

Open Data und Open Access to Data Die ungleichen Geschwister

Dr. Jens Klump

Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum
Zentrum für Geoinformationstechnologie

Open Access Tage 2010, Göttingen
Session 3: Open Access und Open Content

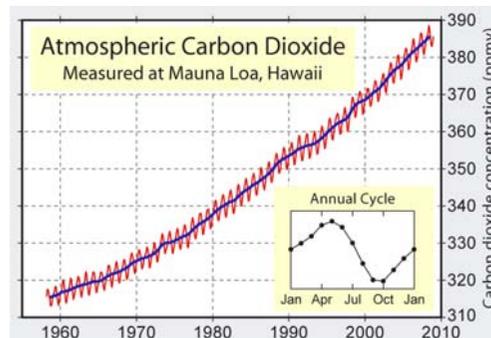
Einleitung

- Open Data wird oft mit Open Access in Verbindung gebracht.
- Worin unterscheiden sich Open Data und Open Access to Data?



Open Data

- Bisher gibt es keine allgemeine Definition von Open Data.
- Open Data Themen:
 - Amtliche Daten (z.B. Verteilung von Agrarsubventionen)
 - Daten staatlicher Stellen (z.B. Vermessungsämter)
 - Daten aus öffentlich geförderter Forschung
 - Daten aus Graswurzel-Initiativen (z.B. Open Street Map)
 - Open Linked Data
 - ...



Motivation für Open Data

- Transparenz
 - Demokratische Kontrolle
 - Partizipation
 - Freie Nachnutzung von Daten, die von den Behörden sowieso erhoben werden
 - ...
-
- Problem: EU-Direktive Public Sector Information (PSI) verlangt, dass Daten kostendeckend und mit Profit abgegeben werden.

Beispiel: Open Street Map

- Gegründet 2004 mit dem Ziel, eine freie Weltkarte zu erschaffen.
- Auslöser war die starre Haltung der staatlichen Vermessungsdienste bei der Lizenzierung von Karten.
- Kartenmaterial war bisher immer an enge Lizenzbedingungen gebunden, meist sehr teuer und oft veraltet.
- OSM war bisher unter CC-BY-SA lizenziert, soll jetzt auf Open Database License (ODbL) umgestellt werden.



Beispiel: Science Commons

*Many scientists today work in relative isolation, left to follow blind alleys and duplicate existing research. Data are fragmented — **trapped behind firewalls, locked up by contracts** or lost in **databases that can't be accessed or integrated**.*

John Wilbanks, Science Commons



SC empfiehlt Public Domain oder CC0-Lizenzierung von Forschungsdaten um unübersichtliche Verkettungen von Lizenzen zu vermeiden.

Widerstände

- **Transparenz:**
 - Politiker befürchten, dass sie durch die Daten einem höheren Rechtfertigungsdruck ausgesetzt werden.
- **Strukturen:**
 - Von den staatlichen Eigenbetrieben wird oft verlangt, dass sie mit den Daten Geld verdienen.
 - Die Behörden existieren auch ohne Open Data.
- **Infrastruktur:**
 - Ohne hoheitlichen Auftrag fehlen oft die Ressourcen.
- **Fazit:** Der Widerstand kommt von oben.

Open Data

- Anders als beim Open Access (to Data) gibt es bei Open Data keine allgemein gültige Definition.
- Der Übergang zu Open Access (to Data) ist fließend (z.B. Science Commons).
- Die meisten Open Data Projekte beziehen sich auf Daten staatlicher Dienste oder Behörden (Kataster, Subventionen, Wahlen, Finanzen, ...)
- Diese Daten werden meist als Teil des Arbeitsauftrags einer Behörde erstellt.
- Ziele sind eine größere Transparenz der Verwaltung und eine Nutzung der Daten durch die Öffentlichkeit.

Open Access to Data

*Establishing open access as a worthwhile procedure ideally requires the active commitment of each and every individual producer of scientific knowledge and holder of cultural heritage. Open access contributions include **original scientific research results, raw data and metadata**, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material.*



Auszug aus: *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003)*

Motivation für Open Access to Data

nature

www.nature.com/nature

Vol 461 | Issue no. 7261 | 10 September 2009

Data's shameful neglect

Research cannot flourish if data are not preserved and made accessible. All concerned must act accordingly.

More and more often these days, a research project's success is measured not just by the publications it produces, but also by the data it makes available to the wider community. Pioneering archives such as GenBank have demonstrated just how powerful such legacy data sets can be for generating new discoveries — especially when data are combined from many laboratories and analysed in ways that the original researchers could not have anticipated.

All but a handful of disciplines still lack the technical, institutional

also the software that will help investigators to do this. One important facet is metadata management software: tools that streamline the tedious process of annotating data with a description of what the bits mean, which instrument collected them, which algorithms have been used to process them and so on — information that is essential if other scientists are to reuse the data effectively.

Also necessary, especially in an era when data can be mixed and combined in unanticipated ways, is software that can keep track of

Weitere gute Gründe ...

The screenshot shows the homepage of stern.de on January 23, 2006. The main article is titled "Norweger hat Krebs-Studie frei erfunden" (Norwegian has invented cancer study). The article text reads: "Ein neuer Skandal erschüttert die Wissenschaft: Ein norwegischer Krebsforscher hat eine Studie frei erfunden und als Beleg für seine Theorien angeführt. 13 Koautoren fielen auf dem Trick." Below the article is a photo of Anders Ekblom, head of the investigation committee, with a caption: "Der Leiter der Untersuchungskommission, Anders Ekblom, schaut sorgenvoll: Jon Sudbø hat seine Studien gefälscht." To the right, a sidebar article titled "Schmerzforscher im Pharmasumpf" (Pain researcher in the pharma swamp) is visible, mentioning Scott Reuben and Edmund Neugebauer.

Die Reaktionen darauf ...

Deutsche
Forschungsgemeinschaft

**Sicherung guter
wissenschaftlicher
Praxis**

Safeguarding Good
Scientific Practice

Denkschrift



Eos, Vol. 90, No. 32, 11 August 2009

FORUM

Documenting Precision and Accuracy in the Open Data Policy Era

PAGE 276

How often do research papers discuss measurement uncertainty or present error bars in observational or statistical analysis figures?

the words "error," "uncertainty," "statistical," and "deviation."

Fifty-five percent of the articles did not mention measurement uncertainty or present error bars in data figures. Of the 47 articles that did mention these metrics,

therefo
ficult to
pender
racy. Ye
or prec
quantit
unacce
tive bel
develo
Stud
include
individ
Several
betwee
any ind
aged q
points 1
to 16

OECD Principles and
Guidelines for Access
to Research Data from
Public Funding



Schwerpunktinitiative “Digitale Information”

der Allianz der deutschen Wissenschaftsorganisationen

- Qualitätsgesicherte Forschungsdaten bilden einen Grundpfeiler wissenschaftlicher Erkenntnis und können unabhängig von ihrem ursprünglichen Erhebungszweck vielfach Grundlage weiterer Forschung sein.
- **Sicherung und Zugänglichkeit**
- **Unterschiede der wissenschaftlichen Disziplinen**
- **Wissenschaftliche Anerkennung**
- **Lehre und Qualifizierung**
- **Verwendung von Standards**
- **Entwicklung von Infrastrukturen**

Open Access to Data

- Open Access to Data bezieht sich auf Forschungsdaten aus öffentlich geförderter Forschung.
- Zielgruppe ist, in erster Linie, die wissenschaftliche Öffentlichkeit.
- Überwiegend „Top-down“ Initiative.

Widerstände

- **Transparenz:**
 - Wissenschaftler befürchten, dass jemand anderes mit ihren Daten eine bessere Idee entwickelt, als sie selbst
- **Struktur:**
 - Wissenschaftler haben wichtigeres zu tun, als sich um die Dokumentation der Daten zu kümmern, denn Daten tragen (noch) nicht zu ihrer Reputation bei.
- **Infrastruktur:**
 - Die Infrastrukturen für den Umgang mit Forschungsdaten sind noch unzureichend.
- **Fazit:** Der Widerstand kommt von der Basis

\$ diff open_data open_access_to_data

Open Data

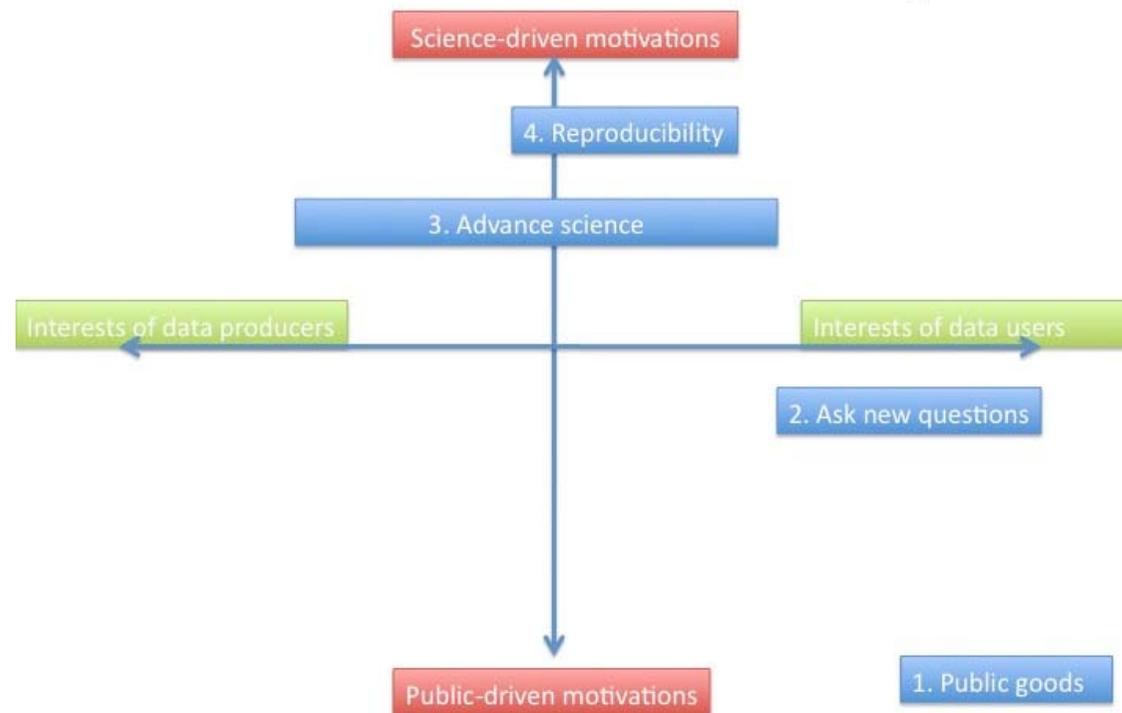
- Ziel: Daten von Datenproduzenten zugänglich machen
- Zielgruppe: Öffentlichkeit
- Initiative: Überwiegend „Grass Roots“

Open Access to Data

- Ziel: Daten als Basis wiss. Arbeit zugänglich machen
- Zielgruppe: wiss. Öffentlichkeit
- Initiative: Überwiegend „Top-Down“

Wen interessiert's?

Motivations and interests in sharing data



C. L. Borgfeld (2010)

Begriffsverwirrung?

- Der Begriff ***Open Data*** wird von den jeweiligen Initiatoren recht frei verwendet, eine einheitliche Definition ist nicht zu erwarten.
- Für den Zugang zu Forschungsdaten im Sinne des ***Open Access*** wird vorgeschlagen von ***Open Access to Research Data*** zu sprechen um diesen Aspekt klar abzugrenzen.

Zusammenfassung

- Open Data und Open Access to Data sind eng miteinander verwandt, aber dennoch zwei verschiedene Dinge.
- Die wichtigsten Unterschiede sind:
 - Angesprochene Datenbestände
 - Initiatoren
 - Motivation
 - Vorgehensweise bei der Umsetzung
- Im Detail kann es zwischen Open Data und Open Access to Data zu Zielkonflikten kommen.
- Bei der Umsetzung ist eine sorgfältige Analyse der jeweiligen disziplinären Praxis ihrer Kulturen notwendig.

Fragen?



Danke für Ihre Aufmerksamkeit!