

ERSCHLIEßUNG HISTORISCHER BESTÄNDE MITTELS CROWDSOURCING: EINE ANALYSE AUSGEWÄHLTER AKTUELLER PROJEKTE.

Simone Waidmann

Badische Landesbibliothek / Bibliotheksakademie Bayern

simone.waidmann@me.com

1. Einleitung

Allerorten entstehen derzeit digitale Sammlungen. Bibliotheken digitalisieren massenhaft ihre Altbestände und stellen sie ins Netz. Der Zugriff wird dadurch erleichtert, er ist jederzeit von jedem Ort der Welt aus möglich, während gleichzeitig die Originale geschont werden. Doch die Auswertung erfolgt immer noch wie seit Jahrhunderten von Hand. Maschinell durchsuchbare Volltexte historischer Bestände stehen in der Regel nicht zur Verfügung. Dies erschwert einerseits deren Auffindung durch Suchmaschinen, andererseits deren inhaltliche Auswertung. Zeitungen sind hervorragende historische Quellen, doch bisher müssen sich Wissenschaftler mühsam durch Massen hindurcharbeiten und auf Zufallsfunde hoffen. Eine systematische Auswertung ist kaum oder doch nur mit sehr viel Zeitaufwand möglich. Bei gedruckten Texten kann mit OCR zumindest teilweise Abhilfe geschaffen werden, doch die automatisch erzeugten Texte sind trotz allem technischen Fortschritt der letzten Jahre immer noch fehlerbehaftet. Bei Handschriften, insbesondere den sehr individuellen neuzeitlichen Handschriften, fällt OCR-Einsatz meist ganz aus. An einer Transkription durch menschliche Bearbeiter führt hier häufig kein Weg vorbei. Doch eine Bearbeitung aller in Bibliotheken vorhandenen historischen Drucke, Handschriften und Autographen durch wissenschaftliches Personal ist weder zeitlich noch finanziell zu leisten. Die Menge des Materials wird zudem vor allem im Bereich

der Nachlässe jeden Tag größer. Eine wissenschaftliche Edition ist aber in den wenigsten Fällen nötig oder sinnvoll. Einfache Transkripte, die zum Teil vielleicht auch fehlerhaft oder unvollständig sind, können schon sehr wertvoll sein.

Erst in den letzten Jahren kam man in Bibliotheken auf die Idee, die Bevölkerung an dieser Aufgabe der Volltexterschließung zu beteiligen. Das noch junge Phänomen des Crowdsourcing, das in der freien Wirtschaft und Industrie bereits einige Verbreitung erfahren hat, erhielt auch in Bibliotheken Einzug und interessierte Laien wurden aufgerufen, sich an der inhaltlichen Erschließung historischer Bestände zu beteiligen; sei es durch die Korrektur von durch OCR erzeugten Texten oder durch originäre Transkription. Historische Bestände sind einerseits besonders gut für Crowdsourcing geeignet, da historische Themen oft auch in der breiten Bevölkerung auf Interesse stoßen, insbesondere, wenn sie einen lokalen oder genealogischen Bezug aufweisen. Andererseits sind sie aber auch besonders schwierig zu bearbeiten, da zum Lesen von Handschriften häufig ein Mindestmaß an paläographischen Kenntnissen nötig ist. Dieser Beitrag beschränkt sich daher auf Projekte, in denen mit neuzeitlichen Quellen gearbeitet wird, da diese im Gegensatz zu mittelalterlichen Handschriften häufig in Volkssprachen abgefasst sind, wodurch zumindest die Sprachbarriere herabgesetzt wird. Während man in deutschen Bibliotheken mit Crowdsourcing noch recht vorsichtig ist, gibt es insbesondere im englischsprachigen Ausland bereits Erfahrungen mit einer ganzen Reihe von unterschiedlichen Projekten. Insgesamt drei Projekte aus Australien, England und Deutschland sollen in diesem Artikel vorgestellt werden. Diese unterscheiden sich sowohl in ihrem Gegenstand (Druck versus neuzeitliche Handschriften) als auch in Umfang, Personal- und Finanzausstattung, wodurch eine möglichst große Bandbreite aufgezeigt werden soll.

2. Was ist Crowdsourcing?

Den Begriff Crowdsourcing führte Jeff Howe erstmals 2006 ein.¹ Es handelt sich bei der Wortschöpfung um eine Kombination der Wörter „crowd“ und „outsourcing“ und meint im Wesentlichen, dass eine Aufgabe oder ein Problem, anstatt an ein Subunternehmen, an eine Masse freiwilliger Internetuser ausgelagert wird. Da es sich um ein noch junges Phänomen handelt, hat sich noch keine allgemein anerkannte Defini-

¹ Howe (2006).

tion herausgebildet. Enrique Estellés-Arolas und Fernando González-Ladrón-de-Guevara haben verschiedenen Definitionen von Crowdsourcing analysiert und diese zu einer umfassenden Begriffsbestimmung integriert:

Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage what the user has brought to the venture, whose form will depend on the type of activity undertaken.²

Besonders hervorzuheben ist der Einsatz des Internets, wodurch sich Crowdsourcing von traditionellem ehrenamtlichem Engagement unterscheidet, der Aufruf zur Beteiligung durch einen „open call“ und besonders im Kulturbereich der gegenseitige Nutzen für alle Beteiligten, wodurch ein Ausnutzen der freiwilligen Beitragenden vermieden wird.

3. Warum Crowdsourcing in Bibliotheken?

Die offensichtliche Antwort auf diese Frage könnte sein, um Ziele zu erreichen, für die Bibliotheken allein keine ausreichenden personellen und finanziellen Kapazitäten haben. Diese Sicht betont den Outsourcing-Aspekt von Crowdsourcing, das Auslagern von Arbeit. Doch in jüngerer Zeit greift immer mehr die Einsicht um sich, dass Crowdsourcing in Kultureinrichtungen mehr ist als das. Bibliotheksbenutzer werden von reinen Konsumenten zu Produzenten und Mitgestaltern. Sie können sich aktiv gerade mit historischen Bibliotheksbeständen auseinandersetzen, anstatt diese einfach nur zu betrachten. Dies stärkt das Gefühl für Kulturgut als Gemeineigentum und die gemeinsame Verantwortung für dessen Bewahrung. Die Nutzer erhalten die Gelegenheit, sich sozial zu engagieren, ihre Zeit und Expertise zum Wohl der Gesellschaft einzubringen und das ohne an bestimmte Orte und Geschäftszeiten gebunden

² Estellés-Arolas & González-Ladrón-de-Guevara (2012, S. 197).

zu sein.³ Indem man digitale Sammlungen mit Crowdsourcing Tools ausrüstet, überlässt man zudem den Nutzern die Entscheidung darüber, was transkribiert wird und es werden die Materialien bearbeitet, die auch tatsächlich genutzt werden. Nicht zuletzt wird die Bindung an die eigene Bibliothek gestärkt. Trevor Owens kommt zu dem Schluss, dass “crowdsourcing is the best way to actually engage our users in the fundamental reason that these digital collections exist in the first place.”⁴ Und Mia Ridge argumentiert für den Museumsbereich: „crowdsourcing helps museums serve their core missions by providing platforms for audiences to follow their own interests and hobbies and connect with communities of practice.“⁵ Bibliotheken haben zudem eine lange Tradition, was die Zusammenarbeit mit Ehrenamtlichen betrifft. Häufig gibt es Freundeskreise oder historische Vereine, die eine Bibliothek unterstützen oder sich speziell für bestimmte Bestände interessieren. Crowdsourcing ist letztlich auch eine Möglichkeit, dieses Engagement weiter auszubauen, den Kreis der Beitragenden zu erweitern und die Arbeitsweise mit Web 2.0 Technologie auf die nächste Stufe zu heben.⁶

Rose Holley von der National Library of Australia konnte in einer Untersuchung verschiedener für Bibliotheken relevanter Crowdsourcingprojekte feststellen, dass die Profile der Beitragenden stark variieren, was Alter oder berufliche Situation anbelangt. Es fanden sich aber immer auch Menschen mit Behinderung oder schwerer Krankheit unter den Beitragenden, die womöglich ans Haus gebunden waren. Diese berichteten, dass die Beteiligung an (nichtkommerziellen) Crowdsourcingprojekten ihrem Tag Struktur verlieh und ihnen das Gefühl gab, zu etwas Sinnvollem beizutragen. Damit erhielten die Projekte eine starke soziale Komponente, ohne dass dies überhaupt intendiert war. Viele Freiwillige hatten vor ihrer Beteiligung zudem nur geringe Computerkenntnisse, eigneten sich diese aber aufgrund von Interesse am Thema oder dem Wunsch zu einem größeren Ziel beizutragen an, womit Crowdsourcing auch eine Bildungskomponente erhält.⁷

³ So auch Proctor (2013).

⁴ Owens (2013, S. 128).

⁵ Ridge (2013, S. 446).

⁶ Vgl. dazu auch Owens (2013).

⁷ Vgl. Holley (2009c, S. 10-11). Gekürzte Version: Holley, R. (2010). Im Folgenden wird nur noch das ausführlichere Research Paper zitiert.

Natürlich geht man mit Crowdsourcing auch Risiken ein; gerade bei der Transkription historischer Bestände könnten Texte mutwillig manipuliert werden, man könnte von Spam überflutet werden, es könnte viel Personal binden oder das Angebot könnte schlicht nicht genutzt werden, womit die Investitionen umsonst gewesen wären.⁸ Diese Befürchtungen haben sich bisher aber nicht bestätigt. Vielleicht die größte Hürde stellt der Verlust von Macht und Kontrolle dar, der mit Crowdsourcing einhergeht, doch gerade darin liegt auch eine große Chance, da Nutzer durch die Möglichkeit der Beteiligung und Mitgestaltung die Bibliothek wieder als positiv und relevant wahrnehmen.⁹ Gerade in Deutschland scheint man sich auch besonders schwer damit zu tun, noch unfertige und nicht perfekte Inhalte zugänglich zu machen. Doch auch hier hat sich gezeigt, dass Nutzern dieses Rohmaterial lieber ist, als lange auf eine wissenschaftliche Edition zu warten oder womöglich gar nichts zu bekommen,¹⁰ und dass sie sich gerne an dessen Verbesserung beteiligen. Ihre Leistungen übertrafen üblicherweise sogar die Erwartungen der Initiatoren.¹¹

4. Was ist zu beachten?

Zur Beantwortung dieser Frage hat Rose Holley eine einfache Checkliste zusammengestellt:

The Thing	The System	The People	The Content
👍 Clear goal	👍 Easy and Fun	👍 Acknowledge	👍 Interesting
👍 Big challenge	👍 Reliable and quick	👍 Reward	👍 New
👍 Progress	👍 Intuitive	👍 Team support	👍 Lots
👍 Results	👍 Options	👍 Trust	👍 History/Science

Abb. 1: Rose Holleys Checkliste für Crowdsourcing. Holley, R. (2009c, S. 14).

⁸ Für eine ausführlichere Liste möglicher Risiken und Strategien zu deren Abschwächung vgl. Holley (2009a, S. 5-6).

⁹ Vgl. Holley (2009c, S. 24).

¹⁰ Vgl. Holley (2009a, S. 26).

¹¹ Vgl. Holley (2009c, S. 11).

Danach sollten Crowdsourcingprojekte ein klares Ziel haben, eine große Herausforderung darstellen und der Fortschritt und das Ergebnis sollten klar kommuniziert werden. Das System sollte schnell und zuverlässig arbeiten, intuitiv und einfach zu bedienen sein, dabei Spaß machen und den Nutzern möglichst verschiedene Optionen, was Aufgabe und Inhalt anbelangt, bieten (z. B. nur taggen, originär transkribieren oder korrigieren). Wichtig ist auch, den Beitrag der Freiwilligen zu würdigen, ihnen Vertrauen entgegen zu bringen und Unterstützung anzubieten. Der Inhalt sollte interessant und neu und es sollte reichlich zu bearbeitendes Material vorhanden sein. Es hat sich zudem gezeigt, dass historische oder naturwissenschaftliche Themen auf besonderes Interesse stoßen.¹²

Bisherige Erfahrungen zeigen, dass Beitragende üblicherweise schon ein hohes Maß an Motivation mitbringen, sie sind aber umso motivierter, je größer und herausfordernder die Aufgabe ist. Motivationssteigernd wirkt sich auch aus, regelmäßig neues Bearbeitungsmaterial bereitzustellen, den Fortschritt des Projekts etwa in Form von Statistiken zu kommunizieren und die Arbeit der Freiwilligen zum Beispiel in Form von Rankings zu würdigen. Die von Holley untersuchten Projekte hatten kein oder kaum Personal, um die freiwilligen Beitragenden zu koordinieren und zu moderieren, was den Unternehmungen aber keinen Abbruch tat. Die Crowd war in der Lage, sich in Foren und über andere Kommunikationsmöglichkeiten selbst zu organisieren und eigenständig Moderationsaufgaben zu übernehmen.¹³ Die Masse der Arbeit wurde zudem üblicherweise von einer kleinen Minderheit der Nutzer geleistet, womit die aktive Crowd also nicht unbedingt besonders groß sein muss.¹⁴ Es ist vielleicht weniger wichtig, besonders viele zu erreichen, sondern nur die Richtigen. Je nach Gegenstand können unterschiedliche Zielgruppen (z. B. Studenten, Wissenschaftler, historische oder genealogische Vereine) von besonderer Bedeutung sein. Diese gilt es dann gezielt anzusprechen.

Mia Ridge weist darauf hin, dass sich erfolgreiche Crowdsourcingprojekte durch besonders einfache Aufgaben auszeichnen. Je komplizierter die Aufgabe ist, desto enger muss die Nutzerführung sein und desto mehr Hilfsangebote und Mode-

¹² Vgl. Holley (2009c, S. 14).

¹³ Vgl. Holley (2009c, S. 12-13).

¹⁴ Vgl. Holley (2009c, S. 10).

ration sind notwendig, damit die Freiwilligen nicht überfordert werden.¹⁵ Große Aufgaben sollten in kleine „microtasks“ atomisiert werden, die schnell zu erledigen sind und damit ein Erfolgserlebnis verschaffen.¹⁶ Dies kann bei der Transkription historischer Bestände mitunter schwierig werden. Eine Möglichkeit könnte sein, anstatt eines kompletten Textes oder einer ganzen Seite nur eine Zeile oder einzelne Wörter transkribieren zu lassen. Wie praktikabel dies ist, hängt sicher auch immer vom konkreten Gegenstand ab, in jedem Fall sollte aber die Möglichkeit bestehen, größere Aufgaben zu unterbrechen und zu einem späteren Zeitpunkt fortzusetzen.

5. Australian Newspapers Digitisation Program

Das Australian Newspapers Digitisation Program (ANDP) ist ein großangelegtes Projekt zur Digitalisierung sämtlicher urheberrechtsfreien australischen Zeitungen, angefangen vom Erscheinen der ersten australischen Zeitung im Jahr 1803 bis einschließlich 1954. Beteiligt sind alle australischen Staats- und Territorialbibliotheken unter der Leitung der National Library of Australia. Ab Ende 2006 wurden dafür 10 Millionen australische Dollar bereitgestellt, digitalisiert wurde von bereits vorhandenen Mikrofilmen, was wesentlich kostengünstiger war als eine Neuaufnahme der gedruckten Originale.¹⁷ Die Digitalisate wurden mittels OCR volltexterschlossen. Sowohl die Digitalisierung als auch die OCR-Bearbeitung erfolgten durch externe Dienstleister. Die elektronisch erstellten Texte wurden dann in ALTO XML Files geliefert und in eine SQL-Datenbank eingespeist.¹⁸ Publiziert werden die Texte unter der Lizenz CC-BY-NC-SA 2.1 AU, auch alle durch die Crowd eingebrachten Korrekturen stehen unter dieser Lizenz.¹⁹

Aufgrund schlechter Druckqualität insbesondere der frühen Zeitungen, variierender Drucktypen und hoher Zeichendichte, war die Qualität der mittels OCR erstellten Texte teilweise sehr schlecht.²⁰ Insbesondere deutsche Zeitungen, die in Fraktur gedruckt wurden, stellen hierbei eine besondere Herausforderung dar. Um

¹⁵ Vgl. Ridge (2013, S. 439).

¹⁶ Vgl. Ridge (2013, S. 437).

¹⁷ Vgl. Holley (2009a, S. 3) und Holley, R. (2009b, S. 285-286). Für eine genaue Beschreibung der technischen Umsetzung und der Workflows vgl. ebenfalls Holley (2009b), insbesondere S. 286-289.

¹⁸ Vgl. Holley (2009a, S. 8) und Holley (2009b, S. 287-288).

¹⁹ Vgl. <http://trove.nla.gov.au/general/termsfuse>.

²⁰ Vgl. Holley (2009b, S. 289).

die Auffindbarkeit zu verbessern, wurden als erste Qualitätssicherungsmaßnahme Titel, Untertitel, Autor und die ersten vier Zeilen eines jeden Artikels manuell nachbearbeitet.²¹ Erst während des laufenden Projekts kam die Idee auf, die Öffentlichkeit in die Textkorrektur mit einzubeziehen.²² Ein Betasystem ging am 25. Juli 2008 online. In den ersten sechs Monaten wurden dort 3,5 Millionen Artikel und die Möglichkeit zur Textkorrektur einzelner Zeilen zur Verfügung gestellt. Aus Sorge, einem massenhaften Nutzeransturm nicht gewachsen zu sein, wurden der neue Service und insbesondere die Möglichkeit zur aktiven Beteiligung nicht beworben. Trotzdem entwickelte sich in diesen sechs Monaten eine aktive Community von ca. 1.300 Textkorrektoren, die in dieser Zeit zwei Millionen Zeilen in 100.000 Artikeln verbesserten.²³ Es gab keinerlei Moderation und auch keine Korrekturrichtlinien. Die Community entwickelte sich völlig autonom. Die bereitgestellten Funktionen waren Textkorrektur sowie das Anbringen von öffentlichen oder privaten Kommentaren und Tags. Da die Nutzer zu diesem Zeitpunkt keine Möglichkeit hatten, sich über persönliche Nachrichten oder in Foren auszutauschen, nutzten sie die Kommentarfunktion, um miteinander in Kontakt zu treten. Aufgrund von Nutzerfeedback wurde die Plattform immer wieder angepasst und erweitert, inzwischen ist sie mit Funktionen eines sozialen Netzwerks ausgestattet. Die registrierten Nutzer können eigene Profile erstellen, sich gegenseitig befreunden, sich persönliche Nachrichten schicken, ihre eigenen Tags und Kommentare verwalten, ihren Korrekturfortschritt verfolgen und in Foren diskutieren. Auch Korrekturrichtlinien wurden inzwischen von den Nutzern selbst erstellt.²⁴ Durch die vielfältigen Funktionalitäten wird die Plattform zu einem echten Arbeitsinstrument für die Nutzer, mit dem sie Inhalte zu ihren Forschungsthemen verwalten können. Dadurch entsteht ein echter Mehrwert für die Beitragenden, der wiederum Motivation bietet, sich am Projekt zu beteiligen.

Das Zeitschriftenportal ist ein Teil von Trove,²⁵ der digitalen Bibliothek Australiens. Ständig werden neue Inhalte hinzugefügt. Die digitalisierten Zeitungen kann man entweder durchblättern oder gezielt nach einzelnen Artikeln suchen. Dabei ist auch eine Suche über Zeilen hinweg möglich, auch bei getrennten Wörtern, sofern

²¹ Vgl. Holley (2009b, S. 287).

²² Vgl. Holley (2009a, S. 4).

²³ Vgl. Holley (2009a, S. 7).

²⁴ <http://trove.nla.gov.au/general/text-correction-guidelines-for-digitised-newspapers/>.

²⁵ <http://trove.nla.gov.au/>.

rierung nicht zwingend notwendig ist, also besonders niedrig. Es hat sich aber gezeigt, dass die meisten Beitragenden, insbesondere wenn sie den Service häufiger nutzen, sich doch registrieren.²⁷

Das Projekt hat nicht den Anspruch einer wissenschaftlichen Edition, es will lediglich die Durchsuchbarkeit und Nutzbarkeit verbessern. Wer es ganz genau wissen will, für den ist immer noch das Original oder ersatzweise der Scan ausschlaggebend. Der von den Freiwilligen korrigierte Text wird daher auch nicht noch einmal durch Projektmitarbeiter überprüft, sondern ist nach dem Speichern sofort für jedermann sichtbar. Allerdings wird nichts überschrieben, alle vorangehenden Textversionen werden weiterhin gespeichert und bleiben durchsuchbar.²⁸

Bis Februar 2014 wurden von Freiwilligen über 120 Millionen Zeilen korrigiert, mehr als 2,6 Millionen Tags vergeben und fast 70.000 Kommentare verfasst. Über 120.000 Nutzer haben sich registriert, von denen zuletzt monatlich durchschnittlich 7.000 bis 8.000 aktiv waren.²⁹ Zur Motivationssteigerung gibt es ein „Hall of fame ranking“, in dem sich jeder registrierte Nutzer anzeigen lassen kann, welchen Platz er gerade in der Rangfolge der Textkorrektoren einnimmt. Die Top 5 der Beitragenden werden auch immer auf der Startseite der digitalen Zeitungen angezeigt, wobei der fleißigste Beitragende allein fast 2,5 Millionen Zeilen korrigiert hat.³⁰ Bis zu 80 % der Arbeit werden von etwa 10 % der User geleistet.³¹

6. Transcribe Bentham

Das Crowdsourcingprojekt Transcribe Bentham ordnet sich in den größeren Zusammenhang des bereits 1958 begonnenen Bentham Projects am University College London ein. Ziel dieses Projekts ist es, eine wissenschaftliche Edition der gesammelten Werke des Juristen, Philosophen und Sozialreformers Jeremy Bentham (1748-1832) herauszugeben. Seine unveröffentlichten Manuskripte umfassen etwa 72.500 Folioseiten. Von 1958 bis 2010 wurde Benthams Nachlass auf herkömmliche Weise durch Projektmitarbeiter ediert. Etwa 20.000 Seiten wurden auf diese Art in 29 von

²⁷ Vgl. Holley, R. (2009a, S. 12) und Holley, R. (2009b, S. 292).

²⁸ Vgl. Holley, R. (2009b, S. 289).

²⁹ Vgl. <http://trove.nla.gov.au/system/stats?env=prod>.

³⁰ <http://trove.nla.gov.au/newspaper?q>

³¹ Vgl. Holley (2009c, S. 10).

angestrebten 70 Bänden veröffentlicht. Über zwei Drittel des Materials waren damit auch nach über einem halben Jahrhundert noch immer gänzlich unbearbeitet.³² Um den Editionsprozess voranzutreiben, beschloss man daher 2010, die Bevölkerung in einem großangelegten Crowdsourcingprojekt um Mithilfe zu bitten. Vom Arts and Humanities Research Council erhielt man Mittel in Höhe von 262.673 Pfund, die ab April 2010 über ein Jahr ausgeschüttet wurden. Der größte Teil davon wurde für die Digitalisierung der Manuskripte benötigt, ein zweiter Teil für die Entwicklung des Editionswerkzeugs. Die dritte Säule bildeten Personalkosten, insbesondere zwei Vollzeitassistenten für zwölf Monate.³³

Transcribe Bentham ist ein äußerst ambitioniertes Crowdsourcingprojekt, allein die Masse ist enorm. Hinzu kommt, dass die Manuskripte zwar von einem einzigen Schreiber stammen, dessen Handschrift mit zunehmendem Alter aber immer unleserlicher wurde. Häufig handelt es sich um Konzepte mit vielen Durchstreichungen, nachträglichen Einfügungen und Marginalien.³⁴ Dies erschwert nicht nur die Lesbarkeit, sondern geht auch mit erhöhten Anforderungen an die Codierung der Transkriptionen einher. Da die Projektergebnisse in eine wissenschaftliche Edition einfließen, sollte der Text in TEI-konformem XML codiert werden. Dadurch werden auch anspruchsvolle Suchanfragen ermöglicht, beispielsweise Suchen nur in den Marginalien oder in gestrichenem Text.³⁵ Es galt also ein Werkzeug zu entwickeln, das diesen komplexen Anforderungen gerecht wird, gleichzeitig aber auch von technischen Laien bedient werden kann. Das Ergebnis ist der Wiki-basierte Bentham Transcription Desk. In der Editions Umgebung wird links das Textfeld für die Transkription, rechts die digitalisierte Manuskriptseite angezeigt. Das Digitalisat lässt sich ganz nach Bedarf vergrößern. In das Eingabefeld gibt man reinen Text ein. Über die Werkzeugleiste kann man Zeilenumbrüche, Seitenumbrüche, Überschriften, Paragraphen, Streichungen und Hinzufügungen einfügen, fragliche oder unleserliche Wörter kennzeichnen, einige Auszeichnungen wie Unterstreichungen oder Hochstellungen vornehmen, fremdsprachlichen Text markieren und Kommentare hinzufügen. Ohne

³² Vgl. Causer, Tonra & Wallace (2012, S. 120).

³³ Vgl. Causer, Tonra & Wallace (2012, S. 121).

³⁴ Vgl. Moyle, Tonra & Wallace (2011, S. 350).

³⁵ Vgl. Causer, Tonra & Wallace (2012, S. 125).

dass der Bearbeiter Code kennen muss, wird der Text über die Werkzeugleiste in TEI-konformes XML gebracht.

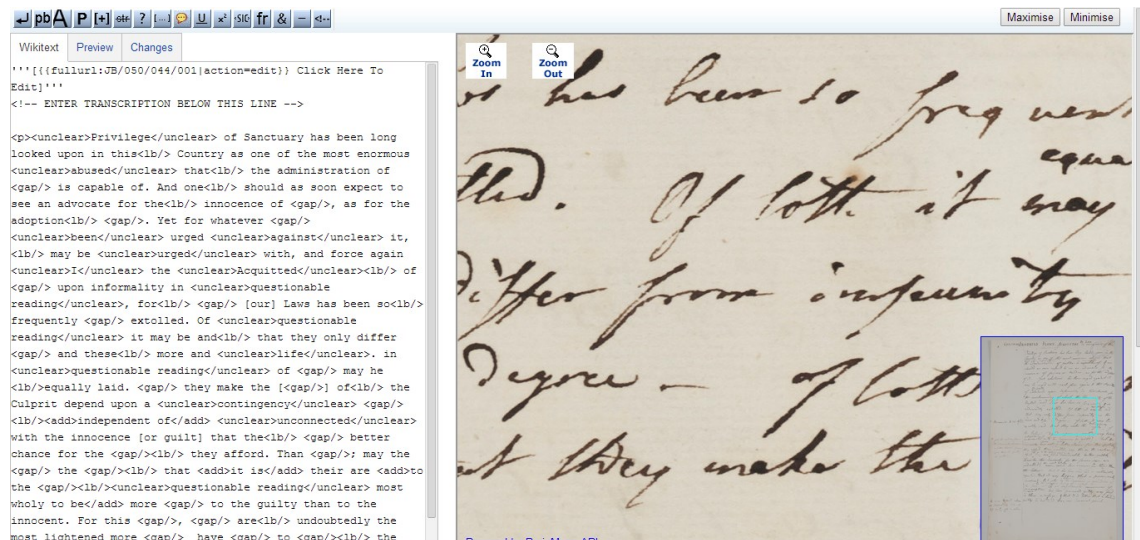


Abb. 3: Bentham Transcription Desk

Im Bentham Transcription Desk werden das Textfeld für die Eingabe und das Digitalisat der Handschrift nebeneinander angezeigt. Die Größe des Digitalisats kann individuell angepasst werden. Über die Werkzeugleiste lässt sich der Text TEI-konform auszeichnen, der Code bleibt für den Bearbeiter aber immer sichtbar. (<http://www.transcribe-bentham.da.ulcc.ac.uk/td/index.php?title=JB/050/044/001&action=edit>)

Der Code wird zwar automatisch erzeugt, bleibt für den Bearbeiter aber sichtbar, was auf einige Nutzer abschreckend wirkt. Die Umstellung des Editors auf ein What-You-See-Is-What-You-Get-Prinzip (WYSIWYG) ist zwar seit längerem geplant, bisher aber nicht realisiert.³⁶ Der Transcription Desk ist Open Source und kann für andere Projekte nachgenutzt werden.³⁷ Die transkribierten Texte werden unter der Lizenz CC-BY-NC-ND 3.0 veröffentlicht.³⁸ Die Verwendung einer NoDerivatives-Lizenz erscheint verwunderlich, da sie Bearbeitungen untersagt, was aber das zentrale Element von Crowdsourcing ist. Auf Nachfrage beim Projekt, weshalb man sich für diese Lizenz entschieden hat, wollte man diese Praxis überdenken.

Nach einer kurzen Onlineregistrierung kann man sofort mit dem Transkribieren beginnen. Um den Einstieg zu erleichtern, gibt es Video-Tutorials, in denen der Umgang mit dem Editionswerkzeug erklärt wird; darüber hinaus Hilfetexte, paläo-

³⁶ Vgl. Causer, Tonra & Wallace (2012, S. 131).

³⁷ Vgl. Moyle, Tonra & Wallace (2011, S. 347).

³⁸ Vgl. http://www.transcribe-bentham.da.ulcc.ac.uk/td/Code_of_conduct.

graphische Informationen, ausführliche Editionsrichtlinien, ein Diskussionsforum und die Möglichkeit Projektmitarbeiter zu kontaktieren. Doch diese Informationsflut kann anfänglich auch abschreckend wirken.

Zu bearbeitende Dokumente können nach Thema, Zeitabschnitt oder Schwierigkeitsgrad ausgewählt werden. Jeder User kann auf alle Dokumente zugreifen und auch Fehler anderer Beitragenden korrigieren. Fast vier Jahre nach Beginn des Projekts ist es gerade als Neueinsteiger allerdings schwierig, noch geeignetes Material zu finden. Die einfachen Stücke sind alle bearbeitet. Die noch unbearbeiteten Dokumente sind einerseits schwer zu finden, andererseits sind sie besonders kompliziert, da sie eine sehr unleserliche Handschrift, viele Durchstreichungen und Anmerkungen aufweisen.

Wie beim australischen Projekt können die Benutzer auch hier ein eigenes Profil anlegen, wie man es aus sozialen Netzwerken kennt; mit Angaben zur Person, Beruf, Interessen und einem eigenen Profilbild. Die Beitragenden können sich gegenseitig als Freunde hinzufügen und sich Nachrichten schicken. Dokumente, deren Fortschritt man verfolgen möchte, können in eine „Watchlist“ aufgenommen werden. Um die Motivation der Teilnehmer zu steigern, werden für verschiedene Aktivitäten Punkte vergeben und daraus ein Ranking der Teilnehmer erstellt. Die Top 20 der Beitragenden werden auf der Startseite angezeigt, ebenso wie der Gesamtfortschritt des Projekts.³⁹

Wenn man der Meinung ist, eine Seite fertig bearbeitet zu haben, schickt man den Herausgebern eine Nachricht. Die Transkription wird dann begutachtet und bei ausreichend hoher Qualität für die weitere Bearbeitung gesperrt. Ist der Herausgeber aber der Meinung, dass durch die Crowd noch weitere Verbesserungen möglich sind, bleibt das Dokument für die Bearbeitung frei.⁴⁰ Insgesamt ist die Qualität der eingereichten Texte sehr hoch.⁴¹

Der Transcription Desk ging am 8. September 2010 online und hatte ab diesem Zeitpunkt noch eine volle Personalausstattung für ein halbes Jahr. In diesen sechs Monaten hatte der Transkription Desk Besucher aus 49 Ländern, 1.207 Nutzer registrierten sich, wobei die Zahlen nach einem Bericht der New York Times am 27.

³⁹ Vgl. http://www.transcribe-bentham.da.ulcc.ac.uk/td/Transcribe_Bentham.

⁴⁰ Vgl. Causer, Tonra & Wallace (2012, S. 128).

⁴¹ Vgl. Causer, Tonra & Wallace (2012, S. 130).

Dezember 2010 enorm stiegen. In diesem ersten Halbjahr wurden 1.009 Manuskripte bearbeitet, von denen 569 (56 %) in diesem Zeitraum auch abgeschlossen wurden. Von den registrierten Nutzern haben im genannten Zeitraum allerdings nur 21 % auch tatsächlich transkribiert. Die sieben aktivsten Nutzer leisteten zusammen 70 % der Arbeit. Die wirklich aktive Crowd, welche die Hauptlast trägt, ist also auch hier relativ überschaubar. In einer Benutzerumfrage sollte herausgefunden werden, warum viele registrierte Nutzer dann doch vor der tatsächlichen Bearbeitung der Manuskripte zurückschrecken. Als Gründe wurden die umfangreichen und komplizierten Anleitungen genannt, Schwierigkeiten, unbearbeitetes Material zu finden oder die Handschrift zu entziffern sowie eine abschreckende Wirkung des verwendeten XML-Codes. Feedback und Moderation seitens der Projektmitarbeiter waren daher besonders wichtig, um diese Hemmnisse abzubauen. Als es aber nach der sechsmonatigen Testphase keine volle Personalausstattung mehr gab, stellten viele Freiwillige daraufhin die Arbeit ein. Die Testphase von sechs Monaten war für eine so komplexe Aufgabe einfach zu kurz, um eine sich selbst moderierende Community aufzubauen.⁴²

Was den Erfolg des Projekts anbelangt, so hat sich zumindest die Geschwindigkeit der Bearbeitung nicht erhöht. Hätte man das wissenschaftliche Personal, das mit organisatorischen und Moderationsaufgaben betraut war, transkribieren lassen, so hätten diese im ersten Jahr etwa den zweieinhalbfachen Output der Crowd produziert. Hätte man mit dem Geld für die Digitalisierung zwei wissenschaftliche Mitarbeiter für drei Jahre angestellt, hätten diese etwa die Hälfte des noch unbearbeiteten Materials transkribieren könne. Doch dafür hätte man einerseits keine Geldgeber gefunden, andererseits berücksichtigt eine solche Rechnung wiederum nur den reinen Outsourcing-Aspekt, nicht aber den eingangs erwähnten Mehrwert von Crowdsourcing. Es ist zudem davon auszugehen, dass mit Einführung eines WYSIWYG-Editors die Effizienz gesteigert wird und sich mit der Zeit auch eine Community aufbauen lässt, die eigenständig Moderationsaufgaben übernimmt, wodurch der Personalaufwand gesenkt würde.⁴³

Das Projekt wird in einem eigenen Blog, auf Facebook und auf Twitter beworben. Zudem gibt es Kommunikationsstrategien für Schulen, akademischen Sek-

⁴² Vgl. Causer, Tonra & Wallace (2012, S. 125-127,132).

⁴³ Vgl. Causer, Tonra & Wallace (2012, S. 130-131).

tor und Hobby-Historiker.⁴⁴ Aktuell (Stand 20.03.2014) sind 8.333 Nutzer registriert, 7.210 Folios wurden transkribiert, womit das Projekt zu 38,15 % abgeschlossen ist.⁴⁵

7. Edition des Nachlasses Franz Brümmer

Sehr viel bescheidener als die beiden vorangehenden Projekte kommt die digitale Edition des Nachlasses Franz Brümmer daher, insbesondere was finanzielle und personelle Ausstattung angeht. Ziel des Projekts ist eine reine Online-Edition, die nach und nach entsprechend dem Transkriptionsfortschritt freigeschaltet wird. Das Projekt startete 2006 und hat keine bestimmte Laufzeit. Es entstand aus einer Kooperation der Humboldt-Universität und der Staatsbibliothek zu Berlin – Preußischer Kulturbesitz. In Praxisseminaren sollten Germanistikstudenten die Möglichkeit haben, sich mit archivalischen Primärquellen zu beschäftigen und durch deren Transkription und Edition philologische Schlüsselkompetenzen erwerben.⁴⁶ Mit einer kleinen Anschubfinanzierung, die zu gleichen Teilen von der Staatsbibliothek und der HU getragen wurde, wurde daher 2006 bei der 3-point concepts GmbH die Entwicklung des Refine!Editors und der Webseite zur Publikation der digitalen Edition in Auftrag gegeben. Hauptamtliches Personal stand zu keinem Zeitpunkt zur Verfügung.⁴⁷ Inhaltlicher Gegenstand des Projekts ist der Nachlass Franz Brümmer, der in der Staatsbibliothek zu Berlin verwahrt wird.

Franz Brümmer (1836-1923) war Lehrer und nebenberuflich Herausgeber des Lexikons der deutschen Dichter und Prosaisten des 19. Jahrhunderts. Zu diesem Zweck rief Brümmer in literarischen Zeitschriften dazu auf, ihm biographische und bibliographische Informationen zu deutschsprachigen Literaten zukommen zu lassen. Unter den Zusendungen befinden sich auch etwa 6.000 eigenhändige Autobiographien deutscher Schriftsteller seiner Zeit.⁴⁸ Brümmers Nachlass ist damit eine Fundgrube, vor allem für nicht kanonisch gewordene Autoren.⁴⁹ Das Lexikon er-

⁴⁴ Vgl. Moyle, Tonra & Wallace (2011, S. 354).

⁴⁵ Vgl. http://www.transcribe-bentham.da.ulcc.ac.uk/td/Transcribe_Bentham.

⁴⁶ Vgl. Martin & Thomas (2008, S. 208). Siehe auch Weber (2013, S. 30).

⁴⁷ Thomas (2014a).

⁴⁸ Vgl. Martin & Thomas (2008, S. 207) und <http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/>. Siehe auch Middell (2007, S. 305-306).

⁴⁹ Vgl. http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/franz_bruemmer/nachlass.html.

schien erstmals 1885 und letztmals 1913 in 6. Auflage.⁵⁰ Brümmer sammelte bis zu seinem Tod 1923 weitere Informationen und Autobiographien, die aber nicht mehr veröffentlicht wurden. Gerade diesen nicht mehr ausgewerteten Materialien gilt das vorrangige Interesse der digitalen Edition.⁵¹ 2006/2007 wurden von den teilnehmenden Studierenden in zwei Semestern ca. 100 Nachlassmappen transkribiert und gleichzeitig in der Nachlassdatenbank Kalliope erschlossen.⁵²

Obwohl als Studentenprojekt gestartet, steht es auch anderen Interessierten frei, sich an der Transkription zu beteiligen. Der Aufruf zur Mitarbeit findet sich allerdings etwas versteckt auf der Projektwebseite unter Kontakt⁵³ sowie auf den jeweiligen Seiten noch nicht transkribierter Mappen, die über das Autorenregister zugänglich sind. Damit erfüllt das Projekt das Kriterium des „open calls“, wenngleich dieser etwas leise ist. Es ist daher nicht verwunderlich, dass die Crowd mit bisher 67 Personen⁵⁴ relativ klein ausfällt, doch wie bereits bei den beiden vorangehenden Projekten gesehen, ist die Größe der Crowd nicht entscheidend, der Hauptteil der Arbeit wird ohnehin von einer kleinen aktiven Minderheit getragen. Die Anmeldung erfolgt über ein Kontaktformular. Anders als beim Australian Newspapers Digitisation Program oder bei Transcribe Bentham, wo man nach einer Onlineregistrierung direkt mit der Bearbeitung beginnen kann, muss man hier unter Umständen mehrere Tage warten, bis einer der ehrenamtlichen Moderatoren die Zeit gefunden hat, einen Benutzeraccount einzurichten und dem Beitragenden seine persönlichen Zugangsdaten zukommen zu lassen. Die Anmeldung erfolgt für eine bestimmte Mappe des Nachlassmaterials, weitere unbearbeitete Mappen kann man im internen Redaktionsbereich durch einen Klick auf „Status“ für sich reklamieren und dann auch bearbeiten. Eine gegenseitige Kontaktmöglichkeit für registrierte Beitragende besteht nicht.

Der Refine!Editor wurde für Mozilla Firefox optimiert. Dort steht neben einer Textbox, die nach dem WYSIWYG-Prinzip arbeitet, auch eine Werkzeugleiste zur Verfügung, mit der die Beitragenden den Text TEI-konform formatieren können, ohne mit der Darstellung des Codes belastet zu werden. Folgende Elemente stehen in der Bearbeitungsleiste zur Auswahl: Seitenumbruch einfügen, Sonderzei-

⁵⁰ Vgl. http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/franz_bruemmer/lexikon.html.

⁵¹ Vgl. Thomas (2009, S. 619).

⁵² Vgl. Martin & Thomas (2008, S. 208).

⁵³ <http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/kontakt/>.

⁵⁴ <http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/projekt/mitarbeiter.php>.

chen einfügen, Fettdruck, Kursivierung, Unterstreichung, Hochstellung, Tiefstellung, Textausrichtung linksbündig, zentriert oder rechtsbündig, Formatierung entfernen, Verknüpfung zum Index herstellen (indexiert werden Personen, Institutionen, Orte, Absendedatum, Periodika und Werke), Indizes der Mappe verwalten, Anzeige in HTML. Mit anderen Browsern wie Microsoft Internet Explorer oder Google Chrome funktioniert dies allerdings nicht. Die Textbox zeigt in diesen Browsern HTML-Code an, die Werkzeugleiste fehlt ganz. Da nirgends auf der Projektwebseite auf diesen Umstand hingewiesen wird, kann es zu Problemen führen, wenn freiwillige Beitragende standardmäßig nicht mit Firefox arbeiten. Unabhängig vom Browser muss das Digitalisat neben der Editions Umgebung immer in einem zweiten Fenster geöffnet werden, es lässt sich nicht aus dem Editor heraus aufrufen.

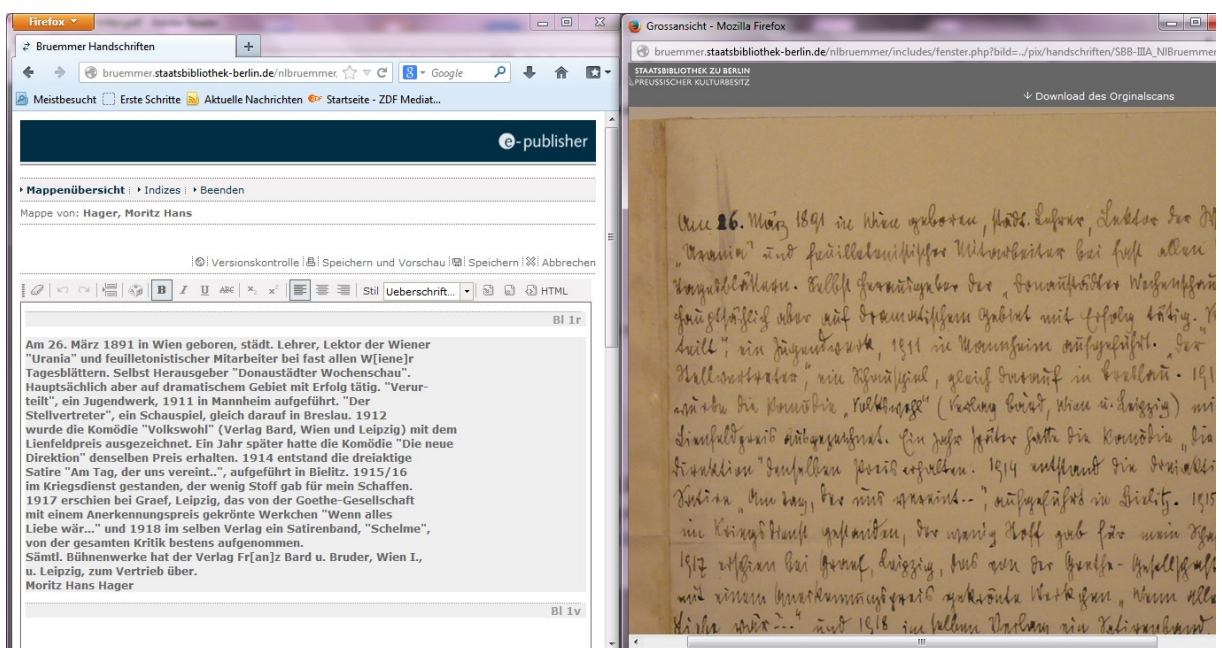


Abb. 4: Refine!Editor mit Digitalisat

Links der Refine!Editor mit Textfeld und Werkzeugleiste, rechts in einem zweiten Fenster das Digitalisat der Handschrift. (http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/redaktion/handschriften_update.php?id=320 (interner Login-Bereich) und <http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/includes/fenster.php?bild=../pix/handschriften/SBB-III-A-NlBruemmer Biographien Suppl I Hager Moritz BI1r.jpg>)

Formale Editionsrichtlinien gibt es nicht. Für die beteiligten Studenten gab es lediglich ein Arbeitspapier, in welchem die Funktionsweise des Editors erklärt und einige Formatierungshinweise gegeben wurden. Externen Beitragenden steht dies

leider nicht zur Verfügung.⁵⁵ Zur Qualitätskontrolle wird jede Transkription noch einmal manuell durch einen Redakteur geprüft. Erst nach erfolgter Kontrolle wird der Text für jedermann sichtbar im Internet freigeschaltet,⁵⁶ davor ist er nur angemeldeten Benutzern im internen Bereich zugänglich.

Der Refine!Editor codiert den transkribierten Text in HTML. Da das Redaktionssystem (Backend) direkt mit der Projektwebseite (Frontend) verknüpft ist, kann nach erfolgter Qualitätskontrolle direkt und ohne Umwege aus dem Editor heraus publiziert werden. Mit einem Zusatzprogramm, das nicht Teil des Editors ist, können zusätzlich zur HTML-Ausgabe auch XML-Dokumente erzeugt werden, die dem TEI-Standard entsprechen. Refine! ist Open Source und kann für andere Projekte nachgenutzt und angepasst werden, was in der Staatsbibliothek zu Berlin mit der Edition des Briefnachlasses Friedrich Nicolais auch schon erprobt wurde.⁵⁷ Statistiktools stehen im Refine!Editor nicht zur Verfügung. Entsprechend gibt es auf der Projektwebseite auch keine Anzeigen über den Fortschritt der Transkriptionsleistungen oder Rankings von Top-Beitragenden. Eine Lizenz wird für die veröffentlichten Transkriptionen momentan nicht vergeben, die Vergabe einer CC-BY Lizenz wird aber diskutiert.⁵⁸

Die digitale Edition des Nachlasses Franz Brümmers ist im Gegensatz zu Transcribe Bentham ein kleines Projekt, das lediglich eine minimale Anschubfinanzierung und keinerlei hauptamtliche Personalausstattung hatte. Die Crowd und deren Output sind dementsprechend eher klein, trotzdem wurde einiges erreicht. Auf die Frage, ob sich der Aufwand gelohnt hat, antwortete Projektmitarbeiter Christian Thomas mit einem klaren Ja.⁵⁹ Auch wenn jeder von Freiwilligen transkribierte Text noch einmal überprüft werden muss, ist der dadurch entstehende Aufwand viel geringer, als alles selbst zu transkribieren. Die beteiligten Studenten konnten durch das Projekt berufspraktische Erfahrungen sammeln, für manche war es gar der Einstieg ins Berufsleben. Durch dieses Projekt kann der Nachlass Brümmer der Forschung zwar langsam, aber stetig im Volltext zur Verfügung gestellt werden. Mit dem Refine!Editor wurde zudem ein simples, aber leistungsstarkes Editionswerkzeug ge-

⁵⁵ Thomas (2014a).

⁵⁶ Vgl. Martin & Thomas (2008, S. 210).

⁵⁷ Vgl. Martin & Thomas (2008, S. 210-211) und Falk (2013, S. 38-40).

⁵⁸ Thomas (2014b).

⁵⁹ Thomas (2014a).

schaffen, das durch das WYSIWYG-Prinzip technisch ungeübten Laien eine Beteiligung ermöglicht und auch die einfache Erstellung von Indizes unterstützt. Durch ein zusätzliches Software-Modul im Frontend können außerdem TEI-konforme Dokumente zur flexiblen Weiterverarbeitung und Archivierung erzeugt werden.

8. Vergleich der Projekte

Die Projekte unterscheiden sich ganz erheblich in Bezug auf Gegenstand, Finanz- und Personalausstattung, Nutzungsbedingungen und die eingesetzten Tools. Das Australian Newspapers Digitisation Program hat die meisten aktiven Nutzer und produziert rein zahlenmäßig den größten Output. Dies hat einerseits damit zu tun, dass die gebotenen Inhalte für eine breite Masse (z. B. Hobbyhistoriker und Familienforscher) relevant und die Aufgabe zugleich einfach ist. Gedruckter Text lässt sich leichter lesen als Handschriften, zudem ist bereits Text vorhanden, der nur an einzelnen Stellen korrigiert und nicht komplett abgetippt werden muss, was ein schnelles Vorankommen und damit Erfolgserlebnisse sichert. Das Projekt kam daher ganz ohne Marketing aus und nur durch die Bereitstellung geeigneter Werkzeuge und Kommunikationsmöglichkeiten entwickelte sich ganz autonom eine sich selbst moderierende Community. Eine Qualitätskontrolle durch wissenschaftliches Personal ist daher (und auch weil der ursprüngliche Text durchsuchbar bleibt) nicht notwendig.

Projekte wie Transcribe Bentham oder der Nachlass Brümmer sind dagegen nur für einen eingeschränkteren Personenkreis interessant. Marketing ist hier besonders wichtig, um den ohnehin beschränkten Nutzerkreis auch zu erreichen, was dem Bentham Projekt gut gelungen ist. Die Komplexität der Aufgabe kann außerdem abschreckend wirken. Allein die nötigen Paläographiekenntnisse dürften viele von einer Beteiligung abhalten. Hinzu kommen Berührungsängste mit der Technik, insbesondere wenn Nutzer mit Code konfrontiert werden und kein WYSIWYG-Editor verwendet wird. Eine enge Nutzerführung, Hilfsangebote und Moderation sind hier besonders wichtig. Qualitätskontrolle ist bei Handschriftentranskriptionen sinnvoll und, besonders wenn die Ergebnisse in eine wissenschaftliche Edition einfließen sollen, nahezu unabdingbar. Dies alles geht wiederum mit erhöhtem Personalaufwand einher. Dass es sich dabei nicht unbedingt um hauptamtliches Personal handeln muss, hat das Projekt zur Edition des Brümmer Nachlasses gezeigt und auch,

dass Crowdsourcing mit äußerst geringen Finanzmitteln möglich ist. Zwar ist der Fortschritt hier langsamer, doch das Ergebnis ist ebenso professionell. In allen drei Projekten wurden zudem Werkzeuge für die verschiedensten Bedürfnisse entwickelt, die nun anderen Bibliotheken als Open Source zur Verfügung stehen.⁶⁰ Durch deren Nachnutzung in anderen Crowdsourcingprojekten können Kosten gespart und somit die Effizienz gesteigert werden.

Nicht nur bei den hier vorgestellten, sondern auch bei anderen Crowdsourcingprojekten lässt sich ein klarer Trend erkennen, die Plattformen mit Strukturen sozialer Netzwerke auszustatten. Dies ist insbesondere dann wichtig, wenn wenig Personal zur Verfügung steht und die Freiwilligen selbständig Moderationsaufgaben übernehmen sollen. Außerdem ist davon auszugehen, dass die Nutzerbindung dadurch gestärkt wird. Je nachdem mit welchen Funktionalitäten sie ausgestattet sind, können die Plattformen sogar zu echten Arbeitsinstrumenten für die Nutzer werden, wie das australische Beispiel gezeigt hat. Das Brümmer-Projekt zeigt, dass es grundsätzlich aber auch ohne solche Funktionen geht. Zur Motivationssteigerung haben sich Fortschrittsbalken und Rankings der aktivsten Benutzer als sinnvoll erwiesen und weitgehend etabliert.

In keinem der Projekte haben sich irgendwelche Befürchtungen, wie bewusste Manipulation der Texte oder das Ausbleiben von Freiwilligen, bewahrheitet. Es hat sich im Gegenteil herausgestellt, dass Nutzer sich gerne aktiv beteiligen und auch nicht perfekte Inhalte schon sehr wertvoll finden.

9. Fazit und Ausblick

An den vorgestellten Beispielen wurde deutlich, dass Crowdsourcing zur Volltexterschließung historischer Bestände geeignet ist. Insbesondere können dadurch auch Bestände erschlossen werden, die aufgrund ihres Umfangs oder ihrer Bedeutung für eine wissenschaftliche Edition im herkömmlichen Sinne ungeeignet sind.

Crowdsourcing ist aber kein Allheilmittel, vor allem ist es nicht automatisch schneller und billiger als herkömmliche Erschließungsmethoden. Je nach Komplexi-

⁶⁰ Neben den hier vorgestellten gibt es noch weitere Editionswerkzeuge, die je nach Zielsetzung eines Projekts speziellen Anforderungen genügen. Jens Brokfeld hat einige davon in seiner Masterarbeit untersucht: Brokfeld (2012).

tät der Inhalte kann Crowdsourcing ebenso personal- und kostenintensiv sein wie wenn die Arbeit durch wissenschaftliches Personal erledigt würde. Crowdsourcing in Kultureinrichtungen ist aber mehr, als das Auslagern von Arbeit an eine anonyme Masse. Nutzer erhalten die Möglichkeit, sich intensiv mit Kulturgut auseinanderzusetzen, neue Fähigkeiten zu entwickeln und sich ehrenamtlich zu engagieren. Durch die aktive Beteiligung fühlen sie sich stärker mit ihrer Bibliothek verbunden.

Meine Vision für das Bibliothekswesen ist es daher, dass jede digitale Sammlung ganz selbstverständlich auch mit einem Crowdsourcing Tool zur Transkription bzw. Textkorrektur ausgestattet ist. Bibliotheksnutzer werden dadurch von reinen Konsumenten zu verantwortungsvollen Beitragenden. Zudem können die Ergebnisse der mühsamen Arbeit, die viele Forscher (womit explizit auch Laien gemeint sind) auch bisher schon im Privaten geleistet haben, dann allen zur Verfügung gestellt werden. Die Bestände werden somit Stück für Stück im Volltext durchsuchbar und dadurch wiederum einem größeren Nutzerkreis zugänglich.

Anhang: Weitere interessante Crowdsourcingprojekte

Frühes Beispiel eines Open Science Crowdsourcingprojekts

Galaxy Zoo

<http://www.galaxyzoo.org/>

Internetuser sind aufgerufen Galaxien zu klassifizieren. Eine Forumdiskussion, die unter einigen interessierten Laien aufkam, führte zur Entdeckung einer völlig neuen Art von Galaxie.

Kommerzielles Crowdsourcing

Amazon Mechanical Turk

<https://www.mturk.com/mturk/>

„Clickworkers“ führen „Microtasks“ gegen Bezahlung aus.

Crowdsourcing zum Aufbau von Sammlungen

Australia in Pictures (früher: Picture Australia)

http://www.flickr.com/groups/pictureaustralia_ppe/

Internetuser weltweit wurden gebeten, Bilder zu Australiens Geschichte und Gegenwart bereitzustellen. Dazu wird mit Flickr eine bereits bestehende und vielen Nutzern bekannte Plattform genutzt. Über einen Tag mit „Picture Australia“ werden die Bilder gefunden und in die nationale Bilddatenbank in Trove integriert.

Erster Weltkrieg in Alltagsdokumenten

<http://www.europeana1914-1918.eu/de>

Privatpersonen sind aufgerufen, private Dokumente und Erinnerungsstücke aus der Zeit des Ersten Weltkriegs zu scannen und einer digitalen Sammlung hinzuzufügen.

Erschließung von Kartenmaterial

British Library, Georeferencer Project

<http://www.bl.uk/maps/georefabout.html>

Erschließung von Bildbeständen

Civil War Faces

http://www.flickr.com/photos/library_of_congress/sets/72157625520211184/

Ein Projekt der Library of Congress. Eine Sammlung von Fotos aus dem amerikanischen Bürgerkrieg wurde auf Flickr bereitgestellt. Die Öffentlichkeit wurde gebeten, die Personen auf den Bildern zu identifizieren.

British Library, Bilder aus Büchern des 17. bis 19. Jahrhundert

<http://britishlibrary.typepad.co.uk/digital-scholarship/2013/12/a-million-first-steps.html>

Die British Library hat eine Million Abbildungen aus Büchern des 17. bis 19. Jahrhunderts gescannt und auf Flickr hochgeladen, zusammen mit der Bitte an die Öffentlichkeit, diese inhaltlich zu beschreiben.

Erschließung mittelalterlicher Handschriften

Carolingian Canon Law

<http://ccl.rch.uky.edu/>

Erschließung neuzeitlicher Handschriften

What's on the Menu?

<http://menus.nypl.org/>

Die New York Public Library lässt historische Speisekarten transkribieren.

1846 Mexican War Diary

http://www.southwestern.edu/live/news/6475-collaborative-transcription-project*

Ein Projekt der Southwestern University, Georgetown, Texas zur Transkription eines einzelnen Tagebuchs.

Family Search Indexing

<https://familysearch.org/indexing/>

Geburts-, Sterbe- und Heiratsregister, Kirchenbücher, Passagierlisten, Zensusaufzeichnungen und andere genealogisch relevanten Dokumente aus aller Welt werden transkribiert bzw. Personennamen indexiert.

Erschließung von Drucken

Wikisource

<http://de.wikisource.org/wiki/Hauptseite>

Sammlung und Transkription von Quellentexten, die entweder urheberrechtsfrei sind oder unter einer freien Lizenz stehen.

Erschließung von Zeitungen

California Digital Newspaper Collection

<http://cdnc.ucr.edu/cgi-bin/cdnc?a=p&p=home&c=-----en--20--1--txt-txTA-%22Mary+O%27Neil+Hartman%22-----Mary+O%27Neil+Hartman>

Historic Cambridge Newspaper Collection

Cambridge Public Library, Cambridge, Massachusetts

<http://cambridge.dlconsulting.com/>

Literatur

- Brokfeld, J. (2012). Die digitale Edition der „preußischen Zeitungsberichte“: Evaluation von Editorenwerkzeugen zur nutzergenerierten Transkription handschriftlicher Quellen. <http://opus4.kobv.de/opus4-fhpotsdam/frontdoor/index/index/docId/331> (abgerufen am 15.02.2014).
- Causser, T., Tonra, J. & Wallace, V. (2012). Transcription maximized; expense minimized? Crowdsourcing and editing The Collected Works of Jeremy Bentham. *Literary and Linguistic Computing*, 27 (2), 119-137. [doi:10.1093/lc/fqs004.]
- Estellés-Arolas, E. & Ganzáles-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38 (2), 189-200. [doi:10.1177/0165551512437638.]
- Falk, R. (2013). Crowdsourcing: Möglichkeiten der (Zusammen-)Arbeit an Brief-Editionen im Internet. In A. Bohnenkamp & E. Richter (Hrsg.), *Brief-Edition im digitalen Zeitalter* (S. 35-42). Berlin & Boston: De Gruyter.
- Holley, R. (2009a). Many hands make light work: Public collaborative OCR text correction in Australian historic newspapers. https://www.nla.gov.au/ndp/project_details/documents/ANDP_ManyHands.pdf (abgerufen am 20.02.2014).
- Holley, R. (2009b). A success story – Australian Newspapers Digitisation Program. *Online Currents* 23 (6), 283-295. http://eprints.rclis.org/14176/1/j21_v023_OLC_pt06_Holley.pdf (abgerufen am 20.02.2014).
- Holley, R. (2009c). Crowdsourcing and social engagement: Potential, power and freedom for libraries and users. http://eprints.rclis.org/13968/1/Rose_Holley_PRDLA_Crowdsourcing_Nov_2009_Final_version.pdf (abgerufen am 23.02.2014).
- Holley, R. (2010). Crowdsourcing: How and why should libraries do it? *D-Lib Magazine*, 16 (3/4). [doi:10.1045/march2010-holley.]
- Howe, J. (2006). The rise of crowdsourcing. *Wired*, 14 (6), 176-183.
- Märting, B. & Thomas, C. (2008). Das Wuchern der Archive. Die digitale Edition des Nachlasses Franz Brümmer mit dem Refine!Editor. *Editio. Internationales Jahrbuch für Editionswissenschaften*, 22, 204-212. [doi:10.1515/edit.2008.012.]
- Middell, G. (2007). Sich den Literaturbetrieb nach Hause holen. Ein Einblick in die digitale Edition des lexikographischen Nachlasses Franz Brümmer. *B.I.T.online*, 10 (4), 305-308.
- Moyle, M., Tonra, J. & Wallace, V. (2011). Manuscript transcription by crowdsourcing: Transcribe Bentham. *Liber Quarterly*, 20 (3/4), 347-356.
- Owens, T. (2013). Digital cultural heritage and the crowd. *Curator. The Museum Journal*, 56 (1), 121-130. [doi:10.1111/cura.12012.]
- Proctor, N. (2013). Crowdsourcing – an introduction: From public goods to public good. *Curator. The Museum Journal*, 56 (1), 105-106. [doi:10.1111/cura.12010.]
- Ridge, M. (2013). From tagging to theorizing: Deepening engagement with cultural heritage through crowdsourcing. *Curator. The Museum Journal*, 56 (4), 435-450. [doi:10.1111/cura.12046.]
- Thomas, C. (2009). Ein Supplement zur Geschichte der deutschsprachigen Literatur. Die digitale Edition „Nachlass Franz Brümmer“. *Zeitschrift für Germanistik*, 19 (3), 618-626.

Thomas, C. (2014a). Telefoninterview vom 14. Februar 2014.

Thomas, C. (2014b). E-Mail vom 19. März 2014.

Weber, J. (2013). Briefnachlässe auf dem Wege zur elektronischen Publikation. Stationen neuer Beziehungen. In A. Bohnenkamp & E. Richter (Hrsg.), *Brief-Edition im digitalen Zeitalter* (S. 25-34). Berlin & Boston: De Gruyter.

<http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/> (abgerufen am 15.02.2014).

http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/franz_bruemmer/nachlass.html (abgerufen am 15.02.2014).

http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/franz_bruemmer/lexikon.html (abgerufen am 15.02.2014).

http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/includes/fenster.php?bild=../pix/handschriften/SBB-III_A_NlBruemmer_Biographien_Suppl_I_Hager_Moritz_B11r.jpg (abgerufen am 15.02.2014).

<http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/kontakt/> (abgerufen am 15.02.2014).

<http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/projekt/mitarbeiter.php> (abgerufen am 15.02.2014).

http://bruemmer.staatsbibliothek-berlin.de/nlbruemmer/redaktion/handschriften_update.php?id=320 (abgerufen am 15.02.2014).

<http://trove.nla.gov.au/> (abgerufen am 23.02.2014).

<http://trove.nla.gov.au/general/termsfuse> (abgerufen am 20.03.2014).

<http://trove.nla.gov.au/general/text-correction-guidelines-for-digitised-newspapers/> (abgerufen am 23.02.2014).

<http://trove.nla.gov.au/ndp/del/article/84203950> (abgerufen am 19.02.2014).

<http://trove.nla.gov.au/newspaper?q> (abgerufen am 23.02.2014).

<http://trove.nla.gov.au/system/stats?env=prod> (abgerufen am 23.02.2014).

http://www.transcribe-bentham.da.ulcc.ac.uk/td/Code_of_conduct (abgerufen am 20.03.2014).

<http://www.transcribe-bentham.da.ulcc.ac.uk/td/index.php?title=JB/050/044/001&action=edit> (abgerufen am 16.02.2014).

http://www.transcribe-bentham.da.ulcc.ac.uk/td/Transcribe_Bentham (abgerufen am 20.03.2014).