

Opening Sources – modulare Wege zur Quellenbereitstellung und -edition

Burckhardt, Daniel

burckhardt@ghi-dc.org
Deutsches Historisches Institut Washington

Hörnschemeyer, Jörg

hoernschemeyer@dhi-roma.it
Deutsches Historisches Institut in Rom

König, Mareike

MKoenig@dhi-paris.fr
Deutsches Historisches Institut Paris

Schulz, Julian

Schulz@MaxWeberStiftung.de
Geschäftsstelle der Max Weber Stiftung, Bonn

Grallert, Till

till.grallert@hu-berlin.de
Humboldt-Universität zu Berlin

Keck, Jana

keck@ghi-dc.org
Deutsches Historisches Institut Washington

Hintergrund

Das Erstellen von digitalen Quelleneditionen gehört mit zu den etabliertesten und einflussreichsten Praktiken und Methoden im Bereich der Humanities Computing bzw. Digital Humanities (Burnard 2014). Die Vielzahl praktischer Erfahrungen und eine fundierte theoretische Reflexion führten zu einer Konvergenz, die sich sowohl an der Stabilität und weiten Verbreitung technischer Standards wie der TEI, in der zunehmenden Akzeptanz von Kriterien für die Bewertung digitaler Editionen (z.B. Sahle 2014) und einem breiten Konsens zu den methodischen, organisatorischen und sozialen Rahmenbedingungen ihrer Erstellung (Fritze et al. 2022) ablesen lässt.

Neben dieser Konvergenz zeigt sich in der Praxis allerdings ein sehr breites Spektrum digitaler Bereitstellungsarten, von Formaten und Zugängen zu historischen Quellen, die wir im Panel „Opening Sources“ adressieren möchten. Im Sinne der Open Humanities steht dabei der offene Zugang zu einem bestimmten Quellenbestand als Hauptziel im Zentrum des editorischen Tuns. Anhand von vier exemplarischen editorischen Projekten,

angesiedelt an den Instituten der Max Weber Stiftung (MWS), soll diese Vielfalt von Formen und Zugängen verdeutlicht und als Ausgangspunkt für eine umfassendere Diskussion genutzt werden.

Die Mehrheit dieser Projekte erfüllt die Kriterien an eine wissenschaftliche Edition gemäß des erwähnten Kriterienkataloges nicht oder nicht ausreichend.¹ Auch fallen sie nicht unbedingt unter den Begriff born digital editions, den Patrick Sahle dahingehend definiert, dass solche Projekte nicht ohne Verlust an wesentlichen Informationen und Funktionen in eine herkömmliche Papierform übertragen werden können (2013, Bd. 2, 149). Dabei teilen die Projektbeteiligten diese Kriterien durchaus. Jedoch erlauben die vorhandenen organisatorischen, finanziellen oder personellen Ressourcen nur Teilschritte der Quellenpublikation und -edition. Dank der Konvergenz von Formaten für Metadaten (MODS/teiHeader), Digitalisate (TIFF/PNG/JPEG), Transkripte und kritische Apparate (TEI) sowie deren Verschränkung (METS, IIIF) handelt es sich – so die erste These des Panels – beim datenzentrierten Ansatz, der den verschiedenen digitalen Zugangsformen zugrunde liegt, weniger um klar voneinander abzugrenzende Genres, sondern um ein gestuftes Editionsmodell, dessen Modularität wir ausgehend von vier kurzen Input-Referaten zur Diskussion stellen. Denn – so die zweite These des Panels – sehr viele Editionsprojekte dürften im Spannungsfeld von Anspruch und Wirklichkeit vor ähnlichen Herausforderungen stehen, sodass ein gemeinsames Nachdenken über Lösungsansätze lohnend erscheint (vgl. Sayers 2017).

Leitfragen

Die folgenden Punkte stehen im Zentrum der Input-Referate und sollen in der Diskussion im Panel und mit den Teilnehmer_innen vertieft werden:

• Linearität des Editionsprozesses

Wie weit deckt sich das informationswissenschaftliche Paradigma eines linearen Vorgangs von der Erfassung der Metadaten und Bereitstellung des Digitalisats über die Erstellung des Transkripts und dessen Annotation mit der Praxis des Edierens in den Geisteswissenschaften? Oberbichler et al. (2021) zeigen am Beispiel von Zeitungsarchiven, dass eine forschungsgeleitete digitale Hermeneutik immer wieder die prozessuale Linearität durchbricht. Nicht allein die Auswahl der zu edierenden Quellen, auch deren Datierung und Transkription setzt häufig Wissen voraus, das erst beim *close-reading* bei der Quellenannotation entsteht. Ein zeitlich gestrecktes oder institutionell verteiltes Verfahren droht den Rückfluss neuer Erkenntnisse in einen vorhergehenden Arbeitsschritt zu erschweren, so dass auch bei digitalen Editionen eine Abkehr vom linearen Wasserfall-Modell hin zu agilen, iterativen Projekten zu beobachten ist (Ferraro et al. 2018). Und würden solche Teilschritte überhaupt von Forschungsförderern Finanzierung erhalten?

An welchen Stellen lässt sich der Publikationsprozess unterbrechen und zu einem späteren Zeitpunkt weiterführen? Was sind Mindestanforderungen im Sinne der Open Humanities für dieses Vorgehen (freie Li-

zenzen, breit akzeptierte Standards, offene Schnittstellen, lückenlose Dokumentation)? Sind das bloß notwendige, oder auch hinreichende Bedingungen? Welche Parameter müssen von vornherein festgelegt werden (z.B. Browsing-, Such- und Auswertungsmöglichkeiten), welche können nachträglich ohne Nacharbeiten an Metadaten und Auszeichnung eingebracht werden?

- **Vom Digitalisat zum Text**

(Putnam 2016) hat die fundamentale Bedeutung der Findbarkeit von (Voll-)Texten für den digitalen Wandel in den Geisteswissenschaften herausgearbeitet. „Digitale Suche bietet eine vermittlungsfreie Entdeckung“. Deshalb ist die einfache Durchsuchbarkeit für eine breite wissenschaftliche Rezeption eines Quellenbestandes meist wichtiger als dessen minutiöse Annotation. Der Durchbruch bei der automatisierten Handschriftenerkennung (HTR) in den letzten Jahren ermöglicht eine Automatisierung, die zuvor Druckerzeugnissen in lateinischen Zeichensätzen vorbehalten war. Da je nach Schreibhand passende Trainingsdaten die Voraussetzung sind, kann sich hier die Verschränkung mit Citizen-Science-Ansätzen zum Crowd-Sourcing anbieten.

- **Vom Text zur Edition**

Die zentrale Frage beim Übergang zur Edition ist die Frage des *Edendum*. Soll und kann der Gesamtbestand an Transkripten inhaltlich annotiert und umfassend kommentiert werden? Falls nicht, sind Kriterien wie „Repräsentativität“, die „Vollständigkeit in Teilen“ beispielsweise für bestimmte Textgattungen oder Zeiträume oder gar „Popularität“ (häufig zitiert oder im Textarchiv oft aufgerufen) zielführend? Ist es auch in diesem Schritt hilfreich, automatisierte Verfahren oder Bürgerwissenschaftler_innen zur Annotation und Verlinkung mit Normdaten in Erwägung zu ziehen?

- **Verlässlichkeit**

Eine schrittweise Edition ist keineswegs eine „Edition light“. Die Streckung des Prozesses kann nur gelingen, wenn jeder einzelne Schritt passende Qualitätssicherungsverfahren und die dauerhafte Sicherung der Zwischenergebnisse gewährleistet. Im Vordergrund stehen hier die FAIR-Prinzipien, die vor allem die Zugänglichkeit und die spätere Weiternutzung gewährleisten. Auch bei Quellenbeständen aus nicht-indigenen Kontexten, sollten zusätzliche Aspekte wie *Collective Benefit*, *Responsibility* und *Ethics* aus den CARE-Prinzipien berücksichtigt werden.

Die Dauerhaftigkeit bezieht sich auf die langfristige Sicherung der Forschungsdaten wie auf die einfache Zugänglichkeit in einer niedrigschwellig verfügbaren Präsentationsumgebung (vgl. Fritze et al. 2022, Abschnitt 12).

Format

Nach der Einführung durch die Organisator_innen (5 Min.) wird jede_r Panelist_in das eigene Projekt mit Bezug auf diese Leitfragen kurz vorstellen (je 5 Minuten). In einem zweiten Teil suchen wir Antwortvorschläge, zunächst in einer Diskussion unter den Panelist_innen (30

Min.) und anschließend gemeinsam mit den anwesenden Kolleg_innen (30 Min.). Die Publikation der Ergebnisse in einem Blogbeitrag ist geplant.

Statements

Einleitung und Moderation

Daniel Burckhardt (Deutsches Historisches Institut Washington) und Julian Schulz (Geschäftsstelle der Max Weber Stiftung, Bonn)

Wenn Technik und Bewusstsein voranschreiten: die drei Leben der Korrespondenz der Constance de Salm (1767-1845)

Mareike König (Deutsches Historisches Institut Paris)
Bisweilen verhindern zögerliche Projektpartner die freie Bereitstellung von Digitalisaten, begrenzen die Mittel das Maximalziel bei Digitalisierung und Erschließung von Quellen und eröffnet die technische Entwicklung neue Möglichkeiten. So geschehen beim Projekt der Digitalisierung und Inventarisierung der Korrespondenz von Constance de Salm. Zu Projektbeginn war weder die besitzende Institution bereit, die Digitalisate ohne Anmeldung und ohne Wasserzeichen online zu stellen, noch gab es die Mittel für eine zweisprachige Erschließung geschweige denn für eine vollständige Transkription der Briefe. Der Impulsbeitrag zeichnet exemplarisch nach, wie dennoch in den oben genannten Einzelstufen – vom Digitalisat zum Text und vom Text zur Edition – trotz Pausen die Korrespondenz der Constance de Salm schrittweise ediert werden konnte: durch einen langen Atem beim Verhandeln mit Partnern, Vernetzung mit ähnlichen Projekten, Nachnutzung von Workflows und Verfolgen kleiner Etappenziele.

Quellen ohne Ressourcen: Ansätze des minimal computing für die Erschließung arabischer Periodika

Till Grallert (Humboldt-Universität zu Berlin, ehem. Orient-Institut Beirut)

„Open Arabic Periodical Editions“ ist ein Projekt, das die Rahmenbedingungen für die digitale Erschließung von Quellen in und aus Gesellschaften des Globalen Südens mit den Ansätzen des *minimal computing* in der Praxis adressiert (Gil und Ortega 2016, Risam 2019). Hierbei geht es um mehrere, miteinander verwobene Schichten der Unzugänglichkeit, die DH-Projekte des Globalen Südens konstant verhandeln müssen (Grallert 2022): Hegemonie des anglophonen Globalen Nordens über die Infrastrukturen der Wissensproduktion (Wissensorganisation, Forschungsfelder, Fördergelder, Arbeitssprachen, vgl. Fiorimonte 2021); computationale Methoden und Werkzeuge der digitalen Editorik, die für nicht-lateinische Schriften und Sprachen des Globalen Südens nicht erprobt oder verfügbar sind (z.B. OCR, NER, vgl. Auddy 2022); begrenzter Zugang zur technischen Grundversorgung mit Strom, Internet und Hardware.

The Wisdom of the Crowd: Quellen öffnen mit Citizen Scholars

Jana Keck (Deutsches Historisches Institut Washington)

„Migrant Connections“ ist eine digitale Forschungsinfrastruktur (Omeka S) des DHI Washington, welche Zugang zu diversen Quellenmaterialien wie Briefen, Tagebüchern, und Zeitungsartikeln zur deutschen Migration in die USA bietet. Gemeinsam mit Citizen Scholars werden kontinuierlich Quellen gesammelt, digitalisiert, transkribiert, übersetzt, Metadaten erstellt und kontextualisiert. Durch die Expertise der Citizens aus Deutschland und den USA konnten bisher hunderte von Quellen erschlossen werden, die Einblicke vor allem in die Alltagsgeschichte von bisher wenig beachteten, „gewöhnlichen“ historischen Akteuren geben. Der Beitrag möchte die Potentiale einer schrittweisen Edition bei offener Kollaboration mit Citizen Scholars im Forschungsprozess beleuchten, dabei die Frage der Qualitätssicherung thematisieren und gleichzeitig hervorheben, wie diese Offenheit uns als Wissenschaftler_innen dazu bringt, Vertrauen zu schenken und Kontrolle abzugeben.

„Königsweg“ Digitale Edition. Offen, aber für wie lange?

Jörg Hörnschemeyer (Deutsches Historisches Institut Rom)

Der Impulsbeitrag geht der Frage nach, wie offen und nachhaltig die „letzte Stufe“ des eingangs skizzierten Modells der Bereitstellung digitaler Quellen, die Digitale Edition, wirklich ist. „Ferdinand Gregorovius – Poesie und Wissenschaft. Gesammelte deutsche und italienische Briefe“ ist eine nach „allen Regeln der Kunst“ erarbeitete Digitale Edition, die über XML/TEI-annotierte, text- und sachkritisch erschlossene Brieftranskriptionen verfügt und mit umfangreichen Normdatenverknüpfungen, einem projektspezifisch entwickelten User Interface und einer REST-API versehen wurde. Trotz (oder auch gerade wegen?) dieser komplexen Architektur ist absehbar, dass die Edition in 10-20 Jahren vor der zentralen Frage stehen wird, ob und wie ein offener Zugang über die verschiedenen Zugänge auch weiterhin gewährleistet werden kann und wie neue Forschungsergebnisse in eine als abgeschlossen angesehene Edition einfließen können. Könnten z.B. Semantic Web-Technologien (Wettlaufer 2018), Social Editing (Crompton et al. 2016) und kollaborative, (Fach)community-getriebene Ansätze Antworten auf diese beiden Fragebereiche liefern? Und welche Ressourcen benötigt dann eine Langzeitstrategie für Editionen auf institutioneller Ebene?

Fußnoten

1. Vgl. IDE-Richtlinien, 1. Geltungsbereich, Definitorik, wonach Digitale Editionen „nicht nur in digitaler Form publiziert [werden], sondern ... in ihrer Methodologie einem digitalen Paradigma [folgen]“.

Bibliographie

Auddy, Purbasha. 2022. „Mining Verbal Data from Early Bengali Newspapers and Magazines: Contemplating the Possibilities.“ In *Global Debates in the Digital Humanities*, hg. von Domenico Fiormonte, Sukanta Chaudhuri, und Paul Ricautre, 117-26. Debates in the Digital Humanities. Minneapolis: University of Minnesota Press.

Burnard, Lou. 2014. *What is the Text Encoding Initiative? How to add intelligent markup to digital resources*. Marseille: OpenEdition Press. DOI: <https://doi.org/10.4000/books.oep.426>.

Crompton, Constance, Daniel Powell, Alyssa Arbuckle, Ray Siemens, Maggie Shirley und Devonshire Manuscript Editorial Group. 2016. „Building a Social Edition of the Devonshire Manuscript“ In *Digital Scholarly Editing: Theory and Practice*, hg. von Matthew James Driscoll und Elena Pierazzo, 19-39. DOI: <https://doi.org/10.11647/OBP.0095>.

Ferraro, Ginestra und Anna-Maria Sichani. 2018. „Design as Part of the Plan: Introducing Agile Methodology in Digital Editing Projects“ In *Digital Scholarly Editions as Interfaces. Schriften des Instituts für Dokumentologie und Editorik* 12, hg. von Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber und Gerlinde Schneider, 83-105. Norderstedt. <https://kups.uni-koeln.de/9113/>; urn:nbn:de:hbz:38-91132 (zugegriffen: 3. August 2022).

Fiormonte, Domenico. 2021. „Taxation against Overrepresentation? The Consequences of Monolingualism for Digital Humanities.“ In *Alternative Historiographies of the Digital Humanities*, hg. von Dorothy Kim und Adeline Koh, 333-76. Earth: punctum books. DOI: <https://doi.org/10.53288/0274.1.00>.

Fritze, Christiane et al. 2022. „Manifest für digitale Editionen“. <https://dhd-blog.org/?p=17563> (zugegriffen: 3. August 2022).

Gil, Alex und Élika Ortega. 2016. „Global Outlooks in Digital Humanities: Multilingual Practices and Minimal Computing.“ In *Doing Digital Humanities: Practice, Training, Research*, hg. von Constance Crompton, Richard J Lane und Ray Siemens, 22-34. Abingdon: Routledge.

Grallert, Till. 2022. „Open Arabic Periodical Editions: A Framework for Bootstrapped Scholarly Editions Outside the Global North.“ In *Digital Humanities Quarterly* 16, Bd. 2, "Minimal Computing" hg. von Roopika Risam und Alex Gil. <http://digitalhumanities.org/dhq/vol/16/2/000593/000593.html> (zugegriffen: 3. August 2022).

Oberbichler, Sarah, Emanuela Boros, Antoine Doucet, Jani Marjanen, Eva Pfanzer, Juha Rautiainen, Hannu Toivonen und Mikko Tolonen. 2021. „Integrated interdisciplinary workflows for research on historical newspapers: Perspectives from humanities scholars, computer scientists, and librarians.“ *Journal of the Association for Information Science and Technology* 73, Issue 2, 225-239. DOI: <https://doi.org/10.1002/asi.24565>.

Putnam, Lara. 2016. „The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast“ *The American Historical Review* 121, 377-402. DOI: <https://doi.org/10.1093/ahr/121.2.377>.

Risam, Roopika. 2019. *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy*. Evanston: Northwestern University Press. <https://doi.org/10.2307/j.ctv7tq4hg>.

Sahle, Patrick. 2013. *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 2: Befunde, Theorie und Methodik*. Norderstedt: BoD. <https://kups.ub.uni-koeln.de/5352/urn:nbn:de:hbz:38-53523> (zugegriffen: 3. August 2022).

Sahle, Patrick. 2014. *Kriterienkatalog für die Besprechung digitaler Editionen*. 1.1. Weitere Schriften. Institut für Dokumentologie und Editorik. <https://www.i-d-e.de/publikationen/weitereschriften/kriterien-version-1-1/> (zugegriffen: 3. August 2022).

Sayers, Jentery, Hrsg. 2017. *Making Things and Drawing Boundaries: Experiments in the Digital Humanities*. Debates in the Digital Humanities 3. Minneapolis: University of Minnesota Press.

Wettlaufer, Jörg. 2018. "Der nächste Schritt? Semantic Web und digitale Editionen" In *Digitale Metamorphose: Digital Humanities und Editionswissenschaft. (Sonderband der Zeitschrift für digitale Geisteswissenschaften 2)*, hg. von Roland S. Kamzelak und Timo Steyer. DOI: https://doi.org/10.17175/sb002_007.