Research Data Management at Humboldt-Universität zu Berlin
– Status Quo and Perspectives

Séminaire DRTD-SHS
« Les données de la recherche dans les humanités numériques »

Maxi Kindling
Berlin School of Library and Information Science
Humboldt-Universität zu Berlin
Agenda

• Who are we and what do we do?
• What are research data?
• Why research data management?
• Which services does HU Berlin offer?
• Research data management for PhD projects?
• Project re3data.org
Who are we and what do we do?
BSLIS Information Management Department

Topics:
Electronic Publishing
Open Access
Research Data
Information Management

Focus:
Information Infrastructure
(Facilities and Services)

Cooperation:
Computer and Media Service
and University Library
Working Group on Electronic Publishing
Research data management

https://www.cms.hu-berlin.de/de/dataman/

Maxi Kindling, MESHS Lille, 2 February 2015
What are research data?
Online survey 2013, response rate: 499 participants resp. ~ 24 % of target group from all disciplines

Participants with respect to university departments (n=499)
Where does your research data derive from? Please indicate your main sources.

- Observations: 134
- Experiments: 132
- Simulations: 66
- Images: 67
- Surveys and interviews: 152
- Statistics and reference data: 120
- Logfiles and usage data: 20
- Text documents: 62
- Other (please specify): 305
Please indicate data types more specifically.

- Images: 164
- Audio recordings: 55
- Texts: 101
- Multi-dimensional visualisations and models: 96
- Video recordings: 199
- Spreadsheets: 219
- Databases: 103
- Programmes and applications: 91
- Data specific for your field or instrument: 18
- Other (please specify): 18
Please indicate specific data types you work with.

- Statistic analysis: 50
- Topographic data: 11
- Satellite imagery: 16
- Measurement series: 21
- Text-corpora / annotations: 55
- GIS data: 4
- Remote sensing: 15
- Satellite imagery: 4
- Climate modelling: 23
- Measurement series: 6
- Other (please specify): 4

Maxi Kindling, MESHS Lille, 2 February 2015
“Living“ definition from LIS perspective:

“[…] research data is defined as digital data being a (descriptive) part or the result of a research process. This process covers all stages of research, ranging from research data generation, which may be in an experiment in the sciences, an empirical study in the social sciences or observations of cultural phenomena, to the publication of research results. Digital research data occur in different data types, levels of aggregation and data formats, informed by the research disciplines and their methods. With regards to the purpose of access for use and re-use of research data, digital research data are of no value without their metadata and proper documentation describing their context and the tools used to create, store, adapt, and analyze them.”

(Pampel et al., 2013 based on Kindling/Schirrmacher, 2013)
Short version:

“Research data means all data that was created in the course of research or results from it. Since the generation or collection of research data depends on a variety of research questions and methods, followed by processing, analysis and publishing and/or archiving, the data may involve a range of media, aggregation levels and formats. Sharing and reusing research data requires documentation about the original context and about tools that were used.”

(HU Berlin Research Data Policy, 2014)
Definition of research data

Pragmatic approach:

Research data is all digital data that results from your PhD project or you make use of. It can be measurement or experimental data, statistical data, text, audio, images and the like.

(Survey on PhD theses and research data, 2015)
Definition of research data: Aspects

Term has an undefined extension:

• Interdisciplinary, disciplinary, community-related, individual approaches

→ Information infrastructure services and facilities for many disciplines

• Significance of RD in scholarly communication

→ Common understanding of „publication“, but „RD“?
Why research data management?
Why RDM?

• Technology makes it possible.
  • Communication networks enabling distribution and networks
  • Tools for data processing and analysing
  • „Data-driven science“

• Research integrity and transparency requires research data sharing.

• Re-use and visibility are encouraged and accelerated by Open Access.
Why RDM?

• Technology makes it possible.
  • Communication networks enabling distribution and networks
  • Tools for data processing and analysing
  • „Data-driven science“

• Research integrity and transparency requires research data sharing.

• Re-use and visibility are encouraged and accelerated by Open Access.
Recommendation 7: Safeguarding and Storing of Primary Data

Primary data as the basis for publications shall be securely stored for ten years in a durable form in the institution of their origin.

“[…] deposit in a research data repository […]”

“A DMP […] outlining how research data will be handled during a research project […]”

presented in scientific publications as soon as possible;

(ii) other data, including associated metadata, as specified and within the deadlines laid down in the data management plan (see Annex I);

A DMP as a document outlining how research data will be handled during a research project, and after it is completed, is very important in all aspects for projects participating in the Horizon 2020 Open Research Data Pilot as well as almost any other research project.
7. Sharing of Data, Materials, and Software

Publication is conditional upon the agreement of the authors to make freely available any materials and information described in their publication that may be reasonably requested by others.

Data Availability

„PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception.“

Acceptable data-sharing methods:

Data deposition (strongly recommended). All data and related metadata underlying the findings reported in a submitted manuscript should be deposited in an appropriate public repository, unless already provided as part of the submitted article. Repositories may be either subject-specific (where these exist) and accept specific types of structured data, or generalist repositories that accept multiple data types, such as Dryad. Guidance on acceptable repositories is included below. The Data Availability Statement must specify that data are deposited publicly and list the name(s) of repositories along with digital object identifiers or accession numbers for the relevant datasets. In some cases authors may not be able to obtain DOIs or accession numbers until the manuscript is accepted; in these cases, the authors must provide these numbers at acceptance. In all other cases, these numbers must be provided at submission.

http://www.plosone.org/static/policies.action#sharing
“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.”

http://openaccess.mpg.de/Berlin-Declaration
HU researchers are committed to the secure storage, processing, documentation and preservation of their research data for the long term. Responsibility for ensuring these processes lies with HU staff in charge of research projects or programmes."
“All HU researchers are encouraged to process research data resulting from their research activities according to the conventions and standards of their respective scientific community. They should document the complete research lifecycle including tools and procedures that they used.”

1. HU researchers are committed to the secure storage, processing, documentation and preservation of their research data for the long term. Responsibility for ensuring these processes lies with HU staff in charge of research projects or programmes.

„HU researchers should take responsibility for deciding at what time and on what legal terms research data may be accessed. [...] recommends making research data as well as scholarly publications publicly available in a timely manner. [...]”
„Research data underlying scholarly publications should be archived for the long-term and/or published in an appropriate trustworthy data archive or repository. They count as scholarly research output at Humboldt-Universität zu Berlin.“
Which services does HU Berlin offer?
Humboldt-Universität zu Berlin has committed itself to establishing the means to enable this policy to be fulfilled.
What support or services do you wish to have at HU?

- general issues
- citing and publishing data
- technical issues
- legal issues
- specific issues
- data management plan
- Secured and backed-up storage
- no need for support or services
- Other (please specify)
RDM initiative at HU Berlin

• Since 2012: Cooperation between
  • Service Center for Research (SZF),
  • Computer and Media Service (CMS),
  • University Library (UB);
  • Led by Vice President for Research
  • and supported by BSLIS
  • 1 FTE at CMS
RDM initiative at HU Berlin

• 2013: Survey on RDM (Report available in DE, RD available in DE and EN)
• 2013-2014: 17 follow-up interviews (Report available in DE)
• 2014: Research Data Management Policy and guidelines (Available in DE and EN)
• 2014: Guidance for creating DMP (Following DMPOnline, translation in DE)
• 2015: Pilot services for PhD theses and RD (Survey results will be available)
Research data management for PhD projects?
Survey: PhD thesis and research data

Preliminary results (n=185)! (Online survey from 19-31 January 2015)

Do you plan to publish your thesis electronically on edoc? (abbr.; n=185)

- Yes: 52%
- No: 48%
Survey: PhD thesis and research data

Preliminary results! (Online survey from 19-31 January 2015)

Do you have an archiving concept? (abbr.; n=185)

- Yes: 19%
- No: 81%
Survey: PhD thesis and research data

Preliminary results (n=185)! (Online survey from 19-31 January 2015)

Do you want to publish your data? (abbr.; n=76)

- Yes, but don't know how. 62%
- Yes, and already use an archive. 7%
- No, due to legal reasons. 17%
- No, I don't want to. 14%
RDM initiative at HU Berlin

- **Ongoing: Institutional concept** (Internal report)
  - Support in legal questions
  - Training, workshops, tutorials
  - Data deposit / data publication
Who is currently responsible for storing, back up or archiving your research data?

- **Myself**: 434
- **My assistant**: 37
- **PhD student**: 22
- **Project or group manager**: 20
- **Special staff within my working group or department**: 50
- **CMS staff**: 39
- **Library staff**: 1
- **External service provider**: 11
- **Other (please specify)**: 28
„Where can we archive and publish our research data?“
re3data.org: Registry of Research Data Repositories
re3data.org: Registry of Research Data Repositories

The research data repository was reviewed by the re3data.org Team.

The research data repository is either certified or supports a repository standard.

The research data repository uses a persistent identifier system to make its provided data persistent, unique and citable.

The research data repository provides additional information on its service.

The research data repository provides a policy.

The terms of use and licenses of the data are provided by the research data repository.

The research data repository provides open/restricted/closed access to its data.
re3data.org: Registry of Research Data Repositories
Ongoing...

- Awareness and engagement of researchers
- Drivers and barriers of data sharing
- Acknowledgement and scientific credit
- Metadata description and data documentation
- Data curation and longterm preservation
- Building own repositories or use other services?
- ...
Practice what you preach...

Humboldt-Universität zu Berlin Research Data Management Survey Results

Simukovic, Elena; Kindling, Maxi; Schirmbacher, Peter

This spreadsheet represents anonymized summary results of the research data management survey at Humboldt-Universität zu Berlin.
Merci beaucoup pour votre attention!

www.cms.hu-berlin.de/dataman

References: http://bit.ly/1msuDfB

@maxi_ki