

PDB18 - Der deutsche Brief im 18. Jahrhundert. Aufbau einer Datenbasis und eines kooperativen Netzwerks zur Digitalisierung und Erforschung des deutschen Briefes in der Zeit der Aufklärung

Baumann, Baptiste

baptiste.baumann@germanistik.uni-halle.de
Interdisziplinäres Zentrum für die Erforschung der europäischen Aufklärung, Deutschland
ORCID: 0000-0003-2063-6199

Décultot, Elisabeth

elisabeth.decultot@germanistik.uni-halle.de
Interdisziplinäres Zentrum für die Erforschung der europäischen Aufklärung, Deutschland

Dumont, Stefan

dumont@bbaw.de
Berlin-Brandenburgische Akademie der Wissenschaften, Deutschland
ORCID: 0000-0002-6923-0950

Fischer, Katrin

katrin.fischer@tu-darmstadt.de
Universitäts- und Landesbibliothek Darmstadt, Deutschland

Sander, Ruth

ruth.sander@bbaw.de
Berlin-Brandenburgische Akademie der Wissenschaften, Deutschland
ORCID: 0009-0008-3678-9070

Stäcker, Thomas

thomas.staecker@tu-darmstadt.de
Universitäts- und Landesbibliothek Darmstadt, Deutschland
ORCID: 0000-0002-1509-6960

Das Projekt PDB18

Das als Kooperation zwischen dem IZEA, der BBAW und der ULB Darmstadt entwickelte Projekt verfolgt das Ziel, eine repräsentative Datenbasis zu Korrespondenzen des 18. Jahrhunderts im deutschsprachigen Raum zu schaffen, die als digitale und wiederverwendbare Sammlung sowohl von Briefmetadaten als auch von Imagedigitalisaten und Volltexten aus Editionen frei zugänglich gemacht wird. Das Portal www.pdb18.de bietet Zugang zu den bereits bestehenden digitalen Infrastrukturen *correspSearch* (BBAW) und *ZEiD/TUeditions* (ULB Darmstadt), in denen die erschlossenen Metadaten resp. Volltexte recherchierbar und durchsuchbar sind. Das in der dreijährigen Projektlaufzeit zu erarbeitende Korpus¹ schließt ältere wie neuere Druckeditionen, moderne digitale Editionen, sowie Datensätze aus der elektronischen Handschriftendatenbank *Kalliope* ein. Es umfasst Briefe von Verfasserinnen und Verfassern, die zwischen 1640 und 1780 geboren sind, und die im deutschen Sprachraum entstanden oder dorthin verschickt worden sind. Durch die Bereitstellung umfassender und normierter Daten öffnet das Projekt neue Möglichkeiten zur Erforschung edierter Briefe und Korrespondenzen aus diesem Zeit- und Sprachraum anhand von digitalen Methoden (Netzwerkanalyse, Text Mining u.a.).

Workflow und neueste Entwicklungen

Seit 2022 arbeitet das Projektteam an einem Workflow zur Digitalisierung, Volltexterstellung und Metadatenerfassung von Briefeditionen aus dem 18. bis zum 20. Jahrhundert. Digitalisate der gedruckten Editionen werden von der ULB Darmstadt bereitgestellt. Die Text- und Layouterkennung der auf *TUeditions* präsentierten Editionen werden mit der OCR-Software *Transkribus* gewährleistet. Konvertiert werden die Transkripte in TEI-XML, auf der Grundlage des vom Deutschen Textarchiv bereitgestellten Formats. Die Präsentationsoberfläche zeigt Faksimile und Volltext nebeneinander unter Verwendung des Framework *wdbplus*.

Die Briefmetadaten werden nach dem Correspondence Metadata Interchange Format (CMIF) erfasst und in *correspSearch* aggregiert. Während die Erfassung noch weitgehend manuell erfolgt, erkundet das Projekt innovative Methoden für die automatische Extraktion von Briefmetadaten aus Drucktexten. Von großer Bedeutung hierfür sind die neuen, vom Projektteam eigens an edierten Briefen trainierten Field Models in *Transkribus* zur Layoutanalyse und damit zur Abbildung komplexer Strukturen. Auch wird ein Plugin für *Transkribus* zum Computational Double Keying mittels verschiedener Engines entwickelt, das die Qualitätssicherung bei der maschinellen Volltexterstellung und dadurch das spätere Modelltraining erleichtert.

In *correspSearch* lassen sich die Daten durchsuchen und nach verschiedenen Kriterien filtern. Via APIs (TEI-XML,

TEI-JSON, CSV) können die Daten außerdem automatisiert abgefragt und nachgenutzt werden.

CorrespSearch wurde im Rahmen von PDB18 bisher um zwei Funktionen erweitert: so können zum einen vom correspSearch-Team "Datensets" angelegt werden, denen CMIF-Dateien zugewiesen werden. Damit können größere kooperative Projekte wie *PDB18* oder *NorKorr - Norwegian Correspondences*, die Daten in größerem Umfang retrodigitalisieren und/oder zusammentragen, besser dargestellt und eigens durchsucht werden. Zum anderen wurde die neue Filterfunktion "Im Brief (hauptsächlich) verwendete Sprache(n)" entwickelt und implementiert. Derzeit noch in Entwicklung ist *csRegistry*, durch die einerseits für einen Brief auf "Werkebene" eine eindeutige URI vergeben werden kann, andererseits URIs der unterschiedlichen Editionen dieses Briefes damit verknüpft werden können. Dadurch werden verschiedene Editionen desselben Briefes im Suchergebnis sicht- und auch filterbar – was u.a. wichtig für die Netzwerkanalyse ist. Das Mapping selbst kann dabei zwar technisch unterstützt, aber nicht gänzlich automatisiert vorgenommen werden.

Anwendungsmöglichkeiten

Neben den digitalen Voraussetzungen der Datenbasis PDB18 weisen die DH-Komponenten des Erschließungsprojekts auch auf vielfältige künftige Nutzungs- und Anwendungsmöglichkeiten in der geisteswissenschaftlichen Forschung hin. Zum Projektende wird das so erschlossene Korpus durch eine Download-Funktion in seiner Gesamtheit verfügbar gemacht. Die umfassende Sammlung an Briefdaten ermöglicht einerseits großformatige Modellierungen und quantitative Analysen im Bereich der historischen Netzwerkforschung. Anhand von mixed methods-Ansätzen und explorativen Forschungsperspektiven können briefgeschichtliche, prosopographische, kulturhistorische und linguistische Forschungsfragen durch Auswertung des Korpus beleuchtet werden. Die Bereitstellung retrodigitalisierter, durchsuchbarer Volltexte eignet sich andererseits für pointierte textbezogene Untersuchungen zu einzelnen Briefschreiberinnen und -schreibern, zu Briefgenres sowie weiteren Forschungsfragen der Briefforschung, der Geschichte der Editionsphilologie oder der Kanonbildung.

Dabei deckt das Projekt nicht zuletzt Probleme auf, die bei der digitalen Bearbeitung und Analyse eines festen Quellenkorpus von edierten Korrespondenzen auftreten können. Im Fokus der Diskussion steht der kritische Umgang mit dem überlieferten literarischen Kanon als theoretische Voraussetzung jeglicher Datenmodellierung; weitere damit verbundene Desiderate der bisherigen DH-Editorik lassen sich benennen, wie etwa die normative Stoßrichtung von Normdatenbanken (bspw. GND) oder die Identifizierung fehlerhafter Briefdaten.

Das Poster

Ziel des Posters ist, das Kooperationsprojekt PDB18 vorzustellen und alle Schritte des Workflows visuell darzustellen. Besonderer Fokus wird auf die Projektergebnisse gelegt, die mit dem Poster erstmals der weiteren Community präsentiert werden.

Fußnoten

1. Als Bibliographie auf der Projektseite einsehbar unter: <https://www.pdb18.de/korpus>

Bibliographie

Berlin-Brandenburgische Akademie der Wissenschaften (Ed.). 2022. DTABf. Deutsches Textarchiv – Basisformat. <https://www.deutschestextarchiv.de/doku/basisformat/brief.html> (zugegriffen: 23. Juli 2024).

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

Dumont, Stefan, Sascha Grabsch, Jonas Müller-Laackman, Ruth Sander und Steven Sobkowski (Hg.). 2024. „correspSearch – Briefeditionen vernetzen (3.0.0)“ [Webservice]. Berlin-Brandenburgische Akademie der Wissenschaften. <https://correspSearch.net> (zugegriffen: 23. Juli 2024).

Fischer, Katrin, Silke Kalmer, Dario Kampkaspar, Sophie Müller, Marc Scheffer, Melanie E.-H. Seltmann und Kevin Wunsch. 2022. „From facsimile to online representation. The Centre for Digital Editions in Darmstadt. An Introduction“ <https://doi.org/10.5281/zenodo.7113193>

Kampkaspar, Dario. 2018. W. Digitale Bibliothek (wdbplus). [Software]. <https://github.com/dariok/wdbplus>

„Portal Der deutsche Brief im 18. Jahrhundert“. <https://www.pdb18.de/> (zugegriffen: 23. Juli 2024).

TEI Correspondence SIG. 2018. Correspondence Metadata Interchange Format (CMIF). <https://github.com/TEI-Correspondence-SIG/CMIF>

Universitäts- und Landesbibliothek Darmstadt. 2024. „Zentrum für digitale Editionen in Darmstadt (ZEiD)“. https://www.ulb.tu-darmstadt.de/die_bibliothek/einrichtungen/zeid/index.de.jsp (zugegriffen: 23. Juli 2024).