

Repräsentativität in digitalen Archiven

Dziudzia, Corinna

corinna.dziudzia@ku.de

Forschungszentrum Gotha der Universität Erfurt

Hall, Mark

mark.hall@work.room3b.eu

The Open University, United Kingdom

Die „Kulturen des digitalen Gedächtnisses“ operieren vor dem Hintergrund eines Spannungsfelds, das man auf eine Dichotomie von Kanon und Archiv zuspitzen kann. Das digitale Archiv umfasst u.a. die digitalisierten Artefakte und darin nimmt der Kanon der aus verschiedenen Gründen besonders bewahrenswerten Artefakte einen spezifischen Stellenwert ein. Sie sind Bestandteil des kulturellen Gedächtnisses (vgl. Assmann, 2010). Das digitale Archiv umfasst eine epistemische Ebene, insofern darüber, auch für eine breitere Öffentlichkeit jenseits der klassischen Bildungsinstitutionen, Wissen bereitgestellt wird. Gleichermaßen liefert das digitalisierte Archiv auch die Grundlage für historisch arbeitende Wissenschaften, was etwa zur Entwicklung und Erprobung neuer wissenschaftlicher Methoden dient (Malazita et al., 2020). Das Archiv als digitalisiertes Gedächtnis stellt einerseits altes Wissen dar und dient andererseits auch als Grundlage, von der aus, und mit der, neues Wissen produziert werden kann.

Das digitale Archiv reduziert dabei per se die Barrieren, die den Zugriff auf die Artefakte im physischen Archiv erschweren. Zugleich fügt das digitale Archiv zusätzliche Abstraktions- und Verarbeitungsschichten zwischen den physischen Artefakten und den digitalen Repräsentationen ein. Historische Artefakte müssen zuerst digitalisiert werden, dann mit Metadaten versehen und schlussendlich auffindbar gemacht werden. Wenn auf den digitalisierten Inhalten des digitalen Archivs aufbauend gearbeitet wird, dann repräsentiert jeder dieser Schritte einen Eingriff in die Datengrundlage mit dem Potential, die Arbeitsergebnisse zu verzerren. Welche Artefakte werden im Digitalisierungsprozess priorisiert? Wer erzeugt wie die Metadaten für die digitalisierten Artefakte? Wie werden die gesuchten Artefakte im digitalen Archiv gefunden? Die Antwort auf jede dieser Fragen beeinflusst die Repräsentativität des Archivs insgesamt und die Verlässlichkeit der darauf aufbauenden Ergebnisse.

Die Frage der Repräsentativität wird insbesondere im Rahmen der Intersektionalitätsforschung aktuell stark thematisiert, nicht selten verbunden mit einer Kritik, u.a. an den Digital Humanities (vgl. Risam, 2015), wobei die zugrundeliegende Debatte zum Kanon und der Kritik daran, dass er zu 'weiß' und zu 'männlich' sei, bis in die 1960er Jahre zurückreicht. Das Bewusstsein wächst dabei bei verschiedenen Akteuren, dass Auswahlen, und insbesondere auch Digitalisierungsprozesse, Ergebnisse hegemonialer Machtstrukturen sein können und es wird mit entsprechenden Erklärungen und Selbstverpflichtungen darauf reagiert (z.B. Ziegler, 2019). Das wird etwa auch vor dem Hintergrund des postcolonial computing thematisiert (Irani et al., 2010; Chan, 2017; Kizhner et al., 2021).

Der zentrale Ansatzpunkt der Kritik am möglichen Quellenbias der Archive und Digitalisate (Hall, 2020; Inwood & Stewart, 2020; Kizhner et al., 2021) ist, dass in den meisten Fällen viel zu wenig über die Zusammensetzung und die Auswahlkriterien in

den digitalen Archiven bekannt ist. Wenngleich positive Beispiele digitalisierender Institutionen existieren, die darüber Auskunft geben, was sie digitalisieren, warum und in welchen Bereichen sie wie weit fortgeschritten sind oder was sie woher kommend darüber hinaus bereitstellen, ist Transparenz generell kaum vorhanden.

Die Komplexität in dieser Frage ergibt sich nicht zuletzt daraus, dass die Interessen und Anforderungen einer Reihe von Akteuren zusammentreffen, insofern es sich bei der Digitalisierung um ein ‚boundary object‘ handelt (Star & Griesmer, 1989). Geisteswissenschaftler_innen, die mit den digitalen Archiven arbeiten wollen, Archivar_innen, die Digitalisierungs- und Metadatenprozesse leiten und durchführen, Geldgeber_innen, die Mittel für die Digitalisierungsprozesse bereitstellen, und Informatiker_innen, die Archivierungs- und Suchsoftware entwickeln – jede dieser Gruppen hat unterschiedliche Sichtweisen darauf, was Repräsentativität für sie bedeutet, aber was fehlt, ist ein Forum, in dem ein Austausch über diese Fragen über die Gruppengrenzen hinweg möglich ist.

Zu der Problematik haben wir im März 2021 den internationalen Workshop „Digital Archive and Canon“ organisiert. Dieser stieß auf reges Interesse mit Beiträgen von Forscher_innen und Archivar_innen und der daraus resultierende Austausch ermöglichte ein tieferes Verständnis für die Wünsche und Positionen der jeweils anderen Gruppe. Aus der Abschlussdiskussion ergab sich, dass großes Interesse an einem Forum für einen regelmäßigen Austausch über die Repräsentativitätsfrage besteht, insbesondere in einer Struktur, welche die Interaktion zwischen den interessierten Gruppen ermöglicht. Der hier vorgeschlagene ganztägige Workshop für die DHd 2022 stellt einen ersten Schritt in diese Richtung dar, mit zwei geplanten Outputs: einem Positionspapier und der Gründung einer DHd-Arbeitsgruppe „Repräsentativität“. Es bestehen zwar schon akteursspezifische AGs (Datenzentren, DH Theory, OCR), die AG „Repräsentativität“ zielt allerdings darauf, Akteur_innen über Gruppengrenzen hinweg zusammenbringen, was unsere Vorarbeit als notwendig identifiziert hat.

Themen und Fragen

Im Rahmen des Workshops sollen eine Reihe von Themen und Fragen diskutiert werden, wobei diese an die Interessen der Teilnehmer_innen angepasst werden. Die zentralen Themen, welche den Workshop potentiell strukturieren werden, sind:

- **Repräsentativität in der Digitalisierung:** Was ist der Status quo hinsichtlich der Digitalisierung? Werden besonders kanonische Artefakte digitalisiert? Werden perspektivisch alle Artefakte digitalisiert und die gesamten Bestände aller bewahrenden Institutionen? Wer entscheidet, welche relevant sind, welche aber nicht, und auf Grundlage welcher Kriterien erfolgt dies? Wer stellt diese Artefakte zur Verfügung, wer digitalisiert sie? Nach welchen Kriterien erfolgt die Priorisierung der Reihenfolge?
- **Repräsentativität in der Annotation:** Werden die Metadaten aus dem physischen Archiv übernommen oder neu erstellt? Wenn Metadaten übernommen werden, wie wird gewährleistet, dass diese nicht historische Voreingenommenheiten widerspiegeln?
- **Koordination der Digitalisierung:** Sollten Digitalisierungsarbeiten über Länder- und Archivgrenzen hinaus koordiniert werden? Wie beeinflusst das Schwerpunktsetzung, Auswahl und Aspekte der Nachhaltigkeit (Barats et al., 2020)? Könnte die Digitalisierung kanonischer Werke zwischen Archiven

aufgeteilt werden? Wie wäre das praktisch umsetzbar? Sind Erfahrungen aus dem WorldCat System hier anwendbar?

- **Digitale Systeme:** Welche Verarbeitungsschritte werden für die Bereitstellung digitaler Archive über Suchsysteme angewandt? Insofern Suchhilfen wie Topic-Modelle bereitgestellt werden, wie können Verzerrungen in diesen vermieden bzw. minimiert werden? Wie kann eine Suche über die Archivgrenzen hinweg ermöglicht werden, im Sinne der Deutschen Digitalen Bibliothek und der Europeana?
- **Finanzierung:** Wie sollen Fragen der Repräsentativität in den Geldgeberprozess eingebunden werden? Wie viel sollte explizit durch die Geldgeber vorgegeben werden, ähnlich den Anforderungen bezüglich der Lizenzierung der Digitalisate?
- **Dokumentation:** Was muss während aller Prozesse beachtet und dokumentiert werden, um die Repräsentativität eines Archivs zu verbessern? Welche Informationen braucht der/die Nutzer_in des Archivs mindestens? Welche Informationen sind für welche Nutzergruppen relevant?

Durchführung

Vor dem Workshop

Im Vorfeld des Workshops werden Teilnehmer_innen gebeten ein Statuspapier zuzusenden, in dem sie darlegen, was die jeweils eigene Perspektive ist, was ihre Interessen an dem Thema sind, und welche Aspekte für relevant erachtet werden. Unmittelbar vor dem Workshop werden diese Papiere, sowie gegebenenfalls weitere Materialien, für alle Teilnehmer_innen zur Verfügung gestellt.

Programm Workshop

Der Workshoptag selbst wird in zwei große Blöcke geteilt.

1. Am Vormittag werden die aus den Statuspapieren identifizierten Themen innerhalb der jeweiligen Interessensgruppen (Archivar_innen/Bibliothekar_innen, Wissenschaft, Technik/IT, Förderer_innen/Geldgeber_innen, ...) in separaten Gruppen besprochen. Jede Gruppe verständigt sich über den Status quo: was ist der Ist-Stand, was ist die zukünftige Perspektive, was wäre wünschenswert. Die Ergebnisse dieser Gruppenarbeit werden gesammelt und dann im Plenum besprochen. Die sich aus dieser Diskussion herauskristallisierenden, zentralen Themen definieren dann den Inhalt des Nachmittags.

2. Im Nachmittagsblock werden die Gruppen neu zusammengestellt, diesmal mit dem Ziel, dass in jeder Gruppe Teilnehmer_innen aus allen Interessensgruppen vertreten sind. In dieser Arbeitsphase werden die am Vormittag identifizierten Themen diskutiert, mit dem Ziel erste skizzenhafte Positionspapiere je Gruppe zu entwickeln. Diese Entwürfe werden dann im Plenum zusammengeführt, mit dem Ziel ein gemeinsames Positionspapier zu formulieren.

Abschließend wird am Ende des Workshoptages in einer optionalen Session die formale Gründung der DHd AG besprochen und beschlossen.

Nach dem Workshop

Nach dem Workshop wird das gemeinsame Statuspapier mit den Teilnehmer_innen geteilt und kollaborativ fertiggestellt. Zusätzlich werden die formalen Schritte der DHd AG Gründung durchgeführt.

Call for Position Statements

Im Rahmen des DHd 2022 findet ein ganztägiger Workshop statt, der sich der Frage der Repräsentativität im digitalen Archiv widmet. Diese Frage wird insbesondere im Rahmen der Intersektionalitätsforschung stark thematisiert, nicht selten verbunden mit einer Kritik, u.a. an den Digital Humanities. Welche Artefakte werden im Digitalisierungsprozess priorisiert? Wer erzeugt wie die Metadaten für die digitalisierten Artefakte? Wie werden die gesuchten Artefakte im digitalen Archiv gefunden? Das Bewusstsein wächst dabei, dass Auswahlen, und insbesondere auch Digitalisierungsprozesse, Ergebnisse hegemonialer Machtstrukturen sein können, wie es u.a. das postcolonial computing thematisiert. Der zentrale Ansatzpunkt der Kritik am möglichen Quellenbias der Archive und den Digitalisaten ist zudem, dass oftmals viel zu wenig über die Zusammensetzung, die Auswahlkriterien usw. einzelner Korpora in digitalen Archiven bekannt ist und es an Transparenz mangelt. Die Komplexität ergibt sich nicht zuletzt daraus, dass die Interessen und Anforderungen verschiedener Akteur_innen zusammentreffen, die jeweils unterschiedliche Sichtweisen darauf haben, was Repräsentativität für sie bedeutet.

Der Workshop "Repräsentativität in digitalen Archiven" bietet ein Forum um Fragen hinsichtlich Auswahl, Annotation, Metadaten, Koordination, Nachhaltigkeit, Suchsystemen, Finanzierung, Dokumentation, Transparenz usw. vor diesem Hintergrund übergreifend und aus den verschiedenen Perspektiven (Archiv/Bibliothek, IT, Wissenschaft, Förderung usw.) zu thematisieren. Dafür wird vorab um kurze Positionspapiere von interessierten Teilnehmer_innen gebeten (max. 500 Worte), die mind. folgende Informationen enthalten sollten: was ist die jeweils eigene Perspektive, was sind die Interessen an dem Thema, welche Aspekte werden für relevant erachtet. Im Verlauf des Workshoptags wird ein gemeinsames Positionspapier erarbeitet und eine DHd-Arbeitsgruppe gegründet, um den Austausch und die Arbeit zum Thema "Repräsentativität" zu verstetigen.

Organisation

Die Teilnehmer_innenzahl ist vor Ort begrenzt, die Veranstaltung wird allerdings hybrid geplant. Die benötigte technische Ausstattung beschränkt sich auf einen Laptop. Spezifische Vorkenntnisse sind nicht nötig. Bedingung zur Teilnahme ist die Einreichung des Statuspapiers via Mail an die Organisator_innen bis zum 09.02.2022.

Corinna Dziudzia

Corinna.dziudzia@uni-erfurt.de
Forschungszentrum Gotha der Universität Erfurt
Schlossberg 2
99868 Gotha

Corinna Dziudzias Arbeitsschwerpunkte liegen im Bereich der Begriffs- und Wissenschaftsgeschichte, insbesondere zur Tradierung von literarischem Wissen und Kanonbildung. Wesentlich ist dafür die Digitalisierung und die dabei beobachtbaren Effekte.

Mark Hall

mark.hall@open.ac.uk
 School of Computing & Communications
 The Open University
 Milton Keynes, MK7 6AA (UK)

Mark Halls Interessenschwerpunkte liegen auf der Öffnung des digitalen Kulturguts für die interessierte Öffentlichkeit und methodischer Klarheit in den Digital Humanities. Aus beiden Schwerpunkten ergibt sich ein Interesse an der Repräsentativität des digitalen Archives.

Bibliographie

Assmann, Aleida(2010): “Canon and Archive”, in: Erll, Astrid / Nünning, Ansgar (eds.): *A Companion to Cultural Memory Studies*. Berlin / New York: De Gruyter 97-107.

Barats, Christine(2020): “Fading Away...The Challenge of Sustainability in digital studies”, in: *digital humanities quarterly*, Volume 14 Number 3, <http://www.digitalhumanities.org/dhq/vol/14/3/000484/000484.html>[letzter Zugriff 10.11.2021].

Chan, A.S.(2018): “Decolonial computing and networking beyond digital universalism”, in: *Catalyst: Feminism, Theory, Technoscience*, 4(2), 1-5.

Hall, Mark M. (2020): “Opportunities and risks in digital humanities research”, in: Carius, Hendrikje / Prell, Martin / Smolarzski, René (eds.): *Kooperationen in den digitalen Geisteswissenschaften gestalten*, Göttingen: Vandenhoeck & Ruprecht 47-66.

Inwood, Kris / Maxwell-Stewart, Hamish(2020): “Selection Bias and Social Science History”, in: *Social Science History*44, Nr. 3: 411-416 <https://doi.org/10.1017/ssh.2020.18>.

Irani, L. / Vertesi, J. / Dourish, P., Philip K. / Grinter, R.E.(2010): “Postcolonial computing: a lens on design and development”, in: *Proceedings of the SIGCHI conference on human factors in computing systems*:1311-1320.

Kizhner, Inna [et.a.] (2021): “Digital cultural colonialism: measuring bias in aggregated digitized content held in Google Arts and Culture“, in: *Digital Scholarship in the Humanities*36, Nr. 3: 607-40. <https://doi.org/10.1093/lc/fqaa055>.

Malazita, James H. / Teboul, Ezra J. / Rafah, Hined(2020): “Digital Humanities as Epistemic Cultures: How DH Labs make Knowledge, Objects, Subjects”, in: *digital humanities quarterly*, Volume 14 Number 3, <http://www.digitalhumanities.org/dhq/vol/14/3/000465/000465.html>[letzter Zugriff 10.09.2021].

Risam, Roopika(2015): “Beyond the Margins: Intersectionality and the Digital Humanities”, in: *digital humanities quarterly*, Volume 9 Number 2, <http://digitalhumanities.org:8081/dhq/vol/9/2/000208/000208.html>[letzter Zugriff 10.09.2021].

Star, Susan Leigh / Griesemer, James R. (1989): “Institutional Ecology, ‘Translations’ and Boundary Objects: Amateurs and Professionals in Berkeley’s Museum of Vertebrate Zoology, 1907-39“, in: *Social Studies of Science*19, Nr. 3: 387–420. <https://doi.org/10.1177/030631289019003001>.

Ziegler, S.L.(2019): “Digitization Selection Criteria as Anti-Racist Action”, in: *Code{4}lib Journal*, Issue 45 <https://journal.code4lib.org/articles/14667>[letzter Zugriff 10.09.2021].