Research Data
- Basics and Results of ETD’s

Prof. Dr. Peter Schirmbacher
http://www.schirmbacher.de
schirmbacher@cms.hu-berlin.de
Research Data
- Basics and Results of ETD’s

Agenda:
1. Introduction of Humboldt-University and its aims in the field ETD’s
2. ETD-Collection at HU and in Germany
3. Research Data – what does it mean
4. Requirements for data sharing
11 Departments

38,700 students
> 14% foreign students
224 study courses
373 professorships
47 Junior professorships
36 Guest Professorships
2,500 staff
ETD in Germany

ETD’s at National Library in Germany

Source: German National Library
ETD in Germany

ETD's at National Library in Germany

Anteil der Online Ressourcen an den Dissertationen und Habilitationen insgesamt nach Erscheinungsjahr im Bestand der Deutschen Nationalbibliothek (Stand: 18.02.2012)

Source: German National Library
Visibility of Repositories

Research Cycle & Repositories

- Repository Infrastructure
- Pre-research documents
- Grey literature?
- Raw data
- Simulations
- Processed data
- Research documents
- Patent documents
- Pre-prints
- Published reports
- Theses
- Secondary publications
- Books, reviews, etc.
- Learning materials

Aus: e-SciDR-Lisbon Workshop Discussion Paper
1. **Part** - With which types of data do researchers in your discipline work?

→ all types of data in the D-MAVT (Mechanical and Process Engineering)

Source: Töwe / Scheid; ETH Zurich
The **Large Hadron Collider** (LHC) has delivered over twice as many collisions to the **ATLAS** and **CMS** experiments this year as it did in the whole of 2011.

Last year, ATLAS and CMS each recorded a total of around 5.6 inverse femtobarns of data. This measure of accelerator performance is equivalent to about 560 trillion proton-proton collisions. [Datasupply a year of 15 Petabytes (=15 Million Gigabytes)]
Definition of "Research Data"

Data can be classified as:

**Raw or primary data:** Information recorded as notes, images, video footage, paper surveys, computer files, etc., pertaining to a specific research project

**Processed data:** Analyses, descriptions, and conclusions prepared as reports or papers

**Published data:** Information distributed to people beyond those involved in data acquisition and administration

**Research Data**

**Definition**

The University of Melbourne:

- **Data are** facts, observations or experiences on which an argument, theory or test is based.
- **Data may be** numerical, descriptive or visual.
- **Data may be** raw or analysed, experimental or observational.
- **Data includes**: laboratory notebooks; field notebooks; primary research data (including research data in hardcopy or in computer readable form); questionnaires; audiotapes; videotapes; models; photographs; films; test responses.
- **Research collections** may include slides; artefacts; specimens; samples.
- **Provenance information** about the data might also be included: the how, when, where it was collected and with what. The software code used to generate, annotate or analyse the data may also be included.

Research Data

Research Cycle

Source: http://www.dcc.ac.uk/resources/curation-lifecycle-model
Research Data

Research Lifecycle

- Ideas
- Partners
- Proposal writing
- Research process
- Publication
- Share data
- Simulate, experiment, observe
- Analyse data
- Manage the data

Virtual Research Environments

Source: Dallmeier-Tiessen, Sünje; draft of Thesis

Prof. Peter Schirmbacher,
Berlin School of Library and Information Science, Humboldt-Universität zu Berlin

13.09.2012
Research Data

Data lifecycle

Source: http://www.ddialliance.org/
Graph 1: The Data Publication Pyramid, developed on the basis of the Jim Gray pyramid, to express the different manifestation forms that research data can have in the publication process. See Chapter 1 for a full explanation.

EU founded project: Opportunities of Data exchange (ODE)
Report on Integration of Data and Publications

Criteria to Research Data

- Availability
- Findability
- Interpretability
- Reusability
- Citability
- Curation
- Preservation

Source: ODE-ReportOnIntegrationOfDataAndPublications-1_1.pdf
Definition

Research Data Repositories

source: Andrew Treloar; http://oa.helmholtz.de/index.php?id=258
ETD’s and Research Data

**Conclusion**

**For doctoral candidates**

- Develop sustainable realistic data management plan
  - Create metadata and follow the standards
  - Use persistent identifier
  - Agree a convention for data citation and re-use

- Collaboration with institutional or public archives
  - Curate and preserve datasets
  - Help the data curator to understand the reuse-conditions
  - Help to develop knowledge base about data and their context

- Collaboration inside the community
  - Ensure common metadata and citation practices
  - Active engagement in the community for reuse of data

Source: modify the ODE-project paper
For libraries (or institutional or public archives)

- Develop data management plan for the institution or the specific discipline
  - Define role of responsibility and control
  - Agree a convention for data citation and re-use
  - Develop a service portfolio
- Provide service around the data
  - Ensure the storage of the data
  - Offer of using persistent identifier
  - Install viewer and searching facilities
- Collaboration with authors and users
  - Curate and preserve datasets
  - Reduce the barrier to store, share and reuse data

Source: modify the ODE-project paper
Thank you for your attention!

http://www.alliancepermanentaccess.org/
http://cerneu.web.cern.ch/cerneu/eu_projects/fp7/#ODE
http://www.re3data.org/

schirmbacher@cms.hu-berlin.de

Prof. Dr. Peter Schirmbacher
Unter den Linden 6
10099 Berlin
030 209370010