



Open Humanities

Vorlesung *Einführung in die Digital Humanities*

Prof. Dr. Christof Schöch
Wintersemester 2017/18

Einstieg

Semesterüberblick

- 23.10.: Digital Humanities im Überblick
- 30.10.: Digitalisierung: Text und Bild
- 06.11.: Grundbegriffe des Programmierens
- 13.11.: Datenmodellierung 1: Modellierung
- 20.11.: Datenmodellierung 2: Datenbanken
- 27.11.: Datenmodellierung 3: Text, Markup, XML
- 04.12.: Digitale Edition
- 11.12.: Geschichte der Digital Humanities
- 18.12.: Informationsvisualisierung
- 22.12.-2.1.: *Weihnachtspause*
- 08.01.: Computerlinguistik / Natural Language Processing
- 15.01.: Quantitative Analyse 1: Stilometrie
- 17.01.: Quantitative Analyse 2: Superv. Machine Learning
- 29.01.: **Open Humanities**
- 05.02.: Klausurtermin

Sitzungsüberblick

1. Was ist mit "open" gemeint?
2. Publikationen: Open Access (OA)
3. Daten: Open Data
4. Software: Open Source
5. Lehre: Open Educational Resources (OER)

1. Was ist mit "open"
gemeint?

Definition von Open Science

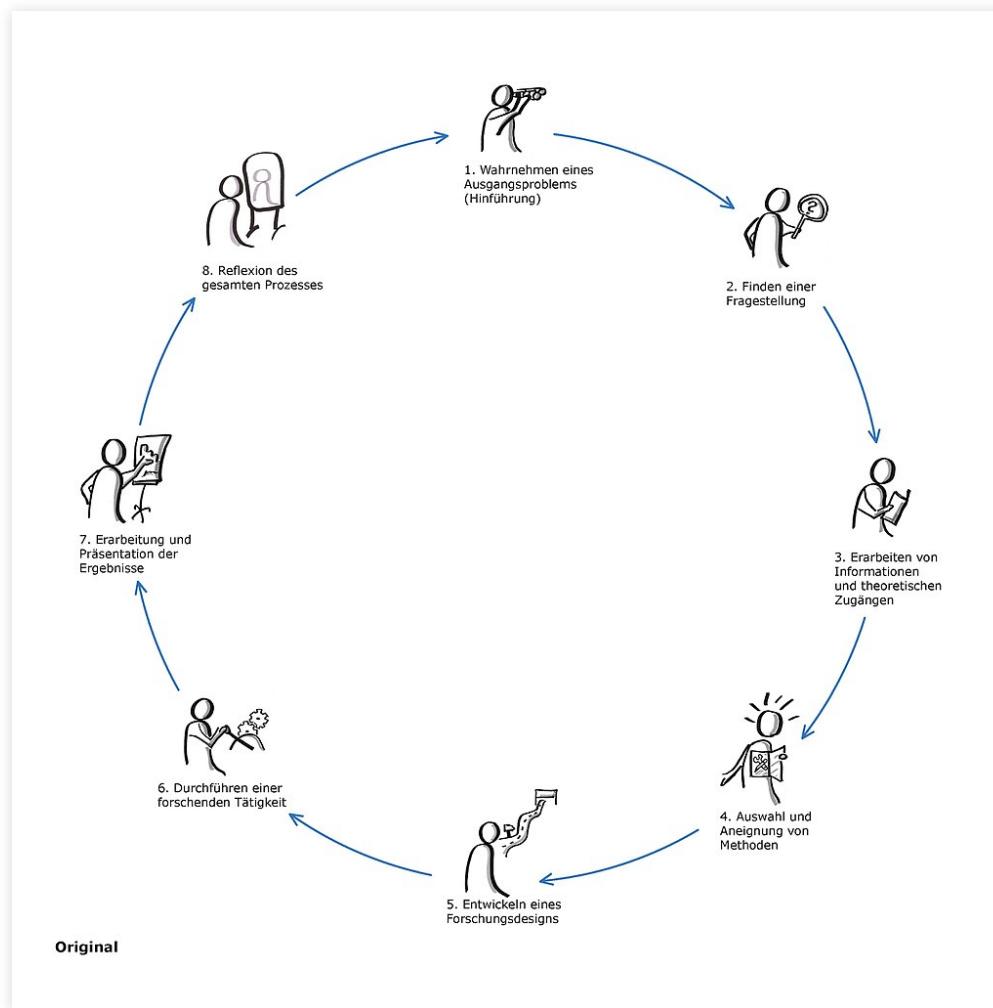
„Open Science öffnet den wissenschaftlichen Prozess von der ersten Idee bis zur finalen Publikation, um diesen möglichst nachvollziehbar und für alle nutzbar zu machen.“

(Quelle: <http://www.openscienceasap.org>)

Forschungsprozess

- Stand der Forschung und Idee
- Fragestellungen / Hypothesen
- Datenerhebung und -erschließung
- (qualitative / quantitative) Datenanalyse
- Auswertung und Interpretation
- Publikation der Daten und Ergebnisse
- und: Vermittlung in der Lehre

Forschungsprozess



Monika Sonntag et al. "Idealtypische Anordnung der Phasen eines Forschungsprozesses im Forschenden Lernen nach Huber", Wikimedia Commons, 2016, <https://commons.wikimedia.org/wiki/File:Forschungsprozess.jpg>, Lizenz: CC-BY-SA 4.0

Gegenstände

- Datenerhebung: „open data“
- Datenanalyse: Software / „open source“
- Ergebnisse: Publikationen / „open access“
- Lehre: „open education“

Aspekte

- Etablierte vs. digitale Abläufe (Digitalisierung)
- rechtliche Rahmenbedingungen (Urheberrecht, Lizenzen)
- technische Rahmenbedingungen (bspw. Standards)
- Geschäftsmodelle und Modelle von Wissenschaft
- kostenlos (free) vs. offen (libre)

2. Publikationen: Open Access (OA)

Definition (Wikipedia)

Open access (OA) refers to research outputs which are distributed online and free of cost or other barriers, and possibly with the addition of a Creative Commons license to promote reuse. Open access can be applied to all forms of published research output, including peer-reviewed and non peer-reviewed academic journal articles, conference papers, theses, book chapters, and monographs.

(Quelle: "Open Access", Wikipedia, https://en.wikipedia.org/wiki/Open_access, Jan. 2019, Lizenz: CC-BY-SA)

Budapest OA Initiative

By 'open access' to this literature, we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited.

(Quelle: Budapest Open Access Initiative, 2001, <https://www.budapestopenaccessinitiative.org/>, Lizenz: CC-BY)

Zeitschriftenpublikation: Prozess

- Forscher/in reicht Artikel bei Zeitschrift ein
- Zeitschrift organisiert den „peer review“
- Annahme oder Ablehnung des Artikels
- Teilweise: Forscher/in bezahlt Publikationsgebühr
- Forscher/in tritt alle Verwertungsrechte an den Verlag ab
- Artikel wird von der Zeitschrift lekoriert
- Zeitschrift veröffentlicht und bewirbt den Artikel
- Bibliotheken kaufen die Zeitschrift
- Andere Forscher/innen lesen den Artikel

Zeitschriftenpublikation: Problem

- Die Forscher/innen machen die meiste Arbeit (Forschen, Schreiben, Review)
- Die Forscher/innen werden mit Steuergeldern bezahlt
- Die Publikationsgebühr wird mit Steuergeldern bezahlt
- Die Zeitschriften werden mit Steuergeldern gekauft
- Die Steuerzahler/innen haben keinen Zugang zu den Artikeln
- Andere Menschen, in vielen Ländern, haben keinen Zugang
- Manche Verlage verdienen sehr viel Geld (Elsevier: 36% Gewinnmarge)

Open Access als Lösung?

- Die wichtigsten Schritte werden vorab finanziert (staatlich oder privat)
- Alle Menschen haben Zugang zu wiss. Erkenntnis
- Zugangsbeschränkung ist zusätzlicher Aufwand
- Aber: auch Korrekturlesen, Layouten, Drucken/Server kosten Geld

Paradigmenwechsel

- **Print:** *Erst filtern, dann publizieren*
(Kosten für Herstellung und Distribution sind hoch)
- **Digital:** *Publish first, filter later*
(Publikation ist billig, aber Aufmerksamkeit ist knapp)

Open Access: Varianten

- Green Open Access: Selbst-Archivierung
- Gold Open Access: OA-Zeitschrift



Green Open Access

- Artikel werden klassisch publiziert (print/digital)
- Parallel passiert ein OA-Deposit:
 - Artikel wird in Repository online publiziert
 - Varianten: *pre-print / post-print*
 - evtl. mit Embargo-Zeit (*moving wall*)
- Klassisches Geschäftsmodell bleibt weitgehend erhalten

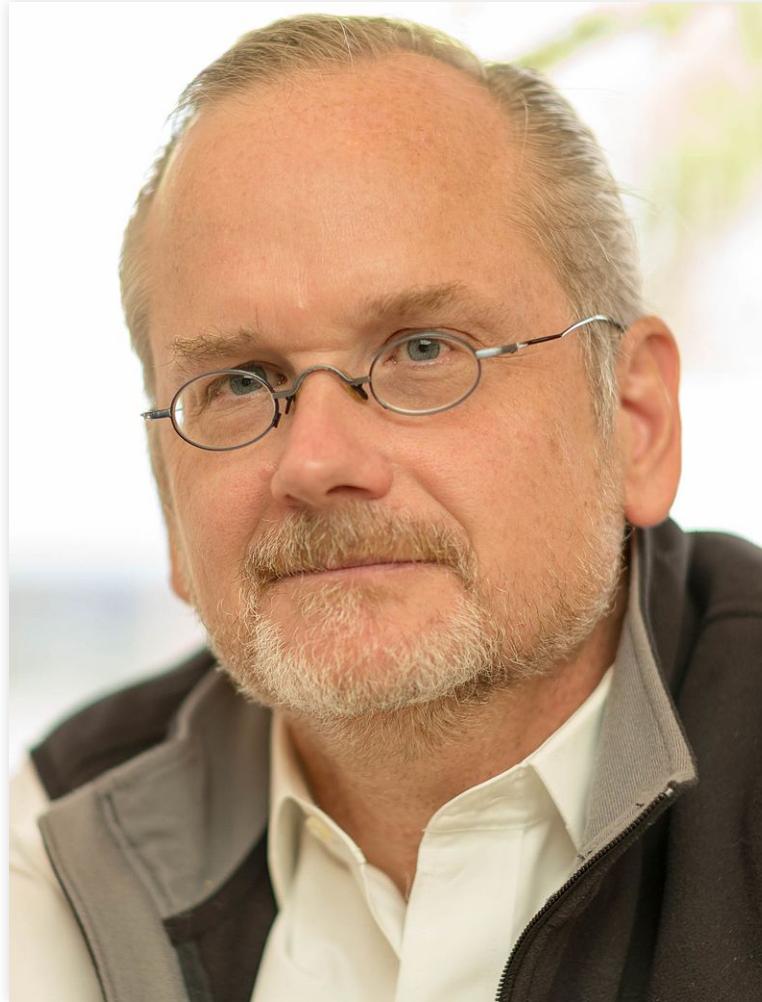
Gold Open Access

- Artikel wird direkt in einer Open Access-Zeitschrift
- Häufig: „Article Processing Charges“ (Publikationsgebühr) statt Subskriptionsgebühr (DEAL: durch nationale Pauschale abgedeckt)
- Neue Geschäftsmodelle:
 - *freemium*: Basisangebot vs. Zusatzleistungen
 - Institutionelle Mitgliedschaft

Creative Commons-Lizenzen (CC)

- 2001 gegründet von Lawrence Lessig
- Lizenzen beschreiben die anderen eingeräumten Rechte
- Ein einfaches, standardisiertes Mittel, um unter präzisen Bedingungen die Nutzung von intellektuellen / künstlerischen Werken zu gestatten
- Urheberrecht und Autorschaft bleibt gewahrt!
- Siehe: <http://creativecommons.org/licenses/>

Lawrence Lessig



(Quelle: Joi Ito, "Lawrence Lessig in May 2017",
https://en.wikipedia.org/wiki/Lawrence_Lessig#/media/File:Lawrence_Lessig_May_2017.jpg, Lizenz: CC-BY 2.0)

Creative Commons Varianten



- CC-0 (Zero): keinerlei Beschränkung der Nutzung
- CC BY (Attribution): Urheber und Lizenz mit Link nennen
- CC BY-SA (Share Alike): Wie CC BY, zudem keine Änderung der Lizenz
- CC BY-NC (Non-Commercial): Wie CC BY-SA, zudem keine kommerzielle Nutzung
- CC BY-ND (No Derivatives): Wie CC-BY-SA, zudem keine Änderungen

Weitere Themen: Wissenschaftskommunikation

- Bloggen: Alternative Publikationsform
 - Kleinere Beiträge aus dem Forschungsalltag
 - Selbstpublikation (kein Filter, keine Qualitätssicherung)
 - Stilistische und thematische Freiheit
 - Plattform für wiss. Bloggen: de.hypotheses.org
- Twitter (Mikroblogging)
 - Ankündigungen, Bemerkungen, Konferenzbegleitung
 - persönliche Perspektive, Diskussionen
 - "Digitale Identität" und Echoboden

Weitere Themen: Impact

- Klassisches Modell
 - Maße für Wichtigkeit einer Publikation: Anzahl der Zitationen durch andere Forscher/innen
 - Verkaufszahlen (Bücher)
- Open Metrics
 - Alternative, „offene“ Maße für die Wichtigkeit einer Publikation
 - Merkmale für „impact“: Downloads, Twitter, Facebook, Blogs, Bibliographien, uvm.
 - Anbieter: Altmetrics.com, ImpactStory, Repositories

Eintrag bei Zenodo.org

The screenshot shows a Zenodo.org page for a research preprint. At the top, the Zenodo logo and the tagline "Research. Shared." are visible. The navigation bar includes links for Search, Communities, Browse, Upload, Get started, and a user account dropdown for c.schoech@gmail.com.

The main content area displays the following details:

- Date:** 15 October 2014
- Type:** Preprint | Open access
- Title:** The Open Science Peer Review Oath
- Authors:** Aleksic, Jelena ; Alexa, Adrian ; Attwood, Teresa K ; Bolser, Dan ; Dahlö, Martin ; Davey, Robert ; Dinkel, Holger ; Förstner, Konrad ; Grigorov, Ivo ; Hériché, Jean-Karim ; Chue Hong, Neil ; Lahti, Leo ; MacLean, Dan ; Markie, Michael L ; Molloy, Jenny ; Schneider, Maria Victoria ; Scott, Camille ; Smith-Unna, Richard ; Vieira, Bruno Miguel
(show affiliations)
- Abstract:** One of the foundations of the scientific method is to be able to reproduce experiments and corroborate the results of research that has been done before. However, with the increasing complexities of new technologies and techniques, coupled with the specialisation of experiments, reproducing research findings has become a growing challenge. Clearly, scientific methods must be conveyed succinctly, and with clarity and rigour, in order for research to be reproducible. Here, we propose steps to help increase the transparency of the scientific method and the reproducibility of research results: specifically, we introduce a peer-review oath and accompanying manifesto. These have been designed to offer guidelines to enable reviewers (with the minimum friction or bias) to follow and apply open-science principles, and support the ideas of transparency, reproducibility and ultimately greater societal impact. Introducing the oath and manifesto at the stage of peer review will help to check that the research being published includes everything that other researchers would need to successfully repeat the work. Peer review is the lynchpin of the publishing system: encouraging the community to consciously (and conscientiously) uphold these principles prior to publication should help to improve published papers, increase confidence in the reproducibility of the work and, ultimately, provide strategic benefits to authors and their institutions. Future incarnations of the various national Research Excellence Frameworks (REFs) will evolve away from simple citations towards measurable societal value and impact. The proposed manifesto aspires to facilitate this goal by making transparency, reproducibility and citizen-scientist engagement with the knowledge-creation and dissemination processes, the default parameters for performing sound research.
- Statistics:** 48 (with a circular icon)
- Sharing:** See more details | Tweeted by 84 | On 2 Facebook pages | Mentioned in 3 Google+ posts
- Publication date:** 15 October 2014
- DOI:** DOI 10.5281/zenodo.12273
- Keywords:** open science | peer review | reviewer | oath
- Grants:** ALLBIO - Broadening the Bioinformatics Infrastructure to unicellular, animal, and plant science (289452)
- Collections:** Communities > FP7 FOSTER, Facilitating Open Science in European Research | Communities > Ocean Literacy | Publications > Preprints | Open Access
- License (for files):** Creative Commons Attribution Share-Alike
- Uploaded by:** danmaclean (on 16 October 2014)

A preview section at the bottom left shows a small thumbnail of the document and various viewing options like search, zoom, and print.

Open Access in den DH

- *Digital Scholarship in the Humanities* (EADH/ADHO, OUP): Subskription / OA
- *Digital Humanities Quarterly* (ADHO): Open Access
- *Journal of Digital Humanities*: OA (post-publ. peer review)
- *Computers and the Humanities* (-2006): Subskription
- *Frontiers in Digital Humanities*: Open Access (APC)
- *Zeitschrift für digitale Geisteswissenschaften*: Open Access, kostenfrei

Open Access-Publikationen finden

- DOAJ (Directory of Open Access Journals): <http://doaj.org/>
- OpenDOAR (Directory of Open Access Repositories):
<http://www.opendoar.org/find.php>
- BASE (Bielefeld Academic Seach Engine): <https://de.base-search.net/>

3. Daten: Open Data

Definition

Open Data bedeutet die freie Verfügbar- und Nutzbarkeit von, meist öffentlichen, Daten. Sie beruht auf der Annahme, dass vorteilhafte Entwicklungen unterstützt werden wie Open Government, wenn Daten für jedermann frei zugänglich gemacht werden und damit mehr Transparenz und Zusammenarbeit ermöglichen. Dazu verwenden die Ersteller Lizenzmodelle, die auf Copyright, Patente oder andere proprietäre Rechte weitgehend verzichten.

(Quelle: Artikel "Open Data", Wikipedia, http://de.wikipedia.org/Open_Data)

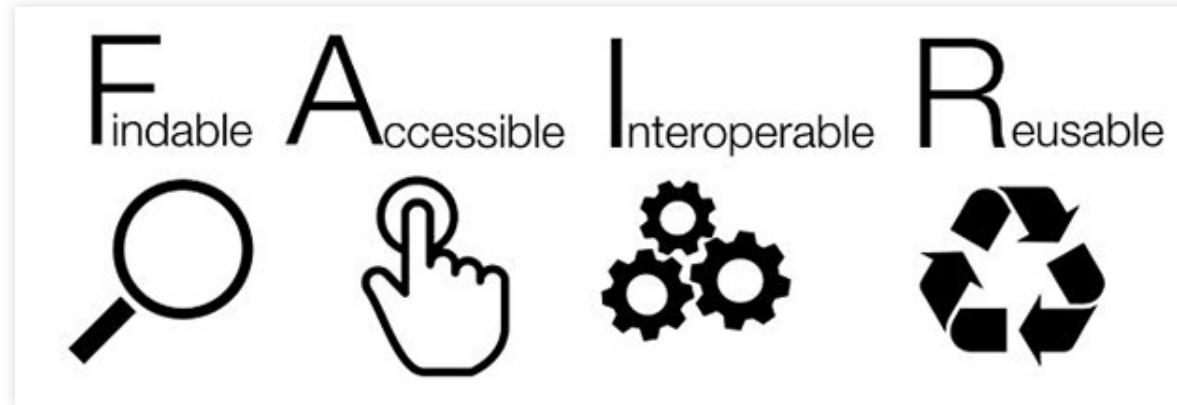
Begrifflichkeit

- Open Data in Science
- Open Government Data
- Linked Open Data
- Data vs. Metadata (Metadaten sind auch Daten!)

Warum "Open Data"?

- Transparenz ("accountability")
- Reproduzierbarkeit
- Kollaboration
- Abgeleitete Services
- Fairness

FAIR Data Principles



- **Findable:** bspw., durch Metadaten, Katalogeinträge
- **Accessible:** bspw., ohne technische Hürden (Repository)
- **Interoperable:** bspw., offene Standards und Formate
- **Re-usable:** bspw., offene Lizenz; Dokumentation

(Quelle: Go FAIR, <https://www.go-fair.org/fair-principles/>, 2016)

Open Data: Rechtliches

- Datenbanken sind rechtlich geschützt
- Nach Urheberrecht: Besondere Struktur der Datenbank; "Schöpfungshöhe" = persönliche kreative Leistung
- Nach Datenbankherstellerrecht: "wesentliche Investition" (fast immer der Fall)
- Hinweise
 - Beides gilt auch, wenn die Inhalte urheberrechtsfrei sind!
 - Ziel ist der Investitionsschutz
 - Ausnahmen für wissenschaftliche, nicht-kommerzielle Zwecke

Open Data: Lizenzen

- Lizenzen sind für Software oder Publikationen entwickelt worden
- Es gibt spezielle Lizenzen für Daten: Open Data Commons: ODC-Lizenzen, <http://opendatacommons.org/licenses>
- Aber auch Creative-Commons-Lizenzen sind geeignet und besser verständlich

Open Data: Standards

- Standards: "codified rules and guidelines for the creation, description and management of digital resources" (Gill and Miller, 2002)
- Beispiele
 - Zeichenkodierung: UTF-8 / Unicode
 - Text: HTML, XML-TEI, etc.
 - Bilder: TIFF, JPEG, SVG, etc.
 - Audio: MP3, FLAC, etc.
 - Metadaten: Dublin Core, MARC, etc.

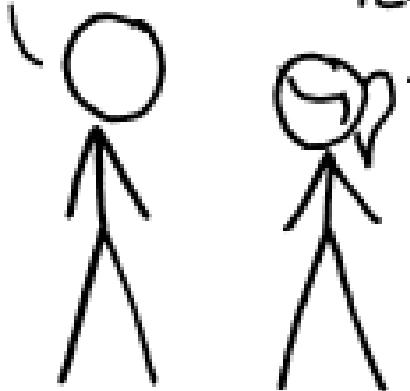
How Standards Proliferate

HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

SITUATION:
THERE ARE
14 COMPETING
STANDARDS.

14?! RIDICULOUS!
WE NEED TO DEVELOP
ONE UNIVERSAL STANDARD
THAT COVERS EVERYONE'S
USE CASES.

YEAH!

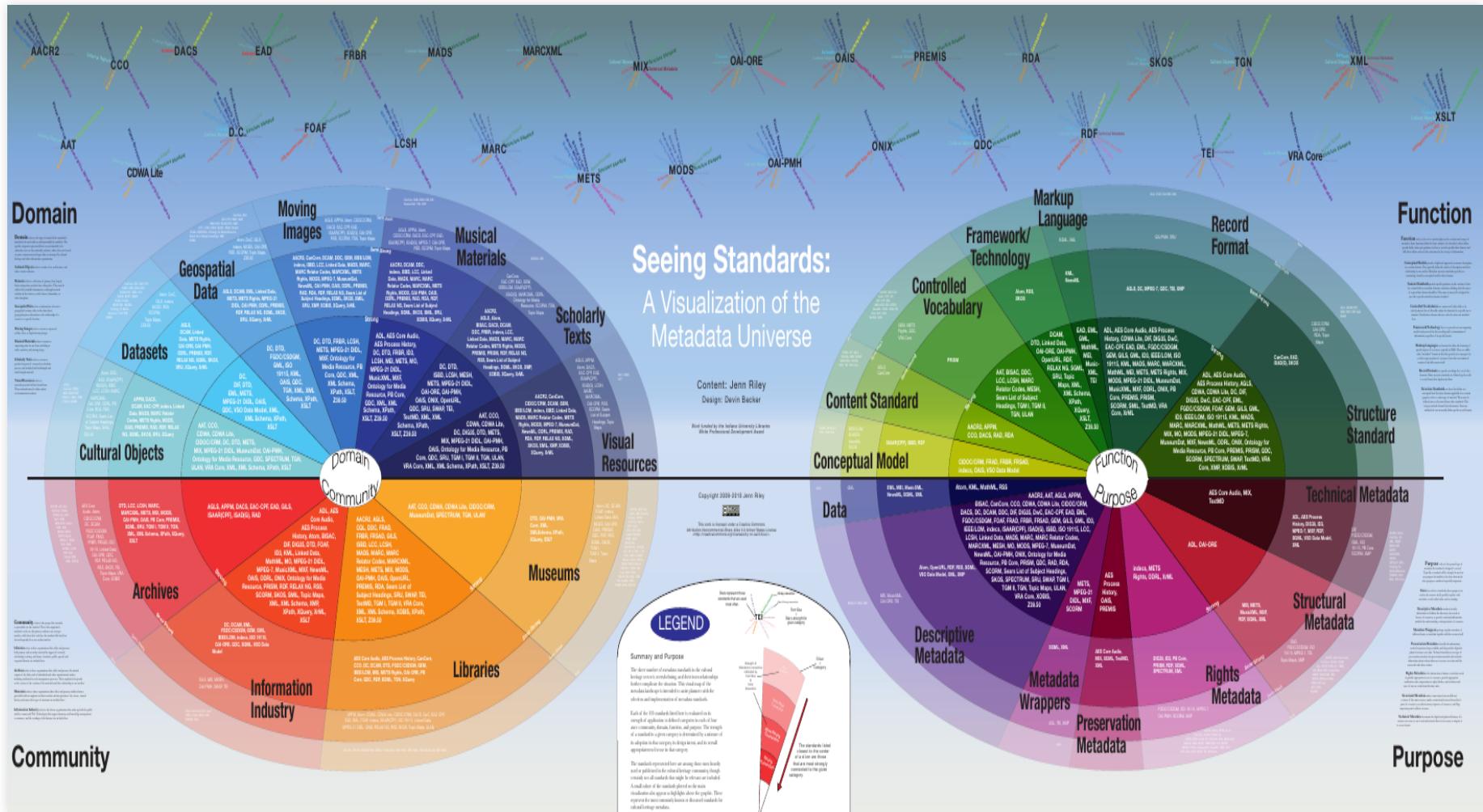


SOON:

SITUATION:
THERE ARE
15 COMPETING
STANDARDS.

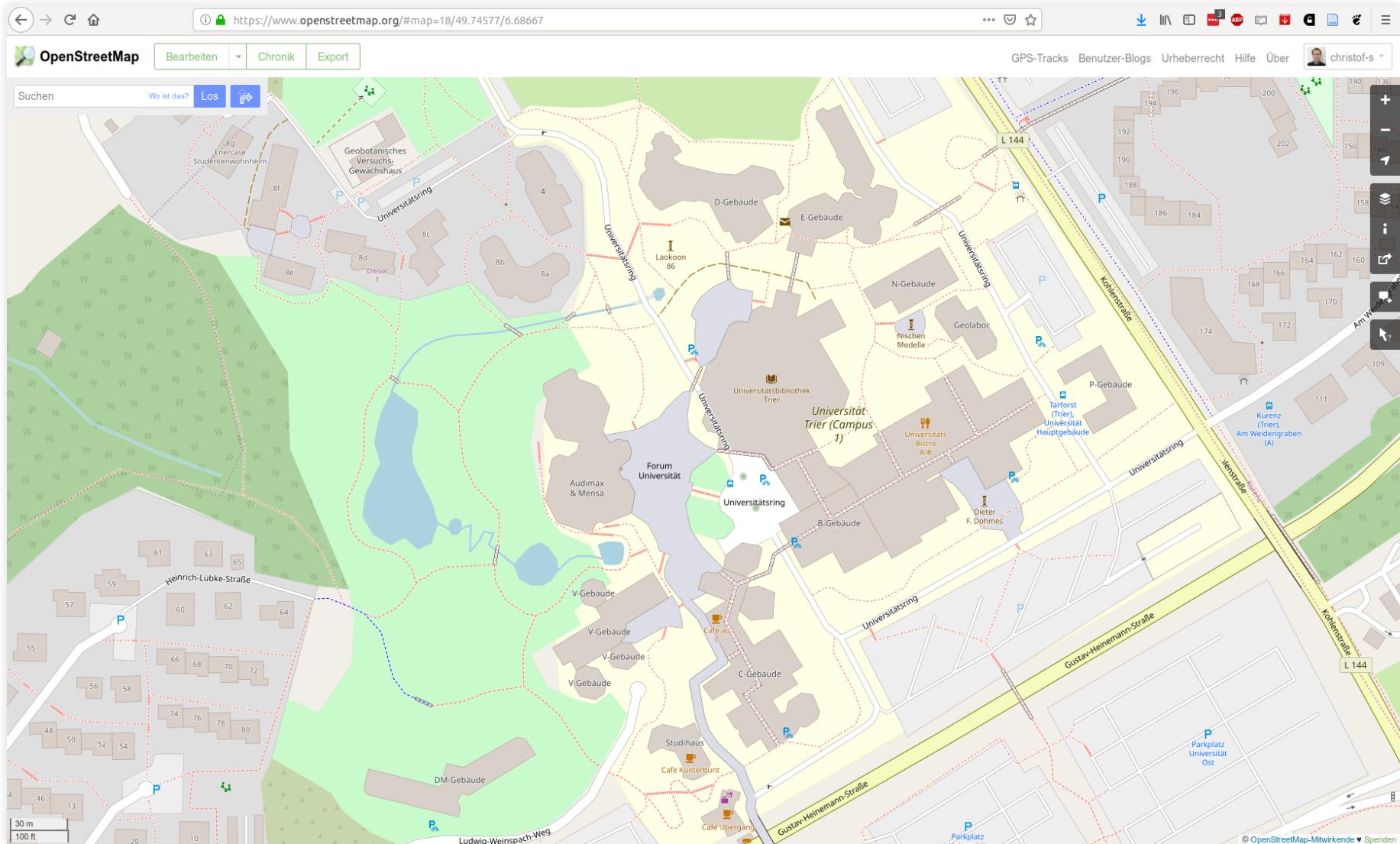
Quelle: xkcd, "How Standards Proliferate", URL: <https://xkcd.com/927/>, Lizenz: CC BY-NC 2.5

Metadata Map



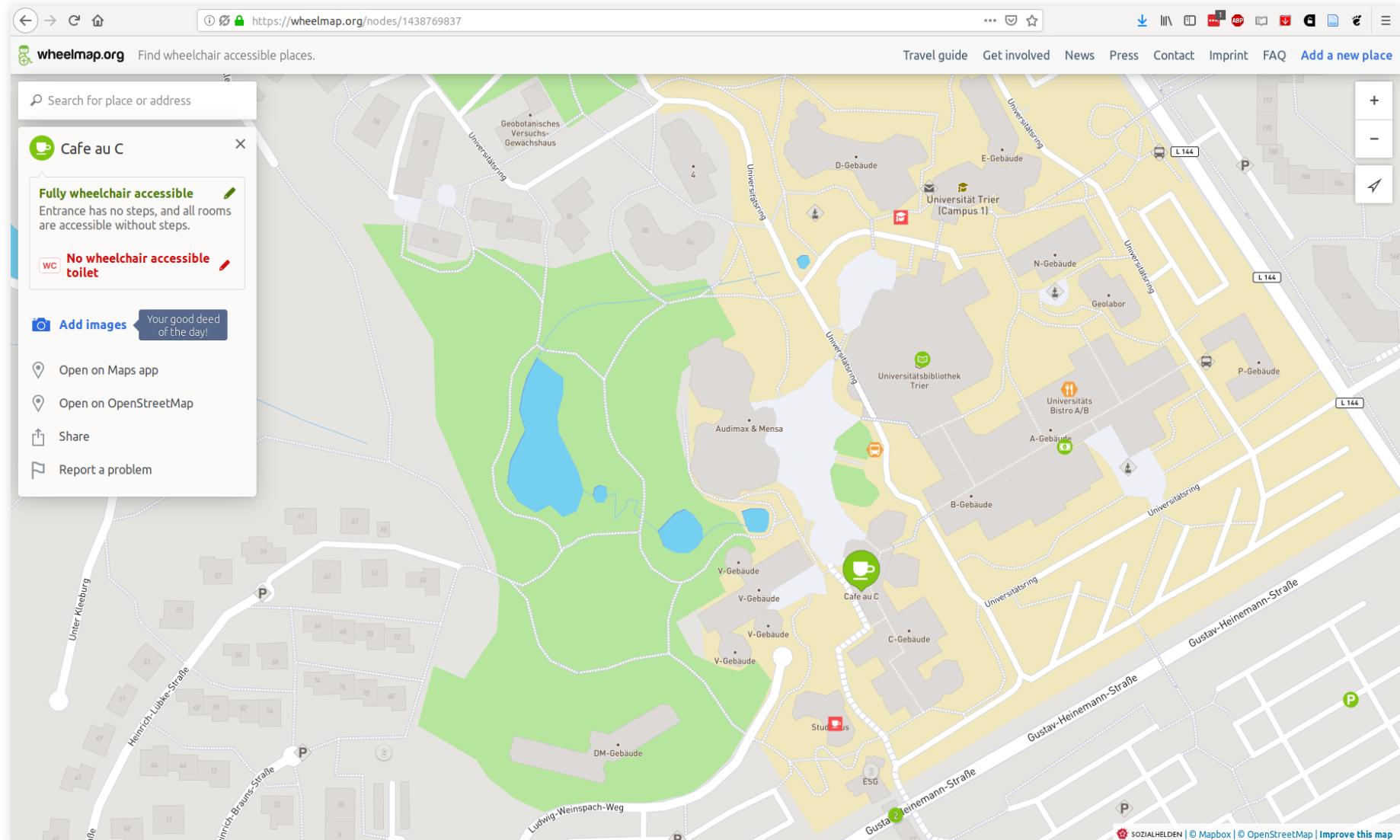
(Quelle: <http://jennriley.com/metadatamap/>, CC-BY-NC-SA 3.0)

Beispiel "Open Street Map"



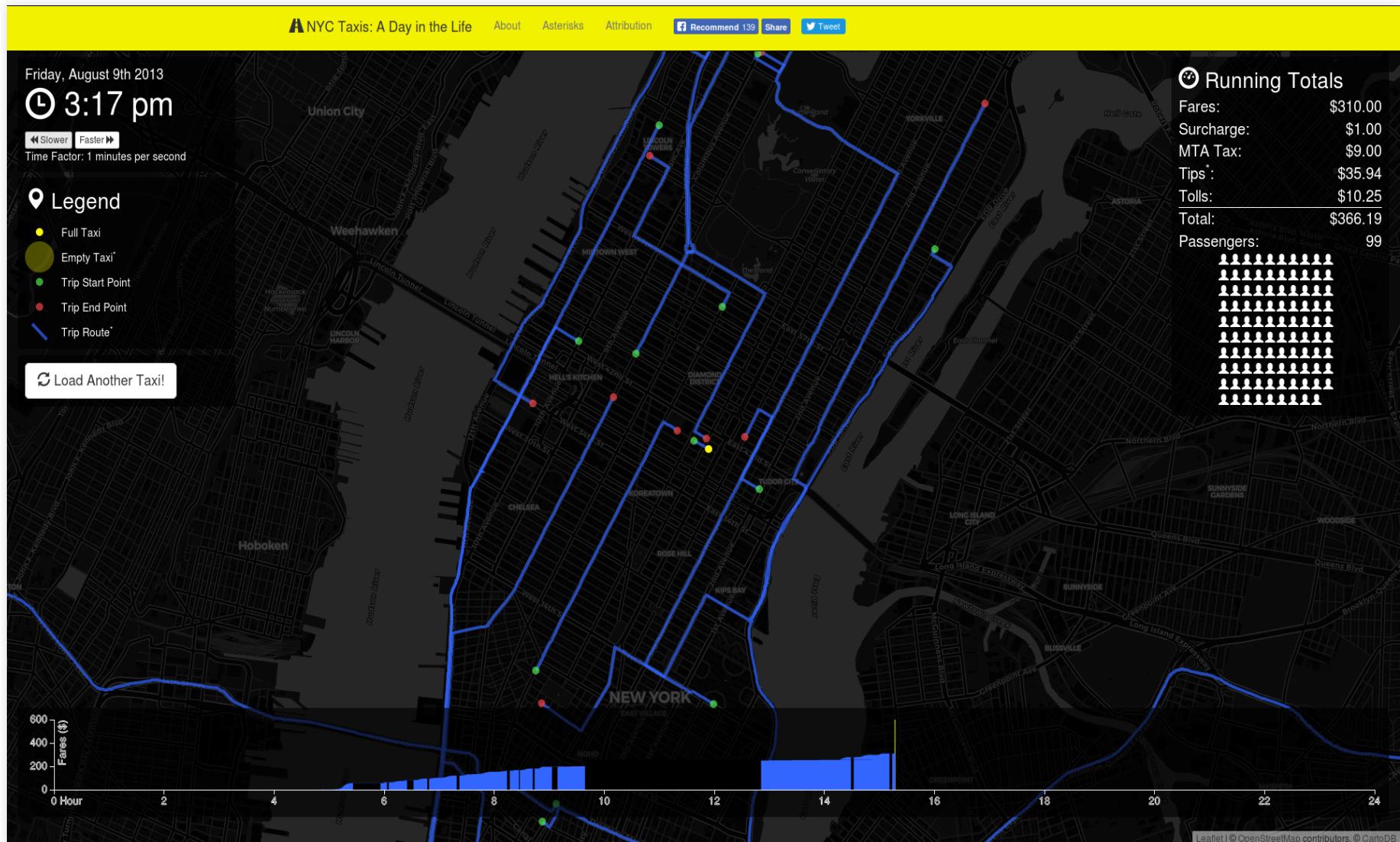
<http://www.openstreetmap.org>

Beispiel "Wheelmap.org"



<http://www.wheelmap.org>

Beispiel "NYC Taxis"



<http://chriswhong.github.io/nytaxi/>

Open Data: Initiativen

- Open Data Foundation: <http://odaf.org/>
- Open Knowledge Foundation: <http://oknf.org>
- Linguistic Linked Open Data: <http://linguistic-lod.org/>
- Journal of Open Humanities Data:
<https://openhumanitiesdata.metajnl.com/>

Open Data: Repositories

- Datenportal für Deutschland: <https://www.govdata.de/>
- Offene Daten Berlin, <http://daten.berlin.de/>
- Freebase, <http://www.firebaseio.com/>
- DBpedia, <http://dbpedia.org/About>
- Flickr Commons (Bilder), <https://www.flickr.com/commons>
- Musicbrainz (Musik-Metadaten), <https://musicbrainz.org/>
- Wikimedia Commons (Bilder uvm.), <http://commons.wikimedia.org>
- Open Culture: <http://www.openculture.com/>
- EU Open Data Portal: <http://open-data.europa.eu/en/data/>
- UNDP Human Development Data: <http://hdr.undp.org/en/data>

4. Software: Open Source

Open vs. Free Software

open source: Quellcode liegt in einer für den Menschen lesbaren Form vor darf beliebig kopiert, verbreitet und genutzt werden darf verändert und in der veränderten Form weitergegeben werden (Quelle: Art. "Open Source", Wikipedia, https://de.wikipedia.org/wiki/Open_Source)

free software: "When we call software free, we mean that it respects the users' essential freedoms: the freedom to run it, to study and change it, and to redistribute copies with or without changes. This is a matter of freedom, not price, so think of free speech, not free beer." (Richard Stallman)

Lizenzen

- Proprietäre Lizenzen (nicht open source)
 - Typisch: EULA = End-User License Agreement)
 - Recht, die Software zu nutzen (aber nicht Besitz der Software)
 - Strenge Bedingungen
- Offene Lizenzen
 - „permissive licenses“
 - der Code kann beliebig verwendet werden
 - auch unter einer restriktiven Lizenz weitergegeben werden
 - Beispiel: MIT license
- „copyleft“ Lizenzen
 - Code kann beliebig verwendet werden (nutzen, ändern)
 - muss unter den gleichen Bedingungen weitergegeben werden
 - Beispiel: GPL

Beispiele

Bereich	Proprietär	Open Source
Betriebssysteme	Windows	Linux
Webservices	Google	Framasoft
Statistik	SPSS/MATLAB	R/Python
GIS	ArcGIS	QGIS
Konferenzen	Conftool	Open Conference Systems
Zeitschriften	Manuscript Central	Open Journal Systems

Initiativen

- Free Software Foundation (FSF, seit 1985!):
<http://www.fsf.org/>
- Open Source Initiative (OSI):
<http://opensource.org/>

Code-Repositories

- GitHub, <https://github.com/>
- Sourceforge, <http://sourceforge.net/>
- Bitbucket (Atlassian): <https://bitbucket.org/>
- Google Code, <https://code.google.com/>

5. Lehre: Open Educational Resources (OER)

Definition: OER

Open educational resources (OER) are freely accessible, openly licensed text, media, and other digital assets that are useful for teaching, learning, and assessing as well as for research purposes. [...] The term OER describes publicly accessible materials and resources for any user to use, re-mix, improve and redistribute under some licenses.

Quelle: "Open Educational Resources", Wikipedia, https://en.wikipedia.org/wiki/Open_educational_resources

Was sind MOOCs

- „Massive Open Online Courses“
- Mehrere Komponenten der Wissensvermittlung
- Video-Vorträge mit Zwischenfragen
- Übungsaufgaben mit automatischer oder verteilter Korrektur
- Diskussionsforen (mit Moderatoren)
- Prüfungen (und Urkunden)

Aspekte

- Man erreicht sehr viele Studierende
- Extrem hohe „Abbrecherquote“ (90-98%)
- Überwiegend kostenlos, aber nur teils „offen“ lizenziert
- Geschäftsmodelle:
 - freemium (Kurs kostenlos, Urkunde gegen \$\$\$)
 - Universität bezahlt die Plattform (Sichtbarkeit gegen \$\$\$)

OER-Ressourcen

- Khan Academy: <https://www.khanacademy.org/>
- Directory of Open Educational Resources,
<http://doer.col.org/>
- DARIAH-DE „Schulungsmaterialien“
(Geisteswissenschaften):
<https://de.dariah.eu/schulungsmaterial>
- DARIAH Teach: <https://teach.dariah.eu/>
- CLARIN „Teaching and Learning Materials Collection“
(Linguistik): <https://fedora.clarin-d.uni-saarland.de/hub/>

Abschluss

Bilanz (Aspekte)

- Ökonomisch:
 - Wie finanzieren, wenn Nutzung kostenfrei ist
 - positive makroökonomische Effekte (Innovation, etc.)
- Politisch:
 - öffentlich finanziert, öffentlich zugänglich
 - Entwicklungs- und Bildungspolitik

Bilanz (Aspekte)

- Rechtlich:
 - Urheberrecht wahren, wenn alles frei publiziert wird
 - Ausgleich der Interessen von Anbietern und Nutzern
- Praktisch:
 - Filtern, wenn alles publiziert wird (Aufmerksamkeitsökonomie)

Bilanz (Aspekte)

- Mediengeschichtlich:
 - jede Medienrevolution bringt neue Umgangsweisen mit Wissen mit sich: konservativ oder progressiv
 - aus vergangenen Medienrevolutionen lernen
 - aus anderen Bereichen lernen (Stichwort DRM bei Musik)
- Wissenschaftlich:
 - Ideal der „freien Zirkulation des Wissens“
 - „Zwerge auf den Schultern von Riesen“
 - Kollaborative, internationale Forschung
 - Reproduzierbarkeit und Transparenz

Fragen?

Lektürehinweise

Referenztext

- <https://open-access.net/informationen-zu-open-access/>

Weitere Empfehlungen

- Eric Raymond, "The Cathedral and the Bazaar", 1997.
<https://archive.org/details/CathedralAndTheBazaar>
- Christian Heise: *Von Open Access zu Open Science. Zum Wandel digitaler Kulturen der wissenschaftlichen Kommunikation.* meson press, 2018. <http://meson.press/books/von-open-access-zu-open-science/>
- Herb, Ulrich. (2012). „Offenheit und wissenschaftliche Werke: Open Access, Open Review, Open Metrics, Open Science & Open Knowledge. In U. Herb (Hrsg.), *Open Initiatives: Offenheit in der digitalen Welt und Wissenschaft*. Saarbrücken: universaar, S. 11-44. <http://universaar.uni-saarland.de/monographien/volltexte/2012/87/> URN: urn:nbn:de:bsz:291-universaar-873
- Suber, Peter. 2012. *Open Access*. Cambridge, Mass: MIT Press. <https://mitpress.mit.edu/books/open-access>



Christof Schöch, 2019
<http://www.christof-schoech.de>

Lizenz (Text): Creative Commons Attribution 4.0