

Nachhaltigkeit technischer Lösungen für digitale Editionen. Eine kritische Evaluation bestehender Frameworks und Workflows von und für Praktiker_innen

Andorfer, Peter

peter.andorfer@oeaw.ac.at
ACDH, Österreich

Durco, Matej

matej.durco@oeaw.ac.at
ACDH, Österreich

Stäcker, Thomas

staecker@hab.de
HAB, Deutschland

Thomas, Christian

thomas@bbaw.de
BBAW, Deutschland

Hildenbrandt, Vera

hildenbr@uni-trier.de
TCDH, Deutschland

Stigler, Hubert

hubert.stigler@uni-graz.at
ZIM Uni Graz, Österreich

Söring, Sibylle

soering@sub.uni-goettingen.de
SUB Göttingen, Deutschland

Rosenthaler, Lukas

lukas.rosenthaler@unibas.ch
DH Lab, Schweiz

Stand der Dinge

Digitale Editionen machen in den Digital Humanities das „Brot- und Buttergeschäft“ aus. Doch während sich der methodisch-theoretische Hintergrund digitaler Editionen zusehends konsolidiert und sich diese neue Form der Publikation von Forschungsergebnissen im

(fach)wissenschaftlichen Diskurs bereits etabliert hat, fehlt es nach wie vor an umfassend dokumentierten und selbstkritisch reflektierten Best-Practice-Beispielen von Frameworks und Workflows zur Erstellung und / oder Publikation von digitalen Editionen, welche als Blaupausen für künftige digitale Editionsprojekte herangezogen werden können. Das Resultat ist so bekannt wie unerfreulich und kann – nur geringfügig überspitzt – auf folgende Formel gebracht werden: So gut wie jedes Projekt erfindet das Rad – das technische Grundgerüst der Edition – wieder neu.

Die wichtigsten Gründe für diese Entwicklung lassen sich rasch benennen:

- Digitale Editionen sind nach wie vor eine sehr junge Publikationsform und darüber hinaus abhängig von der raschen Weiterentwicklung gegenwärtiger Webtechnologien und -standards.
- Digitale Editionen erfordern technische Kompetenzen, welche jene traditioneller Geisteswissenschaftler_innen meist übersteigen, weshalb Kooperationen mit Entwickler_innen notwendig sind.
- Digitale Editionen werden häufig in Form von Einzelprojekten realisiert, weshalb nicht auf bestehende Lösungsansätze anderer Institutionen zurückgegriffen wird.
- In so gut wie allen Fällen fehlen die Ressourcen, manchmal wohl leider auch der Wille, die Eigenentwicklung, das geschaffene technische Grundgerüst hinreichend gut zu dokumentieren und in einer Form zu veröffentlichen, so dass andere den Quellcode, Workflows, Stylesheets etc. nachnutzen zu können.
- Digitale Editionen sind – wenigstens in der Selbstwahrnehmung – häufig hochgradig speziell und unterscheiden sich in Inhalt, Form und Funktion von allen bereits bestehenden Editionsprojekten, womit eine völlige Neuentwicklung des technischen Grundgerüsts gerechtfertigt wird.

Ablauf:

Im Rahmen des Panels sollen einige der aktivsten Institutionen aus dem Bereich der digitalen Editionen an einen Tisch gebracht werden. Diese erhalten im Vorfeld der Tagung einen Fragebogen zur Vorbereitung einer kurzen (pro Teilnehmer ca. fünf Minuten) Vorstellung ihrer Systeme, wobei darin der Fokus auf dem Thema Reusability der in den Projekten verwendeten Technologien und Workflows liegen sollte. Konkret sollen die Teilnehmer_innen des Panels auf folgende Punkte eingehen:

- Kurze Vorstellung der eigenen Frameworks und Workflows, vor allem hinsichtlich einer Einschätzung über die Stärken und Schwächen der eigenen Lösungsansätze, aus welcher Tradition/Disziplin (z. B.

Philologie, Geschichtswissenschaften) sie kommen und welche konkreten Projekte damit realisiert wurden.

- Gibt es ein weitgehend standardisiertes Prozedere im Falle von Kooperations- bzw. Nachnutzungsanfragen (inklusive der dafür notwendigen Ressourcen)?
- Wer soll die Angebote nutzen; gibt es fachlich, institutionell, qualitativ, budgetär, regional, national, zeitlich oder anderweitig konstituierte Zielgruppen?

Ein Ziel dieser Vorstellungsrunde soll es sein, potenziell interessierten Nutzer_innen im Auditorium einen kompakten Überblick über bestehende Angebote zur Erstellung und / oder Veröffentlichung von digitalen Editionen zu vermitteln.

Podiumsdiskussion (ca. 30 Minuten)

Im Anschluss an diese Kurzvorstellung erfolgt eine moderierte Podiumsdiskussion, worin folgende Punkte weiter thematisiert werden:

- Dokumentation von Technologie und Workflows: Wie gut sind die technischen Aspekte dokumentiert? Ist es für Dritte möglich, anhand dieser Dokumente ähnliche Projekte zu realisieren? Welche technische Infrastruktur ist dafür notwendig?
- Veröffentlichung von Code: Ist der für das System geschriebene oder adaptierte Code für andere nachnutzbar veröffentlicht (z. B. auf GitHub)?
- Wird das entwickelte Framework auch als Service angeboten?
- Ressourcen und Organisation der Entwicklung: Wie groß war / ist der Aufwand der Entwicklung des technischen Grundgerüsts, eventuell in Personenmonaten. Woher stammt das verwendete Know-How (Eigenentwicklung oder Adaption bestehender Konzepte)?
- Wie generisch ist das verwendete technische Framework bzw. wie aufwändig sind die Änderungen, die bei der Adaption an ein neues Projekt notwendig werden? Sprich: Wie leicht und wie weit kann das System an projektspezifische Bedürfnisse angepasst werden (unterstützte Datenformate, Funktionalitäten)? Welche Kompetenzen sind notwendig, um Anpassungen auf unterschiedlichen Ebenen (Anzeige, Projektstruktur usw.) vorzunehmen?
- Wie groß ist der laufende Aufwand für Wartung einzelner Projekte bzw. des Frameworks an sich?

Es sollen gemeinsame Problemfelder identifiziert und reflektiert werden. Auf dieser Basis kann dann über mögliche (gemeinsame) Lösungen diskutiert werden.

Publikumsdiskussion (ca. 30 Minuten)

Im letzten Drittel des Panels wird die Diskussion zum Publikum hin geöffnet werden. Dabei sollen vor allem potentielle Nutzer_innen die Möglichkeit bekommen, gezielt konkrete und ggf. eigene Projekte betreffend Fragen zu stellen und direkt mit möglicherweise zukünftigen Projektpartnern ins Gespräch zu kommen.

Teilnehmer

Bei der Auswahl der Teilnehmer wurde einerseits darauf geachtet, vornehmlich etablierte Institutionen anzusprechen, die sich als Dienstleister im Bereich digitaler Editionen profiliert haben, deshalb an möglichst generischen Lösungen zur Erstellung und Publikationen von digitalen Editionen interessiert sind und dafür selbst Frameworks und Workflows entwickelt haben. Außerdem wurde versucht, bei der Auswahl der Teilnehmer möglichst den gesamten deutschsprachigen Raum abzudecken.

ACDH-ÖAW

Matej Durco und Peter Andorfer

Das ACDH verwendet ein eXistdb-basiertes Framework zur Veröffentlichung digitaler Editionen namens cr-xq-mets. cr-xq-mets basiert lose auf SADE. Die Idee von cr-xq-mets ist die konsequente Trennung von Code und einzelner Projekt, mit dem Ziel einen hohen Grad an generischer Projektentwicklung bei gleichzeitig geringem Aufwand an Projektmaintainance zu erreichen.

Das ACDH übernimmt auch die Organisation und Moderation des Panels.

Herzog August Bibliothek

Thomas Stäcker

Die Herzog August Bibliothek (HAB) hat für ihre digitalen Editionsprojekte die Reihe Editiones Electronicae Guelferbytanæ gegründet (<http://opac.lbs-braunschweig.gbv.de/DB=2/SET=1/TTL=16/FAM?PPN=580258351>), in der bisher 20 Werke erschienen sind. Hinzu kommen zahlreiche kleinere und umfangreichere Editionen, die außerhalb dieser Reihe erschienen sind. Hervorzuheben sind die beiden im Langzeitförderprogramm der DFG (je 12 Jahre) erscheinenden Editionen der Kritische Gesamtausgabe der Schriften und Briefe Andreas Bodensteins von Karlstadt und die Digitale Edition und Kommentierung der Tagebücher des Fürsten Christian II. von Anhalt-Bernburg (1599-1656). Die Kodierung erfolgt in TEI (nach den übergreifenden Festlegungen der HAB). Die Anzeige der

Webdarstellung und Suchfunktionalitäten basieren auf PHP sowie eXistdb.

TextGrid

Sibylle Söring

TextGrid ist eine Virtuelle Forschungsumgebung für die text- und quellenbasierten Geisteswissenschaften, die u. a. die Erstellung digitaler Editionen mithilfe Open Source-basierter Tools und Dienste unterstützt. Neben der Software, dem TextGrid Laboratory, bietet TextGrid mit dem TextGrid Repository die Möglichkeit, vielfältige Forschungsdaten - u. a. XML / TEI-kodierte Texte, Bilder und Datenbanken - langfristig zu speichern sowie nach internationalen Standardformaten zitierbar zu publizieren und zur Nachnutzung zur Verfügung zu stellen, wie etwa zur Recherche und Visualisierung. Mit TextGrid wurden und werden verschiedene Editionsprojekte umgesetzt, so u. a. die Digitale genetisch-kritische Edition von Theodor Fontanes Notizbüchern und die Bibliothek der Neologie .

BBAW – DTA, CLARIN-D

Christian Thomas

Das DFG-geförderte Projekt Deutsches Textarchiv der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) erstellt bzw. publiziert digitale Editionen von derzeit (September 2015) 1665 Werken im DTA-Kernkorpus unter Verwendung eines selbst entwickelten, auf Open-Source-Software basierenden Frameworks. Zu diesem Framework gehört mit dem Modul (DTA-Erweiterungen) ein elaborierter Workflow zur Integration hochwertiger Textressourcen aus externen Editions- und Forschungsprojekten. Über DTAE konnten als Ergänzungen des DTA-Kernkorpus bislang weitere 1097 Einzelwerke sowie der gesamte Bestand zweier herausragender Zeitschriften des 19./20. Jahrhunderts, das von J. G. Dingler begründete *Polytechnische Journal* (346 Bände, 1820–1931) und die Zeitschrift *Die Grenzboten* (270 Bände, 1841–1922) in das DTA integriert werden. Der Textbestand aus DTA und DTAE umfasst ca. 200 Mio. Tokens und ca. 1,2 Mrd. Zeichen. Die DTA-Korpora sind einheitlich gemäß dem TEI-basierten und ausführlich dokumentierten DTA-Basisformat (DTABf) kodiert und werden mit Hilfe computerlinguistischer Werkzeuge automatisch annotiert, was unter anderem spezialisierte Suchanfragen nach bestimmten Metadatenfeldern, Wortarten, grammatischen Kategorien, X-Pfaden etc. ermöglicht. Zudem wird der historische Textbestand automatisch in Richtung moderner Orthographie 'normalisiert', was schreibweisentolerante Suchen über das gesamte Korpus ermöglicht (siehe allgemein zur Suche im DTA www.deutschestextarchiv.de/doku/DDC-suche_hilfe). Die Qualitätssicherung sämtlicher Korpusressourcen geschieht kollaborativ in der webbasierten Umgebung DTAQ , in der derzeit

866 registrierte Nutzer mögliche Fehler auf der Text-, Annotations- und Metadatenebene melden und, je nach Rechtstatus, auch direkt online korrigieren können.

Trier Center for Digital Humanities

Thomas Burch, Vera Hildenbrandt

Das TCDH kann inzwischen auf eine mehr als langjährige Erfahrung in der Planung, Durchführung und Betreuung von Projekten im Bereich der Digital Humanities verweisen. Neben dem Schwerpunkt im Bereich der Erstellung und Publikation digitaler Editionen verfügt das Zentrum über eine ausgewiesene Expertise in der Entwicklung von Software-Umgebungen für geisteswissenschaftliche Großvorhaben. In mehreren von der DFG, dem BMBF sowie im Rahmen des Akademienprogramms geförderten Projekten entstanden und entstehen am TCDH digitale Editionen wie z. B. das Heinrich-Heine-Portal , das Christian-Dietrich-Grabbe-Portal , das Cusanus-Portal , die elektronische Publikation der Korrespondenz August-Wilhelm Schlegels sowie die digitale Rekonstruktion der Textgenese und Entstehungsgeschichte von Wolfgang Koeppens ‚Jugend‘. Im Bereich der Softwareentwicklung konzipiert, betreut und entwickelt das Team des TCDH virtuelle Arbeitsumgebungen für Projekte mit hohen Anforderungen an Workflow und grafische Benutzerschnittstellen. Hervorgehoben seien hier Systeme wie das internetbasierte Artikelredaktionssystem für die Produktion und Publikation von Wörterbüchern in dezentralen Arbeitsstellen, das gemeinsam mit dem Forschungszentrum Europa und dem Sonderforschungsbereich 600 entwickelte "Forschungsnetzwerk und Datenbanksystem" (FuD 2015), die Redaktions- und Publikationsplattform zur Europäischen Geschichte Online oder das interaktive Werkzeug "Transcribo" zur Erstellung von Transkriptionen.

ZIM – ACDH Graz

Hubert Stigler

Das ZIM hat im Rahmen einer Vielzahl von Editionsprojekten forschungsgetrieben ein objektorientiertes Framework auf Basis von FEDORA Commons und weiteren Open-Source-Projekten (Apache Cocoon, Blazegraph u. a.) entwickelt, das Aspekte der Publikation von Digitalen Editionen mit jenen der Langzeitarchivierung von Forschungsdaten zu verbinden sucht. Als österreichischer Beitrag zu DARIAH steht es einer breiten Öffentlichkeit zur Nachnutzung zur Verfügung.

DH Lab Basel

Lukas Rosenthaler

Während die anderen Teilnehmer vor allem an Lösungen für XML / TEI-basierte digitale Editionen arbeiten, legt SALSAH (System for Annotation and Linkage of Sources in Arts and Humanities) den Schwerpunkt (zurzeit noch) auf die Verknüpfung und Verlinkung von vornehmlich digitalen Faksimiles.