



Prof. Dr. Erhard Hinrichs, Spokesperson
Text+
Leibniz-Institut für Deutsche Sprache
R-5, 6-13
68161 Mannheim

REINHARD ALTENHÖNER
STÄNDIGER VERTRETER DER
GENERALDIREKTORIN

Potsdamer Straße 33
10785 Berlin

Telefon: +49 30 266-43 14 00
Telefax: +49 30 266-33 13 01

reinhard.altenhoener@sbb.spk-berlin.de

www.staatsbibliothek-berlin.de

Bearbeiter: Reinhard Altenhöner
GD/SV

Berlin, 11. September 2020

Letter of Support

Contribution(s) of SBB to NFDI Text+

Dear Mr. Hinrichs,

Thank you for giving me the opportunity to explain our willingness to actively support Text+.

The State Library (SBB) ranks among a number of collections of worldwide importance with regard to the quality and significance of its holdings for research. The digitization and preparation of these collections for (re)use in science and research according to the FAIR principles is an important element of SBB's strategic orientation. At the same time, the SBB ensures the supply of information and enriched / aggregated content in a national perspective with an international focus to more regionally and technically delimited scientific communities and carries out development projects in this context.

So SBB is involved in a plenty of projects, development assets and services that serve the development and expansion of appropriate infrastructures that can be used in cooperation.

As part of these activities and its existing expertise, SBB will tax specific expertise in the following areas of work and contribute its increases and services to the provision and use of research data in Text+:

CrossAsia

CrossAsia, as the research infrastructure for Asia related studies in conjunction with the Specialised Information Service Asia (FID Asien), a cooperative Project of Staatsbibliothek zu Berlin – PK together with Heidelberg University Library, strongly supports the current project proposal for an NFDI consortium Text+ and is willing to actively participate in this initiative.

CrossAsia as a research infrastructure provider for Asia related research in the fields of humanities and social sciences has long experience in hosting and making accessible huge amounts of copyright restricted text and image data as well as those of the public domain. One of the main goals is – in line with the Text+ initiative – to make those data equally accessible for digital humanities' research in accordance with the FAIR principles. By participating in the Text+ initiative, CrossAsia and the FID Asia would like to bring in its huge amounts of digital full-text research data as well as enquire in the field of authentication and authorisation new, alternative ways for providing access to copyright restricted data.



CrossAsia and the FID Asia are committed to support the NFDI consortium Text+ by acting as intermediary between the consortium, infrastructure and researchers in Asian studies and their scientific communities. We are willing to participate in Text+ activities such as workshops and meetings to communicate our own requirements, as well as those of the academic community. At the same time, we will communicate the idea and goal of the Text+ initiative as well as its current developments and new services to the Asian studies community.

OCR-D

Scholars increasingly require full-text of digitized collections for purposes of text mining and data analysis, but the quality of the OCR introduces errors and biases the impact of which are difficult to assess with regard to the research question. This is where the work of OCR-D sets in. OCR-D provide a modular, scalable and open source set of software tools specifically tailored to OCR challenges for historical printed material. The ongoing project platform establish standards, workflows and best-practices in the area of OCR and Ground Truth for historical printed material and aims to build an interdisciplinary community of practitioners from the GLAM sector, Digital Humanities and Computer Science to further increase knowledge exchange and build capacity for OCR for historical printed material.

In the next funding phase, SBB will cooperate with steering members of the Text+ consortium in Göttingen in order to facilitate and speed up the development. Therefore, we see strong and concrete opportunities to address research needs in the field of OCR and NER in order to facilitate the specialized provision of machine-processed text corpora.

SBB-LAB

In addition, we like to share and offer experiences from a broader variety of data services grouped in the so called SBB-Lab - approach (<https://lab.sbb.berlin/>), which we see as an intermediate step for larger cooperative structures in the field of tool and method development for data analysis in the humanities.

Sincerely Yours