265763>. In addition to the MoU, the following will be jointly addressed as cross-cutting topics: data literacy, authority data, provenance, rights and ethics.

Text+ has signed the Berlin Declaration (DOI: 10.5281/zenodo.3457212) and the Leipzig-Berlin Declaration (DOI: 10.5281/zenodo.3895208) on NFDI Cross-Cutting Topics and is prepared to enter into close NFDI-wide consultation and coordination on the Research Data Commons.

Textual resources are used in a variety of scientific disciplines. For example, scholarly communication in virtually every scientific discipline is language- and text-based. Therefore, interfaces to cross-disciplinary activities like standardization and interoperability of metadata, terminologies and authority data, and services such as text data mining are planned. A cooperation on quantitative and qualitative language- and text-based research methodologies will be established with KonsortSWD, where Text+ focuses on metadata and text resources while KonsortSWD focuses on anonymization procedures. NFDI4Ing and Text+ will collaborate in scoping out and leveraging the potential of text mining techniques for the extraction of research data from pertinent digital scholarly literature and documents. Text+ has also identified common interfaces with other NFDI consortia, e.g. through the workshop on metadata interoperability run by NFDI4Ing in July 2020, as diverse as NFDI4Earth or NFDI4BioDiv.

In addition, we are in contact with other NFDI initiatives who intend to submit proposals in the submission round of the year 2021.

4 Cross-cutting topics

- Please identify cross-cutting topics that are relevant for your consortium and that need to be designed and developed by several or all NFDI consortia.

Cross-cutting topics relevant for Text+ and other NFDI consortia concern, beside the above-mentioned joint topics of the humanities' and cultural studies' consortia: data, service and software quality frameworks, search and interoperability solutions, standardization and certification, and technical infrastructure services, particularly for ensuring compliance with FAIR principles. Technical infrastructure services include authentication and authorisation services, persistent identifiers for research data, certification services for data centres, as well as easy access to data storage and archiving facilities and high performance computing.

- Please indicate which of these cross-cutting topics your consortium could contribute to and how.

Text+ will contribute to cross-cutting topics by sharing expertise and practical experience with other NFDI consortia in the following areas:

Data literacy: Text+ will organize training events and summer schools for junior researchers at the national and international level, provide and contribute to learning and teaching support platforms, as well as contribute to curriculum development.

Authority data: Text+ will actively support the usage and enrichment of cross-disciplinary terminology and authority data, especially the Gemeinsame Normdatei GND.

Provenance: Text+ will promote and help develop solutions for integrating provenance information into data models and curation practices, including its manifold aspects from technical to legal.

Rights and ethics: Text+ will provide consulting on licensing, contributions to data governance solutions and raising awareness for data bias issues.

Data, service and software quality frameworks: Text+ will actively contribute to the development of curation criteria and quality standards for research data in the Humanities and related quality management processes, quality management of software and services, lifecycle management.

Search and interoperability solutions: Text+ will offer a federated metadata and data infrastructure. Relevant services include hosting, migration, integration and synchronization as well as conversion from and to RDF and linking with resources in the Linked Open Data cloud.

Standardization and certification: Text+ will actively promote common reference models, help develop interoperable solutions for metadata application profiles, and support the CTS certification of data centres.

Technical infrastructure services: Text+ will offer authentication and authorisation services, persistent identifiers for research data, digital long-term preservation of research data, development and deployment of Text Data Mining techniques for language- and text-based research data.