arbido

2015/3 GLAM und Wikimedia

Kreyenbühl Elias, Staatsarchiv Basel-Stadt

Tags: Knowledge Management Open Access, Weiterbildung,

Lessons learnt? Eindrücke eines Archivmitarbeiters vom Swiss Open Cultural Data Hackaton

Als ich auf Twitter von #GLAMhack erfuhr, wusste ich: Da will ich dabei sein. Nämlich am ersten Swiss Open Cultural Data Hackathon. Als Mitarbeiter des Staatsarchivs Basel-Stadt beschäftige ich mich momentan in einem mehrjährigen Projekt mit der Digitalisierung und Mikroverfilmung von Archivgut. Dabei ist die kreative Verwendung von Kulturdaten und das Ausprobieren neuer Darstellungsformen ein Dauerthema.

Denn unsere umfassenden Anstrengungen, das analoge und digitale Kulturgut zu erhalten, sollen dessen Nutzbarkeit sicherstellen. Und diese Nutzbarkeit wird künftig primär, behaupte ich, digital unterstützt erfolgen. Als Archivare können wir aber nicht selber die Nutzungsformen der Zukunft entwickeln. Ideen und Umsetzungen müssen von den Nutzerinnen und Nutzern kommen, ihren Ansprüchen und Zwecken entsprechen. Als Dateneigner und Datenanbieter sollten wir Archive vielmehr unsere eigenen Möglichkeiten ausschöpfen, indem wir unseren Nutzerinnen und Nutzern Zugang zu unseren Daten in geeigneter Form ermöglichen. Archive sind – mit Blick auf die Nutzung von Archivgut – Enabler, nicht Researcher.

Digitalisierung 2.0

Also meldete ich mich als Vertreter des Staatsarchivs zum Hackathon an und begann, geeignete Daten zu suchen und vorzubereiten. Schon das war nicht ganz einfach. Am Hackathon selber kam dann die zweite Ernüchterung. Von einer computergestützten Nutzbarkeit ihrer Daten war meine Institution noch weit entfernt.

Der Hackathon führte mir vor Augen, was es für eine zeitgemässe Nutzung von Primärdaten (also digitalen Archivguts, nicht nur der Metadaten) tatsächlich braucht. Zuvor war ich ziemlich stolz gewesen auf die über 50 000 digitalisierten Bilder und Pläne sowie auf die halbe Million Aktenseiten, die das Staatsarchiv Basel-Stadt der Öffentlichkeit online zur Einsicht bietet. Doch dann begriff ich: Es reicht schon bald nicht mehr, Daten nur in einer für Menschen möglichst benutzerfreundlichen Form anzubieten.

Denn wir stehen momentan vor einer tief greifenden Transformation, was das digitale Angebot anbelangt. Digitale Nutzungsformen, wie sie sich derzeit in den Digital Humanities laufend entwickeln, verlangen sowohl nach Metadaten als auch nach Primärdaten, die an offenen Webschnittstellen abgegriffen werden können. Nur so wird ein Analysieren und Vergleichen mit anderen Datenquellen möglich sein.

Ein praktisches Beispiel: Das Lesen von Handschriften macht noch viel mehr Spass, wenn man mehrere Handschriften miteinander vergleichen kann – selbst wenn diese in verschiedenen Bibliotheken aufbewahrt und präsentiert werden. Moderne Plattformen wie e-Codices.ch machen dies dank der IIFSchnittstelle möglich. Solche Möglichkeiten entspringen nicht kauzigen Entwicklerlaunen, sondern den Bedürfnissen einer digitalen Philologie oder Komparatistik. Archive und Bibliotheken müssen sich kurz oder lang auf derartige digitale Nutzungsformen einstellen. Im Staatsarchiv Basel-Stadt verfolgen wir die Vision eines für Menschen wie Maschinen gleichermassen digital nutzbaren Archivs unter dem Banner «Digitales Archiv 2.0».

Daten und Hoffnungen

Zusammen mit dem Verantwortlichen der Archivinformatik investierte ich ziemlich viel Zeit in die Vorbereitung von Daten für den Hackathon. Wir überlegten uns, welche Daten sich gut eignen würden und wie ergänzend zum Archivkatalog ein alternativer Zugriff auf das Archivgut geschaffen werden könnte.

Die Ortsdeskriptoren, mit denen unsere Bildersammlung verschlagwortet ist, könnten dies leisten. Über den Ortsdeskriptor «Rheingasse» findet man beispielsweise 23 Verzeichniseinheiten. Um diese Toponyme für geobasierte Apps nutzbar zu machen, fehlte noch die Georeferenzierung oder die Anbindung an ein Normdatenverzeichnis wie die GND oder VIAF. Wir stellten uns vor, dass ein Genie diese Verbindung am Hackathon zustande bringen könnte.

Der zweite Datenbestand barg unserer Erwartung nach eine grössere entwicklerische Herausforderung und hätte daher reissenden Absatz finden sollen. Wir stellten nämlich die Tonprotokolle des Grossen Rats (des Parlaments) als Audiofiles und die dazugehörigen digitalisierten Protokollbände mitsamt OCR zur Verfügung. Mit Methoden der automatisierten Spracherkennung sollte eine Verknüpfung des Originaltons mit dem Text des Protokolls zu leisten sein. Dachten und hofften wir.

Leider wurde weder das eine noch das andere Datenset am Hackathon bearbeitet. Es mag sein, dass gerade keine Spezialistin der Spracherkennung zugegen war. Für solche Unterfangen empfiehlt es sich, bereits im Vorfeld Allianzen zu schaffen und gemeinsam den Hackathon zum Erstellen eines Prototyps zu nutzen. Vielleicht lag es auch an psychologischen Faktoren, warum niemand unsere Daten aufgegriffen hatte. Oft vermögen Katalogdaten nicht die notwendige kreative Fantasie freizusetzen, auf die Entwicklerinnen und Entwickler ansprechen.

So oder so, ich musste einsehen: Ein Hackathon ist kein Ort, wo jemand darauf wartet, meine Metadaten anzureichern. Ich nutzte die zwei Tage schliesslich, indem ich mich einem Team anschloss, das sich mit der Georeferenzierung von historischen Karten beschäftigte. Und ich lernte, wie sich ganz alte Karten grob mit Geokoordinaten versehen lassen.

Erkenntnisse

Am Hackathon wurde mir ein weiteres Mal bewusst, dass beim Veröffentlichen von Kulturdaten kein Weg am Fünf-Sterne-Prinzip herumführt. Dieses Prinzip von Tim Berners-Lee liest sich wie eine Roadmap zum digitalen Archiv 2.0.

Ganz am Anfang steht der Beschluss einer Institution, ihre Daten wirklich zu öffnen und nutzbar zu machen, und sie nicht bloss in einer Glasvitrine zu präsentieren. Hand aufs Herz: 90 Prozent der digitalisierten Akten und Bilder aus Archiven fallen in die Public Domain, zu Deutsch Gemeinfreiheit. Archive müssen nicht nur mit Schutzfristen den Datenschutz sicherstellen – das müssen sie nach wie vor – sondern auch die urheberrechtlichen Folgen der Digitalisierung berücksichtigen. Mit der Digitalisierung wird ein unermesslicher Mehrwert geschaffen. Dieser Mehrwert, den die Digitalisierung im Unterschied zur Mikroverfilmung schafft, liegt in der massiv erleichterten Nutzung. Die Digitalisierung erzeugt jedoch, obwohl sie teuer und aufwendig ist, kein neues Urheberrecht auf dem Original. Die Auszeichnung der Digitalisate als gemeinfrei gehört deshalb zur Veröffentlichung im Netz dazu.

Ist diese erste Hürde einmal geschafft, bekommt das Archiv den zweiten und dritten Stern gratis geschenkt, da in der Regel die Verzeichnisinformation gut ausgewiesen und kaum proprietäre Formate verwendet werden. Der vierte und fünfte Stern bezeichnen das Themengebiet, das die Archive heute beschäftigt. Während die meisten Bibliotheken ihre digitalen Angebote bereits mit DOIs (Digital Object Identifier) ausgestattet haben, verfügen Archive noch kaum über universell referenzierbare Adressen auf Objektebene. Und auch das Verwenden von global vernetzten Normdaten ist ein klares Desiderat an die Adresse der archivischen Erschliessung. Wer bisher mit lokalen Thesauri oder Deskriptoren gearbeitet hat, kann diese relativ einfach mit den grossen Normdatenverzeichnissen abgleichen (Diese Konkordanz kann auch über einen externen Resolverdienst geschehen, wie ihn metagrid.ch (www.metagrid.ch) anbietet.

Elias Kreyenbühl

Elias Kreyenbühl ist Mitarbeiter des Staatsarchivs Basel-Stadt. Er ist im Moment in einem mehrjährigen Projekt mit der Digitalisierung und Mikroverfilmung von Archivgut tätig.

Abstract

Français

Elias Kreyenbühl est collaborateur aux Archives d'État de Bâle-Ville en charge de projetsde numérisation et de copies sur microfilms. Convaincu que les idées sur l'utilisation du patrimoine doivent venir des utilisateurs eux-mêmes et que les archivistes sont un facilitateur à l'accès à ces données, quel que soit leur forme, il a décidé de participer au premier Swiss Open Cultural Data Hackathon dès qu'il a eu connaissance du #GLAMhack sur Twitter. Il nous décrit ici ses attentes et réflexions avant le Hackathon, comment il a préparé des jeux de données des Archives d'État avec les informaticiens, comment ces données n'ont pas été utilisées lors du Hackathon, et ce qu'il en a retiré en matière de géoréférencement de cartes historiques. Il conclut sur la nécessité de s'appuyer sur l'échelle de qualité des données ouvertes (1–5 étoiles proposées par Tim Berner-Lee) pour publier des données culturelles.