Research traditions and emerging expectations: PhD students and their research data management

Kerstin Helbig
Humboldt-Universität zu Berlin

ETD 2016 “Data and Dissertations”
Session *Uptake and awareness*
July 12, 2016

Photo: Creative Commons CC0

Unless otherwise stated this work is licensed under
Creative Commons Attribution 4.0 DE License

Research data management in Germany

1998 / 2013 German Research Foundation (DFG)
Safeguarding Good Scientific Practice

2009 German Research Foundation
Recommendations for Secure Storage and Availability of Digital Primary Research Data

2010 Alliance of German Science Organisations
Principles for the Handling of Research Data

2014 German Rectors' Conference (HRK)
Management of research data - a key strategic challenge for university management

2015 German Research Foundation
DFG Guidelines on the Handling of Research Data

DFG Guidelines on the Handling of Research Data

Research data is an essential foundation for scientific work. The diversity of this data reflects the wide range of different scientific disciplines, research interests and research methods. Research data might include measurement data, laboratory values, audiovisual information, texts, survey data, objects from collections, or samples that were created, developed or evaluated during scientific work. Methodical forms of testing such as questionnaires, software and simulations may also produce important results for scientific research and should therefore also be categorised as research data. The long-term archiving and accessibility of research data contributes to the traceability and quality of scientific work and enables researchers to carry on work begun by others. The Alliance of Science Organisations in Germany voiced its support for the long-term archiving of research data, open access to it and compliance with the conventions of individual disciplines in the "Principles for the Handling of Research Data", adopted in 2010[1]. The "Guidelines on the Handling of Research Data" put the framework stipulated by the Principles into a concrete form in the DFG's funding arrangements.

The following general guidelines apply for applicants submitting proposals to the DFG:

1. Project planning and submission of proposals
Applicants should consider during the planning stage whether and how much of the research data resulting from a project could be relevant for other research contexts and how this data can be made available to other researchers for reuse. Applicants should therefore detail in the proposal what research data will be generated or evaluated during a scientific research project. Concepts and considerations appropriate to the specific discipline for quality assurance and the handling and long-term archiving of research data should be taken as a basis. The relevant explanations must contain information about data types, discipline-specific standards (if applicable) and the choice of suitable repositories, if these are available for a given research area or particular data types. Details should also be provided on any third-party rights affected and preliminary planning for the data publication schedule.

2. Accessibility
Assuming that the publication of research data from a DFG-funded project does not conflict with the rights of third parties (in particular data protection or copyright), research data should be made available as soon as possible. Data should be made accessible at a stage of processing that allows it to be usefully reused by third parties (raw data or structured data). To make sure this is the case, it must be ensured that access to the data is still guaranteed when, through publication, the rights of use relating to research data are transferred to a third party, usually a publishing house.

3. Long-term archiving
In accordance with the rules of good scientific practice, research data should be archived in the researcher's own institution or an appropriate nationwide infrastructure for at least 10 years.

The DFG offers the following assistance for the implementation of the guidelines:

1. Support and advice
The reuse of research data is playing an increasingly important role in nearly all scientific disciplines. At the same time, there are considerable differences between disciplines with regard to the current state of discussion and the available infrastructures. To help with the planning of research projects, information and suggestions for the handling of research data have been compiled here:
http://www.dfg.de/en/research_funding/proposal_review_decision/applicants/submitting_proposal/research_data/index.html
A directory of research data repositories is available at: http://www.re3data.org/
A summary of other, quality-assured infrastructures is available from the information portal RISources: http://risources.dfg.de/

2. Costs for data preparation and for the use of existing infrastructure
Applicants may request funding for project-specific costs that arise in connection with a scientific project, for the preparation of research data for subsequent reuse and/or the transfer of research data to existing infrastructures as part of a proposal to the DFG. It is also possible to request funding for costs associated with the use of relevant infrastructures. Financial support is available for staff costs, project-related hardware and software, and usage fees.

[1] http://www.allianzinitiative.de/en/core-activities/research-data/principles.html

CC BY

Research Data Management Initiative at Humboldt-Universität zu Berlin

- **2012**: foundation of the initiative though Computer and Media Service, Research Service Centre, University Library and Vice President for Research
- **2013**: survey among researchers of HU regarding their research data handling and desired support[1]
- **2014**: 17 face-to-face semi-structured interviews[2]
- **2014**: adoption of a research data management policy[3] as well as supplementing guidelines[4]
- Since January **2015** training courses

CC BY

Key results of survey and interviews

„A central backup service and storage space for research data is required.“ [2, p. 30] (Ø 20-50 GB)

„A legal clarification of fundamental issues is [...] absolutely essential.“ [2, p. 30] (→ Advice)

Services are necessary to „integrate data with metadata, so that they are more discoverable and accessible” [2, p. 20] (→ Advice)

CC BY

Services of the initiative

- Advice on handling of research data
- Communicating with and transfer to internal and external services
- Support for data management planning
- Workshops on research data management
- Information events within teaching for PhDs and students
- Information material
- Legal advice regarding research data in cooperation with experts

CC BY

Principles of Humboldt-Universität zu Berlin

Policy[3]

HUMBOLDT-UNIVERSITÄT ZU BERLIN

Humboldt-Universität zu Berlin Research Data Management Policy

Preamble

Handling of research data responsibly is critical for confirming research results, for the advancement of science and for the dissemination of knowledge. This policy is addressed to all researchers at Humboldt-Universität zu Berlin both in their role as independent researchers, and as teachers and supervisors. Their obligation includes instructing students and doctoral candidates about handling of research data properly, as well as imparting subject-specific skills and standards.

What is research data?

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.

Policy

To further the Proposals for Safeguarding Good Scientific Practice of the German Research Foundation, and the Rules of the Humboldt-Universität zu Berlin for Safeguarding Good Scientific Practice and Investigating Allegations of Scholarly Misconduct (Satzung der

Guidelines[4]

HUMBOLDT-UNIVERSITÄT ZU BERLIN

Guidelines

A supplement to the Humboldt-Universität zu Berlin Research Data Management Policy

Preface

Humboldt-Universität zu Berlin acknowledges significance of research data as a key basis for scientific and scholarly knowledge. In light of the "Rules of the Humboldt-Universität zu Berlin for Safeguarding Good Scholarly Practice and Investigating Allegations of Scholarly Misconduct" (Satzung der Humboldt-Universität zu Berlin zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit Vorwürfen wissenschaftlichen Fehlverhaltens, in German) the Academic Senate adopted the "Humboldt-Universität zu Berlin Research Data Management Policy" (Grundsätze zum Umgang mit Forschungsdaten an der Humboldt-Universität zu Berlin, in German). Following guidelines complement this document and provide practical advice for implementing this policy.

There is a need for the University to provide central support to implement the policy. Under the aegis of the Vice President for Research, central services including the Computer and Media Service, the University Library and the Research Service Centre will cooperate on a strategic roadmap for the development of research data management services. These services will be integrated into existing service infrastructure and will be coordinated with academic units. International cooperation with other higher education institutions has guided the development of these. The Research data management initiative, which started in the summer of 2012, has coordinated this issue.[1]

CC BY

Humboldt-Universität zu Berlin
Research Data Management Policy[3]

Policy

To further the Proposals for Safeguarding Good Scientific Practice of the German Research Foundation, and the Rules of the Humboldt-Universität zu Berlin for Safeguarding Good Scientific Practice and Investigating Allegations of Scholarly Misconduct (Satzung der Humboldt-Universität zu Berlin zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit Vorwürfen wissenschaftlichen Fehlverhaltens, in German) the University adopts following policy:

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.
2. 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.
3. HU researchers should take responsibility for deciding at what time and on what legal terms research data may be accessed. The Open Access Declaration of the HU (Open-Access-Erklärung der Humboldt-Universität zu Berlin, in German) recommends making research data as well as scholarly publications publicly available in a timely manner. Protection of personal data, copyrights and the legitimate interests of third parties must also be ensured.
4. 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.

Humboldt-Universität zu Berlin has commited itself to establishing the means to enable this policy to be fulfilled.

- Storage/Back-up
- Processing according to discipline-specific standards
- Documentation of the complete research lifecycle including tools and procedures used
- Long-term preservation (at least 10 years) in an appropriate data archive/ repository
- Access to data in a timely manner, especially combined with publication – depending on data protection and copyright

CC BY

Technical tools and services at HU Berlin for research data management

edoc Publication Server

Backup

Persistent Identifier (DOI/handle/URN)

SAN

Database Service

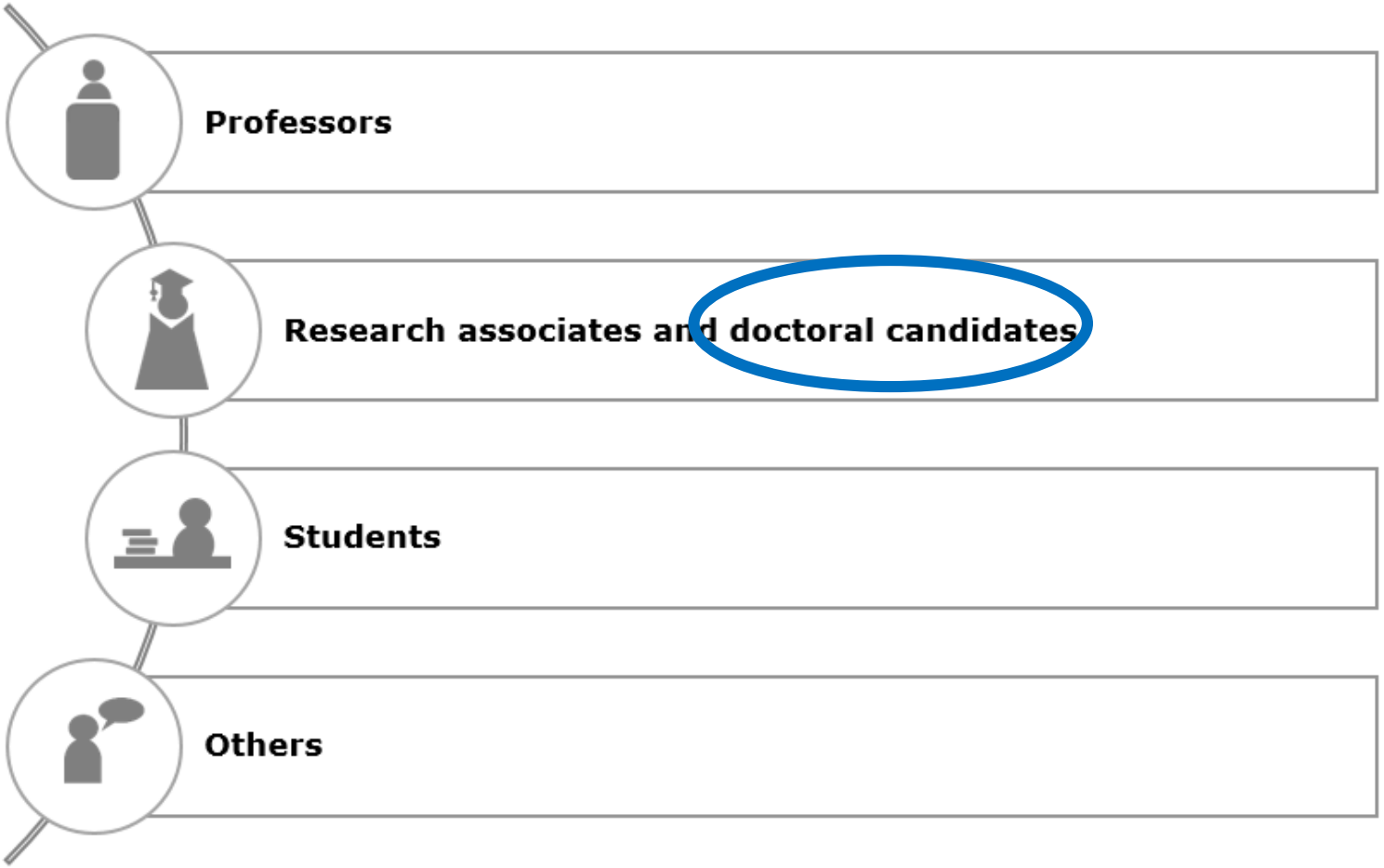
HU-Box (Seafile)

Media-Repository

Photo: Creative Commons CC0

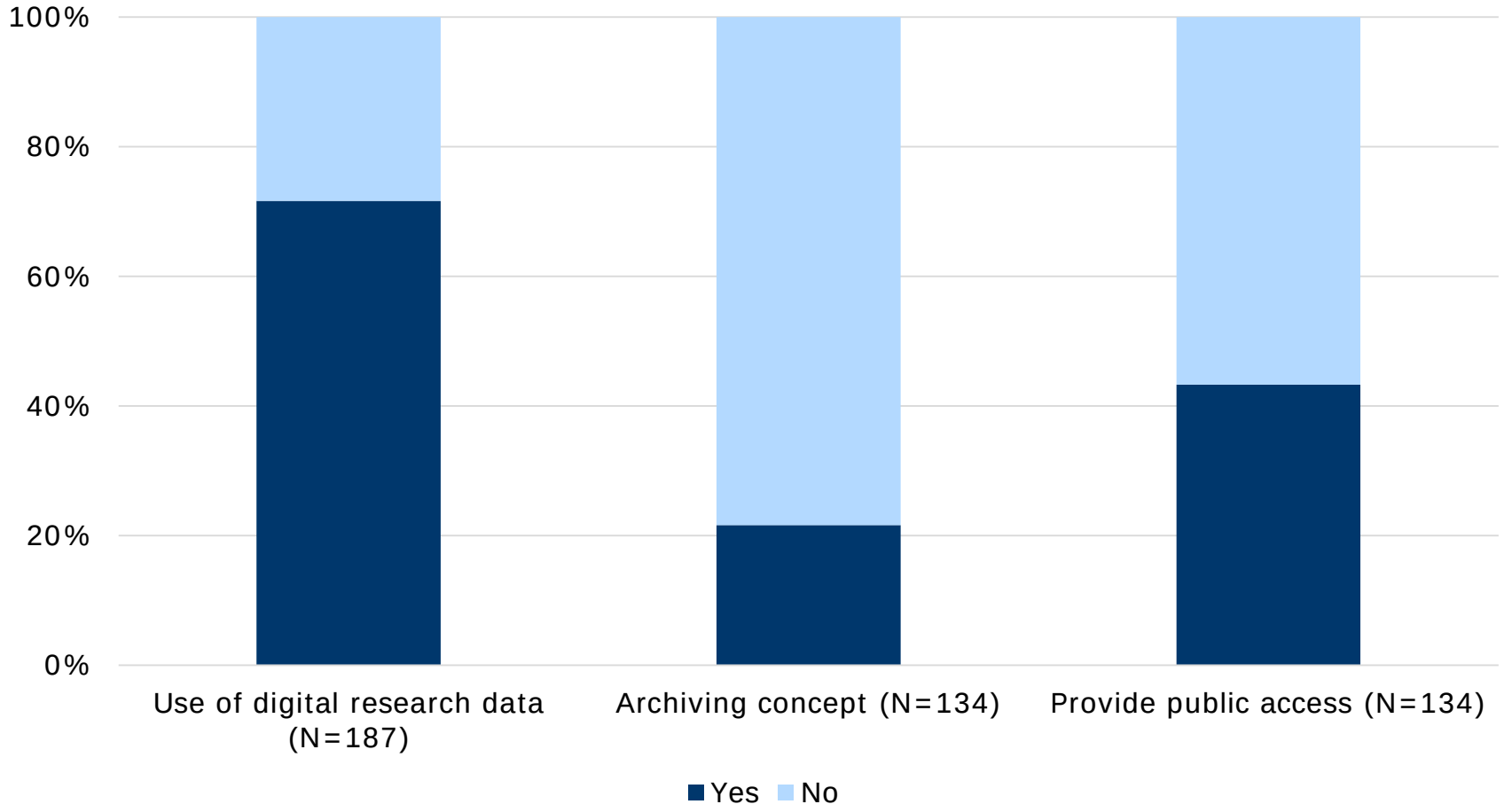
CC BY

Target groups



CC BY

Online survey among doctoral candidates



Source: Research data management of doctoral candidates at Humboldt-Universität zu Berlin; own calculation

CC BY

Training

- Discipline-specific workshops
  e.g. Workshop “Research Data Know-How for Geographers”[5]
- Topical workshops
  e.g. Workshop "Data management plans for EU, DFG and BMBF"
- Audience-specific workshops
  e.g. Workshop “Research data management for PhDs”
- Outreach and information sessions
  e.g. “Research data management for librarians”

We also offer workshops for small groups on demand

CC BY

Workshop „Research Data Management”

- Audience-specific workshop for PhDs
- Aspects of good scientific practice with regard to research data
- Introductory training (120 minutes)
- Group work
- Workshop is offered in English

Workshop Research Data Management for PhDs

The University Library and Computer and Media Service of Humboldt-Universität zu Berlin offer in cooperation with Humboldt Graduate School a workshop with the topic **"Research Data Management"** for PhDs.

The following **topics** will be addressed:

- What is research data?
- Why research data management? What are the requirements of funding organizations such as DFG, BMBF and EU?
- How does successful research data management look like? How do I write a data management plan? Which aspects are important?
- Research data management at Humboldt-Universität zu Berlin (research data policy, services)
- Discipline-specific metadata
- How does the actual work with repositories look like?
- Discussion, question-and-answer session

Photo: © Humboldt-Universität zu Berlin, Maxi Kindling

Previous knowledge is not required. The workshop will be held in English.

Target audience: Candidates in all years of PhD studies

Number of participants: max. 30

Date: Thursday, November 26th, 2015, 16.00 – 18.00 (s.t.)

Place: Humboldt Graduate School, Luisenstraße 56, 10115 Berlin

Registration via website of Humboldt Graduate School.

CC BY

Workshop agenda

- What is research data?
- Why research data management?
- How does good research data management look like?

In the groups:

- Brainstorming about own data management
- How to describe data and what needs to be kept in mind for storage/archiving?
- Which repositories/data archives are relevant and how to find and use them?

CC BY

Answers made by doctoral candidates during group session

Source: Word cloud of answers made by doctoral candidates during the group session; own calculation

CC BY

Specifics of PhD research data management

- Time pressure
- Funder requirements vs. discipline-specific expectations
- Publication pressure
- Missing credit
- Diverse knowledge base (depending on discipline)
- Unclear legal grounds
- Practical help and support
- Easy to use and straightforward tools

CC BY

Recommendations for other higher education institutions

- ✓ Cooperate with strategic partners
- ✓ Offer technical solutions
- ✓ Include research data into electronic thesis consultation
- ✓ Start training as early as possible
- ✓ Continuously offer advice and information material
- ✓ Try to respond to discipline-specific needs
- ✓ Raise awareness
- ✓ Offer consultation for conflict situations
- ✓ Advocate recognition of a transparent research process
- ✓ Create guidelines and standards

CC BY

New in 2016

- Workshop „Data Management Plans“ for PhDs
- Topical workshop, practice-oriented
- Structure of data management plans
- Varying funder requirements
- Practical work with DMPonline tool

Workshop:
Data management plans

The University Library and Computer and Media Service of Humboldt-Universität zu Berlin in cooperation with Humboldt Graduate School offer an application-oriented workshop with the topic **"Data management plans"** in summer semester 2016.

Research funders like German Research Foundation, Federal Ministry of Education and Research or European Commission increasingly expect a statement on the handling of research data. The workshop teaches skills for the independent creation of a data management plan. Participants will be introduced to the tool DMPonline that assists with plan creation. In addition, HU services that facilitate and support data management will be a topic. A first data management plan is created during the workshop.

The following topics will be addressed:

- What are data management plans?
- Why do I need a data management plan?
- How does a good data management plan look like? Which aspects should be considered?
- Practical work with the tool DMPonline (https://dmponline.dcc.ac.uk/)
- Discussion, Q&A session

The workshop language is English. Previous knowledge is not required.

Target audience: Doctoral candidates of all departments

Number of participants: max. 20
Date: Thursday, June 2nd, 2016 16.00 – 18.00 (s.t.)
Place: Jacob-und-Wilhelm-Grimm-Zentrum Geschwister-Scholl-Straße 3 Room 1'701

Registration via Humboldt Graduate School:
https://humboldt-graduate-school.de

CC BY

Further activities

- German video tutorials
- Cooperation projects with researchers
  (third-party funding)
- Project eDissPlus
  (see presentation and paper by Niels Fromm)

CC BY

Thank you very much for your interest!

Kerstin Helbig, research data management coordinator
kerstin.helbig@cms.hu-berlin.de

Further information about research data management as well as our workshops and training:
https://hu.berlin/dataman

Follow us on Twitter @DatawomenHUB

CC BY

Bibliography

[1] E. Simukovic, M. Kindling, P. Schirmbacher: *Umfrage zum Umgang mit digitalen Forschungsdaten an der Humboldt-Universität zu Berlin.* 2013. Retrieved from: http://nbn-resolving.de/urn:nbn:de:kobv:11-100213001.

[2] E. Simukovic, R. Thiele, A. Struck, M. Kindling, P. Schirmbacher: *Was sind Ihre Forschungsdaten? Interviews mit Wissenschaftlern der Humboldt-Universität zu Berlin* 2014. Retrieved from: http://nbn-resolving.de/urn:nbn:de:kobv:11-100224755.

CC BY

Bibliography

[3] Humboldt-Universität zu Berlin: *Humboldt-Universität zu Berlin Research Data Management Policy.* Berlin, 2014. Retrieved from: https://www.cms.hu-berlin.de/de/ueberblick/projekte/dataman/hu-rdm-policy/view.

[4] Humboldt-Universität zu Berlin: *Guidelines. A supplement to the Humboldt-Universität zu Berlin Research Data Management Policy.* 2014. Retrieved from: https://www.cms.hu-berlin.de/de/ueberblick/projekte/dataman/hu-rdm-guidelines/view.

CC BY

Bibliography

[5] K. Helbig: Research Data Management Training for Geographers: First Impressions. In: *ISPRS International Journal of Geo-Information* 5 (2016), S. 40. doi:10.3390/ijgi5040040.

CC BY