

Hacken erwünscht

Autor : Anne-Careen Stoltze

Datum : 16. Juni 2017



In Archiven, Bibliotheken und Museen schlummern grosse Datensätze, die im digitalen Zeitalter aus ihrem Dornröschenschlaf geweckt werden. Wie aus den alten Daten neue Anwendungen entstehen, zeigten Programmiererinnen und Programmierer am „#ZACK – Zürcher Archiv Hackday“. Dafür stellten die beteiligten Archive umfangreiches Text- und Bildmaterial zur Verfügung.

Seit Jahrhunderten sammeln Menschen Akten, Unterlagen, Bilder und andere Artefakte, welche ihr Leben und die Welt um sie herum dokumentieren. Diese Daten und Artefakte werden in Gedächtnisinstitutionen für die kommenden Generationen aufbewahrt und dem interessierten Publikum zugänglich gemacht. Die OpenGLAM-Bewegung engagiert sich im Sinne der Open-Data-Philosophie dafür, dass die Daten und digitalisierten Artefakte aus Gedächtnisinstitutionen („GLAM“ ist Englisch und steht für „Galleries, Libraries, Archives and Museums“) soweit wie möglich im Internet frei verfügbar gemacht werden, damit sie von interessierten Nutzern für diverse Zwecke weiterverwendet werden können. Am Schweizer Archivtag am 9. Juni öffneten die Zürcher Archive daher nicht nur ihre Türen, sondern auch gleich einen Teil ihrer Daten und luden Software Programmierer, Digital Humanists, Wikipedianer und andere Geschichts- und

Kulturinteressierte dazu ein, diese in unterschiedlichsten Formen weiterzuverwenden. Daneben stellten die Archive im Rahmen einer Vortragsreihe verschiedene Pilotprojekte vor.

Dass diese Öffnung der Bestände leider noch nicht überall Common Sense ist, zeigte Beat Estermann, stellvertretender Leiter des Forschungsschwerpunkts Open & Linked Data am [E-Government-Institut der Berner Fachhochschule](#) und Gründer der [OpenGLAM-Arbeitsgruppe Schweiz](#). Auf der Basis einer [Umfrage bei Gedächtnisinstitutionen in neun Ländern](#) untersuchte er, inwieweit die Gedächtnisinstitutionen ihre Daten und Digitalisate heute bereits zur freien Weiterverwendung bereitstellen. Die Unterschiede zwischen den einzelnen Ländern sind zum Teil beträchtlich, wobei die Niederlande generell die Nase vorn hat, während sich die Schweiz im hinteren Mittelfeld tummelt. Allerdings ist der Anteil an Inhalten, die durch Dritte frei genutzt werden dürfen, im Vergleich zu den Gesamtbeständen heute in allen Ländern noch gering. Binnen der nächsten fünf Jahre werde sich diese Situation jedoch grundlegend ändern, ist sich Estermann sicher, denn viele Archive gäben an, künftig einen substantiellen Teil ihrer Bestände für die Weiterverwendung freigeben zu wollen.

Er formuliert fünf Prinzipien, die Institutionen erfüllen müssten, damit OpenGLAM funktioniert:

1. Metadaten sollen ohne Nutzungsbeschränkungen online veröffentlicht werden.
2. Auch bei gemeinfreien Werke, welche online publiziert werden, sollen keine Nutzungsbeschränkungen geltend gemacht werden.
3. Bei der Publikation von urheberrechtlich geschützten Inhalten soll mittels standardisierter Lizenzen deutlich gemacht werden, wie die Inhalte weiterverwendet werden dürfen.
4. Um die Weiterverwendung zu erleichtern, sind Daten und Inhalte in maschinenlesbarer Form und in Open-Source-Dateiformaten zu veröffentlichen.
5. Die neuen Partizipationsmöglichkeiten, die sich durch das Internet ergeben, sollen gefördert werden.

Als Hauptgründe, weshalb sich viele Institutionen bei der Freigabe von digitalisierten Inhalten im Netz noch etwas zurückhalten, nennt Estermann das Urheberrecht, den Extraaufwand und die technischen Voraussetzungen, die es erfordert, analoge Dokumente zu digitalisieren, sowie die Angst vor dem Kontrollverlust seitens der Institutionen.

Eine Vorreiterrolle, wenn es um die Verbreitung der Open-Data-Philosophie in der öffentlichen Verwaltung geht, nimmt [Open Data Zürich](#) ein. Die Dienststelle ist bei der Stadt Zürich für die Umsetzung von Open Government Data zuständig und macht seit einigen Jahren einen wachsenden Teil von Daten der Zürcher Stadtverwaltung unter offenen Lizenzen zugänglich und unterstützt auch immer wieder Hackathons, um deren Nutzung zu fördern. „Wir verfügen über grosse Datenmengen über die Bürgerinnen und Bürger der Stadt Zürich, von den Beförderungszahlen der Zürcher Verkehrsbetriebe über Eheschliessungen und Taufen bis hin zu den registrierten Hunden und wie sie heissen“, sagt Marco Sieber, Projektleiter Statistik bei der Stadt Zürich. Er gab in seiner Präsentation einen anschaulichen Einblick über den Datenschatz. Zudem veröffentlichte Open Data Zürich eigens für den Hackday Datensätze, darunter die Ehe- und Taufbücher des Grossmünsters von 1708 bis 1732, das Bildarchiv des [Schweizerischen Arbeiterhilfswerks](#) sowie Fahrpläne der Tösstalbahn.

Auch andere Archive und Initiativen wie [Histhub](#) oder [TOPOterm](#) stellten ihre Projekte und Daten vor. [Dodis](#) verfügt etwa über die Diplomatischen Dokumente der Schweiz und stellt diese in gedruckter und digitalisierter Form zur Verfügung. Eine Teilnehmerin des letztjährigen [Schweizer Kulturhackathons](#) stellte zudem eine Online-Applikation mit Daten des [Verbandes Schweizerischer jüdischer Fürsorgen VSJF](#) aus dem Archiv für Zeitgeschichte vor, welche die Bewegungen jüdischer Flüchtlinge in der Schweiz im Laufe des 20. Jahrhunderts visualisiert.

Die meisten Institutionen besitzen Originaldokumente und suchen nach einer technischen Lösung, wie sie diese in geeignete Dateiformate umwandeln können. Ein besonders kniffliges Beispiel, die Handschriftenerkennung beim Projekt [READ \(Recognition and Enrichment of Archival Documents\)](#), stellte Tobias Hodel vom [Staatsarchiv Zürich](#) vor. Mittels Software werden historische Aufzeichnungen wie gedruckte Texte eingelesen und in Dateien umgewandelt. Dabei setzen die Forscher auf neuronale Netzwerke. „Allerdings müssen die Maschinen noch viel lernen; neu gibt es eine Software, die Frakturschrift erkennen kann“, erläuterte Hodel. Derzeit sei die Fehlerquote noch relativ hoch, sinke aber, je länger die Texte seien: Bei 150 Seiten liege die Quote bei 10 Prozent, bei 1000 Seiten um 3 Prozent. Das Problem sei, dass handschriftliche Texte meist keinen gleichbleibenden Schemen folgten. Randnotizen und Fussnoten erschweren die Erkennung weiter. Als Beispiel zeigte Hodel einen Brief des Dichters Rainer Maria Rilke, den der Algorithmus in eine Datei umwandelte – der Text war aufgrund der vielen Erkennungsfehler jedoch unlesbar.

Das Staatsarchiv stellte denn auch einen Datensatz von Rilkes Aufzeichnungen zur Verfügung, an dem ein Hackerteam arbeitete und die dazugehörigen Metadaten anreicherte und visualisierte. Andere Teams programmierten unter anderem ein Tool, das die Eheschliessungen in Zürich visualisiert, welche grafisch einer Spinne ähnlich sehen. Thomas Weibel, [Multimediaproduzent an der BFH](#) und an der [HTW Chur](#), verwandelte Fotos des Usterner Unternehmers und Fotografen [Julius Guyer](#) (1893-1909) in den [Prototypen eines Memorys](#).

Hintergrund:

Der [Zürcher Hackday](#) ist ein Gemeinschaftsprojekt von Deutschschweizer Archiven, der Arbeitsgruppe OpenGLAM des Vereins [opendata.ch](#) und [Wikimedia CH](#). Organisiert wurde der Anlass von Tobias Hodel vom Staatsarchiv Zürich und Karin Beck vom [Stadtarchiv Zürich](#) in Zusammenarbeit mit Vertretern von vier weiteren Archiven ([Sozialarchiv](#), [Stadtarchiv Uster](#), [Stadtarchiv Winterthur](#), [Schweizerisches Literaturarchiv](#)), welche für den Anlass eigene Daten bereitstellten. Der Hackday ist einer der Pre-Events zum [Swiss-Open-Cultural-Data-Hackathon](#), der am 15.-16. September an der [Universität Lausanne](#) stattfindet.