

Unveiling Research Data Stocks: A Case of Humboldt-Universität zu Berlin

Elena Simukovic¹, Maxi Kindling¹ and Peter Schirnbacher¹

¹ Humboldt-Universität zu Berlin

Abstract

Research data as an integral part of the scholarly record is increasingly attracting attention of all stakeholder groups. Higher education institutions, funding agencies, policy makers as well as the public at large see benefits in accelerating science through opening access to research data. More specifically, this aims at better re-usability and verification of research findings. The latter is particularly of great interest for higher education and other research institutions, as they embody scholarly scrutiny and trust. Once established as an university of a new type that should unify research and teaching, Humboldt-Universität has recently created a new job position to develop an institutional concept for research data management. In this paper we present the initial situation along with preliminary survey findings and draw the consequences for multidisciplinary higher education institutions by taking the example of Humboldt-Universität zu Berlin (HU).

Keywords: research data, research data management, survey, survey results, higher education institution, responsible conduct of research, research integrity

Citation: Simukovic, E., Kindling, M., & Schirnbacher, P. (2014). Unveiling Research Data Stocks: A Case of Humboldt-Universität zu Berlin. In *iConference 2014 Proceedings* (p. 742–748). doi:10.9776/14351

Copyright: Copyright is held by the authors.

Research Data: Simukovic, Elena et al (2013). Humboldt-Universität zu Berlin Research Data Management Survey Results. doi:10.5281/zenodo.7448 & Simukovic, Elena et al (2013). Humboldt-Universität zu Berlin Research Data Management Survey Results. Comparing respondent groups “Professor” and “Research associate”. doi:10.5281/zenodo.7449

Contact: elena.simukovic@cms.hu-berlin.de, maxi.kindling@ibi.hu-berlin.de, schirnbacher@cms.hu-berlin.de

1 Introduction

Since the Budapest Open Access Initiative in 2002 and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities in 2003, the research community strives strongly for unrestricted access to scientific outputs. In the light of green and gold open access ways the scholarly literature is being published more and more in an "open" mode of access. Some studies even state "that open access is reaching the tipping point" (European Commission, 2013). Access to research data is still in its infancy, though. After recent high-level data manipulation scandals universities are getting more conscious about the importance of research data, which is often fundamentally underlying presented results.¹ Moreover, some funding agencies are fostering the cultural change in a more progressive manner for example by asking for a 'data management plan' when applying for funds.

In Germany, the German Research Foundation (DFG) as a major funding agency plays a significant role in setting general framework for research practice. Its recommendations for responsible conduct of research including preservation of underlying data for at least ten years were passed already in 1998 (and have been updated recently). These recommendations have been incorporated as good scientific practice by a large majority of German universities. However, most of them do not actively promote research data management activities (RDM) in their respective communities. Being presumably the first German university HU surveyed current research data holdings and researchers needs for institutional support in issues pertaining to research data management to such extent. We explored different practices across scientific domains and academic career levels, the problematic notion of "research data" itself and an institutional role in supporting good research data management.

¹ See for instance Tilburg University (2012)

2 Methods

Following the idea of re-usable research results we decided to take advantage of other Higher Education Institutions (HEI) work in surveying research data management in their respective communities. Thanks especially to the expertise of the Digital Curation Centre² and numerous projects in the course of JISC Managing Research Data Programme³ we were able to adopt the questionnaires of the University of Glasgow, Imperial College London and University of Cambridge ("Data Asset Framework" and "Incremental" projects). Furthermore, surveys that have been done at Swiss Federal Institute of Technology (ETH Zürich) and in the PARSE.Insight project benefited our efforts. We then adapted it to local circumstances at Humboldt-Universität zu Berlin (HU) and evolved a final questionnaire consisting of 24 questions.

The survey targeted academic staff across all disciplines and departments at HU excluding service or administrative personnel. It was based on the assumption that exactly this target group does produce or process research data in its daily work. A special mailing list was then created and an open source survey application used to build an online questionnaire. After a short pre-test the survey has officially started and run for six weeks in January-March 2013.⁴

Simultaneously, a hands-on seminar for master students was run during the winter semester 2012-13 at the Berlin School for Library and Information Science. During this period we analyzed different approaches in organizing research data management support in HEIs in UK, US, Australia, Germany and Switzerland. Based on such international comparison, a list of recommendations for next steps at HU was produced for university's management board. A comprehensive report together with principal survey results was presented at Berlin Library Science Colloquium in the end of May 2013 (Kindling et al., 2013).

3 Key Findings

3.1 Response Rate

As compared with analogous surveys at universities in UK and Switzerland (herein before mentioned), a relatively high response rate of 499 responses in total, i.e. approximately 24% was achieved. This was even more encouraging as researchers from all faculty departments have participated allowing a detailed analysis and comparison of results to be made. Additionally, over 70 participants expressed their interest in an individual interview going beyond the questionnaire.

The different participation rate within respondent groups "Professor" (ca. 29 % of all professorships) and "Research associate" (ca. 13 % of all research associates at HU) was a little unexpected. Presumably, these numbers reflect the role of senior researchers being more accountable for RDM issues as well as some senior researchers responding on behalf of the whole working group. Our further analysis also showed a less marked familiarity with regulating policies and possibilities to publish or re-use research data among younger researchers, indicating a need to communicate related information more efficiently (see e.g. Figure 1).

² <http://www.dcc.ac.uk/>

³ <http://www.jisc.ac.uk/whatwedo/programmes/mrd/outputs.aspx>

⁴ For full survey report (currently only in German) see Simukovic et al. (2013a)

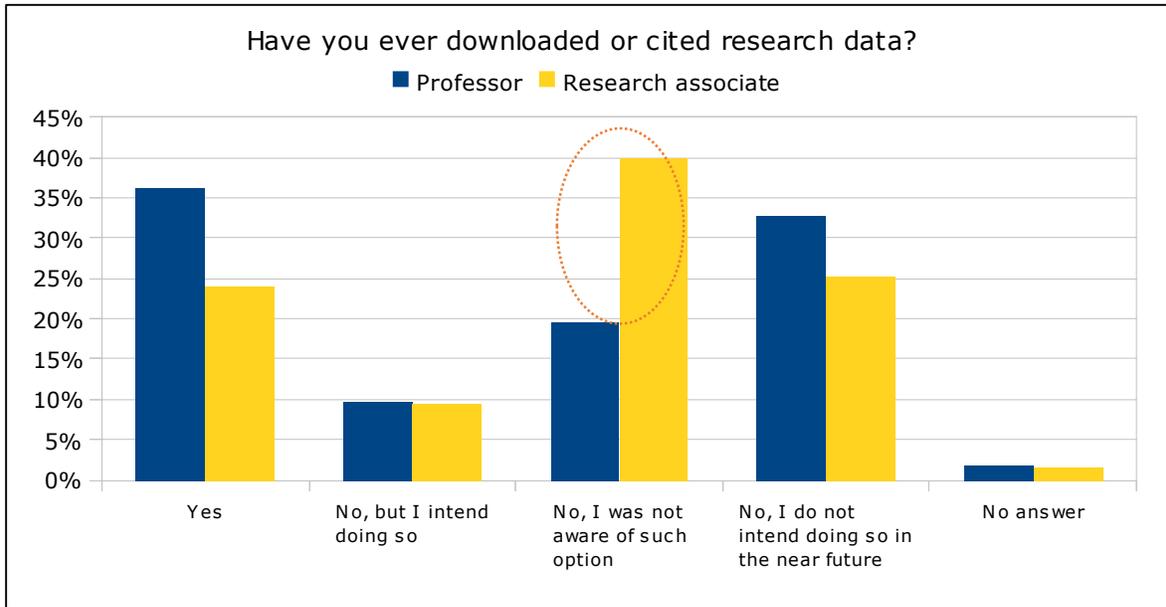


Figure 1: Have you ever downloaded or cited research data? (Comparison) (based on Simukovic et al. (2013c))

3.2 Current Research Data Practice

3.2.1 Data sources and types

The nature of research data among departments and research institutes at HU have proved to be very heterogeneous. As a common university-spanning source and type of research data, text documents were identified, together with wide distribution of databases, spreadsheets and images (see Figure 2 and Figure 3). More specifically, measurement series, statistic analysis, spectra, patient data and surveys were often in place.

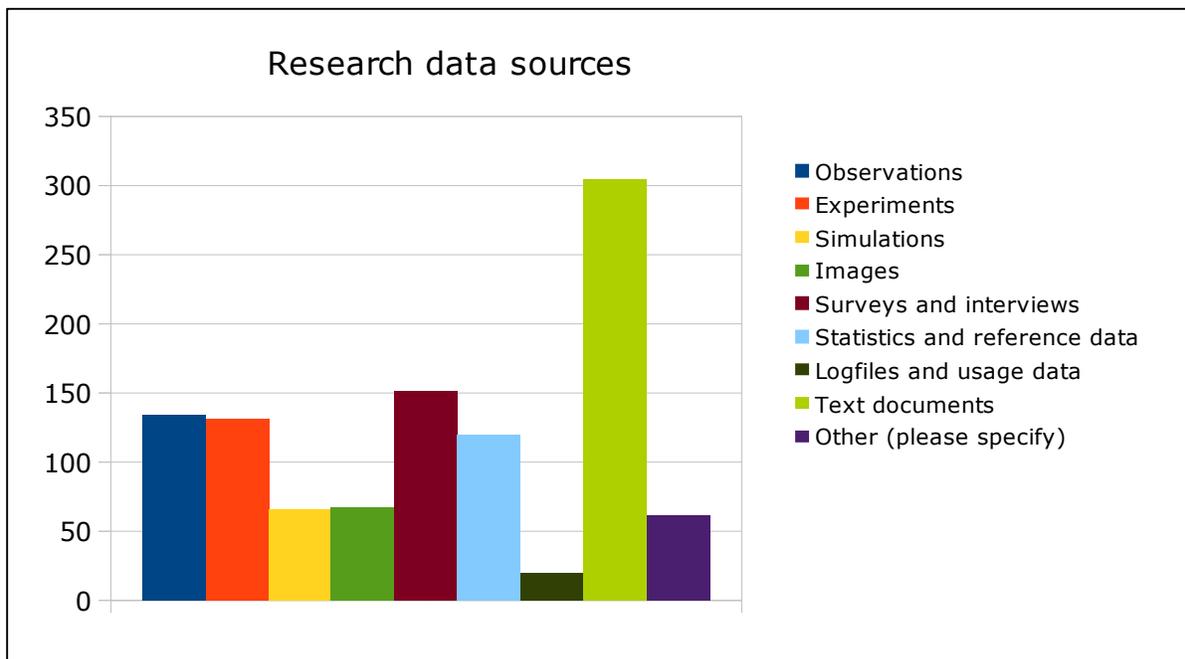


Figure 2: Research data sources (Simukovic et al. (2013b))

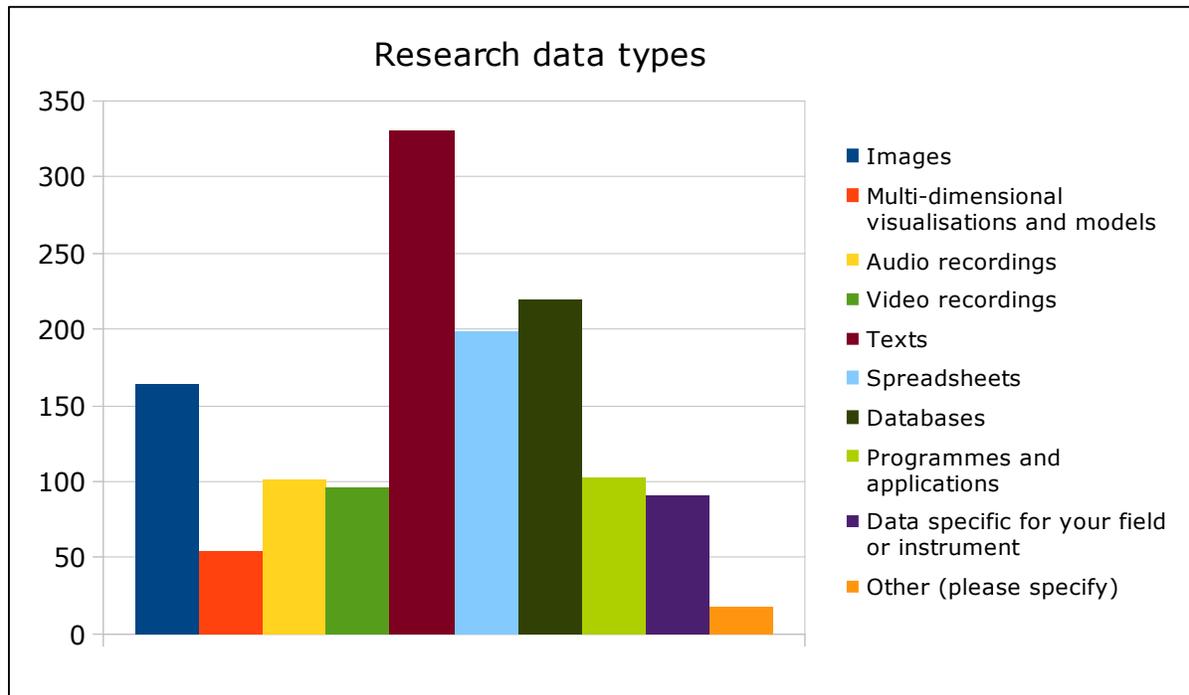


Figure 3: Research data types (Simukovic et al. (2013b))

3.2.2 Storage

Primary storage places as well as back up on further media indicated predominant use of local options: most respondents entrust their research data to hard disk drives of their PCs or laptops and to external hard drives or USB flash drives. Even so, back up was carried comparatively frequently, mostly on a daily or weekly basis. An interesting incidental finding resulted from free text comments, marking commercial cloud storage usage as very common when collaborating with other partner institutions. This highlighted the need for an alternative academic cloud service, as researchers increasingly cooperate and exchange/share materials across institutional borders and physical locations.

3.2.3 Good scientific practice

Apart from several further characteristics on how research data is currently being dealt with at HU, the most controversial matter was safeguarding "Good Scientific Practice". Respondents were presented with an excerpt of HU policy, saying they are committed to preserve primary data underlying scholarly publications for at least ten years. We then asked respondents for a statement, if they do take these rules into account and, optionally, to describe common practice in a comment field. Although already passed in 2002, 20 % of respondents stated they didn't know about these rules and further 17 % were not familiar with the current state of implementation in their working groups. Only half of all respondents (56 %) do preserve their data as required by the policy. The information provided in comment fields has shed light on problematic aspects of these rules. Among most widespread comments researchers considered this obligation as not suitable for arts and humanities and therefore not applicable in their research field. Some respondents have preserved their data without knowing of these rules followed by another group asking for IT support due to short-life media. More importantly, some respondents argued they were not able to guarantee for a ten years period due to prevalently short-term projects and job contracts. Sound arguments like these made clear on the one hand, that the university has to communicate its expectations for researchers more efficiently, and on the other hand to provide institutional support in order to comply with these rules.

3.3 Prospects for Institutional Services

3.3.1 Support and services needed

Regarding institutional support at HU, researchers were asked to indicate services they would like to have. Among the answer options offered, the most desired support was "Secured and backed-up storage for my research data" receiving 277 responses. This was followed by "Advice & guidance on legal issues (e.g. access restrictions, sensible data, licensing)" (256 responses) and "Advice & guidance on technical issues (e.g. metadata, standards, long-term preservation)" (237 responses) (see Figure 4).

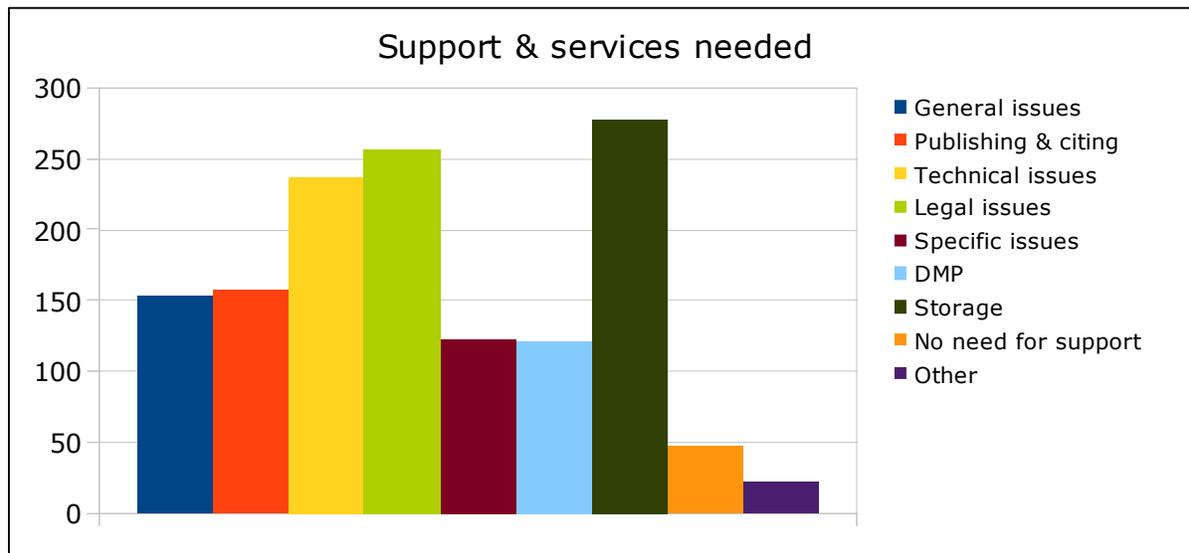


Figure 4: Support and services needed (Short version) (Simukovic et al. (2013b))

The results emphasized researchers need for pragmatic support in clarifying fundamental issues pertaining to RDM first. As funding agencies in Germany do not explicitly ask for a 'data management plan' when applying for funds we observed no strong demand in this area as compared to similar surveys in universities in UK. Accordingly, "Support on compiling a data management plan if requested by a research funder" received only 122 responses and would be ranked seventh if counted from top.

3.3.2 Supporting data sharing and archiving

Furthermore, respondents were asked which type of data archive or repository they would most likely choose to deposit or share their research data. A data archive within the researchers department appeared to be the most popular choice (216 responses), followed by an institutional data archive at HU (144 responses) and an international subject-specific data archive (142 responses). Interestingly, a national subject-specific data archive showed up as less favourable (68 responses). Although a general preference for institutional solutions was prevailing at a first glance, different use cases emerged. On the one side, researchers want to share their working materials including research data in a timely and flexible manner. As it is a matter of unpublished results, access to such materials should be restricted to cooperation partners and ideally administrated by their home institution. As no sufficient academic solution exists yet, respondents often stated to use commercial cloud services such as Dropbox. On the other side, researchers want to disseminate final results in their communities. As research is increasingly being conducted on a global scale, researchers strive to be read and cited internationally. This explains the preference for international subject-specific data archives when it comes to reaching their research community.

4 Conclusion

Summing up, we observed a general willingness of researchers to share research data with their respective communities or the public in the broader sense. However, there are major concerns that were raised in the survey. Amongst them are the protection of privacy, confidentiality or access restrictions that have to be eliminated before granting access to sensitive data. Considerable effort is also needed to prepare data for re-use e.g. in providing appropriate data documentation or metadata. At this point, individual and institutional roles and responsibilities have to be defined prior to setting mandatory policies.

At Humboldt-Universität we are currently developing an institutional concept to support researchers in good RDM. In the course of this a strong preference is given to the most required services as revealed by the three top positions according to survey results. Meanwhile we drafted a RDM policy in addition to the “rules of good scientific practice” at HU (to be passed by the Academic Senate in the near future) in order to reinforce commitment to a common strategy from all parties involved including university library, IT services and research support office.

When developing RDM services, HEIs need to find the right balance by taking different issues into account. It starts with keeping in mind the local setting of their own institution and the global nature of the research being performed at the same time. Establishing a shared view of all different stakeholders is another challenge. Moreover, as most HEIs are home to several scientific disciplines and research fields, different ways of working have to be taken into account. We have learned a lot about research practice at HU and gained invaluable insights into researchers daily work through direct contacts, most notably through follow-up interviews. We also observe increasing inquiry from other scholarly institutions in Germany interested in providing particular support or conducting a similar survey in their own environment. While learning by doing we hope to convince more scholarly institutions to provide pragmatic RDM support and to benefit from collective efforts.

5 References

- Deutsche Forschungsgemeinschaft. (1998). *Safeguarding Good Scientific Practice*. Retrieved from http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf
- European Commission (2013). *Open access to research publications reaching 'tipping point'* (Press Release IP/13/786). Retrieved from http://europa.eu/rapid/press-release_IP-13-786_en.htm
- Freiman, L., Jones, S., Molloy, L., Snow, K., Ward, C. (2010). *Incremental Project Scoping study and implementation plan*. Retrieved from http://www.lib.cam.ac.uk/preservation/incremental/documents/Incremental_Scoping_Report_170910.pdf
- HATII at the University of Glasgow & Digital Curation Centre (2009). *Data Asset Framework Implementation Guide*. Retrieved from www.data-audit.eu/docs/DAF_Implementation_Guide.pdf
- Kindling, M., Schirnbacher, P., Simukovic, E. (2013). "It's all about data." – Stand und Entwicklung im Umgang mit digitalen Forschungsdaten an der HU Berlin. Berliner Bibliothekswissenschaftliches Kolloquium [Berlin Library Science Colloquium], 21st of May 2013. Retrieved from https://www.cms.hu-berlin.de/ueberblick/projekte/dataman/bbk-pdf/at_download/file
- Kuipers, T., van der Hoeven, J. (2009). *PARSE.Insight: Insight into digital preservation of research output in Europe*. (Survey report, Deliverable D3.4.). Retrieved from http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf
- Simukovic, E., Kindling, M., Schirnbacher, P. (2013a). *Umfrage zum Umgang mit digitalen Forschungsdaten an der Humboldt-Universität zu Berlin. Umfragebericht, Version 1.0*. Retrieved from <http://nbn-resolving.de/urn:nbn:de:kobv:11-100213001>

- Simukovic, E., Kindling, M., Schirmbacher, P. (2013b). Humboldt-Universität zu Berlin Research Data Management Survey Results. ZENODO. Retrieved from <http://dx.doi.org/10.5281/zenodo.7448>
- Simukovic, E., Kindling, M., Schirmbacher, P. (2013c). Humboldt-Universität zu Berlin Research Data Management Survey Results. Comparing respondent groups "Professor" and "Research associate". ZENODO. Retrieved from <http://dx.doi.org/10.5281/zenodo.7449>
- Tilburg University. (2012). *The flawed science surrounding Diederik Stapel. Fraud as a catalyst in discussion on integrity.* (Press Release of 28th November 2012). Retrieved from http://uvtapp.uvt.nl/fsw/spits.npc.ShowPressReleaseCM?v_id=5199042271779297
- Töwe, M., Scheid, S. (2011). *User Expectations in Archived Research Data.* Paper presented at the Digital Preservation Summit 2011, Hamburg. Retrieved from <http://www.digitalpreservationsummit.de/presentations/toewe.pdf>

6 Table of Figures

Figure 1: Have you ever downloaded or cited research data? (Comparison) (based on Simukovic et al. (2013c))	744
Figure 2: Research data sources (Simukovic et al. (2013b))	744
Figure 3: Research data types (Simukovic et al. (2013b))	745
Figure 4: Support and services needed (Short version) (Simukovic et al. (2013b))	746