

correspSearch v2.2 – Search historical correspondence

Dumont, Stefan

dumont@bbaw.de
Berlin-Brandenburg Academy of Sciences and Humanities,
Germany

Grabsch, Sascha

grabsch@bbaw.de
Berlin-Brandenburg Academy of Sciences and Humanities,
Germany

Müller-Laackman, Jonas

jonas.mueller-laackman@sub.Uni-Hamburg.de
Staats- und Universitätsbibliothek Hamburg

Sander, Ruth

ruth.sander@bbaw.de
Berlin-Brandenburg Academy of Sciences and Humanities,
Germany

2014 saw the development of the correspSearch prototype ¹: a web service that aggregates metadata of letters from scholarly editions and other publication types, and presents it for subsequent reuse and research (Dumont 2016). Rather than collecting the letter metadata centrally, the process occurs decentralized by the various editions, projects, and institutions, who then provide the data online in the Correspondence Metadata Interchange Format (CMIF) (Stadler 2014; Stadler, Illetschko, and Seifert 2016; Baillet and Busch 2021, 182–85; Kurz 2021). The CMIF in turn was developed collaboratively as part of the Special Interest Group “Correspondence” of the Text Encoding Initiative (TEI) (TEI Correspondence SIG 2014; Dumont et al. 2019). Within the CMIF, people and places are accompanied with URIs from authority files to ensure that these are identifiable and searchable across different projects (Stadler 2012).

Funded by the Deutsche Forschungsgemeinschaft, correspSearch was rebuilt with a new software architecture over the past few years. This short paper sets out to present this new architecture and features of correspSearch v2. A brief outlook on future developments is also given.

The modules ² that build up the version 2 of the web service have clearly distinctive functions: harvesting (eXistdb-App), ingest (Python), search engine (Elasticsearch), and UI (due.js). Elasticsearch in particular will prepare correspSearch for exponentially increasing data sets. Version 2 of correspSearch came online in the summer of 2021, offering new functionalities with it. Among these, the faceted search offers multiple combinations of filters ³ and searching for people, places or date is simplified by an autocomplete.

Additionally, the aggregated data is enriched with information from authority files. Every person who is identifiable via a URI from the Integrated Authority File (Gemeinsame Normdatei - GND) ⁴ is allocated further information, both from the GND and Wikidata. This data makes new filters such as gender and occupations possible. Places that are identified with GeoNames-URIs ⁵

are enriched with geo coordinates. This is used for another search method, the map-based search ⁶. Here, regions can be freely drawn on the map and be combined with dates. The resulting search finds letters from the specified region and time period. The map search also offers the selection of predefined historical state boundaries. These are supplied by HistoGIS (Andorfer et al. 2019), a web service by the Austrian Academy of Sciences and Humanities. The inclusion of external data and web services is conceptually intentional: correspSearch understands itself as a “lightweight infrastructure” that takes care of letters - no more, no less ⁷.

As previously mentioned, correspSearch depends on the metadata of scholarly editions, projects, and institutions. Correspondence metadata is aggregated from a wide array of disciplines, some include the humanities, philology, music, literature. This is possible because correspSearch does not set any temporal, geographical or thematic restrictions. To date (4.11.2022) over 180,000 edited or otherwise published letters have been provided by the community and can be found in correspSearch. The number continues to rise. Detailed documentation and tools aim to aid with the data preparations. Alongside the classical documentation, FAQs, and written instructions, correspSearch now offers two video tutorials ⁸ giving insight into the web service and data provision process.

The primary tool offered is the CMIF Creator. Interested parties can use the CMIF Creator ⁹ to write TEI-XML based files, no technical prior knowledge required (Müller-Laackman 2022). Additionally, further CMIF-workflows have been developed by the community, for example at the Saxon Academy of Sciences and Humanities (Rettinghaus 2021; Kretschmer and Scheideler 2021). Successfully created CMIF files can be verified using one of two online tools: CMIF Preview ¹⁰ and CMIF Check ¹¹. CMIF Preview allows for a visual review of the CMIF file, similar to a search result with an additional timeline based on the dates provided. In contrast CMIF Check validates the CMIF files against a RNG-Schema and Schematron, as well as testing all provided GeoNames-IDs for their correct assignment to feature classes. Both testing-tools were developed based on the experiences gathered in the past nine years of correspSearch and its data aggregation. Despite the standardized and documented interchange format, the process is not altogether effortless and requires intensive community support. From data acquisition, via conceptual discussions about CMIF to the data provision: community is key to correspSearch..

Notes

1. <https://correspSearch.net>
2. See <https://github.com/correspsearch> for already published modules.
3. E.g.: <https://correspsearch.net/en/search.html?s=http://d-nb.info/gnd/118554700>
4. https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.html
5. <https://www.geonames.org/>
6. <https://correspsearch.net/en/map.html>
7. Cf. discussion in (Dumont 2022)
8. <https://www.bbaw.de/mediathek/archiv-2022/introduction-to-correspsearch>
9. <https://correspsearch.net/en/cmif-creator.html>
10. <https://correspsearch.net/services/cmif-preview.xql?url=https://correspsearch.net/storage/forster/Band13.xml>
11. <https://correspsearch.net/services/check/check.xql?url=https://correspsearch.net/storage/forster/Band13.xml>

Bibliography

Andorfer, Peter, Matthias Schlögel, Antonia Dückelmann, Peter Paul Marckhgott-Sanabria, and Anna Piechl. 2019. *His-toGIS. A Geographical Information System, Workbench and Repository to Retrieve, Collect, Create, Enrich and Preserve Historical Temporalized Spatial Data Sets [Webservice]*. Wien: Austrian Centre for Digital Humanities and Cultural Heritage / Österreichische Akademie der Wissenschaften. <https://histogis.acdh.oeaw.ac.at/>.

Baillet, Anne, and Anna Busch. 2021. 'Editing for Man and Machine. Digital Scholarly Editions and Their Users'. *Variants. The Journal of the European Society for Textual Scholarship*, no. 15–16 (July): 175–87. <https://doi.org/10.4000/variants.1220>.

Dumont, Stefan. 2016. 'CorrespSearch – Connecting Scholarly Editions of Letters'. *Journal of the Text Encoding Initiative* 10 (December). <https://doi.org/10.4000/jtei.1742>.

———. 2022. 'Briefeditionen vernetzen'. In *Digitale Literaturwissenschaft. Beiträge des DFG-Symposiums 2017*, edited by Fotis Jannidis. Stuttgart: Metzler. <https://doi.org/10.5281/zenodo.6257224>.

Dumont, Stefan, Ingo Börner, Dominik Leipold, Jonas Müller-Laackman, and Gerlinde Schneider. 2019. 'Correspondence Metadata Interchange Format'. In *Encoding Correspondence. A Manual for TEI-XML-Based Encoding of Letters and Postcards*, edited by Stefan Dumont, Susanne Haaf, and Sabine Seifert. Berlin. <https://encoding-correspondence.bbaw.de/v1/CMIF.html>.

Kretschmer, Uwe, and Lars Scheideler. 2021. 'Ba[Sic?]'. Leipzig: Sächsische Akademie der Wissenschaften. <https://github.com/saw-leipzig/basic.app>.

Kurz, Stephan. 2021. 'Correspondence Metadata Interchange Format (CMIF)'. In *KONDE Weißbuch*, edited by Helmut W. Klug. Graz: Universität Graz. hdl.handle.net/11471/562.50.42.

Müller-Laackman, Jonas. 2022. 'CMIF Creator - Creating a Digital Index of Letters'. Vue. <https://doi.org/10.5281/zenodo.4075014>.

Rettinghaus, Klaus. 2021. 'CSV2CMI'. Leipzig: ba[sic?] is a tool for searching, identifying and connecting named entities with authority data. <https://github.com/saw-leipzig/csv2cmi>.

Stadler, Peter. 2012. 'Normdateien in der Edition'. *editio* 26: 174–83.

———. 2014. 'Interoperabilität von Digitalen Briefeditionen'. In *Fontanes Briefe Ediert*, edited by Hanna Delf von Wolzhagen, 278–87. Fontaneana 12. Würzburg: Königshausen & Neumann.

Stadler, Peter, Marcel Illetschko, and Sabine Seifert. 2016. 'Towards a Model for Encoding Correspondence in the TEI: Developing and Implementing <correspDesc>'. *Journal of the Text Encoding Initiative [Online]* 9. <https://dx.doi.org/10.4000/jtei.1742>.

TEI Correspondence SIG, ed. 2014. 'CMIF - Correspondence Metadata Interchange Format'. GitHub. <https://github.com/TEI-Correspondence-SIG/CMIF>.

(All links accessed on 27.04.2023)