Doctoral theses’ research data

Doctoral theses’ research data
and metadata documentation

ETD 2013 Hong Kong
16th International Symposium on Electronic Theses and Dissertations
25.09.2013
Doctoral theses’ research data

Agenda

- Motivation
- Research data
- Context
- Examples
- Survey results
- Aspects
Doctoral theses’ research data

Motivation

• Research data curation and sharing in times of Open Science
• Supplementary material & enhanced publications
• Less examples
• HU Berlin concept to archive and publish research data
Doctoral theses’ research data

Survey on Research Data Management at HU Berlin

• 6 weeks in early 2013
• Target group was academic staff at HU Berlin (~2000 persons)
• Overall response rate of ~24 % (499)
• 117 PhD students from most disciplines (departments and institutes)
  – Chemistry (13)
  – Psychology (11)
  – Social sciences (9)
  – History (8)
  – Biology (7)
• Important: Participants were not asked to rely answer to a specific type of research data, e.g. only research data based on ETDs
a) Where does your research data derive from? Please indicate your main sources.

- Text documents
- Surveys and interviews
- Observations
- Simulations
- Statistics and reference data
- Experiments
- Images
- Other
- Logfiles and usage data
b) Please indicate data types more specifically.

![Data Types Bar Chart]

- Texts
- Spreadsheets
- Databases
- Programmes and applications
- Images
- Data specific for your field or instrument
- Audio recordings
- Multi-dimensional visualisations and models
- Video recordings
- Other
c) Please indicate specific data types you work with.

- Measurement series
- Statistic analysis
- Spectra
- Surveys
- Other
- Patient data
- GIS data
- Remote sensing
- Satellite imagery
- Topographic data
- Text-corpora / annotations
- Climate modelling

25.09.2013
“[… ] 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.” (Kindling & Schirmbacher, 2013)
Doctoral theses’ research data

Relevance and context: Political Strategies

COUNCIL OF THE EUROPEAN UNION

Brussels, 17 July 2013
11985/13
(OR. en)
PRESSE 319

Agreement on "HORIZON 2020": the EU’s research and innovation programme for the years 2014 to 2020

The Committee of Permanent Representatives today endorsed the agreement reached last June between the Presidency of the Council and the European Parliament representatives on the "Horizon 2020" programme for research and innovation for the years 2014 to 2020. The agreement paves the way for the formal adoption of the "Horizon 2020" legislative package by the European Parliament and the Council through a vote in the coming months.

Horizon 2020 will replace the EU’s 7th Research Framework Programme (FP7), which runs until the end of 2013. Compared with FP7, the new programme is expected to further eliminate fragmentation in the fields of scientific research and innovation.

Horizon 2020, which has a budget of around 70 billion euros, will underpin the objectives of the Europe 2020 strategy for growth and jobs, as well as the goal of strengthening the scientific and technological bases by contributing to achieving a European Research Area in which researchers, scientific knowledge and technology circulate freely.

Horizon 2020 focuses on three priorities, namely generating excellent science in order to strengthen the Union’s world-class scientific excellence and make the Union research and innovation system more competitive, fostering industrial leadership to speed up the development of technologies that will support businesses and innovation, including for small companies, and tackling societal challenges in order to respond to the priorities identified in the Europe 2020 strategy by supporting activities covering the entire chain from research to market.


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

http://oa.mpg.de/lang/de/berlin-prozess/berliner-erklarung/
Doctoral theses’ research data

Relevance and context: „Academic Fraud“


Doctoral theses’ research data

Relevance and context: Research Integrity

The European Code of Conduct for Research Integrity

[Image with code conduct information]

http://www.esf.org/index.php?eID=tx_nawsecuredl&u=0&file=fileadmin/be_user/CEO_Unit/FO_FORA/FOFORUM_ResearchIntegrity/Code_Conduct_ResearchIntegrity.pdf&t=1367499587&hash=ac6e154c2fed65fa0d65467ffaf0d9ef44d

Empfehlung 7

Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für zehn Jahre aufbewahrt werden.

Erläuterungen

Ein wissenschaftliches Ergebnis ist in aller Regel ein komplexes Produkt vieler einzelner Arbeitsschritte. In allen experimentellen Wissenschaften entstehen die Ergebnisse, über die in Veröffentlichungen berichtet wird, aus Einzelbeobachtungen, die sich zu Teilergebnissen summieren. Beobachtung und Experiment, auch numerische Rechnungen, sei es als eigenständige Arbeitsmethode, sei es zur Unterstützung der Auswertung und Analyse, produzieren zunächst „Daten“. Vergleichbares gilt in den empirisch arbeitenden Sozialwissenschaften.

7. Sharing of Materials, Methods, and Data

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 for the purpose of academic, non-commercial research.

Availability of data and materials. PLOS is committed to ensuring the availability of data and materials that underpin any articles published in PLOS journals. PLOS’s ideal is to make all data relevant to a given article and all readily replaceable materials immediately available without restrictions (while not compromising confidentiality in the context of human-subject research).

We appreciate, however, that this ideal is not yet the norm in all fields. We are therefore collaborating with a number of subject-specific initiatives in order to develop relevant policies. In the meantime, authors must comply with current best practice in their discipline for the sharing of data through databases: for example, deposition of microarray data in ArrayExpress or GEO; deposition of gene sequences in GenBank, EMBL, or DDBJ; and deposition of ecological data in Dryad. We encourage all authors to comply with available field-specific standards for the preparation and recording of data; for more information, see below (section 8, Reporting Guidelines for Specific Study Designs). Where no field-specific database exists, authors can deposit data in Dryad.

Failure to comply with this policy will be taken into account when publication decisions are made. PLOS journal editors encourage researchers to contact them if they encounter difficulties in obtaining data or materials from articles published in PLOS journals. PLOS reserves the right to post corrections on articles, to contact authors' institutions and funders, and in extreme cases to withdraw publication, if restrictions on access to data or materials come to light after publication of a PLOS journal article.

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

http://www.nature.com/scientificdata/for-authors/data-deposition-policies/
Doctoral theses’ research data

Relevance and context: Impact on universities

• Increasing impact on universities worldwide
• Research data management services and support are ahead in UK, USA and Australia (evaluation in HU seminar)
• Since some years also a „hot topic“ in German LIS domain on university level
• Some German universities already have research data archiving included in their policy
• HU Berlin did the first extensive survey on research data management at German universities
• Interviews showed that some departments at HU are planning to mandate research data archiving for ETDs
Doctoral theses’ research data

Research data features

- Research object
- Research result
- Scholarly communication (published data)
- **Proof**
- **Retracability/Verifiability**
- **Impact**
- ETDs: Review and evaluation (exam)
- Re-use
- Innovation
- Financial benefits
Doctoral theses’ research data

Research data metadata and documentation

General aspects

• Interpretation
• Reproducability
• Re-Use
• Less motivation to data documentation
• Metadata standards for most disciplines

ETDs

• Research Integrity (exam)
• Citability (PID)
• Findability
• Extensive data description (part of ETDs), but diverse
e) Who is currently responsible for storing, back up or archiving your research data?

- Myself
- Special staff
- CMS staff
- My assistant
- Other
- External service provider
- PhD student
- Project manager
- Library staff
g) Have you ever deposited your research data in a data archive or repository?

- No, I was not aware of such option.
- No, I do not intend doing so in the near future.
- No, but I intend doing so.
- Yes.
- No answer
Doctoral theses’ research data

HU Berlin Phd candidates storing data in...

A sample of journals and data repositories
- www.plosone.org/
- http://www.myexperiment.org/
- Dalton Transaction
- Organometallics
- SHARE European Social Survey
- Scifinder, web of knowledge
- DRYAD
- Dropbox (3)
- prl.aps.org
- Polylog - Zeitschrift für Interkulturelle Philosophie, Deutsche Zeitschrift für Philosophie
- gesis
- SESS
f) Would you be generally willing to deposit particular research data to an archive or to share it?

- Not likely.
- Less likely.
- I have to consider this option more carefully.
- More likely.
- Most likely.
h) What support or services do you wish to have at HU?

- Secured and backed-up storage for my research data.
- Advice & guidance on legal issues (e.g. access restrictions, sensible data, licensing).
- Advice & guidance on technical issues (e.g. metadata, standards, long-term archiving/preservation).
- Advice & guidance on citing and publishing (your own) data.
- Advice & guidance on specific issues (e.g. when submitting your research data to a journal along with a manuscript).
- Advice & guidance on general research data management issues.
- Support on compiling a data management plan if requested by a research funder.
- I have no need for support or services.
- Other

0 10 20 30 40 50 60 70 80
### Dissertation

<table>
<thead>
<tr>
<th>Autor(en):</th>
<th>Carolin Höfler</th>
</tr>
</thead>
<tbody>
<tr>
<td>Titel:</td>
<td>Form und Zeit – computerbasiertes Entwerfen in der Architektur</td>
</tr>
<tr>
<td>Gutachter:</td>
<td>Horst Bredekamp; Jörg Friedrich</td>
</tr>
<tr>
<td>Erscheinungsdatum:</td>
<td>19.09.2011</td>
</tr>
<tr>
<td>Volltext:</td>
<td><a href="">pdf</a></td>
</tr>
<tr>
<td></td>
<td><a href="">pdf</a></td>
</tr>
<tr>
<td>Fachgebiet(e):</td>
<td>Künste, Bildende Kunst allgemein</td>
</tr>
<tr>
<td>Schlagwörter (ger):</td>
<td>Computer, Architektur, Form, Zeit</td>
</tr>
<tr>
<td>Schlagwörter (eng):</td>
<td>Architecture, Computer, Form, Time</td>
</tr>
<tr>
<td>Einrichtung:</td>
<td>Humboldt-Universität zu Berlin, Philosophische Fakultät III</td>
</tr>
</tbody>
</table>

**Lizenz:**

**Zitationshinweis:** Höfler, Carolin: Form und Zeit – computerbasiertes Entwerfen in der Architektur, 19.09.2011, [urn:nbn:de:kobv:11-100194924](urn:nbn:de:kobv:11-100194924)

**Metadatenexport:**
- Endnote
- Bibtex

**Print on demand:**
- gedruckt bestellen bei [epubli](http://edoc.hu-berlin.de/docviews/abstract.php?lang=&id=38578)

**Diese Seite taggen:**
- [CC](http://edoc.hu-berlin.de/docviews/abstract.php?lang=&id=38578)

**Abstract (ger):**

[Full abstract](http://edoc.hu-berlin.de/docviews/abstract.php?lang=&id=38578)
Doctoral theses’ research data

SULB

3D magnetic resonance microscopy of dehydrated biological specimens

3D-Magnetresonanz-Mikroskopie dehydrierter biologischer Materialien

Mietchen, Daniel

Dokument 1.avi (3.644 KB)  Dokument 2.avi (4.068 KB)  Dokument 3.avi (3.444 KB)
Dokument 4.avi (7.759 KB)  Dokument 5.avi (1.829 KB)  Dokument 6.avi (3.047 KB)
Dokument 7.avi (1.45 KB)   Dokument 8.avi (8.202 KB)  Dokument 9.avi (8.159 KB)
Dokument 10.avi (8.199 KB) Dokument 11.pdf (8.975 KB)

SWD-Schlagwörter: NMR-Tomographie, Magnetresonanzmikroskopie, Fossil, Paläontologie, Magnetfeldeffekt, Zellteilung, Kryobiologie
Freie Schlagwörter (Englisch): magnetic resonance imaging, magnetic resonance microscopy, fossils, palaeontology, cell division, cryobiology

Institut: Fakultät 7 - Naturwissenschaftlich-Technische Fakultät II
Fakultät: Naturwissenschaften
DDC-Sachgruppe: Dissertation

http://scidok.sulb.uni-saarland.de/volltexte/2009/2416/
Doctoral theses’ research data

Quocosa (Saxony)

Henrik Philipp Bringmann

Experiments concerning the mechanism of cytokinesis in Caenorhabditis elegans embryos

Dokumente und Dateien

Volltextdokument (PDF) - 13.60 MByte - MD5 SHA512
bringmann_01 - 16.98 MByte - MD5 SHA512
bringmann_08 - 16.66 MByte - MD5 SHA512
bringmann_03 - 29.47 MByte - MD5 SHA512
bringmann_02 - 72.86 MByte - MD5 SHA512

Hinweis

Bitte nutzen Sie beim Zitieren immer folgende URL:

http://nbn-resolving.de/urn:nbn:de:usi14:11724872008822:56919
Biogenic barite as a proxy of paleoproductivity variations in the southern Peru-Chile current

Autor(en): Knump, Jens
Jahr: 1999
Bemerkung: Zugl.: Bremen, Univ., Diss., 1999
Doctoral theses’ research data

University Bremen & PANGAEA

http://nbn-resolving.de/urn:nbn:de:gbv:46-ep000102570
24. **Lamy, F; Hebbeln, D; Wefer, G (1999): (Table 1) Age model of sediment core GeoB3302-1**

   Related to: Klump, J (1999): Biogenic barite as a proxy of paleoproduction variations in the southern Peru-Chile Current. *Berichte aus dem Fachbereich Geowissenschaften der Universität Bremen*


   Lamy, F; Hebbeln, D; Wefer, G (1999): High-Resolution Marine Record of Climatic Change in Mid-latitude Chile during the Last 28,000 Years Based on Terrigenous Sediment Parameters. *Quaternary Research*

   Size: 30 data points
   
   doi:10.1594/PANGAEA.88195 - Score: 2% - Similar datasets

25. **Lamy, F; Hebbeln, D; Wefer, G (1999): (Table 1) Age model of sediment core GlK17748-2**

   Related to: Klump, J (1999): Biogenic barite as a proxy of paleoproduction variations in the southern Peru-Chile Current. *Berichte aus dem Fachbereich Geowissenschaften der Universität Bremen*


   Lamy, F; Hebbeln, D; Wefer, G (1999): High-Resolution Marine Record of Climatic Change in Mid-latitude Chile during the Last 28,000 Years Based on Terrigenous Sediment Parameters. *Quaternary Research*

   Size: 36 data points
   
   doi:10.1594/PANGAEA.88195 - Score: 2% - Similar datasets
Doctoral theses’ research data

ePIC & PANGAEA

PANGAEA®
Data Publisher for Earth & Environmental Science

<table>
<thead>
<tr>
<th>Data Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Abstract:</strong> In order to reveal the structure of the sparsely known deeper sublittoral hard bottom communities of glacial Kongsfjorden, the macroepibenthos from six depth zones (30-200 m) was analysed. A total of 180 still images derived from 8-h video recorded at the Kongsfjordneset remotely operated vehicle station were assessed quantitatively. Overall 27 mainly suspension-feeding species/taxa were observed. Of these, two-thirds have an arcto-boreal distribution, while the remainder are cosmopolitan. The overall mean epibenthos abundance was 33 ind./m<strong>2</strong> with maximum values at 150 m depth (97 9 ind./m<strong>2</strong>). The majority of the taxa inhabited the entire depth range. Encrusting red algae, an unidentified sponge and the sea anemone Uticaria eques, characterized the assemblage of the shallow zone. The sea anemones Hormathia spp. were important below 30 m, the Serpulid polychaete Protula tubularia was characteristic for the community below 50 m and the demosponge Haliclona sp. was a key taxon between 100 and 200 m depth. Cluster analysis and non-metrical multidimensional scaling based on abundance data showed differences between the assemblages along the bathymetric gradient, but only in the shallower depths in relation to the substratum surface incline. As surface and tidal current impacts attenuate with increasing depth, there is a gradual trend from robust key species towards more fragile ones (e.g. P. tubularia), in line with the 'Physical control hypothesis'.</td>
</tr>
<tr>
<td><strong>Project(s):</strong> Benthic-Pelagic Processes @ AWI (AWI_BPP)</td>
</tr>
<tr>
<td><strong>Coverage:</strong> Median Latitude: 78.979001 * Median Longitude: 11.487010 * South-bound Latitude: 78.977667 * West-bound Longitude: 11.476867 * North-bound Latitude: 78.980583 * East-bound Longitude: 11.498617</td>
</tr>
</tbody>
</table>
Doctoral theses’ research data

ePIC & PANGAEA

Zonierung und Diversität makroepibenthischer Hartbodengemeinschaften im Kongfjorden (Spitzbergen, Svalbard)

General Information:

Citation:

Cite this page as:
hdl:10013/epic.38268

Contact Email:
Juergen.Laudien@awi.de

Download:

PDF (Fulltext)
Orc02102.pdf
Doctoral theses’ research data

University of Bielefeld

Interaktionales Planen in aufgabenorientierten Gesprächen - computervermittelt und face-to-face

Schaller B (2012)
Bielefeld: Universität Bielefeld.
Download: Diss_Schaller_2012.pdf
URN: urn:nbn:de:hbz:361-24869507
Bielefeld Doctoral Thesis | German

Authors: Schaller, Birte
Department: Fakultät für Linguistik und Literaturwissenschaft

http://pub.uni-bielefeld.de/publication/2486950
Interaktionales Planen in aufgabenorientierten Gesprächen - computervermittelt und face-to-face

Schaller B (2012)
Bielefeld: Universität Bielefeld.

Download Diss_Schaller_2012.pdf
URN urn:nbn:de:hbz:361-24869507

Bielefeld Doctoral Thesis | German

Cite this


http://pub.uni-bielefeld.de/publication/2486950
Doctoral theses’ research data

Examples: Summary

• ETD full text including graphs, pictures, tables etc.
• ETD full text file + pdf attachments mostly including graphs, pictures and tables
• ETD full text + attachment (zip-files)
• ETD full text + linked research data, metadata on ETD
• ETD full text + linked research data in institutional repository, metadata on ETD and research data
• ETD full text + linked research data in (multi-)disciplinary repository (with PID and metadata/data documentation), metadata on ETD and research data
• ...

However, at least in Germany less examples showing ETDs interlinked with research data.
Doctoral theses’ research data

Possible university strategies

• Institutional research data repository for archiving
  – With possibility to publish research data or not
  – Mandating research data deposit or offering optional deposit service
  – Zipped data packages
  – Mostly possible, but not often used

• (Multi-)Disciplinary repositories recommended by university
  – But still for a lot of ‘small sciences’ there is no such infrastructure; very likely an institutional infrastructure is needed
  – Cross-linking workflow for enhanced publications
  – Not yet common

• Challenges
  – Disciplinary differences and heterogeneity of research data and research data infrastructures
Doctoral theses’ research data

re3data.org

Registry of Research Data Repositories
Doctoral theses’ research data

Research Data Alliance (RDA)

Research Data Sharing without barriers

Research Data Alliance Second Plenary 16-18 September 2013  View the event details and documentation!

Webcasting URL for the Plenary Sessions on Wednesday, September 18th
Starting at 9:00 AM ET | 15:00 CEST | 11:00 PM EST

Big Data Analytics IG

Status: Planning Action
The Big Data Analytics (BDA) Interest Groups seeks to develop community-based recommendations on feasible data analytics approaches to address scientific community needs of utilizing large quantities of data. BDA seeks to analyze different scientific domain applications and their potential use of various big data analytics techniques.

Agricultural Data Interoperability IG

Big Data Analytics IG

Brokering IG

Certification of Digital Repositories IG

Community Capability Model WG

Data Citation WG

Maxi Kindling
Institut für Bibliotheks- und Informationswissenschaft
Humboldt-Universität zu Berlin

18.09.2013
Doctoral theses’ research data

Digital Curation Center (DCC)

Disciplinary Metadata

While data curators, and increasingly researchers, know that good metadata is key for research data access and reuse, figuring out precisely what metadata to capture and how to capture it is a complex task. Fortunately, many academic disciplines have supported initiatives to formalise the metadata specifications the community deems to be required for data re-use. This page provides links to information about these disciplinary metadata standards, including profiles, tools to implement the standards, and use cases of data repositories currently implementing them.

For those disciplines that have not yet settled on a metadata standard, and for those repositories that work with data across disciplines, the General Research Data section links to information about broader metadata standards that have been adapted to suit the needs of research data.

Search by Discipline
Doctoral theses’ research data

Some aspects...

- Research data definition depending on disciplinary characteristics!
- Workflows considering universities’ settings
  - Mandating Policy
  - Documented process in Phd guidelines
  - Participating committees
  - Self-deposit? Delivery workflow
  - Open Access, Embargo, Access restrictions
Doctoral theses’ research data

Some aspects...

- Metadata and data documentation
  - Cross-linking between ETD and research data (minimum dc:relations, optimal: “semantic“ linking and exposure via e.g. OAI-ORE)
  - ETDs data set provided with PIDs
  - ETDs data description parts
  - DataCite Metadata: Core Element Set (Identifier, identifierType, Creator, creatorName)
  - Visibility via NDLTD Union Catalogue, ProQuest etc.
Doctoral theses’ research data

Some aspects...

• Repository infrastructure
  – Institutional or disciplinary/multidisciplinary?
  – Evaluation of repositories as support service from the library?
  – Long term preservation
  – Trust and sustainability
  – Restrictions of file formats, space, ...
  – Legal aspects
  – Upload/Ingest via data package file (OAIS SIP)?

• Training
  – Awareness
  – Workshops and training tutorials in Graduate Schools and Phd courses
Thank you very much for your attention!
(maxi.kindling@hu-berlin.de, twitter: maxi_ki)

Thanks to my colleagues from

Electronic Publishing Working Group at Humboldt-Universität zu Berlin:
Niels Fromm, Sabine Henneberger Peter Schirmbacher, Elena Simukovic, Paul Vierkant, Dennis Zielke

German repository community:
Götz Hatop, Ulrich Herb, Marten Hoogerwerf, Najko Jahn, Jens Klump, Angela Schäfer, Jochen Schirrwagen, Michaela Voigt, Jan Weiland, Karin Zwiesler
References