Journal of the Text Encoding Initiative

Issue 10 | December 2016 - July 2019

Selected Papers from the 2015 TEI Conference

correspSearch – Connecting Scholarly Editions of Letters

Stefan Dumont

Electronic version

URL: <http://journals.openedition.org/jtei/1742>

DOI: 10.4000/jtei.1742

ISSN: 2162-5603

Publisher

TEI Consortium

Electronic reference

Stefan Dumont, « correspSearch – Connecting Scholarly Editions of Letters », *Journal of the Text Encoding Initiative* [Online], Issue 10 | December 2016 - July 2019, Online since 14 February 2018, connection on 03 July 2019. URL : <http://journals.openedition.org/jtei/1742> ; DOI : 10.4000/jtei.1742

For this publication a Creative Commons Attribution 4.0 International license has been granted by the author(s) who retain full copyright.

correspSearch – Connecting Scholarly Editions of Letters

Stefan Dumont

1. Research Context

- 1 Letters are an essential source for many research fields in the Humanities, including history, literature, musicology, and linguistic studies. In contrast to published works such as books and articles, private letters entice with the possibility of a more intimate and authentic insight into the author's thoughts, since they belong to those text types that are not usually intended to be published. Besides the (more or less presumed) privacy there is one even more important feature of letters: in contrast to publications, the content of a historical letter (either private or official) generally cannot be anticipated by a modern reader before actually reading it, because the content is extremely heterogeneous. Letters may contain information about a variety of people, publications, events, and issues from the author's working and living environment. Scholars therefore can make use of letters not only to study a person or a certain event, but also to investigate more systematic research questions, for instance: "How was the music business organized in the nineteenth century?" That is why letters are an important source that are often collected in scholarly editions, which prepare and provide these texts for subsequent research.

Furthermore, a letter is usually not a “nomadic single object” (Bunzel 2014, 234), but part of a larger correspondence network, allowing insights into connections and networks between correspondence partners. For scholars, questions emerge such as: “Who is in contact with whom and to what extent?,” “Where do circles of correspondence partners evolve and who are the central players within these circles?”

- 2 For editorial and practical reasons, however, scholarly editions of letters are usually focused on *partial* correspondences: either they concentrate on a certain person or they deal exclusively with the correspondence between two specific correspondence partners. The motives for this approach are usually straightforward and reasonable, considering methodology and work efficiency. But for the research questions mentioned above that involve a broader examination of correspondences it is often necessary to perform time-consuming searches across various editions of letters. For example, historians might be interested in the reactions of contemporary people to certain historical events. To study them, they would have to examine multiple scholarly editions and other publications. Furthermore, the number of letters that are available in scholarly editions is so large that often even experts cannot keep track of them—let alone the whole amount of letters preserved as archival documents. Thus, the majority of edited letters remain unused and uncited (Bunzel 2013, 117).
- 3 Since the early 2000s the scientific community has been discussing different approaches to solve this problem with the help of information technology. One early attempt was the BRIEFkasten project of the Institut für Textkritik in Heidelberg, which began in 2005.¹ The aim of BRIEFkasten was to establish a database for printed or edited letters between 1750 and 2000. Some further projects arose around the research subject “Republic of Letters” (ca. 1550–1750), for example, the database Early Modern Letters Online² or the project Frühneuzeitliche Ärztebriefe³ (Physicians’ Correspondences from the German-Speaking Territories), which gather metadata from both scholarly editions and archives to achieve better insights into the early modern scientific community. But these databases have some restrictions: first, they are created to address specific research questions and therefore focus on specific time periods, geographic areas, or topics. On the one hand, this is a legitimate approach considering the research questions they address, but on the other hand it excludes letters that might be interesting for other research. Second, these

databases do not offer an open, standard-based, and documented way to provide and update data. And finally, the data cannot be retrieved via an open Application Programming Interface (API) under a free license for subsequent use.

- 4 A general problem seems to lie with the conception of a single, centralized database, into which all the letter data have to be imported and where the data must be maintained. Considering the huge numbers of letters and potential research focuses, this concept may not be realistically pursuable. This could be one reason, why the BRIEFkasten project was discontinued after a couple of years. In contrast, an approach where decentralized data is simply aggregated via a standardized interchange format seems more promising (Rapp 2009, 205–6; Herres and Neuhaus 2002, 18)—especially if we would like to allow editors to use their own appropriate data model (e.g., in digital scholarly editions, the specific underlying TEI encoding) for their different purposes. This consideration also led Romanticism specialist Wolfgang Bunzel to request

the creation of a decentralized, preferably open digital platform, based on HTML/XML and operating with minimal TEI standards, which is extensible in different directions and allows for existing web portals and websites to contribute at the lowest possible cost. This doesn't require some kind of superstructure which covers the entire amount of letters from the Romantic era (the size of which could not be estimated exactly, anyway) but rather an intelligent linking system, which associates existing documents with one another. The creation of such a nexus will naturally lead to research options reaching from searches for persons and places to specific keyword-based searches ...

(Bunzel 2013, 123, my translation)

- 5 The web service correspSearch takes a step in this direction, collecting the letter metadata from various editions by using a standardized interchange format and making the data available through open interfaces, based on TEI XML and under a free license. The initiative for this web service arose from the workshop “Editions of Letters from around 1800: Finding and Connecting Interfaces,” which was organized in February 2014 by Anne Baillot and Markus Schnöpf.

2. Concept of correspSearch

- 6 The web service correspSearch aims to help scholars by offering a central place to search for edited letters, and by guiding them to the original publications. Only the metadata is aggregated from letter indexes of scholarly editions. Legal problems which could arise while retrieving full texts can thus be avoided; furthermore, scholarly editions found only in print form can be included without expensive digitization of entire publications. CorrespSearch is conceptually open in the sense that there is no focus on a particular research subject, time period, or region. This allows for new kinds of research questions to be explored.
- 7 The web service correspSearch is designed to *aggregate* correspondence metadata from digital indexes of letters which are hosted elsewhere on the web. These digital indexes of letters are created and provided by various scholarly edition projects according to a specific TEI XML-based interchange format (see [section 3](#) below) and were registered to the web service with their URL. The metadata provided via these letter indexes is then retrieved by the web service in periodic intervals. Thus, it is very straightforward and easy to add indexes of letters to the web service and update them if necessary. The data not only can be searched via a website, but also can be queried automatically via an Application Programming Interface (API) that is open, well documented, and supports different formats. As the data gathered is provided under free licenses, the web service offers all aggregated data under a CC-BY 4.0 license as well, thus making the data available for further reuse.
- 8 There already exist a variety of digital infrastructures regarding correspondence. The web service correspSearch is therefore designed to fit into the existing infrastructure landscape and complement existing services in a useful way. For information regarding purely archival resources, for example, there are already digital formats and services available, such as the Encoded Archival Description (EAD) format and the German digital union catalogue [Kalliope](#).⁴ This is why correspSearch focuses more specifically on *edited* letters—correspondence which has undergone scholarly editing, meaning that there is a published abstract or transcription, maybe even including a commentary. With further developments we envision that the data aggregated by correspSearch can be combined with data from other resources.
- 9 The technologies behind correspSearch are the open source software [eXist-db](#)⁵ and the programming languages XQuery and XSLT.

Figure 1. Technical concept of correspSearch.

3. Correspondence Metadata Interchange Format

- 10 In order for correspSearch to be able to aggregate the metadata of letters from diverse sources, these metadata have to be provided in a standardized and machine-readable format. The TEI XML format seemed appropriate in this context, because it had been used in digital editions for several years, and has now become a de facto standard. Furthermore, it is possible to rely on the work of the TEI Correspondence Special Interest Group (SIG) which developed the element `<correspDesc>` (“correspondence description”) for the TEI Guidelines. This extension of TEI provides a tagset for the recording of correspondence-specific metadata within the TEI Header, such as sender, addressee, and place of writing (Stadler, Illetschko, and Seifert 2016; Stadler 2014). After going through a few changes, `<correspDesc>` was integrated into the TEI Guidelines in April 2015.⁶

- 11 It therefore seemed straightforward to base the exchange format needed here on TEI XML and to use the new element `<correspDesc>` in this context (Stadler 2014). In order to allow for the provision of digital indexes of letters, the “Correspondence Metadata Interchange Format” (CMIF) is currently being developed within the TEI Correspondence SIG.⁷ The essence of a CMI document consists of multiple `<correspDesc>` elements, each describing one letter. However, the `<correspDesc>` elements are used in a significantly reduced and restricted manner in order to allow for further automatic processing. For instance, the `<correspAction>` element may only be of types “sent” and “received” and can only contain the elements `<persName>`, `<placeName>`, and `<date>` as children. In addition, if applicable, each `<correspDesc>` element may contain a reference to the corresponding digital edition which is available online. Only a few elements from the TEI Header are relevant for the CMIF besides the `<correspDesc>` element, for example, a bibliographic reference about the corresponding edition of the letters. Using a very reduced and restricted subset of the TEI for the CMIF is necessary in order to enable interoperability without any human intervention (besides registering the URL of the CMI file with correspSearch).⁸ However, this restriction applies only to the CMI file, but not to the TEI encoding of the texts and their metadata in the scholarly edition itself.

Figure 2. Schematic rendering of the different parts of a `<correspDesc>` element in the CMI format. The example originates from an example provided by the TEI Correspondence SIG.⁹

- 12 At this point in the development of the web service, only basic metadata about letters can be recorded: information about sender, addressee, places of writing and receiving, and dates. In the future, it will be possible to record additional metadata such as people or events mentioned in a letter, or whether some references are uncertain, or the type of textual basis of a letter edition

(e.g., manuscript or draft). It will then also be possible to browse through these categories in the correspSearch web interface. We also plan to allow the encoding of the corresponding archival resource in CMIF with the help of official URIs from digital archival catalogues (in Germany, for example, from Kalliope). These further developments raise the question whether CMIF should continue to be fully conformant to the TEI Guidelines or if it might be necessary to stray slightly from the TEI in certain aspects. These and other issues regarding the further development of CMIF are discussed in detail in “Perspectives of the Further Development of the Correspondence Metadata Interchange Format (CMIF)” (Dumont 2015).

4. Authority Files

- ¹³ The CMIF specification strongly recommends the usage of IDs from authority files to identify persons and places across projects and publications.¹⁰ Authority files are databases which were usually created and are still maintained by cooperative consortia of libraries, often headed by a major or national library, such as the Library of Congress or the German National Library. They are designed to “establish forms of names” and other metadata “used on bibliographic records ... to provide uniform access to materials in library catalogs and to provide clear identification of authors and subject headings.”¹¹ For each entity there exists a record with a unique, project-independent and persistent identifier, which can be referenced by everyone and is widely used as a reference in scholarly communities. In the context of correspSearch, the use of authority files avoids the limitations that would be created by referring to specific strings. Names can occur in different variants (e.g., without a certain first name) and vary orthographically. Furthermore, they can be ambiguous (e.g., “John Smith”). Thus, correspSearch searches are based on the IDs from authority files as encoded in the CMI files. The benefit of using authority-controlled IDs was proven by the German scholarly community via the application of the simple BEACON¹² format in diverse encyclopedias or digital editions (Baillot and Busch 2016).
- ¹⁴ The web service correspSearch supports several authority files for persons. Currently, these are the Integrated Authority File (GND) of the German National Library,¹³ the Library of Congress Authorities,¹⁴ the Autorités de la Bibliothèque Nationale de France (BNF),¹⁵ as well as the Web Authorities of the New Diet Library (NDL) in Japan.¹⁶ The Virtual International Authority File (VIAF) hosted by OCLC is also supported.¹⁷ With the help of VIAF, correspSearch maps IDs from one

authority file to the IDs of the others. In the end, it does not matter which one of the five supported authority files is used in a CMI file to identify senders or recipients: the web service automatically looks up all available IDs of a person.

- 15 Using IDs from authority files is not entirely unproblematic. First, in some cases entities may have been recorded multiple times, which means that there are several IDs available for one person. This complicates the selection of the appropriate ID, but orientation can be gained by assessing the validity of the different options: oftentimes, one of the records available is more enriched than the others and should therefore be used. Second, it also happens that entities, especially persons, that are of interest to research in general do not have any record in an authority file. In such cases it is necessary to add new records to an authority file, which usually has to be done by the staff of the maintaining library or organization. Because of limited human resources in libraries, completing this step can take some time after a new record has been suggested. Despite this lack of optimization, authority files remain the best approach currently available to connect information about persons, places, etc., across projects and institutions.¹⁸ Especially in the context of semantic web technologies, the usage of standardized unique identifiers (i.e., URIs from authority files) for common entities is crucial in order to ensure the sharing of research data as linked open data. We can expect the maintenance of authority files to be enhanced in the future.
- 16 For the identification of places, the support of [GeoNames](#) was implemented in correspSearch. GeoNames is a free licensed database, which consists of over ten million geographic names that are categorized in nine “feature classes” and 645 “feature codes.”¹⁹ For each place name, additional data such as geographic coordinates are available. Place names that are not already available in GeoNames can be added by any registered user. If necessary, it is possible to note alternate names for a place, for example, names in other languages or historical names. It is also possible to categorize as a “historical site” a place that no longer exists. Despite these features, the handling of historical places in GeoNames is limited because it is a contemporary gazetteer: that is, it is not possible to record the time period during which a place (or place name) existed. It is a major limitation that the data about state boundaries or the state to which a city belongs only reflect the current situation and have no historical depth whatsoever. This is challenging, but not impossible to address, when it comes to correspondences since the places mentioned in the context of historical correspondence are often names of cities or villages. They may today be part of a

larger municipality but remain addressable because they still exist as a separate data record with a geographic *point* and not with a certain *area*, which can change over time. All in all, the usage of GeoNames in CMIF and correspSearch up to now has been beneficial and raised only limited issues.

5. Data

- 17 As of the beginning of February 2017, correspSearch has gathered the metadata of more than 26,000 letters from almost 110 publications. Letters from digital editions (e.g., *Carl Maria von Weber –Collected Works*,²⁰ *Letters and Texts: Intellectual Berlin around 1800*,²¹ or the project *Alfred Escher's Correspondence*²²) are not only registered to the web service but also directly available to the user via a link. Additionally, many letters from printed scholarly editions have been recorded in CMIF and are thus available through a digital catalogue for the first time: for example, the letters of the natural scientist and anatomist Samuel Thomas Soemmerring or the correspondence between Karl August Varnhagen von Ense and Friedrich de la Motte Fouqué. Besides scholarly editions, the web service records edited letters that are published in appendixes of monographs or in journal articles. Examples are the open access journal *HiN: HiN. Alexander von Humboldt im Netz. International Review for Humboldtian Studies*,²³ where edited letters of the famous scientist are published periodically for the first time, or the monograph *Studien zum geistlichen Werk Otto Nicolais*,²⁴ where Klaus Rettinghaus published a number of edited letters in appendix B.
- 18 The aggregated data is growing and will continue to grow in the future; multiple scholarly editions have announced their upcoming contributions. Such participation is crucial: it is only through the willingness of relevant projects to provide correspondence metadata that a web service like correspSearch can be valuable to the scholarly community. Every scholarly project is welcome to provide its metadata in the CMI format. Interested scholars and projects can find step-by-step instructions, documentation of the CMI format, and a [FAQ section](#) on the website²⁵ as well as templates and examples in the [GitHub repository](#) of the TEI Correspondence SIG.²⁶
- 19 Providing a CMI file has multiple advantages for the participating research projects: First, it enables them to make their edited letters more easily available to the community and to connect their editions of letters (automatically—see below) with other scholarly editions. Second, they help to facilitate research that involves a broader examination of correspondences, as mentioned above.

Furthermore, by publishing their metadata, they can get useful feedback to correct and refine their metadata, as the experience with CMIF and correspSearch shows so far: more users read and use the metadata and therefore errors can be more easily discovered.²⁷

6. APIs

- 20 Researchers—and of course editors working on their own scholarly editions—can search the correspSearch data pool easily via the web interface. The search functionalities are currently restricted to the essential fields, but they will be extended in the future. The aim from the very beginning, however, was to provide the data not only in a human-readable but also in a machine-readable way in order to allow for automated queries and subsequent use. Therefore correspSearch offers multiple application programming interfaces. The APIs support all search parameters available in the web interface.²⁸ The results can be retrieved in several formats. First, correspSearch offers all data as TEI XML in the CMI format. Second, to enable as much reuse as possible, the output is also provided in a JSON serialization. Mixed-content XML elements cannot be well converted into JSON, but such markup almost never occurs in the CMIF. Besides TEI XML and TEI JSON, the web service offers, third, an experimental CSV output, which is intended for individual use of search results (which is why this API is also implemented in the graphical web interface). Finally, a BEACON API enables the cross-linking of search results to a specific person in correspSearch.²⁹
- 21 Via the APIs, scholars can exploit the aggregated data using the technologies and software they favor. Therefore, with sufficiently extended data and suitable software it will be possible to perform research on, for instance, social or correspondence networks based on correspSearch. The developers from LAB110, for example, have imported correspSearch data into their visualization tool “nodegoat” to visualize a correspondence network (figure 3).³⁰

Figure 3. A visualization in “nodegoat” from correspondence metadata gathered by correspSearch.

- 22 Thanks to the API, it is also possible to automatically refer or even link from one digital letter edition to related letters provided by other editions. This function has already been implemented in the *edition humboldt digital*¹ (figure 4). Whenever a user visits a certain letter on the website, the digital edition automatically performs a query via the correspSearch API to find out if there are any other letters from Humboldts’s correspondent (in the pictured case, Samuel Thomas Soemmerring) in *other* scholarly editions. In the pictured case the results from the correspSearch API offer—among others—a hint to a letter from Georg Christoph Lichtenberg, which was sent to Samuel Thomas Soemmerring two days before.

Figure 4. Screenshot of the digital scholarly edition *edition humboldt digital*, which presents among other manuscripts some letters to and from Alexander von Humboldt. In the top right corner, links to letters in other scholarly editions are presented with the help of the correspSearch API. Source: <http://edition-humboldt.de/H0002729>.

- 23 This feature helps researchers avoid methodological problems when interpreting a piece of correspondence: when analyzing a letter they usually consider the preceding and following letters in the correspondence between the sender and addressee as well. Their interpretation often does not include the letters which the correspondents sent to or received from *other* persons within the same context, although the content of those letters may be highly relevant to the letter they are interested in. With this feature, the broader correspondence context of historical letters can be explored in a greater depth than ever before.
- 24 Besides TEI XML, TEI JSON, and CSV, the correspSearch API also offers a preliminary RDF version of the data by using the web service [XTriples](#) developed by the Academy of Sciences, Humanities and Literature in Mainz. With the help of XTriples, the CMIF is converted into RDF statements.³² Providing the aggregated correspondence metadata as RDF enables users to analyze the data with the help of semantic web technologies. Furthermore, the RDF data can be used as Linked Open Data in new contexts and in combination with data from other sources in order to answer

questions beyond the correspondence-specific issues (Grüntgens and Schrade 2016). For instance, with data from multiple universities’ catalogues of professors, lecturers, and students, scholars could research whether or not a correspondence network (or part of it) reflects a certain “school.”

Figure 5. This diagram, created by Torsten Schrade,³³ illustrates how the correspondence metadata aggregated by correspSearch could be used as Linked Open Data (LOD) with information from other resources like authority files (here GeoNames and the Integrated Authority File (GND)). This simple example illustrates the idea of using data from correspSearch with further LOD resources.

7. A First Use Case: Alexander von Humboldt

- 25 Alexander von Humboldt is a good illustration of how a decentralized approach such as the one proposed here can provide a major contribution to research. In the course of his life Alexander von Humboldt wrote about 50,000 letters. Today we know about the content of around 12,400 of them. The number of letters addressed to Humboldt probably exceeded 100,000, even though only 3,400 of those letters are known today (Schwarz 2002, 194–95; Suckow and Schwarz 1998, 119). From the beginning, this extraordinarily large number of documents has deterred Humboldt researchers

from planning an edition of Humboldt's complete letters. They have concentrated instead on providing editions for certain parts of his correspondence.³⁴ The focus on specific correspondents of Humboldt is legitimate from an economical and methodological point of view. However, these selective editions also entail problems (Suckow and Schwartz 1998). First, the choice of criteria applied to guide the text selections is of decisive importance, as well as the choice of whose correspondence to prioritize, since, after all, the aim of completeness will never be achieved, though it still exists in principle.³⁵ Secondly, users of selective editions may lose sight of the fact that individual letters not only form part of a correspondence between two people, but rather are embedded in a much larger network of letters. This is the case in particular when it comes to Humboldt, who wrote various letters to different people at the same time, and often his statements on similar issues or people vary depending on the addressee. Humboldt possessed great diplomatic skills and knew very well that he might risk making enemies had he expressed critical opinions too openly towards the person concerned (Werner 2000, 2).

- 26 Collecting the edited letters of Humboldt in correspSearch (a task that has not yet been completed) helps overcome a methodological disadvantage of selective editions, since it enables the contextualization of individual letters within the entire correspondence of Humboldt: on the one hand for the editor during the creation of a scholarly edition, on the other hand for the user while exploring and analyzing edited letters. By proceeding like this, the practice of selective editions, which is straightforward with regard to other aspects, may be continued. Therefore, the possibility of automatically connecting letters in scholarly editions with the help of correspSearch, as shown above, was implemented within the *edition humboldt digital*.³⁶ The connection is especially useful for this edition project because the selection of letters and texts considered for the edition is partially based on thematic categories. In addition, further development of correspSearch will allow for improved accessibility to the various topics of Humboldt's correspondence.
- 27 However, the web service correspSearch is not only meant to provide renewed access and contextualization for historical persons who are already known. It is also intended to draw attention to less well-known correspondence partners. The aggregation of correspondence metadata from diverse editions enhances the research options and makes it possible to evaluate networks and their presumably central figures. This method not only draws attention to persons

who have not previously been taken into consideration by a scholarly edition, but it also facilitates the evaluation of their importance, because it focuses not only on one or two persons but makes it possible to display a whole network of, for example, academics.

8. Conclusion

- 28 Especially for editions of letters, the digital age has brought significant progress that allows scholars to address a series of known challenges like incompleteness and contextualization of letters, or the discovery and disclosure of correspondence networks. The web service correspSearch is a building block for improved solutions in this context. We plan to develop correspSearch further in order to enhance its search functionalities (e.g., for persons and publications mentioned in the text) and to make use of the extensions planned for the CMI format. In the course of this development, support for further authority files will be implemented. Besides the search functionalities, we plan to simplify the creation of CMI files through diverse tools. The aggregated data should grow further. Recently, new metadata records have been added: among others, records from the online platform of the critical edition of the works of *Richard Strauss Werke*³⁷ and from the digital edition *Briefe der Fruchtbringenden Gesellschaft und Beilagen*.³⁸ The more correspondence metadata is published and may be included in correspSearch, the more useful the web service will become to creators and users of scholarly editions. We welcome additional contributions!³⁹

BIBLIOGRAPHY

- Bauman, Syd. 2011. "Interchange vs. Interoperability." In *Proceedings of Balisage: The Markup Conference 2011*. Balisage Series on Markup Technologies, vol. 7. doi:10.4242/BalisageVol7.Bauman01.
- Bailot, Anne, and Anna Busch. 2016. "Editing for Man and Machine: The Digital Edition *Letters and Texts: Intellectual Berlin around 1800* as an Example." *Users of Scholarly Editions: Editorial Anticipations of Reading, Studying and Consulting*, November 2015, Leicester, United Kingdom. Author version available at <https://halshs.archives-ouvertes.fr/halshs-01233380>.

- Bunzel, Wolfgang. 2013. "Briefnetzwerke der Romantik. Theorie – Praxis – Edition." In *Brief-Edition im digitalen Zeitalter*, edited by Anne Bohnenkamp and Elke Richter, 109–31. Beihefte zu editio, vol. 34. Berlin: de Gruyter.
- . 2014. "Briefe, Briefnetze, Briefnetzwerke. Überlegungen zur epistolaren Interkonnektivität." In *Fontanes Briefe ediert*, edited by Hanna Delf von Wolzogen and Rainer Falk, 232–45. Fontaneana, vol. 12. Würzburg: Königshausen & Neumann.
- Dumont, Stefan. 2015. "Perspectives of the Further Development of the Correspondence Metadata Interchange Format (CMIF)." In *digiversity: Webmagazin für Informationstechnologie in den Geisteswissenschaften* (October 29). <http://digiversity.net/2015/perspectives-of-the-further-development-of-the-correspondence-metadata-interchange-format-cmif>.
- Grüntgens, Max, and Torsten Schrade. 2016. "Data Repositories in the Humanities and the Semantic Web: Modelling, Linking, Visualising." In *WHiSe 2016. Humanities in the Semantic Web*, edited by Alessandro Adamou, Enrico Daga, and Leif Isaksen, 53–64. *CEUR Workshop Proceedings*, vol. 1608. Aachen. <http://ceur-ws.org/Vol-1608/#paper-07>.
- Herres, Jürgen, and Manfred Neuhaus. 2002. "Vorwort der Herausgeber." In *Politische Netzwerke durch Briefkommunikation*, edited by Jürgen Herres and Manfred Neuhaus, 7–25. Berlin: Akademie Verlag.
- Rapp, Andrea. 2009. "Einige Anmerkungen zu Retrodigitalisierungs-Verfahren und Perspektiven digitaler Briefeditionen." In *Digitale Edition zwischen Experiment und Standardisierung. Musik – Text – Codierung*, edited by Peter Stadler and Joachim Veit, 203–6. Beihefte zu editio, vol. 31. Tübingen: Niemeyer.
- Schwarz, Ingo. 2002. "Korrespondenz als Last und Vergnügen. Zum Briefwechsel Alexander von Humboldts." In *Politische Netzwerke durch Briefkommunikation*, edited by Jürgen Herres und Manfred Neuhaus, 193–217. Berlin: Akademie Verlag.
- Stadler, Peter. 2012. "Normdateien in der Edition." *editio* 26, 174–83. doi:10.1515/editio-2012-0013.
- . 2014. "Interoperabilität von digitalen Briefeditionen." In *Fontanes Briefe ediert*, edited by Hanna Delf von Wolzogen and Rainer Falk, 278–87. Fontaneana, vol. 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>." In *Journal of the Text Encoding Initiative* 9. <https://journals.openedition.org/jtei/1433>; doi:10.4000/jtei.1433.
- Suckow, Christian, and Ingo Schwarz. 1998. "Zur Problematik einer auswählenden Briefedition. Beispiel: Die Briefe Alexander von Humboldts." In *Wissenschaftliche Briefeditionen und ihre Probleme*, edited by Hans-Gert Roloff, 119–22. Berlin: Weidler.
- Werner, Petra. 2000. *Casanova ohne Frauen? Bemerkungen zu Alexander von Humboldts Korrespondenzpartnerinnen*. Teil 1: "Kosmos. Entwurf einer physischen Weltbeschreibung." *Berliner Manuskripte zur Alexander-von-Humboldt-Forschung*, vol. 16. Berlin: Alexander-von-Humboldt-Forschungsstelle.

NOTES

- 1 BRIEFkasten, accessed July 31, 2017, <http://www.textkritik.de/briefkasten/bkprojekt.htm>.
- 2 Early Modern Letters Online, accessed July 31, 2017, <http://emlo.bodleian.ox.ac.uk/>.
- 3 Frühneuzeitliche Ärztebriefe des deutschsprachigen Raums (1500–1700), accessed July 31, 2017, <http://www.medizingeschichte.uni-wuerzburg.de/akademie/index.html>.
- 4 Kalliope Union Catalog, accessed July 31, 2017, <http://kalliope-verbund.info/en/index.html>.
- 5 Accessed July 31, 2017, <http://exist-db.org/>.
- 6 TEI P5: Guidelines for Electronic Text Encoding and Interchange, Appendix C Elements, accessed July 31, 2017, <http://www.tei-c.org/Vault/P5/3.3.0/doc/tei-p5-doc/en/html/ref-correspDesc.html>.
- 7 Last updated February 6, 2017, <http://www.tei-c.org/Activities/SIG/Correspondence/>.
- 8 “Interoperability” as defined by Syd Bauman (Bauman 2011). Bauman argues rightly that “since interoperability ... is often bought at the expense of expressivity ... , interoperability is the wrong goal for scholarly humanities text encoding.” This is the reason why the TEI Correspondence SIG decided not to standardize the parts of the TEI header of an edited letter, but conceptualized a format for an *additional* digital index of letters, which can be created (almost) independently from the scholarly edition itself.
- 9 Accessed February 25, 2018, https://github.com/TEI-Correspondence-SIG/CMIF/blob/master/examples/example01_basic.xml.
- 10 On the use of authority-controlled IDs, see Stadler 2012.
- 11 Library of Congress Authorities Help Pages, Frequently Asked Questions, accessed July 31, 2017, <http://authorities.loc.gov/help/auth-faq.htm#1>.
- 12 Specification of BEACON link dump format, GitHub, accessed July 31, 2017, <https://github.com/gbv/beaconspec>.
- 13 Last updated October 21, 2016, http://www.dnb.de/EN/Standardisierung/GND/gnd_node.html.
- 14 Accessed July 31, 2017, <http://authorities.loc.gov/>.
- 15 Accessed July 31, 2017, http://www.bnf.fr/fr/professionnels/autorites_bnf/s.autorite_bnf_presentation_statistiques.html.

- 16 Accessed July 31, 2017, <http://id.ndl.go.jp/auth/ndla>.
- 17 Accessed July 31, 2017, <http://viaf.org/>.
- 18 The project and web service Metagrid (accessed July 31, 2017, <http://www.metagrid.ch/>) has been under development in Switzerland for a few years. It is not an authority file, but a central web service that maps entity IDs from one project to another. This is a different and interesting approach that, however, requires greater involvement and commitment from the participating projects, because it is necessary for the projects to enter their IDs into an access-restricted concordance. Authority files, in contrast, can be used by anyone in any context in order to link data, which also includes private projects.
- 19 Accessed July 31, 2017, <http://www.geonames.org/>.
- 20 *Carl-Maria-von-Weber-Gesamtausgabe* (WeGA), accessed July 31, 2017, <http://www.weber-gesamtausgabe.de/>.
- 21 *Briefe und Texte aus dem intellektuellen Berlin um 1800*, edited by Anne Baillot, accessed July 31, 2017, <http://www.berliner-intellektuelle.eu/>.
- 22 *Digitale Briefedition Alfred Escher*, edited by Jopseh Jung, Zürich 2015. accessed July 31, 2017, <https://www.briefedition.alfred-escher.ch/>.
- 23 Accessed July 28, 2017, <http://www.hin-online.de/>.
- 24 PhD diss., Technische Universität Berlin, 2012, published online 2014, <http://dx.doi.org/10.14279/depositonce-4210>.
- 25 “Participate,” correspSearch, accessed July 31, 2017, <http://correspsearch.net/index.xql?id=participate&l=en>.
- 26 Correspondence Metadata Interchange Format, edited by TEI Correspondence SIG, accessed July 31, 2017, <https://github.com/TEI-Correspondence-SIG/CMIF>.
- 27 For example, the use of the wrong ID from an authority file.
- 28 For the available output formats and search parameters, see the overview of the correspSearch API, accessed July 31, 2017, <http://correspsearch.net/index.xql?id=api#parameter>.
- 29 Which has been done, for example, by the Deutsche Biographie, edited by Historische Kommission bei der Bayerischen Akademie der Wissenschaften, accessed July 31, 2017, <https://www.deutsche-biographie.de/>.

- 30** The visualization is available in an interactive format, accessed July 31, 2017, at <http://correspsearch-test.nodegoat.net/viewer.p/4/136/scenario/1/geo/fullscreen>.
- 31** *edition humboldt digital*, edited by Ottmar Ette. Berlin: Berlin-Brandenburgische Akademie der Wissenschaften, 2017, accessed July 31, 2017, <http://edition-humboldt.de/>.
- 32** XTriples, accessed July 31, 2017, <http://xtriples.spatialhumanities.de/>. A prototype configuration, accessed July 31, 2017, is available at <http://xtriples.spatialhumanities.de/examples.html>. The interaction between the web services correspSearch and XTriples was presented by Torsten Schrade at the DHD conference 2016 in Leipzig: “Geisteswissenschaftliche Fachdatenrepositorien im Semantic Web: Modellierung, Vernetzung, Visualisierung,” accessed July 31, 2017, <http://dhd2016.digitale-akademie.de/presentation/#/step-1>.
- 33** “Digitale Briefeditionen im Semantic Web: Der XTriples Webservice—Semantische Aussagen aus XML,” presentation by Torsten Schrade at the Workshop of the Electronic Publishing working group of the Union of the German Academies of Sciences and Humanities, September 15, 2015, Heidelberg, <http://metacontext.github.io/presentation-correspsearch-xtriples>.
- 34** Besides a complete correspondence between Humboldt and one of his correspondents, scholars also publish individual or small numbers of letters in dedicated journals such as *HiN: Alexander von Humboldt im Netz* (mentioned above).
- 35** The aim of completeness is still a valid ideal for scholarly editions of letters. The problem that “completeness” can never be achieved, exists for selective editions as well as for complete editions. In both cases, digital scholarly editions have the advantage over printed editions that newly discovered letters can be added more easily.
- 36** *edition humboldt digital*, edited by Ottmar Ette. Berlin: Berlin-Brandenburgische Akademie der Wissenschaften, 2017, accessed July 31, 2017, <http://edition-humboldt.de/>.
- 37** *Kritische Ausgabe der Werke von Richard Strauss: Online-Plattform*. München: Bayerische Akademie der Wissenschaften, 2017, accessed July 31, 2017, <http://richard-strauss-ausgabe.de>.
- 38** *Briefe der Fruchtbringenden Gesellschaft und Beilagen. Die Zeit Fürst Ludwigs von Anhalt-Köthen 1617–1650, vol. 1: 1617–1626*. Ed. by Klaus Conermann with the support of Dieter Merzbacher. Wolfenbüttel: Herzog August Bibliothek, 2014. <http://diglib.hab.de/edoc/ed000213/start.htm>.
- 39** “Participate,” correspSearch, <http://correspsearch.net/index.xql?id=participate>.

ABSTRACT

The web service correspSearch collects correspondence metadata from diverse scholarly editions by using the TEI XML-based “Correspondence Metadata Interchange Format” developed by the TEI Correspondence SIG. The web service provides the aggregated data not only on a website, but also via an API in different machine-readable formats and under a free license. The API also makes it possible to automatically refer or even link a letter in one digital edition to related letters provided by other editions. Collecting edited letters in correspSearch and linking them via the API helps overcome certain methodological disadvantages of scholarly editions of letters, since it enables the contextualization and connection of single letters within an entire correspondence network.

INDEX

Keywords: correspondence, letter, web service, interchange, metadata, authority files

AUTHOR

STEFAN DUMONT

Stefan Dumont studied history, public law, and political science at the universities of Mainz and Dijon. Since 2011 he has been working as a researcher and developer for TELOTA, the digital humanities initiative of the Berlin-Brandenburg Academy of Sciences and Humanities.