

“... ich würde keinen Teufel schonen, möcht’ er laborieren oder kollaborieren” – Jean Paul's Letters as Data for Various Research Domains in the Context of the National Research Data Infrastructure Text+

Neuber, Frederike

frederike.neuber@bbaw.de
Berlin-Brandenburg Academy of Sciences and Humanities,
Germany; NFDI / Text+

Hug, Marius

marius.hug@bbaw.de
Berlin-Brandenburg Academy of Sciences and Humanities,
Germany; NFDI / Text+

Wiegand, Frank

wiegand@bbaw.de
Berlin-Brandenburg Academy of Sciences and Humanities,
Germany; NFDI / Text+

Digital editions, text collections and lexical resources each have their own tradition in the (digital) humanities and at the same time, the data domains share common methods and practices of indexing, modeling, and analysis. The Text+ consortium in the German National Research Data Infrastructure (NFDI) ¹ is dedicated to the handling of research data in the three aforementioned research domains and is exploring methods and practices of cross-domain use of datasets and the resulting collaboration between research fields. The poster shows the connections between the three data domains in Text+ using a concrete example: the edition of letters by the German writer Jean Paul (Jean Paul Friedrich Richter, 1763–1825) of the Berlin-Brandenburg Academy of Sciences and Humanities (BBAW).

The edition of Jean Paul's letters was first published in print in nine volumes from 1952 to 1964 (Berend) as part of the Historical-Critical Edition (HKA). These volumes were retro-digitized by the German Text Archive (DTA) in 2017 and in 2018 'retro-converted' as a digital edition by TELOTA (The Electronic Life Of The Academy) which instead of the 'book' puts the 'single letter' in the focus of the edition. The digital edition comprises 5526 letters in full text accessible on the platform "Jean Paul – Sämtliche Briefe digital" (Bernauer et al. 2018–2022). The digital edition offers many advantages over the digitized books and further exploits the potential of the digital medium. For instance, the indexes were linked dynamically to the letter texts as well as

combinable filters and a full-text search are offered. The letters are encoded in XML/TEI, following the guidelines of the DTA Base Format (BBAW 2011–2022). During the development of the edition, standardized metadata (including GND) was recorded for each letter. In addition, the digital edition offers various BEACON APIs for networking with other resources, as well as a CMIF API. The data sets of the letters were published on Github and Zenodo under a CC BY-SA license (Neuber 2022).

In the context of the consortium Text+ in the National Research Data Infrastructure (NFDI, <https://www.text-plus.org/>), the datasets were integrated into the infrastructure of the BBAW's "Zentrum für Sprache" in 2022 (Neuber et al. 2022). For this purpose, the Text+ cluster "Historical Texts" first selected the full-text and dated documents from the digital edition and made them available as a DTA text collection – 5004 documents (54,177 sentences, 985,703 tokens) from the period 1780–1825 – for subsequent use.

The corpus "Jean Paul Letters" (DWDS 2022), in turn, represents an important basis for dictionary work as a lexical resource. Such corpora contain frequencies, variety and change of meaning, different spellings or the (historical or regional) word environment important for collocation information. They thus function as data providers for real voucher examples and are helpful to prove the use of a word. The "Wortgeschichte digital" project of the Centre for Digital Lexicography of the German Language (ZDL 2022) is a corpus-based dictionary based on such digital editions and collections. Another scenario of re-use is the integration of the resource into the infrastructure of the Digital Dictionary of the German Language (DWDS). With its established services for the (computer-)linguistic processing of historical text data, the DWDS enables advanced, spelling-tolerant (also cross-corpus) searches as well as further possibilities for corpus analysis, such as collocation analyses over time with the tool DiaCollo.

Using the letters of Jean Paul as an example, the poster illustrates the life cycle of research data from a digital edition at the BBAW. By making the data available together with enriched metadata in accordance with the FAIR principles, the prerequisites for a variety of reuse scenarios are fulfilled. Such implementation of 'Open Data' is still the exception rather than the rule in the edition context. Greta Franzini's Editions Catalog (2016–2022) shows that of 320 digital editions, only ~27% use CC licenses, ~23% make TEI data available for download, and ~5% offer APIs. In addition, the poster also represents the division of labor – for instance in the realm of Text+ and its data domains editions, collections, and lexical resources – that goes hand in hand with a return to one's own core competencies.

In 1795, Jean Paul wrote in a letter that he "... would not spare any devil, would he like to labor on something or collaborate" ² (see title). We share this enthusiasm for collaboration, and hope that our poster can spark a discussion on how collaboration and resource sharing can be applied in the DH between various research data domains.

Notes

1. <https://www.text-plus.org/>.
2. https://www.jeanpaul-edition.de/brief.html?num=II_79

Bibliography

BBAW (2011–2022): *DTA Base Format*, <https://www.deutschestextarchiv.de/doku/basisformat/> [20.04.2023].

Berend, Eduard (ed. 1952–1964): *Jean Pauls Sämtliche Werke. Briefe an Jean Paul (III. Abt.)*. BBAW.

Bernauer, Markus / Miller, Norbert / Neuber, Frederike (ed. 2018–2022): *Jean Paul – Sämtliche Briefe digital*, <https://www.jeanpaul-edition.de> [20.04.2023].

DWDS (2022): Jean Paul Briefe (Textkorpus), https://www.dwds.de/d/korpora/jean_paul [20.04.2023].

Franzini, Greta (2012–2022): *Catalogue of Digital Editions*, <https://dig-ed-cat.acdh.oeaw.ac.at/> [20.04.2023].

Neuber, Frederike (ed. 2022): *telota/jean_paul_briefe* (v.6.0.). DOI: 10.5281/zenodo.4109517.

Neuber, Frederike / Wiegand, Frank / Hug, Marius (2022): “‘Eine einfältigere Frage an dich gab es wol nie als die, deren Antwort ich für mein Buch nöthig habe: nämlich wie geht das Kinderlied weiter: Nikolaus, fang’ die Maus...’ Task-Area-übergreifende Zusammenarbeit in Text+ am Beispiel der Forschungsdaten der Jean Paul Briefedition, in: Text+ Blog, <https://textplus.hypotheses.org/2095> [20.04.2023].

ZDL (2022): *Wortgeschichten digital*, <https://www.zdl.org/wb/wortgeschichten/> [20.04.2023].