

DHd 2019 Book of Abstracts Hackathon

Andorfer, Peter

peter.andorfer@oeaw.ac.at

Österreichische Akademie der Wissenschaften, Österreich

Cremer, Fabian

Cremer@MaxWeberStiftung.de

Max Weber Stiftung, Bonn

Steyer, Timo

steyer@hab.de

Forschungsverbund Marbach Weimar Wolfenbüttel /
Herzog August Bibliothek Wolfenbüttel

Einleitung

„Auf der Konferenz [DHd 2018; d. Verf.] war zu jeder Zeit an jedem Ort genug Kompetenz versammelt, um in einer Kaffeepause eine digitale Publikation der Abstracts zu bauen, mit Inhaltsverzeichnis, Volltext- und Schlagwortsuche und anderen netten Features, die sonst zum Standard jeder Webpräsentation in den DH gehören.“ So postuliert Fabian Cremer in seinem Blog-Post „Nun sag, wie hältst Du es mit dem Digitalen Publizieren, Digital Humanities?“ (Cremer 2018). Nur, wie Cremer weiter ausführt, niemand habe es gemacht oder irrt der Autor in dieser Aussage, da die Aufgabe nicht so trivial ist, wie auf dem ersten Blick erscheint? Der hier vorgeschlagene Workshop „DHd 2019 Book of Abstracts Hackathon“ soll der DHd-Community den Raum bieten, dieser Frage nachzugehen, eine gemeinsame digitale Publikation der Konferenz-Abstracts zu realisieren und so einen Diskussionsimpuls zur Zukunft des Digitalen Publizierens in den Digital Humanities zu geben.

Digitales Publizieren und Digital Humanities

Das digitale Publizieren hat sich im Rahmen des Kanonisierungsprozesses zu einem etablierten Bestandteil der Digital Humanities entwickelt. Dies umfasst sowohl methodische Überlegungen (Kohle 2017) wie auch die praktische Umsetzungen (DHd-AG 2016). Längst existieren etablierte digitale Publikationsorgane der Digital Humanities, welche die Vielfalt und Potentiale der digitalen Publikationsformate demonstrieren und als Vorbild der Publikationspraxis und Wissenschaftskommunikation gelten.¹ Diesen Entwicklungen zum Trotz werden in der Breite und Spitze der Digital Humanities Forschung

diese Potentiale nicht ausgeschöpft und traditionelle Publikationspraktiken weiterhin gepflegt (Stäcker 2012). Dies gilt für die Ebene der Technologie, so liegt das Book of Abstracts der DHd 2018 als PDF ohne Strukturdaten vor (Vogeler 2018). Aber gleichfalls auch für die offene Zugänglichkeit, so wurde das deutschsprachige Standardwerk zu den DH nicht als Open Access publiziert (Jannidis et al. 2017 vgl. dazu Stäcker 2017). Die Anforderungen an digitale Publikationen sind schon seit längerem formuliert. „Digital publishing is not simply repackaging a book or article as a computer file, although even a searchable pdf has advantages over paper“, bemerkt Borgman in ihrem „Call to Action for the Humanities“ und adressiert hier auch dezidiert die Digital Humanities (Borgman 2010: #p16).

Neben den wissenschaftsökonomischen und wissenschaftspolitischen Vorteilen digitalen Publizierens eröffnen digitale Formate neue wissenschaftliche Methoden zur Weiterverarbeitung. Diese Potentiale sind sich insbesondere die mit digitalen Quellen und Daten arbeitenden Geisteswissenschaftler*innen bewusst und Formulieren entsprechende Ansprüche an die Digitalisierung und Bereitstellung der Untersuchungsgegenstände, wie etwa Volltexte mit standardisierter Strukturierung und Interoperabilität sowie mit Entitäten und komplexen Strukturmerkmalen angereichert (Klaffki et al. 2018: 19-20).² Diese Ansprüche müssten in den Geisteswissenschaften, in der die eigenen Texte in Form einer kritischen Rezeption und Iteration Teil der Informationsquellen sind, auch an die eigene Textproduktion gestellt werden. Folgerichtig lautet die Empfehlung der DHd-AG „Digitales Publizieren“, die semantischen Strukturen zu kodieren, die Dokumente maschinenlesbar und prozessierbar zu machen und PDF nicht als primäres Publikationsformat zu verwenden (DHd-AG 2016). Die TEI-basierten Veröffentlichungen der Digital Humanities Community demonstrieren das Potential der Publikationen als Untersuchungsgegenstand des Faches (Sahle/Henny-Krahmer 2018 und Hanneschläger/Andorfer 2018). Allein die Zusammenführung der strukturierten Datenbasis der Texte mit den vorhandenen Technologien und Methoden neuer Publikationsformate steht häufig noch aus. Diese Lücke adressiert das vorliegende Konzept.

Konzeption des Workshops

Ziele

Der Hackathon liefert einerseits ein Proof-of-Concept für die Implementierung digitaler Technologien in einen Publikationsprozess (der digitalen Geisteswissenschaften) und möchte andererseits ein partizipatives Format zur Unterstützung der DHd-Tagung durch die DHd-Community in kollaborativer Arbeitsform (Hackathon) darstellen. Der Einsatz vorhandener Frameworks aus

der Community demonstriert die Leistungsfähigkeit und das vorhandene Potential. Die erarbeiteten Transformationsskripte und Workflows können die Basis für eine Weiterentwicklung und Nachnutzung in institutionellen Publikationsprozessen darstellen. Eine Analyse des Workshops, der Ergebnisse und der damit verbundenen Rezeption liefert die Grundlage für die Formulierung von Empfehlungen zum möglichen zukünftigen Umgang mit den DHd-Abstracts und deren stetig steigende Relevanz als Publikationsformat. Der Hackathon kann dabei sowohl technologische wie methodische Impulse für das Digitale Publizieren in der DHd-Community liefern.

Vorbereitungen

Die Datengrundlage für den Workshop bilden die zur DHd 2019 eingereichten Abstracts im dhc-Format und idealerweise in einer harmonisierten TEI-Version. Die Workshopleiter vertrauen hier auf die wertvolle Redaktionsarbeit der DHd-Konferenzorganisation, wie es die DHd 2018 vorbildlich umgesetzt hat.³ Um die hier angestrebten Verarbeitungsprozesse und Methoden des Digitalen Publizierens umsetzen zu können, müssen die Daten weiter vorbereitet und angereichert werden. Die vorbereitende Prozessierung soll umfassen:

- Harmonisierung (Sichtung; Vereinheitlichung von Ambiguitäten, Setzen der Mindeststandards)
- Strukturdatenauszeichnung (Überschriften, Absatznummerierung, Metadaten)
- Verknüpfung mit Normdaten (Erkennung und Auszeichnung von zentralen Entitäten wie Personen, Orte, Institutionen, Werke)
- Rahmenwerke (Generierung von Listen und Registern für Schlagworte, Titel, Namen)

Die Aufbereitung und Anreicherung der Datenbasis wird von den Organisatoren in Zusammenarbeit mit dem Austrian Centre for Digital Humanities (ACDH) im Vorfeld der DHd 2019 vorgenommen. Die Workshopleitung wird dahingehend mit der Konferenzorganisation und dem Programmkomitee eine enge Kooperation anstreben.

Themen und Arbeitsgruppen

Die Entwicklungsphasen des Hackathons finden als Gruppenarbeit statt. Die parallel arbeitenden Gruppen widmen sich verschiedenen Repräsentationen und Verarbeitungsprozessen der gemeinsamen Datenbasis. Die hier skizzierten Themen und Arbeitsgruppen sind als Vorschläge von Seiten der Organisatoren zu verstehen und werden abhängig von den Kompetenzen und Interessen der Teilnehmenden zu Beginn des Workshops adaptiert. Die beiden Arbeitsbereiche "Transformation" und "Präsentation" werden für die Ziele des Workshops

als zwingend notwendig erachtet und sind daher gesetzt. Weitere Themen und Arbeitsgruppen können von den Teilnehmenden ausgestaltet werden. Die Organisatoren moderieren die Bildung der Arbeitsgruppen und begleiten beratend die Entwicklungsphasen.

AG "Style und Sheet": Transformation

Die Transformation der Ausgangsdaten (XML/TEI) in verschiedene Zielformate bildet die Grundlage für die verschiedenen Nutzungs- und Rezeptionsformen. Als Zielformate der Stylesheets bieten sich u.a. an: HTML, LaTeX, PDF, MS-Word, Markdown, JATS. Idealerweise können die Teilnehmenden eigene, bereits genutzte Transformationen in der Gruppe diskutieren, erweitern und optimieren.

AG "Web und App": Präsentation

Der Zeitrahmen des Workshops erlaubt keine Neuentwicklung eines Frontends. Für die Realisierung verschiedener Präsentationsschichten müssen die Basisdaten in vorhandene Publikationsframeworks integriert werden. Dabei sind generische Ansätze (z.B. eXist Webapp, GitHub pages, eLife Lens Viewer, Jupyter) ebenso möglich wie spezifische oder institutionelle Lösungen der DH-Community.

AG "Maschine und Modell": Schnittstellen

Die Bereitstellung der Daten über standardisierte maschinenlesbare Schnittstellen (API) ist eine grundlegende Repräsentationsform zukünftiger Publikationspraktiken. Idealerweise wird hier prototypisch ein Protokoll oder eine Spezifikation implementiert (z.B. OAI-PMH, FCS, DTS).

AG "Wolke und Vektor": Textanalyse

Durch Tokenisieren, Lemmatisieren, Vektorisieren werden Wortlisten, Wortfrequenzen, Wortwolken, Topicmodelling realisiert. Die Strukturierungs- und Visualisierungsformen sind die Grundlage für alternative Rezeptionsformen und textinterne Analysen bis zu (korpus)linguistischen Methoden.

AG "Beziehung und Geflecht": Netzwerke

Die mit Normdaten angereicherte Datenbasis bietet die Möglichkeit, die Beziehungen zwischen den Entitäten (Personen, Orte, Institutionen, Werke) zu extrahieren, zu visualisieren und zu analysieren. Zu den Anwendungsfällen gehören Netzwerkanalysen und explorative Navigationsformen.

AG “Medial und Modular”: Multimedia

Die Verwaltung, Adressierung und Einbettung multimedialer Inhalte die sowie Erzeugung und Integration interaktiver Elemente, wie dynamische Visualisierungen, gehören zu den großen Herausforderungen des Digitalen Publizierens. Prototypische Umsetzungen können hierfür Impulse liefern oder Lücken in bestehenden Frameworks aufzeigen.

Agenda:

15' Begrüßung, Organisatorisches, Vorstellungsrunde,
15' Vorstellung der Daten und Aufteilung auf Arbeitsgruppen
60' Entwicklungsphase I
30' Pause
45' Entwicklungsphase II
30' Vorstellung der Ergebnisse
15' Abschlussdiskussion

Organisation

Ergebnissicherung und Outreach

Die ausgearbeiteten Ergebnisse der Arbeitsgruppen (Daten, Skripte, Anwendungen) werden offen lizenziert und frei zur Verfügung gestellt. Ein Workshopbericht fasst die Ergebnisse übersichtlich als Blogpost zusammen. Gemeinsam mit interessierten Teilnehmenden plant die Workshopleitung eine Ausarbeitung der Ergebnisse als Empfehlung für mögliche Umsetzungen der DHd-Abstracts. Die konkreten Implementierungen sollen während der Konferenz online verfügbar gehalten werden und werden von der Workshopleitung (und freiwillig Teilnehmenden) über soziale Medien bekannt gegeben und zur Diskussion gestellt. Weitere erwünschte Kommunikationskanäle (Website, Email) werden mit der Konferenzorganisation besprochen.

Datenmanagement und Infrastruktur

Die Einrichtung einer dedizierten Organisation auf GitHub.com für den Workshop erlaubt das Management der verschiedenen Entwicklungen und kollaborative Arbeitsformen. Die Dokumentation und das Projektmanagement werden ebenfalls über git umgesetzt (GitHubWiki/ GitHub-Projects/Issues). Die Publikation der Ergebnisse erfolgt über GitHub-Repositories via Zenodo. Grundsätzlich stehen auch nichtkommerzielle Gitlab-Instanzen und das DARIAH-DE-Repository zur Sicherung zur Verfügung. Das ACDH kann für das kurz- und mittelfristige Hosting der im Rahmen des Workshops entwickelten Applikationen und Services die erforderliche Server-Infrastruktur zur Verfügung stellen. Weiter ist die

Bereitstellung von mehreren Instanzen der Applikationen Voyant und eXistDB zur Nutzung im Workshop über DARIAH-DE vorgesehen.

Teilnehmer*innen

Die praktischen Arbeitsphasen der Gruppenarbeit erlauben nur kleine Gruppengrößen. Die Zahl der möglichen Teilnehmer*innen beträgt daher maximal 25 Personen. Für die Workshopteilnahme werden zwar keine spezifischen technischen oder methodischen Grundlagen (jenseits des Umgangs mit den TEI/XML-Basisdaten) vorausgesetzt, jedoch erfordert ein Hackathon von den Teilnehmenden selbständiges Arbeiten und die Anwendung vorhandener Kompetenzen auf den Gegenstandsbereich. Der Workshop bietet den Rahmen und Raum für die gemeinsames Arbeiten und offenen Austausch.

Fußnoten

1. Dies zeigt sich am Beispiel der Zeitschrift für digitale Geisteswissenschaften (ZfdG), die als Vorbild für die Zeitschrift Medieval and Early Modern Material Culture Online (MEMO) fungierte.
2. Siehe die Digitalisierungsklassen IV und V für Texte.
3. Die vielgestaltigen TEI-Kodierungen aus den Einreichungen und Transformationen wurden einheitlich angeglichen, dokumentiert auf: <https://github.com/GVogeler/DHd2018>.

Bibliographie

Borgman, Christine L. (2010): “*The Digital Future is Now. A Call to Action for the Humanities*”, in: Digital Humanities Quarterly 3 (4): #p16, <http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html> [letzter Zugriff 26. September 2018].

Cremer, Fabian (2018): “*Nun sag, wie hältst Du es mit dem Digitalen Publizieren, Digital Humanities?*”, in: Digitale Redaktion (Blog), 21.03.2018, <https://editorial.hypotheses.org/113> [letzter Zugriff 26. September 2018].

DHd-AG "Digitales Publizieren" (2016): “*Digitales Publizieren*”, Working Paper, 01.03.2016, <http://diglib.hab.de/ejournals/ed000008/startx.htm> [letzter Zugriff 26. September 2018].

Hanneschläger, Vanessa / Andorfer, Peter (2018): *Menschen gendern? Datenmodellierung zur Erhebung von Geschlechterverteilung am Beispiel der TEI2016 Abstracts App*, DHd 2018, Köln, <https://doi.org/10.5281/zenodo.1182576>

Jannidis, Fotis / Kohle, Hubertus / Rehbein, Malte (2017)(eds.): *Digital Humanities. Eine Einführung*, Stuttgart: J. B. Metzler.

Kohle, Hubertus (2017): “*Digitales Publizieren*”, in: **Jannidis, Fotis et al. (eds.):** *Digital Humanities. Eine Einführung*, Stuttgart: J. B. Metzler 199-205.

Klaffki, Lisa / Schmunk, Stefan / Stäcker Thomas (2018): “*Stand der Kulturgutdigitalisierung in Deutschland. Eine Analyse und Handlungsvorschläge des DARIAH-DE Stakeholdergremiums 'Wissenschaftliche Sammlungen'*”, DARIAH-DE Working Papers 26, Göttingen, URN: urn:nbn:de:gbv:7-dariah-2018-1-3 .

Stäcker, Thomas (2012): “*Wie schreibt man digital humanities?*”, in: DHd-Blog, 19.08.2012, <https://dhd-blog.org/?p=673> [letzter Zugriff 26. September 2018].

Stäcker, Thomas (2017): “*Digital Humanities : eine Einführung / Fotis Jannidis, Hubertus Kohle, Malte Rehbein (Hg.)*”, in: O-Bib. Das Offene Bibliotheksjournal 4(3): 142-148, <https://doi.org/10.5282/o-bib/2017H3S142-148>

Vogeler, Georg (2018)(ed.): *DHd 2018. Kritik der digitalen Vernunft. Konferenzabstracts*, Köln: Universität zu Köln 2018.