

e-manuscripta.ch: Volltext – Crowdsourcing mit Qualitätssicherung

Alexa Renggli, Zentralbibliothek Zürich

Zusammenfassung

Die Plattform für digitalisierte handschriftliche Quellen aus Schweizer Bibliotheken und Archiven, e-manuscripta.ch, wurde erweitert um ein Modul, das die Erstellung von Transkriptionen und Annotationen zu bereits darauf publizierten Dokumenten erlaubt. In enger Zusammenarbeit mit der Firma semantics Kommunikationsmanagement GmbH wurde das für die Applikation Visual Library (Basissystem der Plattform e-manuscripta.ch) bereits existierende Modul um ein formalisiertes Redaktionskonzept (online Zugang, Anmeldung, Überarbeitung, Freischaltung) ergänzt, was den freien Zugang zum Transkribieren unter Einhaltung gewisser Richtlinien erlaubt und gleichzeitig als Grundlage für zukünftiges Crowdsourcing dient. Im vorliegenden Beitrag werden die neu entwickelten Funktionen präsentiert, sowie die verschiedenen möglichen Arten der Mitwirkung und Zusammenarbeit mit interessierten Nutzenden erörtert.

Summary

The platform for digitised manuscript material from Swiss libraries and archives, e-manuscripta.ch, has been extended with a module that allows the creation of transcriptions and annotations for previously published documents. In close cooperation with the company semantics Kommunikationsmanagement GmbH, the existing module of the Visual Library application (the basic system of e-manuscripta.ch) was supplemented with a formalised editorial concept (online access, registration, revision, publication), which allows free access to transcription in compliance with certain guidelines, while serving as a foundation for future crowdsourcing. This paper presents the newly developed functions and discusses the various possible types of participation and cooperation with interested users.

Zitierfähiger Link (DOI): <https://doi.org/10.5282/o-bib/2018H4S58-70>

Autorinnenidentifikation: Renggli, Alexa GND 141905190

Schlagwörter: Transkription; Crowdsourcing

1. Einleitung

Kaum war die von den drei Schweizer Hochschulbibliotheken ETH-Bibliothek, Universitätsbibliothek (UB) Basel und Zentralbibliothek (ZB) Zürich aufgebaute Plattform e-manuscripta.ch im März 2013 online gegangen, kam bei den Projektpartnern auch schon der Wunsch auf, neben der Präsentation von Digitalisaten zusätzlich ein gerade neu entwickeltes Programm zur Erstellung von Transkriptionen mit einzubauen. Es sollte für alle interessierten Nutzenden frei zugänglich sein, damit so viele Beiträge wie möglich die Auffindbarkeit der Dokumente verbessern würden und die mit den Digitalisaten getätigte Forschung zusammen mit den Vorlagen präsentiert werden könnte. Aber wie lassen sich diese gleichzeitig auftretenden Ansprüche verbinden: Einerseits freien und offenen Zugang für alle zu bieten und dabei auch noch die inhaltliche Qualität des Angebots zu gewährleisten? Diese

Bedürfnisse stellten die Ausgangslage dar, als es darum ging, auf der Plattform die Möglichkeit zum Transkribieren anzubieten. Welche Massnahmen getroffen wurden, um diesen Herausforderungen zu begegnen, soll in diesem Beitrag erläutert werden.¹

2. Ausgangspunkt e-manuscripta.ch

Das neue Transkriptionstool stellt nicht ein alleinstehendes Programm dar, sondern wurde als Erweiterung einer bestehenden Präsentationsplattform hinzugefügt. Im konkreten Fall diene e-manuscripta.ch, die Plattform für digitalisierte handschriftliche Quellen aus Schweizer Bibliotheken und Archiven, als Basis.² Entsprechend müssen die Bestände zunächst auf dieser Plattform publiziert sein, um mit Transkriptionen ergänzt zu werden. Sind sie einmal veröffentlicht, dann kann aus sämtlichen Titeln einer frei ausgewählt werden, für welchen eine Transkription erstellt werden soll.

Die Plattform wurde als Kooperationsprojekt der drei Schweizer Hochschulbibliotheken ETH-Bibliothek, UB Basel und ZB Zürich aufgebaut. Sie konnten dabei von den Erfahrungen beim Einrichten der 2010 aufgeschalteten kooperativen Plattform für Drucke des 15. bis 20. Jahrhunderts aus Schweizer Bibliotheken, e-rara.ch, profitieren, zumal mit Visual Library von semantics GmbH auch die gleiche Software zum Einsatz kam.³ Der Aufbau von e-manuscripta.ch wurde finanziert aus Eigenmitteln, und am 13. März 2013 wurde die Plattform freigeschaltet, worauf seither kontinuierlich neue Digitalisate zur Verfügung gestellt wurden sowie weitere Partner und neue Funktionen hinzugekommen sind. Die technische Betreuung und Infrastruktur sind an der ETH-Bibliothek angesiedelt, während die Gesamtkoordination und Geschäftsstelle in der ZB Zürich beheimatet ist.

Für die Präsentation der Dokumente wurden fünf nach Dokumenttyp unterschiedene Kollektionen eingerichtet: Texte, Briefe, Noten, Bilder, Karten und Pläne sowie Varia. Für Transkriptionen dürften die Kollektionen Briefe und Texte am meisten Auswahl bieten. Unter den momentan 31.360 Briefen – wobei es sich je nach Katalogisat um einen einzelnen Brief oder ein Konvolut mit mehreren Stücken handelt – sind solche von Reformatoren wie Huldrych Zwingli oder Theodor de Bèze, vom Naturforscher Johann Jakob Scheuchzer, von der Basler Mathematiker-Familie Bernoulli, von Johann Jakob Bodmer, Johann Caspar Lavater, Richard Wagner, Hans Conrad und Arnold Escher von der Linth, Albert Einstein oder Rainer Maria Rilke zu finden. Die Texte im Umfang von 1 bis 1.000 Seiten beinhalten Rezepte, Selbstzeugnisse, Stammbücher, Rechnungsbücher und chronikalische Aufzeichnungen des Zürcher Ratsherrn Gerold Edlibach oder des Glarner Landvogts und Historikers Aegidius Tschudi sowie die Nachrichtensammlung des Zürcher Chorherrn Johann Jakob Wick.

- 1 Der vorliegende Beitrag beruht auf: Renggli, Alexa: e-manuscripta.ch: Volltext – Crowdsourcing mit Qualitätssicherung (Vortrag auf dem 107. Deutschen Bibliothekartag in Berlin am 13.06.2018), Vortragsfolien. Online: <<http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0290-opus4-36093>>.
- 2 Vgl. allgemein zur Plattform: Renggli, Alexa: E-manuscripta.ch: Aufbau und stetige Weiterentwicklung, in: Keller, Alice; Uhl, Susanne (Hg.): Bibliotheken der Schweiz: Innovation durch Kooperation. Festschrift für Susanna Bliggenstorfer anlässlich ihres Rücktritts als Direktorin der Zentralbibliothek Zürich. Berlin; Boston 2018, S. 390-400. Online: <<https://doi.org/10.1515/9783110553796-027>>.
- 3 Hotea, Meda Diana: E-rara.ch: eine Schweizer Erfolgsgeschichte, in: Keller, Alice; Uhl, Susanne (Hg.): Bibliotheken der Schweiz: Innovation durch Kooperation. Festschrift für Susanna Bliggenstorfer anlässlich ihres Rücktritts als Direktorin der Zentralbibliothek Zürich. Berlin; Boston 2018, S. 381-389. Online: <<https://doi.org/10.1515/9783110553796-026>>.

Kurz, es besteht eine grosse Themenvielfalt aus mehreren Jahrhunderten, was vielen verschieden gelagerten Interessen entgegenkommen dürfte.

Von Beginn an wurde eine nachhaltige Plattformlösung mit langfristiger Ausrichtung angestrebt. So wurde jeder Titel mit einem persistenten Identifier, dem Digital Object Identifier (DOI), adressierbar gemacht. Und die Archivierung der Digitalisate samt Metadaten und Transkriptionen geschieht an einem einzigen Ort in ZIP-Kapseln. Diese werden bei der Freischaltung erzeugt und müssen von den Partnerinstitutionen auf den Servern der ETH-Bibliothek abgeholt werden. Wenn vorhanden, können sie gleichzeitig direkt in ein Langzeitarchivierungssystem eingegliedert werden.

3. SUK P-2-Projekt «e-manuscripta: Volltext»

Schon bald nach der Freischaltung von e-manuscripta.ch begannen sich die teilnehmenden Institutionen für die Möglichkeit zu interessieren, direkt neben den Digitalisaten Transkriptionen zu erstellen.

Nach ausgiebigen Vorabklärungen und ersten Tests mit dem bei Visual Library bereits vorhandenen Modul „Autographenerschliessung“ wurde beim Infrastrukturprogramm SUK P-2 „Wissenschaftliche Information“ von swissuniversities im Februar 2016 ein Projekt-Antrag eingereicht, der das Einrichten eines solchen Programms zum Ziel hatte. Der offizielle Titel lautete „e-manuscripta.ch: Nachnutzung für die Wissenschaft – Der Weg zum Volltext“. Ursprünglich war geplant, den Antrag mit demjenigen für das Projekt zur Einführung von Volltexterkennung mittels OCR bei e-rara.ch zu koordinieren. Aber konkrete Bedarfsabklärungen bei verschiedenen Editionsprojekten nahmen mehr Zeit in Anspruch und führten dazu, dass im August 2015 zuerst nur der Antrag von e-rara.ch eingereicht wurde mit Verweis auf die später folgende Eingabe von e-manuscripta.ch.

Eingereicht wurde das Projekt gemeinsam durch die drei Partnerinstitutionen von e-manuscripta.ch, der ETH-Bibliothek, der UB Basel und der ZB Zürich. Die Projektleitung wurde in der ZB Zürich angesiedelt. Die Laufzeit sollte ein Jahr betragen. Gelder wurden insbesondere für den Kauf der bestehenden Module und deren nötige Weiterentwicklung beantragt sowie während eines Jahres 20% für die Projektleitung, 7% für die technische Umsetzung und einzelne Sitzungstage für drei Vertreter aus den Partnerinstitutionen im Projektteam. Das Projekt wurde im Juni 2016 genehmigt und die Fördermittel unter gewissen Auflagen zugesprochen.⁴ Ab August starteten die Projektarbeiten. Mittlerweile ist die Abschlussphase des Projekts erreicht worden. Am 1. Juni 2018 erfolgte die Freischaltung des Moduls auf dem Produktivsystem. Letzte Texte und Übersetzungen müssen noch eingefügt werden und auch die Redaktionsabläufe sind zunächst einzuüben mit weniger grossen Mengen an Transkriptionen. Daher wurde eine grössere Kommunikationsmassnahme erst für den Herbst zum Semesterbeginn an den Schweizer Universitäten eingeplant.

Vom neuen Modul erhoffen sich die Projektpartner, dass es durch die Transkribierenden zu einer Anreicherung der Texte mittels Eingabe von Anmerkungen und Erläuterungen kommt sowie zu einer

4 Swissuniversities: Wissenschaftliche Information. Laufende Projekte, Hauptstossrichtung Publikationen, <<https://www.swissuniversities.ch/de/organisation/projekte-und-programme/p-5/laufende-projekte/>>, Stand: 28.07.2018.

besseren Auffindbarkeit der Inhalte durch Anzeige der Suchbegriffe auch im Volltext. Zum Transkribieren können sämtliche auf der Plattform veröffentlichte Digitalisate frei ausgewählt werden. Innerhalb eines ganzen Bandes kann demnach auch eine einzelne Seite bearbeitet werden oder es können mehrere Transkribierende gleichzeitig jeweils eine Seite auswählen und daran arbeiten, was als Übung in einem Seminar gut denkbar wäre.

Im Projektantrag wurden als konkrete Ziele angegeben, dass das Modul in die bestehende Plattform integriert werden solle und gleichzeitig das Frontend anzupassen sei, um die Transkriptionen anzuzeigen und danach zu recherchieren. Für den freien Zugang waren als Rahmenbedingungen plattformspezifische Transkriptionsregeln und ein formalisiertes Redaktionskonzept zu entwerfen. Der ursprünglich geplante Ausbau von Export- und Importschnittstellen musste aus Zeitgründen auf später verschoben werden.

4. Transkribieren

Die einzelnen Elemente des Transkriptionstools lassen sich am besten anhand der für die Weiterentwicklung eingerichteten Arbeitspakete präsentieren, die als jeweilige Schwerpunkte die Bereiche „Transkribieren“, „Crowdmanagement“ und „Workflow“ beinhalten.

Zum Erstellen von Transkriptionen enthielt das schon vor dem Projekt vorhandene Modul „Autographenerschliessung“ eine Arbeitsoberfläche, in welcher die Texte unter Verwendung einer Markdown-Syntax eingegeben werden konnten. Basierend darauf sollte der für die Texterfassung zum Einsatz kommende Editor für die breite Öffentlichkeit benutzerfreundlicher gestaltet werden. Entsprechend erhielten die wichtigsten Formatierungselemente und Funktionen direkt anklickende Buttons.

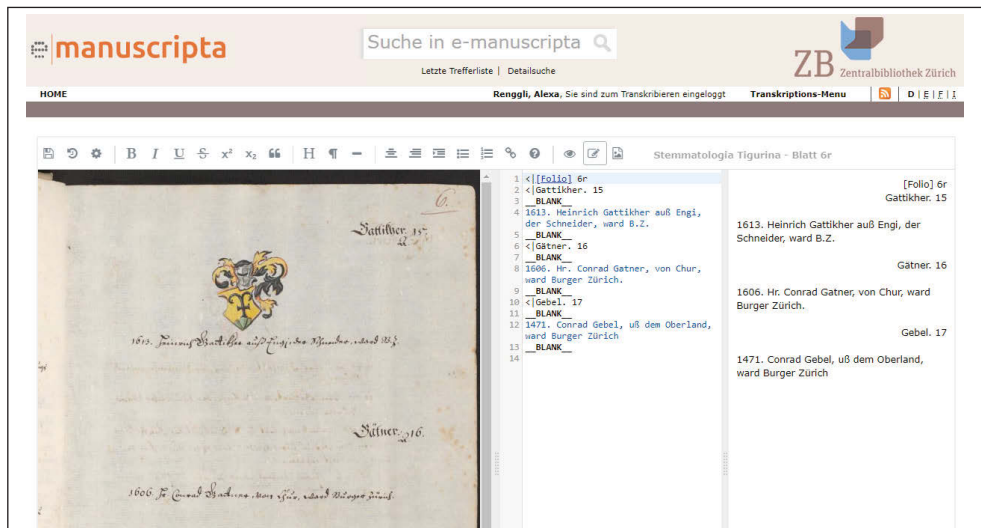


Abb. 1: Editorbereich, wie er sich den Transkribierenden präsentiert

Damit für eine bestimmte Seite der Editorbereich angezeigt werden kann, muss sich die bearbeitende Person zuerst einloggen. Nach der Anmeldung erscheint bei jedem in der Seitenansicht aufgerufenen Bild ein Transkriptionsbutton in der rechten oberen Ecke. Mit Klick darauf wird die Seite in den Editor geladen und die Transkriptionsarbeit kann beginnen. Mit dem ersten Speichern des eingegebenen Textes bleibt die Seite gesperrt für andere und erhält in der Versionsgeschichte den Namen der bearbeitenden Person zugeordnet.

Für die wichtigsten Bearbeitungsfunktionen wurden Buttons in einer Formatierungszeile angeordnet. So können Darstellungselemente wie fett, kursiv, unterstrichen oder durchgestrichen sowie zentriert oder rechtsbündig direkt per Knopfdruck eingerichtet werden. Mit Klick auf den ?-Button lässt sich ein separates Fenster öffnen mit einer ausführlicheren Markdown-Anleitung. Dort können weitere Gestaltungselemente wie zweispaltig, Randnotiz oder Fussnoten gefunden und aktiviert werden. Für die Darstellung von mathematischen Formeln müssen noch bessere Lösungen gefunden werden. Hier wird weiter daran gearbeitet, um entsprechende Möglichkeiten für die bereits zahlreich vorhandenen naturwissenschaftlichen Texte zu bieten. Während es nachvollziehbar ist, dass nicht alle graphischen Elemente direkt in eine Abschrift umgesetzt werden können, sollte es zumindest für wichtige Berechnungsvorgänge eine direkte Übertragung geben, da diese schlichtweg nicht in einer Fussnote zu erläutern sind.

Das Bild der Vorlage bleibt die ganze Zeit im Editor gleich neben dem Feld für die Texterfassung sichtbar; es kann aber auch per Button in einem separaten Browserfenster angezeigt und in dieser Umgebung mit Zoomfunktion vergrößert werden. Die Anzeigeform der Bilder sollte zudem in Zukunft weiter verbessert werden, es sind insbesondere für den Viewer neue Funktionen in Entwicklung, die vermutlich ein direktes Zoomen innerhalb der Abbildung im Editor erlauben werden.

5. Crowdmanagement

Sämtliche Texte auf der Plattform sollen allen interessierten Personen frei zur Transkription zur Verfügung stehen. Um dennoch eine gewisse Qualität zu gewährleisten und grobe Missbräuche zu vermeiden, waren die nötigen Rahmenbedingungen mit einem Redaktionskonzept zu schaffen.

Ausgangspunkt für das Crowdmanagement bildet die Anmeldung mit Namen und E-Mail-Adresse. Diese Identifikation erlaubt die Absicherung, dass nur ernsthaft interessierte Personen die Anwendung nutzen. Auch entsteht damit die Möglichkeit, mit den Transkribierenden in Kontakt zu treten.

Während des Registrierungsprozesses müssen die Nutzungsbedingungen akzeptiert werden. Dieses Dokument regelt die Rechte und Pflichten zwischen dem Anbieter und den Benutzenden. Auch wird darin festgehalten, unter welchen Lizenzen die Transkriptionen weiter zu verwenden sind. Hat die transkribierende Person bei der Anmeldung die Zustimmung zur Nennung des Namens unterhalb der Transkription gegeben, ist dieser entsprechend sichtbar und in einem Zitat aufzuführen, was einer Creative Commons Lizenz CC-BY-SA entspricht. Der gesamte Text der Nutzungsbedingungen wurde im Übrigen gleich auf die am 25. Mai 2018 in Kraft getretene Datenschutz-Grundverordnung

(DSGVO) der EU ausgerichtet,⁵ damit die Informationen von Anfang an auf dem für den europäischen Raum gültigen Stand sind.

Als weitere längere Texte wurden eine Bedienungsanleitung und Transkriptionsregeln erstellt. Die Bedienungsanleitung liefert die Erläuterung der Handhabung und des Workflows. Um eine gewisse Regelmässigkeit im Erscheinungsbild der Transkriptionen zu erreichen, wurden in den Transkriptionsregeln Anhaltspunkte formuliert, wie die Texte etwa aussehen sollten. Aufgrund der grossen Vielfalt an Dokumenten aus verschiedenen Jahrhunderten können die Regeln nicht zu eng gefasst werden. Es geht vielmehr darum, dass innerhalb eines Titels ein einheitliches Vorgehen gewählt wird.

Ist die Transkription einer Seite abgeschlossen, wird sie zur Kontrolle an die Redaktion geschickt. Dieser Vorgang gibt den Anbietern nach der Registrierung eine weitere Gelegenheit, abzusichern, dass die Texte seriös genug sind für eine Publikation. Im Gegenzug erhalten die Transkribierenden auch die Gewähr, dass ihre Beiträge in einem vertrauenswürdigen Umfeld erscheinen und die eigenen Texte nicht jederzeit wieder unbegrenzt von Dritten geändert werden können.

6. Workflow

Aus dem Redaktionsdurchgang vor der Freischaltung wurde ein eigenständiger Workflow entwickelt. Es wurden mehrere Schritte der Interaktion möglich zwischen der Redaktion und den Benutzenden. Das wird wichtig, sobald Korrekturen nötig sind, schlechte Ergebnisse sich abzeichnen oder lange Zeiträume entstehen, während welcher die Seiten gesperrt bleiben für andere.

Nach der ersten Speicherung des eingegebenen Textes bleibt die Seite gesperrt für andere Benutzende und kann bis zur Freischaltung von keiner anderen Person mehr bearbeitet werden. Mit Abschluss der Transkription wird eine Meldung an die Redaktion generiert, dass eine Seite auf die Freischaltung wartet. Wenn der Text den Anforderungen entspricht, kann die Redaktion die Seite direkt freigeben. Braucht es eine Überarbeitung, hat die Redaktion die Wahl, entweder eine kleine Korrektur gleich selber vorzunehmen oder sie schickt die Seite an die transkribierende Person zurück, wobei die Möglichkeit besteht, den Fehler in einem Kommentarfenster zu schildern. Ist der Text einmal freigegeben, erscheint der Transkriptionsbutton wieder beim Bild und erlaubt es allen Nutzenden, die Seite wieder in den Editor zu verschieben und Korrekturen oder Ergänzungen daran vorzunehmen. Erst eine klare Gliederung der Abläufe und die Sperrung während der Bearbeitung machen ein gleichzeitiges Nebeneinander von ganz verschiedenen Nutzern möglich. Es muss sofort erkennbar sein, wer gerade zuständig ist für eine Seite.

Eine wichtige Voraussetzung für eine erfolgreiche Teilnahme ist die Möglichkeit, dass sich die bearbeitenden Personen selber jederzeit einen Überblick über den aktuellen Stand ihrer Arbeiten verschaffen können. Dies geschieht in der Kontoansicht „Transkriptions-Menü“ beim Auswählen

5 Verordnung (EU) 2016/679 des europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung - DSGVO), am 25. Mai 2018 in Kraft getreten. Online: <<https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX%3A32016R0679&from=de>>, Stand: 28.07.2018.

des Buttons „Letzte Änderungen“. Jede Seite, die von dieser Person gespeichert worden ist, wird hier angezeigt und ist markiert mit dem aktuellen Status der Bearbeitung (in Bearbeitung, zur Freigabe, zur Weiterbearbeitung). Auf diese Weise bleiben alle Seiten, die für andere gesperrt sind, im Bewusstsein der bearbeitenden Person, welche die Seite blockiert. Der Status „in Bearbeitung“ wird angezeigt während der ganzen Herstellung einer Transkription, bis sie zum Abschluss kommt und an die Redaktion geschickt wird zur Begutachtung. Sobald die Seite an die Redaktion übergeben wird, erhält sie den Status „zur Freigabe“ und kann nicht mehr im Editor bearbeitet werden. Das ist erst wieder nach der Freigabe für alle Nutzenden möglich. Wenn die Redaktion eine Korrektur durch die transkribierende Person wünscht, dann schickt sie die Seite zurück und sie erscheint im Status „zur Weiterbearbeitung“.

In der Übersicht „Letzte Änderungen“ wird nicht nur der aktuelle Status einer Transkription angezeigt, sondern es sind auch Buttons zu finden, mit welchen die gewünschte nächste Aktion in Gang gesetzt werden kann, sei das die weitere Bearbeitung oder dann direkt die Übergabe an die Redaktion zur Begutachtung. Mit dem Klick auf den Button „Zur Revision an die Redaktion übergeben“ wird die Seite in einen neuen Status gesetzt und gleichzeitig eine E-Mail ausgelöst an die zuständige Redaktion. Jede Seite wird direkt von derjenigen Institution betreut, die auch im Besitz des Originaldokuments ist. Vor dem Abschicken der Mail öffnet sich ein Pop-up-Fenster und die bearbeitende Person hat die Möglichkeit, einen kurzen Kommentar oder Hinweis hineinzuschreiben. Dieser Text wird in den Mailtext integriert und erreicht auf diese Weise die Redaktion.

Wird die Seite zur Begutachtung an die Redaktion abgeschickt, gelangt sie in eine komplett andere Umgebung, nämlich den Redaktionsbereich. Dieser bildet ein ganzes Programm für sich ab und wurde während der allgemeinen Weiterentwicklung des Transkriptionstools immer stärker und detaillierter ausgebaut.

Für jede Partnerinstitution wurde ein eigenes Login eingerichtet und es werden nur diejenigen Seiten angezeigt, die aus dem eigenen Bestand stammen. Erkannt wird das vom Programm automatisch am Institutionskürzel, das jeweils in der URL vorhanden ist. Jedem Redaktionskonto wird eine E-Mail-Adresse hinterlegt, über welche die Interaktion mit den Transkribierenden der entsprechenden Seiten läuft. Das kann durchaus dazu führen, dass eine teilnehmende Person, die mit Dokumenten aus verschiedenen Institutionen arbeitet, von mehreren Redaktionen angeschrieben wird, die nichts von den Aktivitäten der jeweils anderen Institution wissen.

Mit dem zugeteilten Login gelangt die für die Redaktion zuständige Vertretung der Institution auf das Redaktionsfrontend, das etwas anders gestaltet ist als das allgemeine „Transkriptions-Menü“. Die Seiten von Titeln der eigenen Institution sind je nach Bearbeitungsstatus in den Kategorien „Review“ und „Transkripte in Bearbeitung“ gelistet. Der Button „Letzte Änderungen“ zeigt Seiten, die vom Redaktionskonto selber aus abgespeichert und bearbeitet werden.

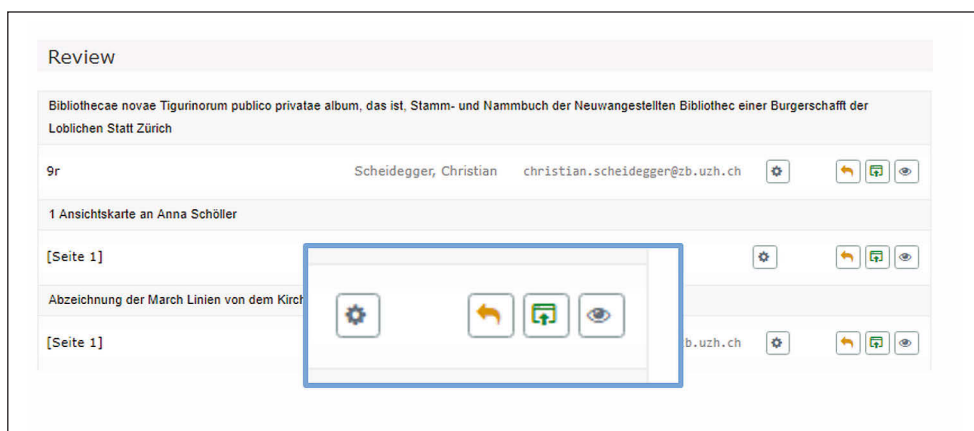


Abb. 2: Seite „Review“ im Redaktionsbereich

Nach Klick auf den Button „Review“ wird eine Übersicht angezeigt mit allen Seiten aus der eigenen Institution, bei welchen der Antrag auf Freigabe gestellt worden ist. Die Angaben zur transkribierenden Person wie Name und E-Mail-Adresse sind hier jedes Mal direkt bei der bearbeiteten Seite sichtbar. Dies erlaubt der Redaktion eine Kontaktaufnahme im Falle von Klärungsbedarf. Bei jeder Seite im Bearbeitungsstatus „Review“ erscheinen die Buttons, die für die nächsten Aktionen verwendbar sind. Es besteht hier beispielsweise die Möglichkeit, die Seite zur Überarbeitung zurückzuschicken. In einem Popup-Fenster kann durch die Redaktion festgehalten werden, was genau zu korrigieren ist. Dieser Text wird in die E-Mail integriert, die aus dem Programm heraus an die transkribierende Person geschickt wird. Ein anderer Button in dieser Auswahl erlaubt es der Redaktion, die Seite direkt freizugeben. Die Transkription erscheint unmittelbar auf der Plattform, samt dem Namen der bearbeitenden Person. In der Praxis dürfte es aber jeweils nötig sein, die Transkription zuerst zu begutachten. Dies ist mit dem Button „Ansehen“ möglich. Er öffnet die Seite in der Vollansicht. Wenn noch genauer kontrolliert werden soll, kann die Seite von hier aus auch in den Editor geladen werden. Dort wäre es für die Redaktion möglich, gleich selber letzte kleine Korrekturen vorzunehmen, bevor sie die Freigabe erteilt.

Wenn das Interesse einer bearbeitenden Person erlahmt oder nach mehreren Anläufen einfach keine korrekte Transkription möglich ist, dann bestehen für die Redaktion verschiedene Möglichkeiten, mit der Seite weiter zu verfahren. Entweder lässt sich die Seite mit kleinen Anpassungen direkt zum Abschluss bringen oder es kommt diejenige Funktion zum Einsatz, mit welcher die Seite an eine andere angemeldete Person weitergeleitet werden kann. Die neue verantwortliche Person erfährt von diesem Transfer via E-Mail, die durch die Aktion ausgelöst wird. Sollen die Spuren des Transkriptionsversuchs komplett gelöscht werden, dann reicht eine Mail an die Kontaktadresse der Plattform und der Text sowie die Zuständigkeit werden spurlos entfernt. Nutzende können generell jederzeit mitteilen, dass sie ihre Teilnahme beenden möchten. Wenn wichtige Gründe vorliegen, können – sofort oder nach Vorankündigung und bei Einhaltung der nötigen Frist – ihrerseits auch die Anbieter ein Konto sperren. Das Konto wird daraufhin deaktiviert.

Mit verschiedenen eingerichteten Aktionen lässt es sich somit vermeiden, dass einmal reservierte und gesperrte Seiten auf lange Zeit blockiert bleiben. Es darf auch etwas ausprobiert werden und der Einstieg soll möglichst unverbindlich und zwanglos sein. Aber das Ziel besteht darin, dass sich nach einer Einarbeitungsphase ein regelmässiges und zuverlässiges Mitwirken entwickelt.

7. Teilnahme

Die Teilnahme soll allen interessierten Personen offenstehen. Um diesen freien Zugang und die Auswahl von allen Seiten zuzulassen, mussten die nötigen Voraussetzungen geschaffen werden zur Sicherung der Qualität der Arbeiten und zur Absicherung des Betriebs der Plattform. Die Anmeldung mit einem Benutzerkonto bildet eine wichtige Basis, um von Beginn an eine direkte Verbindlichkeit zu schaffen und gleichzeitig die Möglichkeit zur Interaktion zu bieten. Der Name der transkribierenden Person wird auch bei jeder Seite nach der Freischaltung angezeigt. Dies ist nicht nur als Zwang zur Personalisierung zu verstehen, sondern bietet den Transkribierenden gleichzeitig die Gewähr, dass sie als Urheber der Transkription zu nennen sind, wenn jemand den Text zitiert. Laut Nutzungsbedingungen erhalten die mit Namen unterzeichneten Transkriptionen eine CC-BY-SA-Lizenz. Für Forschende, die ihre als Nebenprodukt von Quellenarbeit entstandenen Transkriptionen auf der Plattform platzieren, bedeutet dies eine wichtige Möglichkeit, ihre gesamte Forschungsarbeit zu deklarieren. Falls jemand doch um keinen Preis den eigenen Namen publiziert sehen möchte, besteht die Option, bei der Anmeldung anzugeben, dass die Freischaltung als „Anonymus“ geschehen soll.

Bei jeder Seite erkennen die Transkribierenden sofort, ob sie frei zur Bearbeitung ist. Von der ersten Speicherung an ist nur noch eine einzige Person im Besitz der Seite und behält sie reserviert bis zur Freischaltung. Damit ist gewährleistet, dass es jeweils nur eine Version der Seite gibt. Ab dem Zeitpunkt der Freischaltung können auch alle anderen Personen wieder Korrekturen oder Ergänzungen anbringen. Auch der Name derjenigen Person, die eine freigeschaltete Seite überarbeitet, wird nach der Freigabe unterhalb der Transkription hinzugefügt. Mit den verschiedenen Workflow-Funktionen lässt sich entsprechend ein geregelter Ablauf auf einer öffentlich zugänglichen Plattform sicherstellen. Die Redaktionen sollten im Normalfall allein zum Zeitpunkt der Freigabe mit der Transkription in Kontakt kommen.

Um auch die Nachhaltigkeit der erstellten Transkriptionsarbeiten zu gewährleisten, wurde die Archivierung der freigeschalteten Seiten direkt in den ZIP-Kapseln der entsprechenden Titel eingerichtet. In jeder ZIP-Kapsel werden alle TIFF-Bilder zusammen mit den Metadaten archiviert. Jede Änderung im Umfeld eines Titels wird als Aktualisierung der ursprünglichen ZIP-Kapsel hinzugefügt. So auch der transkribierte Text im Moment der Freischaltung. Die Seite besitzt die gleiche Identifikationsnummer wie der zugehörige Scan und lässt sich entsprechend zuordnen.

Bereits während der Bearbeitung einer Seite wird jede Abspeicherung in einer Versionsgeschichte deponiert, die von der transkribierenden Person im Editor abrufbar ist. Wie in einer Wikipedia-Seite sind die Änderungen gegenüber der Vorversion markiert und gut sichtbar. So könnte jederzeit auch auf eine frühere Version wieder zurückgegriffen werden.

Im Verlauf der Abklärungen zu möglichen Einsatzbereichen wurde bald deutlich, dass grosse Editi-onsprojekte lieber direkt mit ihren besonders elaborierten und individuell an die Charakteristik ihres Bestandes angepassten Programmen arbeiten werden. Auch ist die Arbeit nach der Freigabe sofort allgemein zugänglich und kann nicht der Öffentlichkeit vorenthalten werden bis zum Projektabschluss oder der offiziellen Publikation.

Für frühere Phasen des Studiums oder für Transkriptionen als Nebenprodukte von wissenschaftlichen Arbeiten bietet das Modul hingegen eine gute Gelegenheit, die Resultate öffentlich zugänglich zu machen. Auch könnten in einer Seminarübung oder an einem Transkriptionsanlass konkret einzelne Seiten eines Titels auf verschiedene Leute verteilt werden. Gemeinsam liesse sich entsprechend ein grösserer Teil eines Bandes erschliessen.

Um einen grösseren Kreis anzusprechen, käme durchaus ein klassischer Aufruf zum Crowdsourcing in Frage. Dies wäre gut vorzubereiten und würde beginnen mit einem konkreten Vorschlag zu vor-handenen Materialien. Als Thema kämen Dokumente zu einem aktuellen Ereignis oder zu einem vordefinierten Bestand in Frage. Mit dem Datum der Mitteilung kann der Zeitpunkt des Arbeitsbe-ginns gesteuert werden. Im weiteren Verlauf der Aktion dürfte es jedoch nötig sein, immer wieder zu motivieren oder wichtige Erfolge und Meilensteine bekannt zu geben. Es dürfte auch hilfreich sein für die Motivation, wenn jede teilnehmende Person Zugriff auf die eigene Statistik hat und genau verfolgen kann, wie viele Seiten schon freigeschaltet worden sind. Das könnte durchaus dazu anspornen, noch weitere Seiten zu bearbeiten, um eine gewisse Zahl zu erreichen. Zusätzlich liesse sich der Wettbewerbseifer steigern, wenn es ein Ranking gibt, das den Vergleich mit anderen Teil-nehmern erlaubt. Es gibt bereits schöne Beispiele von Crowdsourcing-Communities, die sich nach einem Aufruf aus den treuesten Mitwirkenden gebildet haben. Zum einen entstand eine solche Gruppe für das Projekt „Transcribe Bentham“, bei welchem mittlerweile fast 50% der vorhandenen Texte, die alle von einer einzigen Person stammen, bearbeitet worden sind.⁶ Als weiteres Beispiel ist in Zürich eine eingefleischte Gruppe entstanden während der Identifikation von Luftbildern auf der Plattform „E-Pics“ der ETH-Bibliothek.⁷ Diese werden auch persönlich betreut und stehen seither in regelmässigem Kontakt mit den Verantwortlichen der Aktion.

Der Reiz von e-manuscripta.ch als thematisch und zeitlich breit aufgestellter Plattform liegt sicher darin, dass es an verschiedenen Ecken zu Gruppierungen kommen dürfte. Zu Beginn ist noch kaum absehbar, in welchen Bereichen die Schwerpunkte liegen werden.

Eine gewisse Gruppe von zukünftigen Transkribierenden dürfte die Plattform zudem als Ausgangs-punkt nutzen, um erste Basistexte zu erstellen. Diese können via Schnittstelle jederzeit auch in ein stärker spezialisiertes Programm exportiert und dort vertieft weiter bearbeitet werden. Der

6 Vgl. Causer, Tim; Terras, Melissa: „Many hands make light work. Many hands together make merry work“: *Transcribe Bentham* and crowdsourcing manuscript collections, in: Ridge, Mia (Hg.): *Crowdsourcing our Cultural Heritage*, Ash-gate 2014, S. 57-88. Online: <<http://discovery.ucl.ac.uk/id/eprint/1393567>>, Stand: 28.07.2018. Der aktuelle Stand wird regelmässig hier angezeigt: *Transcribe Bentham*, Transcription Desk, <http://www.transcribe-bentham.da.ulcc.ac.uk/td/Transcribe_Bentham>, Stand: 28.07.2018.

7 Weiterhin wird auf dem Blog über aktuelle Resultate und neue Projekte berichtet: *ETH-Bibliothek, Crowdsourcing*, <<https://blogs.ethz.ch/crowdsourcing/>>, Stand: 02.12.2018.

umgekehrte Weg, also ein Import von anderen Plattformen zurück nach e-manuscripta.ch, dürfte jedoch mit Schwierigkeiten verbunden sein. Primär handelt es sich beim Markdown-Format um einen einfacheren Standard. Da würden beispielsweise aus TEI-Daten einige Details verloren gehen. Damit die Daten aber für die Volltext-Suche erhältlich sind, wäre ein Verlust wohl akzeptabel.

8. Ausblick, mögliche Folgeprojekte

Es zeichnen sich bereits Folgearbeiten ab, die die Funktionalitäten verbessern und ausbauen sollen. Zur besseren Sichtbarkeit des neuen Angebots ist nach der Einführung und den letzten Nachbesserungen auch der Einbau von Anzeigemöglichkeiten im Frontend einzuplanen. Schon jetzt tauchen Begriffe aus den ersten freigeschalteten Transkriptionen als Suchresultate auf und werden gleich angezeigt wie Treffer in der Inhaltsansicht. Sie sind jeweils farblich markiert und die Stellen in der Transkription können direkt angesteuert werden. Die Anzeige der Treffer innerhalb der transkribierten Texte erlaubt demnach, im Inhalt der Texte zu suchen, über was jeweils geschrieben wurde.

Auf verschiedenen anderen Visual Library-Plattformen sind gewisse Anzeigeformen für Volltext bereits vorhanden und können für das Transkriptionstool in modifizierter Form eingebaut werden.⁸ Eine Facette mit der Benennung „Volltext durchsuchbar“ wird es erlauben, alle Titel aufzulisten, die auch schon Transkriptionen enthalten. Diese Anzeige liefert einen deutlich sichtbaren Hinweis, dass überhaupt Volltext vorhanden ist. Und es erlaubt denjenigen Besuchern der Plattform, die weniger geübt sind im Handschriftenlesen, direkt bei solchen Titeln einzusteigen, bei welchen es Unterstützung für das Entziffern der Texte gibt.

Für die Aktivierung der Anzeigefunktionen bleibt noch etwas Zeit. Das ist erst dann sinnvoll, wenn eine gewisse Menge an Seiten freigeschaltet ist.

Eine grössere Menge an Beispieldaten ist ebenfalls nötig für die Abklärungen zu allfälligen weiteren Funktionalitäten im Bereich von Schnittstellen. Speziell für den Import und Export von Plain Text oder LaTeX müssten die konkreten Eigenschaften genau erörtert werden, damit es zu einer erkennbaren Systematik kommen kann. Da wäre auf jeden Fall jeweils noch einiges an Vorbereitungsarbeit zu leisten.

Es dürfte zunächst das folgende Szenario denkbar werden. Im Rahmen des Horizon2020-Projekts READ (Recognition and Enrichment of Archival Documents),⁹ das sich der Verbesserung der maschinellen Erkennung von handschriftlichen Texten widmet, könnten die Scans von Texten samt den zugehörigen Metadaten bezogen werden. Für die ersten Seiten wäre im Programm Transkribus¹⁰ ein Training mit bereits maschinenschriftlich vorliegenden Transkriptionen nötig, wonach die maschinelle

8 Hier sei auf die beiden Plattformen e-rara.ch (Facette „Volltext“), <<https://www.e-rara.ch/nav/index/all>> und Francke-Portal der Franckeschen Stiftungen Halle (Kollektion „Transkribierte Briefe“), <<https://digital.francke-halle.de/fsfp>> verwiesen. Stand: 28.07.2018.

9 o. A.: Recognition and Enrichment of Archival Documents, European Commission: CORDIS, <https://cordis.europa.eu/project/rcn/198756_en.html>, Stand: 16.09.2018.

10 Transkribus, <<https://transkribus.eu/Transkribus/>>, Stand: 16.09.2018.

Erkennung der übrigen Seiten erfolgen würde. Diese Daten im TEI-Format würden anschliessend nach e-manuscripta.ch zurückgeholt. Beim Import in die Markdown-Syntax ist wohl mit Verlust an Details zu rechnen, aber es wären grosse Fortschritte möglich, was die Menge an durchsuchbarem Volltext anbelangt.

Bis es aber soweit ist, dass zu den Bildern aus allen Jahrhunderten und Weltgegenden maschinell generierte Transkriptionen erhältlich sind, wird die Teilnahme von realen Personen am Erschliessen der Texte das zentrale Element darstellen. Auch für die Handhabung durch menschliche Fachkräfte wird es einen stetigen Entwicklungsbedarf geben und noch viel Möglichkeiten zur Optimierung der Abläufe. Mit Stillstand ist hier kaum zu rechnen. Für den Moment soll aber ein kurzer Augenblick der Freude über die Freigabe des Tools und den Beginn der Transkriptionsaktivitäten erlaubt sein.

Literaturverzeichnis

- Causer, Tim; Terras, Melissa: „Many hands make light work. Many hands together make merry work“: *Transcribe Bentham* and crowdsourcing manuscript collections, in: Ridge, Mia (Hg.): *Crowdsourcing our Cultural Heritage*, Ashgate 2014, S. 57-88. Online: <<http://discovery.ucl.ac.uk/id/eprint/1393567>>, Stand: 28.07.2018.
- ETH-Bibliothek, Crowdsourcing, <<https://blogs.ethz.ch/crowdsourcing/>>, Stand: 02.12.2018
- Hotea, Meda Diana: E-rara.ch: eine Schweizer Erfolgsgeschichte, in: Keller, Alice; Uhl, Susanne (Hg.): *Bibliotheken der Schweiz: Innovation durch Kooperation*. Festschrift für Susanna Bliggenstorfer anlässlich ihres Rücktritts als Direktorin der Zentralbibliothek Zürich. Berlin; Boston 2018, S. 381-389. Online: <<https://doi.org/10.1515/9783110553796-026>>.
- Renggli, Alexa: E-manuscripta.ch: Aufbau und stetige Weiterentwicklung, in: Keller, Alice; Uhl, Susanne (Hg.): *Bibliotheken der Schweiz: Innovation durch Kooperation*. Festschrift für Susanna Bliggenstorfer anlässlich ihres Rücktritts als Direktorin der Zentralbibliothek Zürich. Berlin; Boston 2018, S. 390-400. Online: <<https://doi.org/10.1515/9783110553796-027>>.
- Renggli, Alexa: e-manuscripta.ch: Volltext – Crowdsourcing mit Qualitätssicherung (Vortrag auf dem 107. Deutschen Bibliothekartag in Berlin am 13.06.2018), Vortragsfolien, <<http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0290-opus4-36093>>.
- Swissuniversities: Wissenschaftliche Information. Laufende Projekte, Hauptstossrichtung Publikationen, <<https://www.swissuniversities.ch/de/organisation/projekte-und-programme/p-5/laufende-projekte/>>, Stand: 28.07.2018.

- Verordnung (EU) 2016/679 des europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung - DSGVO), am 25. März 2018 in Kraft getreten. Online: <<https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX%3A32016R0679&from=de>>, Stand 28.07.2018.