

Wissenschaftliches Rundgespräch zur Archivgutdigitalisierung

Veranstalter: Archivschule Marburg

Datum, Ort: 26.05.2014, Marburg

Bericht von: Stephanie Oertel, Archivschule Marburg

„Wenn ich mir was wünschen dürfte. Wunsch(t)raum Archiv für NutzerInnen im digitalen Zeitalter“ lautete einer der Vortragstitel auf dem 18. Archivwissenschaftlichen Kolloquium am 26. und 27. November 2013 in Marburg: SYLVIA NECKER (damals IRS, Erkner) ging in ihm auf eine mögliche Beständepriorisierung für die Archivgutdigitalisierung, das Rechercheverhalten und die Möglichkeiten eines digitalen Archivs aus wissenschaftlicher Sicht ein. Ihre Anregungen und auch die der anderen Referenten aus der Wissenschaft, den Bibliotheken und den Museen zeigte ein Potential auf, das innerhalb des DFG-geförderten Produktivpiloten „Digitalisierung von archivalischen Quellen“ genutzt werden soll. Ziel ist es, in den zukünftigen Digitalisierungsstrategien der Archive die Sicht und die Bedürfnisse der Nutzer stärker als bisher zu berücksichtigen. Die Koordinierungsstelle des DFG-geförderten Produktivpiloten „Digitalisierung von archivalischen Quellen“ nahm das wissenschaftliche Interesse am Thema, das im Kolloquium signalisiert wurde, zum Anlass und lud Forscher verschiedener geistes-, sozial- und wirtschaftswissenschaftlicher Sparten zu einem vertieften Austausch ein. Am 26. Mai 2014 fand dazu ein wissenschaftliches Rundgespräch in der Archivschule Marburg statt. Die Wissenschaftler diskutierten gemeinsam mit den Projektpartnern des Produktivpiloten die Möglichkeiten der digitalen Bereitstellung von archivalischen Quellen. Der Austausch fand als offenes Plenum statt. Die wichtigsten Gesprächsthemen werden nachfolgend zusammenfassend dargestellt:

1. Auswahl – Priorisierung

Der enorme Archivalienumfang ist ein möglicher Grund, warum die Digitalisierung in den deutschen Archiven bislang noch nicht umfassend voran geschritten ist. Da aus wirtschaftlichen Gründen eine Totaldi-

gitalisierung nicht möglich ist, muss aus den als archivwürdig bewerteten Beständen eine relevante Auswahl getroffen werden. CLEMENS REHM (Landesarchiv Baden-Württemberg) veranschaulichte dies am Beispiel einer Priorisierungsliste, die lediglich 7 Prozent des Archivguts des Landesarchivs Baden-Württemberg umfasst und für deren Digitalisierung eine Summe von ca. 88 Millionen Euro bereitgestellt werden muss. Grundlage der Priorisierung sind aktuelle Forschungsschwerpunkte. Sie beinhaltet vorrangig Rückgratbestände, hoch frequentierte Bestände, Bestände zu Jubiläen und diejenigen, die von Bestandserhaltungsaspekten betroffen sind. Niedrig priorisiert werden hingegen schutzfristenbehafte Bestände, bspw. Daten von Finanzämtern, die aus rechtlichen Gründen nicht online gestellt werden dürfen. Dabei ist zu berücksichtigen, dass nur mit einer Veröffentlichungsmöglichkeit der Digitalisate eine DFG-Förderung möglich ist. Für schutzfristenbehafte Bestände findet diese Finanzierungsmöglichkeit keine Anwendung.

Von Seiten der Wissenschaft wurde angeregt, die Bestände, die die Grundlagenforschung bedienen, hoch zu priorisieren und möglichst als komplette Bestände online zu stellen. Zudem sollten alle Findmittel digital recherchierbar sein. Die Digitalisate, die bei der digitisation-on-demand Methode und der Sicherungsverfilmung entstehen, sind ebenfalls für die digitale Bereitstellung zu berücksichtigen. Neben ihren eigenen Forschungsschwerpunkten empfahlen die Wissenschaftler auch die digitale Bereitstellung von vollständig unbekanntem Materialien, die wiederum neue potentielle Forschungsmöglichkeiten eröffnen.

Auf die wissenschaftlichen Anregungen folgten zum Teil archivarische Bedenken. Argumentiert wurde, dass die digitalen Auftragsbestellungen (digitisation-on-demand) meist nur wenige Seiten umfassen und eine Digitalisierung des vollständigen Bestandes mit Einbindung der notwendigen Kontextinformationen nicht finanziert werden kann. Zu den Kontextinformationen zählen die Bestandsinformationen im Findbuch und in der Klassifikation, sowie die Tektonik aller Bestände im jeweiligen Archiv und die Metadaten zum Digitalisat. Weiterhin

ist zu berücksichtigen, dass die Digitalisate, die parallel zur Schutzverfilmung entstehen, deutliche Qualitätsverluste beinhalten können.

2. Erschließung und Digitalisierung

2.1 Erschließungstiefe

Der Wunsch nach möglichst vielen digitalen Beständen mit flacher Erschließung wurde von den Forschern aus den Kreisen der Digital Humanities geäußert und als Beispiel die Protokolle der Behörden genannt. Die Archivare wiesen auf die Fehlerquellen hin, die die Umsetzung dieses Wunsches mit sich bringen würde. Werden Digitalisate nicht oder nur rudimentär erschlossen, kann deren Analyse zu Fehlinterpretation führen, da dem Benutzer die Informationen zur Rekonstruktion der ursprünglichen Zusammenhänge im Bestand fehlen. Die Interpretation des einzelnen Digitalisats kann mit dem Image und dem Dateinamen nicht ausreichend durchgeführt werden. Aus gutem Grund ist die Erschließung der Bestände mit der Darstellung der Bestands- und Behörden Geschichte und der Bereitstellung weiterer Informationen eine Fachaufgabe im Archiv. Die meisten Historiker sind sich der Bedeutung der Informationsquelle bewusst. Sie sollte auch allen zukünftigen Wissenschaftlern vermittelt und in digitaler und gut ablesbarer Form bereitgestellt werden.

2.1 Normdaten

Die Normdatenerhebung und damit die Verzeichnung von normierten Begriffen ist bei der Erschließung der Bestände und ihrer digitalen Bereitstellung ein wissenschaftlicher Zusatzgewinn. Das automatische Auslesen durch den Einsatz der OCR (Optical Character Recognition) - Technologie für handschriftliche Quellen zeigt eine hohe Fehlerquelle auf. Aus diesem Grund sind hier die weiteren technischen Entwicklungen abzuwarten.

2.3 Crowdsourcing

Crowdsourcing könnte die Lücke zwischen Erschließungsansprüchen der Forschung und Erschließungsnormen der Archive verkleinern, wenn ein gegenseitiger Nutzen realisiert wird. Den Archivaren und Forschern ist

das große Kooperationspotenzial mit Synergien auf beiden Seiten bewusst. Für die manuelle Anreicherung wird eine bedienerfreundliche Infrastruktur für effektives Arbeiten mit geringem Zeitaufwand angeregt. Dabei sollten Modulationen und semantische Bausteine ebenfalls berücksichtigt werden. Als Mangel wird bislang der Datenaustausch und damit die fehlenden Schnittstellen für die Einbindung der Daten in eine Forschungsumgebung und in das archivische Informationssystem gesehen. Hierfür bietet sich als technische Lösung das Daten-Harvesting zum automatischen Abgleich der Daten zwischen den Institutionen an.

3. Auffindbarkeit der Quellen (Persistent Identifier und Speicherort)

Für den Persistent Identifier empfahl PATRICK SAHLE (Universität Köln) eine „sprechende“ Signatur, die feingranular auf die einzelne Seite verweist, und die Bereitstellung der dazugehörigen technischen Metadaten. Die Umsetzung dieser Anregung beinhaltet enorme Personalkosten, die in einem Digitalisierungsprojekt nicht aufgebracht werden können, bemerkte MARTINA WIECH (Landesarchiv Nordrhein-Westfalen). Hier müssen automatische und halbautomatische Verfahren etabliert und weitere Möglichkeiten gefunden werden, damit Digitalisierungsprojekte realisiert werden können. HUBERT LOCHER (Bildarchiv Foto Marburg) wies darauf hin, wie essentiell ein verlässlicher Speicherort zum Zitieren der elektronischen Quelle sei.

Mit dem digitalen Wandel in der Informationstechnologie sind gerade auch die Begrifflichkeiten zu klären. Was ist die originale Quelle? In erster Hinsicht ist es das Unikat. Ist es aus bestandserhaltenden Gründen nicht mehr lesbar, tritt an dessen Stelle die digitale Kopie mit dem Informationsgehalt. In wissenschaftlichen Arbeiten sind möglichst beide Verweise, sowohl die analogen als auch die digitalen, aufzulisten und sollten daher in der digitalen Bereitstellung ablesbar sein.

4. Verwertung der Daten

Neben manuellen Quellenauswertungen werden heute immer häufiger automatische Auswertungen von Metadaten in der Forschung

eingesetzt, die einen effizienten wissenschaftlichen Erkenntnisgewinn versprechen. Die Aufgabe der Archivare ist dabei die Datenbereitstellung. Das in der Archivwelt etablierte Format EAD (Encoded Archival Description) sollte in seiner Funktion als Austauschformat genutzt und in Forschungsinfrastrukturen eingebunden werden.

Der hohe Aufwand, für die Nachbearbeitung der OCR-Erkennung ist in der Projektplanung für die Digitalisierung von archivalischen Quellen zu berücksichtigen. Der Anspruch ist, die von der Deutschen Forschungsgemeinschaft geforderte Fehlergenauigkeit von 99,98 % für maschinenschriftliche und 98 % für handschriftliche Quellen zu garantieren. STEPHAN HOPPE (Institut für Kunstgeschichte München) bemerkte, dass auch ein fehlerfreies Arbeiten mit 80 %iger Genauigkeit einen Forschungsgewinn erbringt.

5. Vernetzung

Ein möglichst hoher Grad an Transparenz wurde für die Digitalisate angeregt und hierfür eine Ampelstrategie für die digitale Präsentation vorgeschlagen. Auf die Kooperation und den Austausch zwischen Bibliotheken, Museen und Archiven wurde aus wissenschaftlicher Sicht mehrfach hingewiesen. Im digitalen Zeitalter nähern sich die Merkmale der Sammlungstypen generell einander an. Der Mehrwert liegt hier in der Nachnutzung der technischen Errungenschaften und der Ausbildung gemeinsamer Standards innerhalb der Gedächtnisinstitutionen unter Einbindung der Forschung.

Fazit

Das wissenschaftliche Rundgespräch war auf beiden Seiten gewinnbringend und lädt zur Verstärkung des Austauschs ein. Deutlich wurden die verschiedenen Sichtweisen der Wissenschaftler, die zum Teil stark divergierten. Der Wunsch möglichst bald und möglichst viele Bestände online recherchieren zu können, wurde hingegen von allen Wissenschaftlern geäußert. Die Anregungen fließen in den Produktivpiloten ein und werden aktuell unter wirtschaftlichen und rechtlichen Gesichtspunkten auf ihre Umsetzbarkeit geprüft. Die spartenübergreifenden Ge-

spräche werden weiter intensiviert. Das Protokoll zum wissenschaftlichen Rundgespräch ist auf der Projektseite der Archivschule Marburg abrufbar.¹

Übersicht zum wissenschaftlichen Rundgespräch

Irmgard Ch. Becker (Archivschule Marburg): Begrüßung und Vorstellung des Produktivpiloten

Vorstellungsrunde der Teilnehmer

Offenes Plenum:

a) Auswahl – Priorisierung

b) Verhältnis zwischen Erschließung und Digitalisierung

-Erschließung (flach - tief / Quantität - Qualität)

-Normdaten

-Crowdsourcing - Bereitschaft der Forschung

c) Auffindbarkeit der Quellen (Persistent Identifier und Speicherort)

-Zitieren der archivalischen digitalen Quelle

d) Verwertung der Daten

-Auslesen der Metadaten / automatisierte Auswertung (Erschließungsinformationen, digitalisiertes Archivgut)

e) Vernetzung

-Metadaten und Archivalien

-Projekte

Irmgard Ch. Becker (Archivschule Marburg): Zusammenfassung und Ausblick

Tagungsbericht *Wissenschaftliches Rundgespräch zur Archivgutdigitalisierung*. 26.05.2014, Marburg, in: H-Soz-u-Kult 01.09.2014.

¹Protokoll zum wissenschaftlichen Rundgespräch Link < http://archivschule.de/uploads/Forschung/Digitalisierung/Veranstaltungen/Protokoll_wissenschaftlichen_Rundgespraechs_zur_Archivgutdigitalisierung_2014-05-26.pdf> (Stand: 24.07.2014).