

Überlegungen zu einem Crowdsourcing-Konzept des Landesarchivs-Baden-Württemberg

Leicht überarbeitete Fassung des Vortrags gehalten bei der Tagung „Offene Archive 2.1“, Stuttgart, 4. April 2014

Dr. Esther-Julia Howell

Einleitung

Die „Digitale Revolution“ hat auch das Archivwesen erfasst. Der technologische Wandel wirkt sich nachhaltig auf die archivischen Kernaufgaben aus und stellt die Archive vor neue Aufgaben. Herausforderungen begegnen uns dabei nicht zuletzt in Form gewachsener Ansprüche seitens der Nutzer. Im Zeitalter von Google und Wikipedia gehen diese häufig davon aus, dass Archivbestände digital vorliegen, auf Dokumentenebene erschlossen, Einzelstücke und ihre Inhalte damit elektronisch suchbar und durchsuchbar und natürlich online zugänglich sind. Wer seine Bücher über Amazon bestellt, sie in digitaler Form auf einem E-Book-Reader liest und seinen Freunden über Facebook, Pinterest oder in einem eigenen Blog weiterempfiehlt, der möchte auch in einer möglichst interaktiven Netzumgebung nach Originalquellen recherchieren. Die Generation der Digital Natives, denen weder gedruckte Findbüchern noch Zettelkataloge mehr ein Begriff sind, wird die Forderung nach umfassender Zugänglichkeit zu archivischen Informationen und Archivalien deshalb in Zukunft noch vehementer erheben.¹

Die deutschen Archivare stellen sich den Herausforderungen des digitalen Zeitalters und sind bemüht, den steigenden Ansprüchen gerecht zu werden. Nachdem die sogenannte „Retrokonversion“ bereits gut vorangeschritten ist, wird seit mittlerweile über zehn Jahren auch ausgewähltes analoges Archivgut digitalisiert. Mit der Digitalisierung allein ist es jedoch nicht getan. Neben Fragen der Präsentation und dauerhaften Verfügbarkeit der Digitalisate stellt deren Online-Stellung auch neue Anforderungen an die Erschließung und Kontextualisierung der Archivalien. War im analogen Findmittel schon aus praktischen Gründen häufig nur eine Erschließung auf Ebene des Bestandes oder der Archivalieneinheiten – also beispielsweise einer Akte – möglich, so ist dank datenbankgestützter Archivinformationssysteme auch eine Erschließung auf Vorgangs- oder Dokumentenebene denkbar und im Falle bestimmter Archivaliengattungen wie Fotos sogar erforderlich.

Allerdings stehen die technischen Möglichkeiten und die Wünsche der Nutzer in einem Spannungsverhältnis zu den begrenzten und tendenziell sinkenden personellen und finanziellen Spielräumen vieler öffentlicher Archive. Kaum ein Archiv verfügt über die notwendigen Ressourcen, um seine Bestände flächendeckend und in absehbarer Zeit auf Dokumentenebene tief zu

¹ Vgl. David Fricker, Archives – what is our business model for the digital age? Rede gehalten auf der Jahrestagung des International Council on Archives in Brisbane am 23.8.2012, URL: <http://www.naa.gov.au/about-us/media/speeches/ica-director-generals-speech.aspx> (5.5.2013); Hartwig Walberg, Die Erschließung von Archivgut zwischen Tradition und Zukunft. in: Archive in Thüringen, Sonderheft 2003, S. 109-112, hier S. 109; Mechthild Black, Recherche via Internet: Neue Wege zum Archivgut, in: Fundus – Forum für Geschichte und ihre Quellen, Heft 4, S. 215-231, hier S. 217, URL: <http://webdoc.sub.gwdg.de/edoc/p/fundus/4/black.pdf> (5.5.2014); J. Gordon Daines III, Cory L. Nimer, Web 2.0 and Archives (18.5.2009), in: The Interactive Archivist. Case Studies in Utilizing Web 2.0 to Improve the Archival Experience, URL: <http://interactivearchivist.archivists.org/> (5.5.2014); Jörg Fischer, Lost in (Cyber-)Space? Das Stadtarchiv Amberg in der schönen neuen Welt des Web 2.0. Vortrag gehalten auf dem Deutschen Archivtag in Saarbrücken am 26.9.2013, URL: <http://archive20.hypotheses.org/905> (8.10.2013); jüngst auch Bastian Gillner, Archive im digitalen Nutzerkontakt. Virtuelle Lesesäle, soziale Medien und mentale Veränderungszwänge, in: Der Archivar 66 (2013), S. 406-415, hier S. 407.

erschließen. Doch auch eine bessere finanzielle Ausstattung würde angesichts der enormen Mengen des in öffentlichen Archiven lagernden Materials und der bereits bestehenden Verzeichnungsrückstände keine vollumfängliche Einzelstück-Erschließung garantieren.

Gerade im Bereich der Erschließung bringt das Informationszeitalter jedoch nicht nur neue Herausforderungen mit sich, sondern bietet auch enorme Chancen: Das Web 2.0 und die Sozialen Medien eröffnen neue Kommunikationswege mit traditionellen ebenso wie mit neuen Nutzergruppen. Dabei agieren die Nutzer nicht mehr nur als Konsumenten, sondern wollen sich beispielsweise bei der Erschließung aktiv mit ihrem oftmals hochspezialisierten Expertenwissen zu bestimmten Themengebieten einbringen. Zwar haben Archivare auch in der Vergangenheit Feedback und Hinweise von Nutzern erhalten und darauf reagiert, das Internet bietet hier jedoch neue, transparentere Möglichkeiten.² War es in der Vergangenheit wegen des hohen Arbeits- und Kostenaufwands schwierig, analoge Findmittel zu verändern, ermöglichen moderne Online-Findmittelsysteme mittlerweile die laufende Bearbeitung und Ergänzung von Erschließungsinformationen.³ Durch die Einbindung sozialer Navigationswerkzeuge können interessierte Nutzer nun sogar direkt als Co-Produzenten an der Erschließung mitwirken.⁴

Was ist Crowdsourcing und warum ist es interessant für Archive?

Den Rahmen für die gezielte und großangelegte Einbindung der Nutzer bietet das sogenannte „Crowdsourcing“, das seit einigen Jahren auch in archivarischen Fachkreisen diskutiert wird.⁵ Dieser an den Begriff „Outsourcing“ angelehnte Neologismus wurde 2006 von dem Online-Journalisten Jeff Howe geprägt und bezeichnet eine interaktive Art der Leistungserbringung, bei der eine große Zahl an Freiwilligen über Web 2.0-Applikationen an Entscheidungs- oder Kreativprozessen beteiligt wird. Dabei werden Tätigkeiten, die traditionell von dafür ausgebildeten oder geschulten Kräften verrichtet werden, (meist) unentgeltlich von einer (meist) undefinierten und in der Regel großen Gruppe von Leuten übernommen.⁶

Im Zielkonflikt von wachsenden Anforderungen an die Erschließungstiefe bei gleichzeitig anwachsenden Erschließungsrückständen einer- und sinkenden finanziellen und personellen Ressourcen andererseits könnte Crowdsourcing eine Möglichkeit für die Behebung von Erschließungsdefiziten bieten. Dabei können mit Hilfe der „Crowd“ ganz verschiedene Inhalte generiert werden. Zu nennen sind hier vor allem Schlagworte („Tags“), Transkriptionen, Ergänzungen zum Kontext eines Archivale, inhaltliche Beschreibungen (z.B. Regesten), Orts- und Personennormdaten und Georeferenzierungen. Daneben kann Crowdsourcing auch zur Überlieferungsbildung eingesetzt werden, indem die Bürgerinnen und Bürger aufgefordert werden, eigene Unterlagen zu einem bestimmten Thema oder Ereignis beizutragen. So können thematische

² Magia Ghetu Krause; Elizabeth Yakel, Interaction in Virtual Archives: The Polar Bear Expedition Digital Collections Next Generation Finding Aid, in: The American Archivist 70 (2007), S. 282-314, hier S. 299.

³ Ebd. S. 290.

⁴ Vgl. Mario Glauert, Archiv 2.0 – Interaktion und Kooperation zwischen Archiven und ihren Nutzern in Zeiten des Web 2.0, in: Archivpflege in Westfalen-Lippe 79 (2009), S. 29-34; Joachim Kemper, Erschließung im Archiv: Von Richtlinien, Problemen und möglichen Lösungen, URL: http://www.edvtage.de/vortrag.php?kapitel=2010_04 (5.5.2014); Max J. Evans, Archives of the People, by the People, for the People, in: The American Archivist 70 (2007), S. 387-400, hier S. 397; Janusch Carl, Andreas Rutz, Das digitale Historische Archiv Köln und die Zukunft des Kölner Stadtarchivs im Web 2.0, in: Der Archivar 65 (2012), S. 143-153, hier S. 144.

⁵ Vgl. z.B. Joachim Kemper, Jörg Fischer, Katharina Hasenfratz, Thomas Just, Jana Moczarski und Andrea Rönz, Archivische Spätzünder? Sechs Web 2.0-Praxisberichte, in: Der Archivar 65 (2012), S. 136-143; Joachim Kemper, Crowdsourcing, ein „Miniprojekt“, in: Archive 2.0, 8.10.2013, URL: <http://archive20.hypotheses.org/907> (9.10.2013).

⁶ Vgl. die Definition von Jeff Howe auf dessen Blog „Crowdsourcing“, URL: <http://crowdsourcing.typepad.com/> (11.11.2013), sowie die Definition auf Wikipedia: <http://de.wikipedia.org/wiki/Crowdsourcing> (5.5.2014).

Sammlungen, beispielsweise zum Ersten Weltkrieg, entstehen.⁷ Über die Generierung zusätzlicher Erschließungsinformationen und Sammlungen hinaus sind Crowdsourcing-Projekte außerdem ein Mittel, um die Aufmerksamkeit für bestimmte wichtige Bestände und die Wahrnehmung des Archivs insgesamt zu erhöhen und neue Nutzergruppen anzusprechen. Crowdsourcing ist daher teilweise auch den Aufgabenbereichen Öffentlichkeitsarbeit und Bildungsarbeit zuzurechnen. Welche Inhalte generiert werden sollen, hängt vom jeweiligen Bestand und den Zielen des Projekts ab. Alle Arten von Crowdsourcing haben bestimmte Vor- und Nachteile, die häufig in engem Zusammenhang mit den personellen, technischen und finanziellen Rahmenbedingungen der durchführenden Einrichtung stehen. Die Entscheidung für eine oder eine Kombination von mehreren Arten des Crowdsourcing ist daher wichtig für den Erfolg des Projektes und muss bewusst getroffen werden.

Entwicklung eines Crowdsourcing-Konzepts im Landesarchiv Baden-Württemberg

Im Rahmen des DFG-Produktivpiloten „Digitalisierung von archivalischen Quellen“, den das Landesarchiv Baden-Württemberg (LABW) zusammen mit sechs weiteren archivischen Partnern von 2013 bis 2015 durchführt,⁸ hat das LABW u.a. die Aufgabe, Web 2.0-Werkzeuge zur Einbindung externer Nutzer bei der tieferen Erschließung digitalisierter Archivalien zu erproben. Ziel ist die Entwicklung eines archivischen Crowdsourcing-Konzepts, das von Archiven aller Sparten und Größen nachgenutzt werden kann. Die dafür gegründete Arbeitsgruppe hat zunächst eine Bestandsaufnahme nationaler und internationaler Crowdsourcing-Projekte im Bereich von Kulturerbe-Einrichtungen vorgenommen und auf dieser Grundlage die Voraussetzungen und Bedingungen für erfolgreiches archivisches Crowdsourcing diskutiert. Vorläufiges Ergebnis der Diskussion ist der Entwurf eines Kriterienkatalogs für archivische Crowdsourcing-Projekte, der im Folgenden näher vorgestellt wird.

Kriterien für erfolgreiches archivisches Crowdsourcing

Bei der Planung und Durchführung von Crowdsourcing-Projekten sollten – abhängig von den jeweiligen Rahmenbedingungen des durchführenden Archivs – folgende Punkte bedacht werden:

1. Strategische Planung

Idealerweise ist Crowdsourcing Teil einer umfassenden Web 2.0-**Strategie** eines Archivs.

Archive, die keine solche Strategie haben, sollten sich davon aber nicht abhalten lassen. Denn ein gut geplantes Crowdsourcing-Projekt kann durchaus der erste Baustein für eine Web 2.0-Strategie sein. Dabei ist jedoch wichtig, dass ein Engagement in Bereich Web 2.0 kein Selbstzweck ist, sondern immer in engem Zusammenhang mit den Kernaufgaben der Überlieferungsbildung, Erschließung, Zugänglichmachung oder Vermittlung historisch relevanter Unterlagen steht.

2. Projektbetreuung und Ressourcen

Die **Projektbetreuung** muss geregelt sein; personelle **Ressourcen** müssen zur Verfügung stehen. Crowdsourcing bindet in jedem Fall personelle und ggf. auch finanzielle Mittel, die in ihrem Umfang nicht zu unterschätzen sind. Die Bestandsaufnahme hat gezeigt, dass archivische Crowdsourcing-Projekte insbesondere dann erfolgreich sind, wenn gezielt und ausreichend Ressourcen für ihre Konzeption und Durchführung zur Verfügung gestellt werden.

⁷ Eine solche Sammlung entsteht derzeit im Rahmen des Projekts Europeana 1914-1918, bei dem Bürger aus ganz Europa aufgefordert sind, Briefe, Fotos und andere Unterlagen beizusteuern. URL: <http://www.europeana1914-1918.eu/de> (5.5.2014).

⁸ Die Projektpartner sind: Archivschule Marburg – Hochschule für Archivwissenschaft, Generaldirektion der Staatlichen Archive Bayern, Landesarchiv Nordrhein-Westfalen, LWL Archivamt für Westfalen, Sächsisches Staatsarchiv und Stadtarchiv Mannheim – Institut für Stadtgeschichte. Nähere Informationen unter <http://www.landearchiv-bw.de/web/55477#55485> und <http://archivschule.de/DE/forschung/digitalisierung/> (jeweils 5.5.2014).

Personelle oder finanzielle Ressourcen werden zum Beispiel gebraucht für die Konzeption, Koordinierung und Betreuung des Projekts, die technische Umsetzung (z.B. Programmierung und Pflege einer Plattform oder Datentransfer), die Erstellung von Editionsrichtlinien oder anderen Anweisungen für die Teilnehmer, sowie für die Öffentlichkeitsarbeit.

Crowdsourcing gleichsam „nebenbei“ zu betreiben, birgt Risiken und ist weniger erfolgversprechend. Zwar können Crowdsourcing-Projekte auch ohne zusätzliche Mittel durchgeführt werden – in diesem Fall ist die Planung jedoch besonders wichtig. Damit andere wichtige Kernaufgaben nicht unnötig vernachlässigt werden, muss zum Beispiel eine bewusste Priorisierung und Personalplanung erfolgen.

3. Definition von Projektzielen

Die **Projektziele** sollten klar definiert sein.

Wie bereits erwähnt kann Crowdsourcing in verschiedenen Bereichen eingesetzt werden. Umso wichtiger ist es, sich bei der Projektplanung ganz genau zu überlegen, was mit einem bestimmten Projekt erreicht werden soll.

Denkbare Ziele können beispielsweise sein

- Verbesserung der Erschließung des (Teil)Bestands
- Verbesserung der Zugänglichkeit des (Teil)Bestands
- Verbesserung der Wahrnehmung des (Teil)Bestands und/oder des Archivs
- Ansprechen neuer Nutzungsgruppen

Natürlich wird ein Projekt immer alle oder mehrere dieser Ziele haben – aber es kommt auch hier auf die Priorisierung an. Denn ein Projekt, bei dem es vor allem um die Generierung von qualitativ hochwertigen Erschließungsinformationen geht, sollte anders konzipiert sein, als ein Projekt, das in erster Linie der Öffentlichkeitsarbeit dient. **Alle anderen Entscheidungen hängen maßgeblich von den Projektzielen ab!**

4. Beständeauswahl

Es sollten konkrete **Bestände oder Teilbestände** ausgewählt werden.

Eine gezielte Beständeauswahl erleichtert die Definition der Zielgruppe und die Öffentlichkeitsarbeit für das Projekt. Alternativ können auch Teilbestände oder einzelne Stücke aus verschiedenen Beständen in einem Projekt zusammengeführt werden, wenn sich dies thematisch anbietet. Die ausgewählten Bestände oder Stücke müssen aber in jedem Fall eine minimale Erschließung aufweisen und spätestens zu Projektbeginn digitalisiert vorliegen. Zu fragen ist außerdem, ob der Bestand attraktiv für eine potentielle Crowd ist.

5. Rechtsfragen

Es ist wichtig, dass **Rechtsfragen**, zum Beispiel zum Urheberrecht und zum Persönlichkeitsrecht, bei der Planung berücksichtigt werden.

Idealerweise sollten die Rechtsfragen vor Projektbeginn geklärt sein. Dies ist leider nicht immer möglich – in diesem Fall ist es jedoch wichtig, sich zumindest mit der Problematik zu beschäftigen und bei offenen Fragen bewusste Entscheidungen zu treffen.

6. Definition des zu generierenden Inhalts

Der zu generierende **Inhalt** sollte definiert sein.

Dies ist natürlich von den Projektzielen und der Beständeauswahl abhängig. Vorstellbar sind z.B. die Erstellung von Indices, Transkriptionen, Regesten, Normdaten, Kommentaren, Schlagworten oder „Tags“, Geodaten oder auch die Akquise neuer Unterlagen. Je nach Projekt können natürlich auch mehrere Inhalte generiert werden – z.B. Transkriptionen und Regesten. Zu bedenken ist auch, dass die verschiedenen Arten des zu generierenden Inhalts unterschiedliche Anforderungen an die Eingabeoberfläche stellen.

7. Definition der Zielgruppe

Es ist wichtig, die **Zielgruppe** – die „Crowd“ – zu definieren, die mit dem Projekt angesprochen werden soll.

Je nach Projektzuschnitt kann die Zielgruppe eng definiert sein, indem man eine begrenzte Gruppe von Leuten, wie zum Beispiel Studenten eines Lehrstuhls, die Mitarbeiter und Pensionäre einer Firma, oder die Mitglieder eines Heimatvereins gezielt anspricht. Die Crowd kann aber auch ganz groß gedacht werden – dann adressiere ich alle potentiell Interessierten.

8. Entscheidung für eine Plattform

Es sollte eine bewusste Entscheidung für eine **Plattform** getroffen werden.

Je nach Art des Projektes ist zu unterscheiden zwischen Projekten, die innerhalb des eigenen Archivinformationssystems oder auf einer vom Archiv selbst betriebenen Plattform durchgeführt werden und Projekten, die zusammen mit einem Kooperationspartner auf dessen externer und ggf. kommerzieller Plattform angesiedelt sind.

9. Datentransfer

Wird eine externe Plattform gewählt, so muss der **Transfer der generierten Daten** ins eigene Archivinformationssystem technisch sichergestellt werden.

Der Mehrwert nutzergenerierter Daten kommt erst dann richtig zum Tragen, wenn diese in die elektronischen Findmittel integriert sind und somit bei einer Suchanfrage einbezogen werden können.

10. Definition von Qualitätsstandards

Man sollte sich bei der Projektplanung darüber klar werden, welches **Qualitätsniveau** man bei den nutzergenerierten Inhalten anstrebt.

Dies gilt insbesondere dann, wenn die nutzergenerierten Daten ins eigene Archivinformationssystem übernommen werden sollen; ggf. sind redaktionelle Bearbeitungsschritte einzuplanen. Die Entscheidung hängt auch hier wiederum von den Projektzielen, den ausgewählten Beständen und den zu generierenden Inhalten ab. Aber auch von den Ressourcen, die für Qualitätskontrolle und Nachbearbeitung zur Verfügung stehen. Bei der Transkription von Urkunden wird man sicher hohe Qualitätsstandards ansetzen. Dann kann man jedoch unter Umständen nicht mit der Generierung hunderter oder tausender Transkriptionen in kurzer Zeit rechnen.

11. Definition der qualitätssichernden Maßnahmen

Der Umfang der **qualitätssichernden Maßnahmen** muss projektbezogen definiert und die Umsetzung geregelt werden.

Es gibt verschiedene Vorgehensweisen, die von einer engmaschigen und aufwändigen Prüfung der nutzergenerierten Inhalte bis zum bewussten gänzlichen Verzicht auf eine redaktionelle Kontrolle und Bearbeitung reichen können. Generell gilt, dass nutzergenerierte Inhalte als solche gekennzeichnet sein und für die Nutzer eindeutig von den archivarischen Erschließungsinformationen unterscheidbar sein sollten. Art und Umfang qualitätssichernder

Maßnahmen sind immer in Relation zu den Zielen des jeweiligen Projekts und zu dem für dieses Projekt als vertretbar erachteten Ressourcenaufwand zu sehen. Grundsätzlich gibt es verschiedene Möglichkeiten, die Qualität der nutzergenerierten Inhalte positiv zu beeinflussen.

Hier müssen wenige Beispiele genügen: Bei Transkriptionsprojekten empfiehlt es sich, Editions- oder Transkriptionsrichtlinien zu erstellen. Darüber hinaus könnte eine Qualitätssicherung durch das sogenannte „Double-Keying“ – also die Erfassung von Texten durch zwei voneinander unabhängige Teilnehmer – erfolgen. Es kann sinnvoll und arbeitersparend sein, Antworten auf die häufigsten Fragen der Teilnehmer in FAQs zur Verfügung zu stellen. Auch ein Forum, in dem sich die Teilnehmer bei Problemen an die Projektbetreuer wenden, aber sich auch untereinander austauschen können, kann bei größeren Projekten zur Qualitätssteigerung beitragen.

12. Entscheidung über Registrierung

Es ist zu entscheiden, ob sich die Projektteilnehmer registrieren müssen, es Ihnen als eine Option angeboten wird, oder ob auf eine **Registrierung** bewusst verzichtet wird.

Je nach Art des Projekts kann eine Registrierung Anreize zur wiederholten Mitarbeit bieten (vgl. auch Nr. 14). Andererseits kann sie aber auch die Hemmschwelle für eine Teilnahme erhöhen, da spontane Beiträge nicht möglich sind. Eine Registrierung kann auch eine qualitätssichernde Maßnahme sein (vgl. Nr. 11).

13. Definition eines Redaktionsworkflows

Entsprechend der angesetzten Qualitätskriterien und in Abhängigkeit vom Ablauf des ggf. zu leistenden Datentransfers und der zur Verfügung stehenden Ressourcen ist ein **Redaktionsworkflow** zu definieren.

Dabei ist zwischen der Veränderung (z.B. Modifikation des Titels, Berichtigungen, Anlage von Enthält-Vermerken etc.) und der Ergänzung schon vorhandener archivischer Erschließungsinformationen (z.B. Hinzufügen von Normdaten) zu unterscheiden.

14. Entscheidung über ein Anreizsystem

Es sollte darüber nachgedacht werden, **Anreize** für eine (längerfristige) Teilnahme zu schaffen.

Es kann zum Beispiel sinnvoll sein, die Teilnehmer durch die Vergabe von Punkten zu einer wiederholten Teilnahme zu animieren.

15. Öffentlichkeitsarbeit

Der Erfolg eines Crowdsourcing-Projekts hängt auch von der **Öffentlichkeitsarbeit** ab. Es sollte daher ein Werbekonzept für das Projekt geben.

Möglicherweise bietet sich als Aufhänger ein aktuelles Thema oder ein Jubiläum etc. an. Je nach Bestand lassen sich eventuell auch geeignete Multiplikatoren identifizieren (z.B. genealogische Vereine).

Vielen Dank für die Aufmerksamkeit!