

correspSearch. Ein zentraler Service zum Vernetzen von Briefeditionen und -repositorien

Stefan Dumont, Marcel Illetschko, Sabine Seifert, Peter Stadler

Als am 24. Februar 1848 die Revolution in Paris der Herrschaft des ‚Bürgerkönigs‘ Louis-Philippe von Orléans ein Ende setzte und ihre alles verändernden Wogen über Europa aussandte, schrieb Gustave Flaubert an seine Geliebte Louise Colet, er vergnüge sich „höchstlich bei der Betrachtung all der zunichte gewordenen Ambitionen“. Zwar wisse er nicht, ob die neue Form der Regierung und der gesellschaftliche Zustand, der daraus hervorgehen werde, für die Kunst günstig sei. Man könne allerdings kaum bürgerlicher und belangloser als die alte werden. „Und noch dümmere – ist das möglich?“¹ Etwa gleichzeitig meinte Charles Dickens gegenüber seinem Freund, dem Schauspieler William Macready, dass er den neuen französischen Regierungschef für „one of the best fellows in the world“ halte und die große Hoffnung hege, „that great people establishing a noble republic.“² Heinrich Heines Reaktionen auf die republikanischen Umbrüche in Frankreich und im restlichen Europa waren weniger positiv. Seine Mutter ließ er wissen: „Eben weil es jetzt so stürmisch in der Welt und hier besonders tribulant hergeht, kann ich Dir wenig schreiben. Der Spektakel hat mich physisch und moralisch sehr heruntergebracht. Ich bin so entmuthigt, wie ich es nie war. [...] Sollten die Sachen sich hier, wie ich fürchte, noch düsterer gestalten, so gehe ich fort, mit meiner Frau, oder auch allein.“³

Die Februarrevolution 1848 steht hier als ein willkürlich herausgegriffenes Beispiel mit einer zufällig getroffenen Anzahl an Briefen und Briefeditionen. Dabei wird deutlich, dass Briefe zu den wertvollsten Quellen historischer Forschung zählen. Unterschiedliche Themen und Ereignisse aus der Lebenswelt der Korrespondenten werden angesprochen⁴ und soziale Netzwerke abgebildet: „Auf Grund der Tatsache, dass ein Individuum nie nur mit einem einzigen Gegenüber korrespondiert, ist jeder Brief immer auch ein kommunikativer Akt in einem größeren, interpersonalen Zusammenhang. So wie das einzelne Subjekt in seiner Rolle als Schreiber-Adressat Teil eines Beziehungsgeflechts ist, so nimmt jeder einzelne Brief eine Mehrfachposition im kommunikativen Gesamtgefüge epistolaren Austauschs ein [...]“.⁵ Und so stellen sich Fragen an die Briefnetzwerke, wie z.B.: Welche politischen oder künstlerischen Kreise sind abgrenzbar, wie intensiv ist die Kommunikation innerhalb solcher Zirkel, welche Schlüsselfiguren gibt es, wie ist es um die Kommunikation mit anderen Kreisen bestellt? Etc.

¹ Gustave Flaubert. Briefe. Hg. von Helmut Scheffel. Zürich: Diogenes 1977, S. 110.

² Letters of Charles Dickens: 1833-1870. Hg. von Georgina Hogarth und Mary Dickens. Cambridge: Cambridge University Press 2011, S. 182.

³ HSA Bd. 22, S. 270, Brief Nr. 1215, Online-Abruf Heine-Portal, 3.11.2014.

⁴ Einschlägig zum Thema etwa Wolfgang Frühwald et al. (Hg.): Probleme der Brief-Edition: Kolloquium der Deutschen Forschungsgemeinschaft. Boppard 1977. – Irmtraut Schmid, Was ist ein Brief? Zur Begriffsbestimmung des Terminus ‚Brief‘ als Bezeichnung einer quellenkundlichen Gattung. In: editio 2 (1988), S. 1–7. – Reinhard Nickisch: Brief. Stuttgart 1991. – Roloff, Hans-Gert (Hrsg.): Wissenschaftliche Briefeditionen und ihre Probleme. Editionswissenschaftliches Symposium. Berlin 1998. – Waltraut Wiethölter: Der Brief – Ereignis und Objekt. Frankfurt am Main 2010. – Anne Bohnenkamp und Elke Richter (Hg.): Brief-Edition im digitalen Zeitalter (=Beihefte zu editio Bd. 34). Berlin/Boston 2013.

⁵ Wolfgang Bunzel: Briefnetzwerke der Romantik. Theorie – Praxis – Edition. In: Anne Bohnenkamp und Elke Richter (Hg.): Brief-Edition im digitalen Zeitalter (=Beihefte zu editio Bd. 34) Berlin/Boston 2013. S. 109-131, hier S. 113. – Im Folgenden zitiert als: Bunzel (2013).

Allerdings werden meist nur Briefeditionen *eines* Autors oder *einer* speziellen Korrespondenz (z.B. Verlegerbriefwechsel, Familienbriefwechsel) erstellt. Lediglich in wenigen neueren Projekten sind Briefe im Hinblick auf z. B. ein historisches Thema ediert.⁶ Übergreifende Forschungsprojekte, die auf Basis von Korrespondenz(en) größere thematische Zusammenhänge (wie das eingangs genannte Beispiel) oder größere Korrespondenznetzwerke in den Blick nehmen wollten, mussten die Datenbasis meist aufwendig aus vorhandenen gedruckten oder online verfügbaren Editionen zusammentragen und aufbereiten.

Prinzipiell gilt also: „Because of the everyday occurrence of the epistolary exchange, the [...] large stocks of letters render—without the support of computers—a scientific analysis of the correspondence networks spanned within them nearly impossible.”⁷ Bekannt ist diese Problematik in der deutschsprachigen Literaturwissenschaft seit Langem v.a. in Bezug auf Romantikernetzwerke. So meint etwa Wolfgang Bunzel: „Die schiere Masse der überlieferten Briefe ist selbst für Experten nicht zu überschauen, viel weniger für Forscher, die sich nur mit einzelnen Personen beschäftigen oder an Detailspekten interessiert sind“, was zur Folge habe, „dass schon das gedruckt vorliegende Material nur höchst selektiv benutzt wird und die Mehrzahl der Briefe unzitert, meist sogar ungesichtet bleibt.“⁸ Er fordert deshalb

„die Schaffung einer dezentralen, möglichst offenen, auf (bis auf weiteres) html/xml-Grundlage basierenden und mit TEI-Minimalstandards operierenden digitalen Plattform, die nach vielen Richtungen hin erweiterbar ist und es den bereits bestehenden Portalen und Homepages erlaubt, sich mit denkbar geringem Zusatzaufwand daran zu beteiligen. Nötig ist dafür keine Superstruktur, welche die – ohnehin nicht exakt bezifferbare – Gesamtheit aller Briefe der Romantik überwölbt, sondern vielmehr ein intelligentes Verknüpfungssystem, das vorhandene Dokumente in Konnex zueinander bringt. Mit dem Aufbau eines solchen Nexus gehen natürlich Recherchemöglichkeiten einher, die von der Personen- über die Datums- und Orts- bis hin zur gezielten Stichwortsuche (und dies natürlich in beliebiger Kombination der Suchparameter) reichen.“⁹

Wir möchten mit „correspSearch“ einen Webservice vorstellen, der einen ersten Schritt in diese Richtung geht, indem er aus verteilten Editionen und Repositorien die Briefmetadaten aggregiert und über offene Schnittstellen, auf TEI-XML-Grundlage sowie unter einer freien Lizenz zentral zur Verfügung stellt. Die Initiative zu diesem Webdienst entstand im Februar 2014 im Workshop „Briefeditionen um 1800: Schnittstellen finden und vernetzen“, der von Anne Baillot (Nachwuchsgruppe „Berliner Intellektuelle 1800–1830“ an der HU Berlin) und Markus Schnöpf (TELOTA, BBAW) organisiert worden war.

Das Fundament des Webservices ist ein Austauschformat, das auf dem Modul „correspDesc“ (Correspondence Description)¹⁰ für die Richtlinien der Text Encoding Initiative¹¹ basiert. Das Modul

⁶ So z.B. das Projekt „Letters 1916“, <http://dh.tcd.ie/letters1916/>, das Briefe aus der Zeit um den Osteraufstand 1916 in Irland sammelt und transkribiert, oder das Projekt „Vernetzte Korrespondenzen | Exilnetz33“, <http://exilnetz33.de/>.

⁷ Martin Andert, Frank Berger, Paul Molitor and Jörg Ritter: An optimized platform for capturing metadata of historical correspondence. In: *Lit Linguist Computing* (2014), doi: 10.1093/llc/fqu027

⁸ Bunzel (2013), S. 117.

⁹ Ebd., S. 123.

¹⁰ <https://github.com/TEI-Correspondence-SIG/correspDesc>

¹¹ Burnard, Lou; Bauman, Syd (Hg.): TEI P5: Guidelines for Electronic Text Encoding and Interchange. Charlottesville, Virginia, USA 2014. URL: <<http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>>

wurde von der TEI Special Interest Group Correspondence¹² entwickelt, um die Metadaten einer einzelnen Korrespondenz (d.h. insbesondere eines Briefes) standardisiert in TEI-XML notieren zu können. Derzeit befindet sich das Modul „correspDesc“ im Antragsverfahren¹³. Auf der Grundlage dieses TEI-Moduls wurde (und wird) ein Austauschformat entwickelt, das nach Abschluss des Prozesses von der TEI SIG Correspondence empfohlen werden soll. Dieses Austauschformat ist ebenfalls in TEI-XML definiert und bietet die Möglichkeit, das Briefverzeichnis einer Edition standardisiert zu notieren – also Absender, Empfänger, Datums- und Ortsangaben eines jeden Briefes. Neben dieser einheitlichen TEI-XML-Kodierung wird der Austausch auch durch die Verwendung von Normdaten ermöglicht.¹⁴ Absender, Empfänger, Schreib- und Empfangsorte werden durch Normidentifikationsnummern (wie z.B. die GND-Nummer der Deutschen Nationalbibliothek¹⁵) identifiziert. Datumsangaben werden ebenfalls standardisiert erfasst. Dadurch werden die Metadaten eines Briefes über Projekt- und Sprachgrenzen hinweg operabel gemacht. Schließlich wird im Austauschformat auf den einzelnen Brief referenziert. Während es sich bei ausschließlich gedruckten Editionen um die Briefnummer und die bibliografische Angabe handelt, können digitale Editionen zusätzlich die URL hinterlegen. So kann jedes Editionsprojekt sein Briefverzeichnis digital bereitstellen.

Um das Potential dieser digitalen Briefverzeichnisse zu nutzen, wurde von der TELOTA-Arbeitsgruppe an der BBAW in Zusammenarbeit mit der TEI SIG Correspondence und weiteren Wissenschaftler(inne)n der Webservice „correspSearch“ (<http://correspSearch.bbaw.de>) entwickelt, der digitale Briefverzeichnisse aggregiert und abfragbar bereitstellt. Die Verzeichnisse werden dabei jeweils vom Anbieter unter einer CC-BY-Lizenz¹⁶ vorgehalten und vom Webservice in periodischen Abständen neu bezogen. Jedes Verzeichnis muss also lediglich mit seiner permanenten URL im Webservice registriert werden und wird dann automatisch ausgelesen. Dadurch können weitere Briefverzeichnisse ganz leicht dem Webservice hinzugefügt werden. Beim Einlesen der Verzeichnisse wird für jede Norm-ID eines Korrespondenten bei der Virtual International Authority File¹⁷ nach den gängigsten Norm-IDs gefragt, so dass unterschiedliche Normdatensysteme aufeinander abgebildet werden. Derzeit unterstützt der Webservice GND, VIAF, BNF, LC und NDL. Für die Ortsnamen wird „GeoNames“¹⁸ unterstützt.

Die aggregierten Briefverzeichnisse kann man nun nach Korrespondenzpartner (auf Wunsch eingeschränkt auf dessen Rolle als Absender oder Empfänger), nach Schreibort und Datum durchsuchen. Als Ergebnis werden die Kopfdaten der einzelnen Briefe mitsamt bibliografischer Angaben ausgegeben. Briefe aus digitalen Editionen werden zusätzlich direkt verlinkt. Diese Recherchen können zum einen über eine grafische Benutzeroberfläche ausgeführt werden. Zum anderen wurde auch ein Application Programming Interface (API)¹⁹ implementiert, wodurch man den Webservice automatisiert abfragen kann. Das Ergebnis wird dann ebenfalls unter einer CC-BY-Lizenz im beschriebenen Austauschformat ausgegeben, d.h. als TEI-XML, und kann von Programmen zur Anzeige in der eigenen Webapplikation weiter verwendet werden. Mit Hilfe der

¹² <http://www.tei-c.org/Activities/SIG/Correspondence/>

¹³ Antrag: <http://sourceforge.net/p/tei/feature-requests/510/>

¹⁴ Zur Verwendung von Normdaten in Editionen vgl. Stadler, Peter: Normdateien in der Edition. In: editio 26, 2012, S. 174-183.

¹⁵ <http://www.dnb.de/gnd>

¹⁶ Creative Commons Attribution 3.0 Unported: <https://creativecommons.org/licenses/by/3.0/de/>

¹⁷ <http://www.viaf.org>

¹⁸ <http://www.geonames.org/>

¹⁹ <http://correspSearch.bbaw.de/api/tei-xml.xql>

API können zukünftige digitale Editionen daher auch automatisiert auf verwandte Briefe aus anderen Editionsprojekten hinweisen oder direkt auf diese verlinken.

Obwohl schon funktionsfähig und frei zugänglich, befindet sich der Webservice noch in der Aufbauphase. Zum einen werden sowohl das Austauschformat als auch der Funktionsumfang noch weiter entwickelt. Zum anderen fällt der Datenbestand im Moment noch recht klein aus. Derzeit werden digitale Verzeichnisse aus drei Editionen ausgewertet: der Weber-Gesamtausgabe²⁰, der Edition „Briefe und Texte aus dem intellektuellen Berlin um 1800“²¹ sowie dem Soemmerring-Briefwechsel 1792–1805²². Während letzteres ein retrodigitalisiertes Briefverzeichnis einer gedruckt vorliegenden Edition ist, stammen die ersten beiden Verzeichnisse aus digitalen Editionen und wurden direkt aus deren Datenbestand generiert. Mehrere Editionsprojekte haben bereits zugesagt, ein digitales Briefverzeichnis bereit zu stellen, so dass der Datenbestand demnächst weiter wachsen wird.

Trotz der Aufbauphase macht der Webservice schon jetzt deutlich, dass er – mit einer vergrößerten Datenbasis – eine wertvolle Ressource für die weitere Forschung sein kann. So aggregiert er Daten, die ansonsten im jeweiligen Projektkontext verbleiben würden. Im Fall der gedruckt vorliegenden Editionen werden Textinformationen überhaupt erstmalig digital aufbereitet und als Daten der maschinellen Verarbeitung zugänglich gemacht. Zudem beschränkt sich correspSearch weder auf einen thematischen noch auf einen zeitlichen Schwerpunkt, so dass die Daten auch für bisher noch nicht entwickelte Forschungsfragen genutzt werden können.²³ Da correspSearch über eine API verfügt und die digitalen Briefverzeichnisse transparent abgelegt sowie frei nachnutzbar sind, können Forscher den Datenbestand auch mit Technologien abfragen, die neuartig sind oder für die der Webservice selbst keine technische Basis bietet. So wird mit einer ausreichenden Datenmenge und einer entsprechenden Software auch die Erforschung von sozialen Netzwerken möglich sein. Darüber hinaus könnten mit einem weiteren angedachten Ausbau des Webservices auch die thematischen Aspekte eines Netzwerkes untersucht werden: Wie diffundieren Themen und politische oder gesellschaftliche Ereignisse durch persönliche Netzwerke? Welche lokalen Zentren gibt es in Bezug auf bestimmte Fragestellungen? Wie werden veröffentlichte Werke oder Zeitschriftenartikel bewertet und diskutiert? Mit correspSearch wurde nun zumindest der Grundstein dazu gelegt, um diese Fragen eines Tages beantworten zu können.

²⁰ <http://www.weber-gesamtausgabe.de>

²¹ <http://tei.ibi.hu-berlin.de/berliner-intellektuelle/>

²² Dumont, Franz (Hrsg.): Samuel Thomas Soemmerring. Briefwechsel November 1792 – April 1805. Basel 2001 (= Samuel Thomas Soemmerring. Werke, begr. v. Gunter Mann, hrsg. v. Jost Benedum u. Werner Friedrich Kümmel, Bd. 20).

²³ Der Schwerpunkt des derzeitigen Datenbestandes auf der ersten Hälfte des 19. Jahrhunderts ist rein zufällig und der Aufbauphase geschuldet.