Open Science Data

HTW Berlin
May 28, 2014

Michael Hörz – hoerz@michael-hoerz.de - @data_meining

Re: Questions

- Reasons for Companies not providing data
  - Fear of competitive disadvantages
  - Cost argument
  - Usually no legal obligation - yet in some cases
- Data provided by World Bank
  - Transparency of awarded deals etc.
- Examples for enriched corporate data
  - Marketing data
  - Statistical combinations

Science Data: Overview

- Open Scientific Data
- Open Access to Publications
- Open Science
- Related: Open Educational Resources

Open Scientific Data

Re-use of scientific data

- Basis: Berlin Declaration on Open Access by Max-Planck-Gesellschaft (2003): openaccess.mpg.de/286432/Berlin-Declaration
- Overview of Data Policies (German): forschungsdaten.org/index.php/Data_Policies
- Committee on Data for Science: codata.org
- ICSU World Data Centre System: icsu-wds.org

Open Scientific Data II

- National Institutes of Health (NIH) make data sharing a condition for grants above $ 500.000
- German National Library of Science and Technology in charge of Digital Object Identifiers
  tib-hannover.de/en/services/doi-service
  - Project Data Cite: datacite.org
- Helmholtz-Gemeinschaft - New Guideline:
  helmholtz.de/en/artikel/helmholtz-association-commits-to-policy-1983
- Science Commons as a part of CC
  creativecommons.org/science

Open Science Data: Key Aspects

- Correct metadata essential for findability
- Data available on a long-term basis
- In particular Big Science as bioinformatics, geoscience, environmental sciences
  - data collected, analysed and interpreted collectively
- Possibility to verify results /do own tests

Open Scientific Data: Examples

- Registry of Research Data Repositories: re3data.org
- DataBib: databib.org
- British Academic Data: data.ac.uk (hub.data.ac.uk)
- GenBank: ncbi.nlm.nih.gov/genbank
- Protein Databank: rcsb.org/pdb/home/home.do
- Personal Genome Project: personalgenomes.org
- Data Dryad (general purpose): datadryad.org
- Pangea (Earth & Environment): pangaea.de/projects

Knowledge Economy

- (Often) publicly funded research results only accessible through expensive subscriptions
- Subscriptions prices too high for some libraries
- Forced to subscribe to bundles of publications
- Few publishers (i.e. Elsevier) control the market of papers written by scientists
- Big journals are established, editorial boards and peers contribute to their reputation
- Young scientists prefer them because of impact index
- Journals now partly boycotted by scientists

Open Access Journals

- Peer reviewed including board celebrities
- Scientists have to pay for publication costs
- No distribution revenues
- Number of journals still low
- In April 2012 Harvard requested its scientists to no longer publish in paid-for journals anymore
- EU plan: 60 per cent of results published in European countries as open access
- Public funding increasingly linked to OA publication

Open Access Journals

- Approach I: Golden Way (8% of literature)
  - Journals funded through research budgets
  - Less costs through saved subscriptions
  - Access for the entire public
  - Some quality and fraud aspects
- Approach II: Green Way (12% of literature)
  - Scientists organize online access to articles and review process
  - Self-archiving independently of publication

europa.eu/rapid/press-release_MEMO-12-565_en.htm

Open Access Examples

- Public Library of Science (PLOS), PLOS Medicine ($ 1500 per year and research team): plos.org
- Nature Conservation (conservation, biodiversity) pensoft.net/journals/natureconservation
- The Cost of Knowledge - Researchers in protest thecostofknowledge.com

Open Science

- Different emphases - openingscience.org/research
  - Technological aspects (Open Science Infrastructure)
  - Accessibility of knowledge creation
  - Alternative impact measurement (Altmetrics)
  - Access to knowledge (Open Access)
  - Access to data (Open Data)
  - Science Communication (blogs, podcasts etc.)
  - Collaborative research (osl.tib.eu/w/Hauptseite)

Example: offene-doktorarbeit.de

Background: book.openingscience.org

Open Educational Resources

Background

- Traditional Learning Management Software
- Self learning courses (no teacher)
- Digital teaching platforms, i.e. MOOCs

- Open and sustainable access to learning resources
- Remixability - free and open licences (CC-0)
- Strong linkage to Wikipedia

chronicle.com/article/What-You-Need-to-Know-About/133475

Open Educational Resources II

- Also applicable to higher education, i.e. MIT Open Courseware: ocw.mit.edu/about/our-history
- Startup model can mean droping the "open" chronicle.com/blogs/wiredcampus/flat-world-knowledge-to-drop-free-access-to-textbooks/40780
- Resources - OER Commons: oercommons.org

Further Info: Science Data

- Background on Open Access and Data: open-access.net/de_en/general_information/what_does_open_access_mean/open_access_to_data
- Overview on data repositories by Heinz Pampel: blogs.lse.ac.uk/impactofsocialsciences/2013/11/29/how-to-find-an-appropriate-research-data-repository
- Publishing data and databases: wiki.creativecommons.org/Data
- Exhaustive wiki of scientific data repositories: oad.simmons.edu/oadwiki/Data_repositories
- Linked Open Science Data: linkedscience.org/data
- Ulrich Herb (Specialist on Open Access/Research Data): scinoptica.com/pages/en/start.php?lang=EN