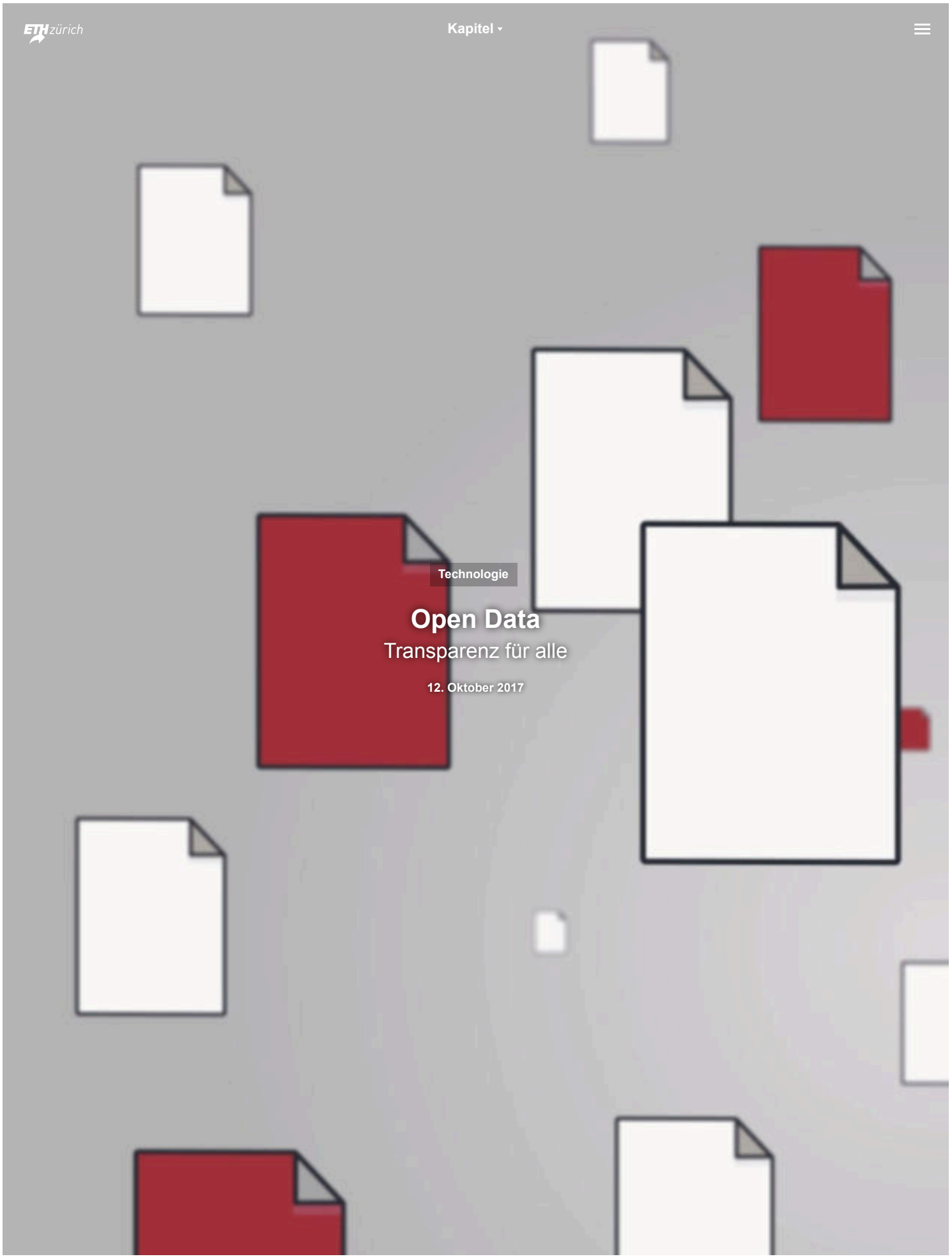


Technologie

Open Data

Transparenz für alle

12. Oktober 2017

Die Nutzung von Open Data ermöglicht neue Einblicke in Wissenschaft, Industrie und Gesellschaft. Wussten Sie, dass es in der Schweiz 21 Berge mit dem Namen „Schwarzhorn“ gibt? Oder dass die Bahnhofstrasse in Zürich 1863 benannt wurde und 1’173 Meter lang ist? Welche Denkmalpflegeobjekte befinden sich in Ihrer unmittelbaren Nachbarschaft? Diese und viele andere Erkenntnisse sind dank Open Data und Open Government Data (OGD), also offenen Behördendaten, möglich. Durch die Verknüpfung verschiedener Datensätze lassen sich beispielsweise Anwendungen[1] entwickeln, die mit verschiedenen Visualisierungen arbeiten. Doch wie macht man Daten offen und was zeichnet ‚offene Daten‘ aus? Gemäss Open Definition[2] lässt sich folgende Formulierung aufstellen:

> *Offene Daten sind Daten, auf die alle frei zugreifen und die frei genutzt, verändert und geteilt werden können.*[3]

Von historischen Bildern über Forschungsdaten bis zu Zürichs Spielplätzen…

Grundsätzlich können alle Daten geöffnet, sprich zugänglich gemacht, werden: Seien dies die eigenen Fotos auf Bildplattformen wie Wikimedia Commons, Kochrezepte in einem Forum oder Reiseberichte im eigenen Blog. Typischerweise handelt es sich bei Open Data aber immer um grössere

Datensets. Denn eine bestimmte Datenmenge ist Voraussetzung für die Entwicklung entsprechender Applikationen und somit auch für die Erzeugung eines Mehrwerts.

Mittlerweile betreiben viele Bibliotheken, Archive oder Museen Online-Plattformen, auf denen „Daten“ offen und frei zugänglich sind. Dies können digitalisierte Bilder sein und/oder auch die Metadaten, also die beschreibenden Informationen, zu diesen Daten. Auf Basis von Metadaten einer historischen Sammlung könnte beispielsweise eine Applikation entwickelt werden, die auf einer Karte anzeigt, aus welchen Orten die Urheber der Objekte stammen. Oder ein User kann erfahren, aus welcher Epoche am meisten Werke stammen. Richtig spannend wird dann die Verbindung der beiden Informationen in einer dynamischen Visualisierung: Ändern sich die Herkunftsorte im Lauf der Zeit?

In der Wissenschaft sind es üblicherweise Forschungsdaten, also gesammelte Daten aus Beobachtungen oder Experimenten, die öffentlich zugänglich gemacht werden. Dies fördert den internationalen Austausch und treibt die Forschung voran.

Auch Open Government Data stehen der breiten Bevölkerung vermehrt zur Verfügung. So können im Open-Data-Katalog der Stadt Zürich[4] Daten unterschiedlichster Natur recherchiert und anschliessend im gewünschten Format heruntergeladen werden. Dies können Informationen zu Verpflegungsbetrieben nach Jahr und Stadtquartier, Standorte von Brunnen, Spielplätzen, Kirchen oder Kinderkrippen in der Stadt Zürich oder auch Treibhausgasemissionen über mehrere Jahrzehnte sein, um nur ein paar Beispiele zu nennen.

Open Data City

Open Data im Web stammen von vielen Produzenten und kommen in unterschiedlichsten Formen vor. Von Bilddateien eines Archivs wie z.B. auf der Plattform E-Pics der ETH-Bibliothek, zum Open Data Portal der Schweiz (opendata.swiss) bis hin zu offenen Datensammlungen (sogenannten Repositorien). Hierbei bestehen freier Zugang und vielfältige Weiterverarbeitungsmöglichkeiten.

Wie viel Offenheit darf es sein? Die fünf Stufen von Open Data

Damit offene Daten mit verschiedenen Programmen kompatibel sind und weltweit auf Rechner heruntergeladen und weiterverarbeitet werden können, braucht es eine allgemein gültige Definition von „Offenheit“.[5] Tim Berners-Lee, der Erfinder des World Wide Web, hat hierzu ein fünfstufiges Open-Data-

Modell vorgeschlagen.[6] Auf der ersten, niedrigsten Stufe werden Daten – egal in welchem Format – unter einer offenen Lizenz[7] zur Verfügung gestellt. Die nächste Stufe erfordert eine Strukturierung der Daten, beispielsweise als Excel-Tabelle. Open Data auf der nächsthöheren Stufe zeichnen sich durch ein nicht-proprietäres, also ein offenes Dateiformat aus. Dies kann zum Beispiel eine CSV-Datei sein. Die zweithöchste Stufe erfordert zudem URIs (Unique Resource Identifiers), sodass Daten einfach und persistent auffindbar sind. Open Data auf der fünften Stufe sind als Linked Open Data mit anderen Daten verlinkt, sodass Kontext ersichtlich wird.

Fünfstufiges Open Data Modell
Tim Berners-Lee

Sichtbar und innovationsfördernd – Vorteile von Open Data

Das Konzept Open Data bringt viele Vorteile mit sich, hier werden die sechs in der Fachliteratur besonders oft genannten aufgeführt.

1. Transparenz
User können sich schnell und unkompliziert ein Bild über Art, Umfang und Inhalt der offenen Daten machen.

2. Zugänglichkeit
Der Zugriff auf Open Data ist einfach und unkompliziert.

3. Beitrag zur weltweiten Informationsinfrastruktur
Die Effizienz wird erhöht, da Daten nur einmal gesammelt werden müssen und keine doppelte Arbeit entsteht.

4. Entwicklung innovativer Anwendungen und Dienste
Open Data vereinfachen und beschleunigen die Entwicklung neuer Dienste, da rechtliche Abklärungen wegfallen.

5. Schaffung neuer Geschäftsmodelle
Dank dem Konzept Open Data ergeben sich neue Geschäftsmodelle.

6. Auffindbarkeit in Suchmaschinen
Offene Daten werden indexiert und somit bei Websuchen in der Ergebnisliste angezeigt.

Nebst den positiven Aspekten und Chancen zu neuen Erkenntnissen bringen offene Daten selbstverständlich auch Herausforderungen mit sich. Allem voran zählt dazu der Datenschutz: Gewisse Daten wie z. B. Kundendaten, dürfen nicht einfach offen zugänglich gemacht werden. Dies betrifft beispielsweise

Patientendaten in der Medizin oder personenbezogene Forschungsdaten. Hier muss die Anonymität zwingend gewährleistet werden können. Es bedarf also grundsätzlich immer einer gründlichen rechtlichen Abklärung, welche Daten überhaupt offen zugänglich gemacht werden dürfen (Data Governance).

Institutionen müssen erhebliche Ressourcen aufwenden, um vorgängig eine Auswahl an Daten zu treffen und die rechtlichen Abklärungen vorzunehmen. Auch eine technische Downloadmöglichkeit muss bereitgestellt werden, damit Entwickler und Programmierer möglichst unkompliziert zu den offenen Daten gelangen. Des Weiteren bedarf es auf Seiten der Nutzerinnen und Nutzer eines gewissen Know-hows, um aus den blossen Daten eine Visualisierung oder Applikation zu programmieren. Sonst nützen die schönsten Datenreihen auf der höchsten offenen Stufe nichts.

Open Data in der Praxis: Visualisierungen und Applikationen

Das Zugänglichmachen von Open Data durch Institutionen und Behörden ist selbstverständlich nur ein Aspekt. Genauso wichtig ist die Weiterverarbeitung der Daten zu Visualisierungen oder Applikationen, um einen Mehrwert aus den offenen Daten zu gewinnen. Ein Beispiel hierfür stellen sogenannte Hackathons dar. Datenlieferanten, Hacker und Interessierte treffen sich und versuchen während zwei bis drei Tagen, auf Basis von offenen Daten neue Anwendungen oder Visualisierungen zu entwickeln. In diesem Zusammenhang wurde auch der Swiss Open Cultural Data Hackathon → im Jahr 2015 ins Leben gerufen. Im September 2017 fand bereits die dritte Durchführung statt. Es entstand eine Vielzahl neuer Projekte, die allesamt auf Open Data basieren.

Auch die ETH-Bibliothek war als Datenlieferantin vor Ort und stellte Metadaten der Korrespondenz von Carl Gustav Jung, die sich in den Beständen des Hochschularchivs der ETH Zürich befindet, zur Verfügung. In einem vor Ort gebildeten Projektteam „Jung-Rilke Correspondence Networks“ fand eine intensive Auseinandersetzung mit den Korrespondenzbeständen von C. G. Jung und jenen von Rainer Maria Rilke (des Schweizerischen Literaturarchivs) statt. Das Ziel war die schrittweise Aufbereitung und Anreicherung der Daten, um sie anschliessend auf unterschiedliche Weise visualisieren zu können. So ermöglichte etwa die Geocodierung der vorhandenen Absendeorte die Darstellung der Korrespondenznetzwerke auf Karten.[8]

Geografische Verteilung der vermerkten Absendeorte der C.G. Jung-Korrespondenz
Visualisierung im Online-Tool Palladio

Heatmap der Korrespondenz von Rainer Maria Rilke
Je grösser und je mehr Rot in den Punkten, desto mehr Briefe wurden von diesem Ort aus versendet. Dargestellt mit Google Fusion Tables.

Korrespondenz von und an C. G. Jung als Netzwerk
Deutlich zu erkennen die beiden Knotenpunkte: Carl Gustav Jung (unten) und sein Sekretariat (oben). Dargestellt mit dem Tool Gephi.

Open Data in der Forschung: Open Research Data

Die Veröffentlichung von Ergebnissen ist ein Grundprinzip moderner Wissenschaft. Sie erlaubt es Wissenschaftlerinnen und Wissenschaftlern, Fehler zu identifizieren sowie Theorien zu unterstützen, abzulehnen oder weiter auszuarbeiten. Die Fähigkeit der modernen Wissenschaft zur Selbstkorrektur und stetigen Verbesserung setzt Offenheit voraus.[9] Ein neuer Gedanke ist die Idee, dass nicht erst Auswertungen und Interpretationen von Daten veröffentlicht werden sollen, sondern bereits blosse Messreihen und Rohdaten von Experimenten. Hiermit sind die sogenannten offenen Forschungsdaten gemeint.

Es gibt gute Gründe, warum auch Forschungsdaten als Open Data publiziert werden sollen. Hierzu zählen unter anderem, dass wissenschaftliche Studien häufiger zitiert werden, wenn die zugrundeliegenden Daten öffentlich zugänglich sind. Auch haben solche Studien eine höhere Glaubwürdigkeit. Zudem können qualitativ hochwertige Forschungsdaten allenfalls zusätzlich in spezialisierten Zeitschriften publiziert werden, und nicht zuletzt kann so die Forschungsarbeit für neue Projekte mit reproduzierbaren Ergebnissen nachgenutzt werden.[10]

Im Internet-Zeitalter sind Open Research Data aktueller denn je. Besonders Daten und Resultate von Forschung, die mit öffentlichen Geldern unterstützt wird, sollen nach Vorgaben der Förderinstitutionen für alle frei zugänglich sein, nicht zuletzt auch, um die Wissenschaft voranzutreiben. Daher verlangt der Schweizerische Nationalfonds zur Förderung der wissenschaftlichen Forschung (SNF) seit Oktober 2017 Datenmanagementpläne (DMP) als Bestandteil von Projektgesuchen. Ein Datenmanagementplan skizziert, wie die Forschungsdaten im Rahmen eines Projekts generiert, gesammelt, dokumentiert, publiziert, respektive öffentlich zugänglich gemacht, und schliesslich digital archiviert werden und unterstützt somit Forschende bei der Planung des Lebenszyklus‘ ihrer Daten.[11]

Lebenszyklus von Forschungsdaten

Die Leitlinien des Schweizerischer Nationalfonds (SNF)[12] für Forschende geben wichtige Kriterien zur Erstellung eines Datenmanagementplans. Zugrunde liegen die FAIR-Prinzipien[13]: Forschungsdaten sollen **F**indable, **A**ccessible, **I**nteroperable, **R**e-Usable sein, also auffindbar, zugänglich, in verschiedenen Systemen verwendbar sowie nachnutzbar. Vier Bereiche müssen mit einem Datenmanagementplan abgedeckt werden:

- Datenerhebung und -dokumentation
- Ethische, rechtliche und sicherheitstechnische Fragen
- Datenspeicherung und -erhalt
- Austausch und Weiterverwendung von Daten

Die Fachstelle Digitaler Datenerhalt → der ETH-Bibliothek bietet umfangreiche Beratung und Unterstützung zu allen Aspekten des Datenmanagements.[14] Forscherinnen und Forscher der ETH Zürich haben zudem die Möglichkeit, ihre Daten auf der Research Collection →, dem Repositorium für Publikationen und Forschungsdaten der ETH Zürich, zu publizieren. Von dort werden sie automatisch zur langfristigen Archivierung ins ETH Data Archive exportiert.[15]

Open Data an der ETH-Bibliothek

Als Informationsdienstleisterin nutzt die ETH-Bibliothek die Möglichkeit von Linked Open Data, indem sie beispielsweise die Einträge im eigenen Katalog → mit dynamischen Verlinkungen zu Wikipedia, der Deutschen Digitalen Bibliothek und anderen Quellen anreichert. Als Lieferantin von offenen bibliografischen Metadaten und Digitalisaten möchte die ETH-Bibliothek einen Beitrag zur Open-Data-Bewegung leisten. Wann immer möglich werden die eigenen Daten ohne rechtliche Beschränkungen für die weitere Verwendung bereitgestellt, das heisst unter einer Public Domain Mark [16] oder einer so genannten CC0-Lizenz[17]. Sind die Voraussetzungen hierfür nicht gegeben, wird eine möglichst offene Lizenz aus dem „Baukasten“ der Creative Commons → gewählt. Konkret stehen bibliografische Metadaten als Metadatensets verschiedener Publikationstypen → zum Download als Paket oder über eine direkte Schnittstelle (Z 39.50) zur Verfügung. Zudem ermöglicht das Bildarchiv der ETH-Bibliothek über eine so genannte BEACON-Datei die Verlinkung zu jenen Beständen, die mit GND[18] -referenzierten Personen verbunden sind.

Mehrere hunderttausend digitalisierte Dokumente – sowohl Text als auch Bild umfassend – können über verschiedene Plattformen eingesehen und heruntergeladen werden:

- E-Pics Bildarchiv Online →: Bildarchiv der ETH-Bibliothek mit Bildern zur Wissenschafts- und Technikgeschichte der Schweiz, Landschafts- und Ortsansichten, Fotoarchive der Swissair und der Fotoagentur Comet Photo AG. Im Blog Crowdsourcing → werden regelmässig frei zugängliche Bilderreihen vorgestellt. Ausgewählte Digitalisate und ‚born digitals‘ werden zur Nachnutzung mit einer offenen Lizenz auch auf Wikimedia Commons → bereitgestellt, womit sich die Einbindung in Wikipedia-Artikel vereinfacht.

- e-rara.ch →: Digitalisierte Drucke aus Schweizer Bibliotheken des 15. bis 20. Jahrhunderts.
- e-manuscripta.ch →: Digitalisierte handschriftliche Quellen aus Schweizer Bibliotheken und Archiven.

Airbus A310-221, HB-IPD „Solothurn" im Flug über dem Mattertal nach Norden
ETH-Bibliothek Zürich, Bildarchiv/Stiftung Luftbild Schweiz / Fotograf: Swissair →

Engelberg
ETH-Bibliothek Zürich, Bildarchiv/Stiftung Luftbild Schweiz / Fotograf: Swissair →

Zürich, Altstadt um St. Peter
ETH-Bibliothek Zürich, Bildarchiv / Fotograf: Georg Mörsch →

Querschiff der Kathdrale von Burgos
Genaro Pérez Villaamil. Patricio de la Escosura: España Artística... →

Waldlaubsänger
The birds of America: from drawings made in the United States and their territories / by John James Audubon →

Der Affe von Maco
Georges-Louis Leclerc de Buffon: Histoire naturelle générale... →

Maschinenbaukunde. Vorlesung von Franz Reuleaux
Vorlesungsnachschrift, ausgearbeitet von Heinrich Berchtold →

Wälle am Simplon
Arnold Escher von der Linth →

Mechanik. von David Lehrer an Kunstschu
Vorlesungs ausgearbei Johannes Hottinger →

Ausblick: Open Data auf dem Vormarsch?

Dass sich das Prinzip von Open Data immer weiter verbreitet und auch Regierungen weltweit ihre Daten mit zunehmender Tendenz öffnen, lässt sich sehr gut am Open Data Barometer[19] ablesen. Seit 2013 wird auf dem Portal verzeichnet, wie Behörden Open Data im Hinblick auf die Rechenschaftspflicht, die Innovation und den Social Impact bereitstellen. Die Schweiz konnte sich zwar jährlich steigern, befindet sich aber aktuell mit einem Wert von 43/100 nur auf Platz 22 von 77 ausgewerteten Nationen. Nur wenn Behörden und Organisationen weiterhin im Sinne von Open Data agieren und auch die breite Öffentlichkeit erfährt, welches Potenzial in Open Data steckt, kann der Trend zu einer offenen Informationsgesellschaft beschleunigt werden. Die ETH-Bibliothek engagiert sich, um diese Entwicklung voranzutreiben und den Gedanken von Open Data weiter zu fördern und bekannt zu machen.

---

1. Eine Auswahl an Open-Data-Anwendungen gibt es auf dem Schweizer Open Government Data Portal: https://opendata.swiss/de/app → (29.08.2017). ↩
2. Open Definition (2017): Offen-Definition. Version 2.0. Verfügbar auf: http://opendefinition.org/od/2.0/de/ → (08.08.2017). ↩
3. Die Hauptmerkmale von Open Data sind: 1. Verfügbarkeit und Zugang (Daten müssen als ganze Einheit und als kostenfreier Download im Internet zur Verfügung stehen). 2. Wiederverwendung und Weitergabe (Daten müssen maschinenlesbar sein und so zur Verfügung gestellt werden, dass eine Wiederverwendung und Weitergabe bedingungslos möglich ist). Universelle Beteiligung (Es muss für jede und jeden möglich sein, die Daten zu nutzen und weiterzuverwenden). Vgl. auch Open Knowledge International (2017): What is open? Verfügbar auf: https://okfn.org/opendata/ → (08.08.2017). ↩
4. Verfügbar auf https://data.stadt-zuerich.ch/ → (11.08.2017). ↩
5. Open Data Handbook (2017): Was ist Open data? Verfügbar auf: http://opendatahandbook.org/guide/de/what-is-open-data/ → (08.08.2017). ↩
6. 5 Stern offene Daten (2012). Verfügbar auf: http://5stardata.info/de/ (08.08.2017). ↩
7. Eine offene, oder freie, Lizenz erlaubt die Nutzung, Weiterverbreitung und Änderung urheberrechtlich geschützter Werke. Ein Beispiel für diese frei zu vergebenden Lizenzen sind jene von Creative Commons: https://creativecommons.org/ → (28.09.2017). ↩
8. Einblick in das Projekt “Jung-Rilke Correspondence Networks» und die vielfältigen Visualisierungsmöglichkeiten gibt das Video der Projekt-Abschlusspräsentation: https://vimeo.com/234627486 → (28.09.2017) sowie die Projektseite im Wiki des Hackathons: http://make.opendata.ch/wiki/project:jung_rilke_correspondance_network → (29.09.2017). ↩
9. The Royal Society (2012): Final report – Science as an open enterprise. Verfügbar auf: https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/ → (28.07.2017). ↩
10. Vgl. auch ETH-Bibliothek, Open Access an der ETH Zürich (2017): Forschungsdaten publizieren. Verfügbar auf: http://www.library.ethz.ch/de/ms/Open-Access-an-der-ETH-Zuerich/Forschungsdaten-publizieren → (28.07.2017). ↩
11. Schweizerischer Nationalfonds (SNF) (2017): Open Research Data: Das sind die SNF-Guidelines für die Datenmanagementpläne. Verfügbar auf: http://www.snf.ch/de/fokusForschung/newsroom/Seiten/news-170511-open-research-data-snf-guidelines-fuer-datenmanagementplaene.aspx → (28.07.2017). ↩
12. Schweizerischer Nationalfonds (SNF) (2017): Daten Management Plan (DMP) – Leitlinien für Forschende. Verfügbar auf: http://www.snf.ch/de/derSnf/forschungspolitische_positionen/open_research_data/Seiten/data-management-plan-dmp-leitlinien-fuer-forschende.aspx → (28.07.2017). ↩

13. Vgl. auch Wilkinson, Mark D. (2016): The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3:160018. Verfügbar auf: http://doi.org/10.1038/sdata.2016.18 → . ↩

14. Vgl. auch ETH-Bibliothek, Digitaler Datenerhalt (2017): Datenmanagement. Verfügbar auf: http://www.library.ethz.ch/de/ms/Digitaler-Datenerhalt-an-der-ETH-Zuerich/Forschungsdaten/Datenmanagement → (28.07.2017). ↩

15. Auch eine direkte Archivierung von Daten im ETH Data Archive ist möglich. Diese wird v. a. bei grossen Datenpaketen, einer regelmässigen und automatisierten Archivierung oder auch bei einer frühzeitigen Strukturierung der Daten vor Ort empfohlen. Eine detaillierte Aufstellung findet sich auf der Website der ETH-Bibliothek → . ↩

16. Vgl. auch https://creativecommons.org/publicdomain/mark/1.0/deed.de → (02.10.2017). ↩

17. Vgl. auch https://creativecommons.org/publicdomain/zero/1.0/deed.de → (02.10.2017). ↩

18. GND ist die Gemeinsame Normdatei und enthält Datensätze für Personen, Familien, Körperschaften, Konferenzen, Geografika, Sachbegriffe und Werktitel: https://wiki.dnb.de/display/ILTIS/Informationsseite+zur+GND → (29.09.2017). ↩

19. http://opendatabarometer.org → (11.08.2017). ↩

Teile diesen Artikel

Newsletter

Eine Erlebniswelt der ETH-Bibliothek. Jetzt den Newsletter abonnieren und keine Geschichte verpassen.

Email Adresse

Anmelden

Text
Deborah Kyburz

Redaktion
Lisa Ott, Kathrin Reith, Matthias Töwe, Stefan Wiederkehr

Wir danken
Hackathon Projektteam Jung-Rilke Correspondence Networks

DOI Link

10.22010/ethz-exp-0004-de →

Dieses Werk unterliegt einer Creative Commons Attribution-ShareAlike 4.0 International Public License.

© 2017 ETH-Bibliothek