



ERWERBEN &  
SICHTBAR  
MACHEN

österreichischer  
bibliothekartag  
2015

wien  
15.-18. september 2015

## Slot 4.4: Neue Zugänge zu historischen Quellen

**Moderation:** Veronika Plöbñig (*Universitäts- und Landesbibliothek Tirol ULB, Innsbruck, Österreich*)

### Crowdsourcing : neue Möglichkeiten und Grenzen für Bildarchive

**Vortragende/r:** Graf, Nicole (*ETH Zürich, ETH-Bibliothek, Schweiz*)

**Zur Person:** Leitung Bildarchiv ETH-Bibliothek seit 2008, Diplombibliothekarin BBS, lic. rer. soc. Univ. Bern (Soziologie), MA Bildwissenschaft, Donau-Universität Krems.

Crowdsourcing ist das Outsourcing von Teilaufgaben von Institutionen oder Firmen an Freiwillige (Crowd), die diese Aufgaben aufgrund von Interesse und spezifischem (Fach-)Wissen in ihrer Freizeit erledigen helfen. Dank Internet wurde diese Art der Mitarbeit technisch vereinfacht. Das wohl bekannteste Projekt dieser Art ist Wikipedia. Auch im kulturellen Sektor der Gedächtnisinstitutionen (kurz: GLAM) hält diese neue Form der Generierung und des Teilens von Wissen langsam Einzug. Einige Crowdsourcing-Projekte grosser Institutionen sind etwa die Library of Congress als Vorreiter mit schlecht beschrifteten Bildern auf Flickr Commons, das Geotagging von alten Kartenmaterialien an der British Library, oder die Text-Korrekturen beim Australian Newspapers Digitisation Program der National Library of Australia.

Im Bildarchiv der ETH-Bibliothek wurde von Dezember 2009 bis Dezember 2013 ein erfolgreiches Projekt mit ehemaligen Mitarbeitenden der ehemaligen Luftfahrtsgesellschaft Swissair durchgeführt. Aus dem Firmen-Bildarchiv der Swissair wurde eine Auswahl von 40.000 Bildern digitalisiert und in Tranchen von 200 Bildern wöchentlich auf der eigenen Bilddatenbank online gestellt. Eine Gruppe von interessierten Swissair-Pensionären half regelmässig mit, diese Bilder online mit Metadaten und Kontextwissen anzureichern, Personen und Ereignisse, Flugzeugtypen und Flugzeuge zu identifizieren sowie die Bilder zu datieren. Im Vortrag werden erstens die im Projekt gemachten Erfahrungen kritisch berichtet: der Redaktionsworkflow, die Nutzerauthentifizierung- und -verwaltung, der (unterschätzte) Betreuungsaufwand und die Öffentlichkeitsarbeit sowie wichtigste Ergebnisse, der allgemeine Nutzen und Mehrwert. Zweitens soll auch ein konzeptueller Ausblick auf künftige Crowdsourcing-Projekte des Bildarchivs der ETH-Bibliothek gegeben werden, unter Berücksichtigung der Fragestellung, ob weitere Online-Plattformen/Tools zum Einsatz kommen sollen oder nicht.

### Das Crowdsourcing-Modul der „Digitalen Landesbibliothek Oberösterreich“ (DLOÖ)

**Vortragende/r:** Neuböck, Gregor (*Oberösterreichische Landesbibliothek, Linz, Österreich*)

**Zur Person:** Studium an der Pädagogischen Hochschule, Studium Medienpädagogik an der Universität Krems, Freiberufliche Tätigkeit im Bereich eLearning, Leiter Service- und Benutzungsabteilung und Digitale Bibliothek der OÖ Landesbibliothek.

Die „Digitale Landesbibliothek Oberösterreich“ (<http://digi.landesbibliothek.at>) ist seit März 2010 online. Viele verschiedene Entwicklungen wurden seither vorangetrieben. Zu eine der wichtigsten zählt die Entwicklung eines Crowdsourcing-Moduls.

Im Juni 2013 hörte ich auf einer Tagung zu unserer Viewer-Software zum ersten Mal von einem Crowdsourcing-Tool, das die Zentral- und Landesbibliothek Berlin gemeinsam mit ihrer technischen Betreuungsfirma entwickelte und dazu benutzte, den OCR-Text von Adressbüchern zu verbessern. Naturgemäß sind verschachtelte Tabellen, wie sie in Adressbüchern Verwendung finden, neben unzureichender Druckqualitäten und Frakturschriften, als Hauptursache für schlechte OCR-Ergebnisse anzusehen. An der Landesbibliothek Berlin versuchte man, die Erkennungsqualität durch den Einsatz eines Crowdsourcing-Moduls zu verbessern. Im

Vorfeld gab es Bedenken wegen möglicher Probleme mit elektronischem Vandalismus. Eine Mitarbeiterin der Landesbibliothek Berlin berichtete davon, dass sich derartige Bedenken nicht bewahrheitet hätten. Ganz im Gegenteil, durch den breiten Einsatz vieler „freiwilliger“ Mitarbeiterinnen und Mitarbeiter wurde die Qualität des OCR-Textes erheblich verbessert.

Gerade im Bereich der Volltexte kann die Verbesserung von Volltexten kaum mehr von den Einrichtungen im Detail geleistet werden und so hat sich auch die OÖ Landesbibliothek dazu entschlossen, ein derartiges Tool zu entwickeln.

Seit Jänner 2014 arbeitete die OÖ Landesbibliothek an eben diesem Crowdsourcing-Modul gemeinsam mit der Betreuungsfirma. Rasch zeigte sich, dass es viele weitere und spezifische Anforderungen von unserer Seite her gab, die bisher noch gar nicht entwickelt waren. Diese standen auch unmittelbar im Kontext der bei uns vorhandenen Materialheterogenität. Neben Monographien können bei uns auch Zeitschriften, alte Drucke und Handschriften im Crowdsourcing-Modul bearbeitet und angereichert werden. In dem nun folgenden Prozess wurden über Monate hinweg vorhandene Funktionen angepasst und neue Features entwickelt. Im Nachhinein darf diese lange Phase der Entwicklung als Glücksfall für die DLOÖ bezeichnet werden, da erst durch diese tiefgehenden Tests und laufende Anpassungen Innovation, Usability und Stabilität ermöglicht wurden.

Um Besucher zu animieren, an unserer Plattform teilzunehmen, wurden insbesondere vorhandene Zugangsbeschränkungen so weit als möglich ausgeräumt. Zur Mitarbeit genügt schon ein Open-ID-Account, wie er in vielen sozialen Netzwerken zum Einsatz kommt, gleichwohl man sich aber auch ganz herkömmlich registrieren kann.

Neben unserem Hauptziel, bessere Volltexte zu erhalten, besteht auch die Möglichkeit, Metadaten zu Personen, Einrichtungen, Adressen und Kommentaren mit unterschiedlichen Eingabefeldern zu erfassen. Mit Jänner 2015 wurde der Probetrieb aufgenommen und seit März dieses Jahres wird unser Modul zur Datenanreicherung im Vollbetrieb geführt.

## **ABSAGE - Volltextsuche in historischen Texten**

**Vortragende/r:** **Johann Klugsberger** (Österreichische Nationalbibliothek, Wien, Österreich)

**Zur Person:** Langjährige Tätigkeit in der Softwareentwicklung, Development Manager, Projektleiter. Seit 2009 an der Österreichischen Nationalbibliothek, seit 2014 Teamleiter für Applikationen und Datenbanken und stv. Abteilungsleiter des Zentralen Informatikdienstes.

Die Österreichische Nationalbibliothek bietet ihren Benutzerinnen und Benutzern über die Katalogrecherche hinaus die Möglichkeit, Volltextrecherchen in historischen Texten an. Dieses Service umfasst neben historischen Druckschriften aus dem Projekt Austrian Books Online Zeitungen, Zeitschriften und Gesetzestexte. Während im allgemeinen Suchportal „QuickSearch“ über alle Medientypen gesucht wird, bieten spezielle Portale wie ANNO eine gezielte Suchmöglichkeit in bestimmten Medien an. Ermöglicht wird dieser Zugang durch die Prozessierung der Digitalisate mittels Optical Character Recognition (OCR) sowie durch ein zentrales Suchservice, welches als zentrale Instanz eine einheitliche Basis und ein definiertes Interface für alle derzeitigen und zukünftigen Applikationen und Volltext-Suchportale der ÖNB bietet. Derzeit sind ca. 270.000 Bücher, 340.000 Zeitungen sowie 13.000 Zeitschriften durchsuchbar – insgesamt etwa 95 Millionen Seiten, bis Ende 2015 werden der gesamte digitalisierte Zeitungsbestand der Österreichischen Nationalbibliothek, mehr als 300.000 Bücher sowie umfangreiche Zeitschriften und Gesetzestexte in die Volltextsuche integriert sein.