A Survey on Research Data at the Faculty of Arts and Humanities of the University of Cologne

Simone Kronenwett
simone.kronenwett@uni-koeln.de
Cologne Center for eHumanities, Germany

Brigitte Mathiak
mathiak@gmail.com
Cologne Center for eHumanities, Germany

Executive summary

In the course of the digitization of information, research data management has become one of the most important new areas of research. Universities have to prepare themselves to provide their academics and researchers with the necessary infrastructures and services. To identify the current demands regarding the handling of research data at the Faculty of Arts and Humanities of the University of Cologne, the Data Center for the Humanities conducted an online survey in 2016 in cooperation with the Office of the Dean of the Faculty of Arts and Humanities as well as the University and Library of Cologne. The enquiry aimed to characterize the present situation and to obtain information on the demands in the sectors research data management and consultation services. Our talk will show ongoing developments at the international and national level in research data management, present the results of the survey and discuss potential conclusions.

Relevance

One of the most important fields of action in research which is developing along with the digitalization of information is research data management (RDM). The universities face the challenge of offering their researchers adequate structures and services. The managing board of the German universities organized in the German Rector's Conference (HRK) has identified this as a key task (HRK 2014 and 2015). Moreover, in the recently published position paper called "Performance by diversity" the German Council for Scientific Information Infrastructures (RfII) makes a series of recommendations concerning how research data should be managed in the future (RfII 2016). The RfII was tasked by Germany's Joint Science Conference (GWK) with formulating broad-based recommendations for the science system in Germany as a whole. In addition, according to estimates by the German Research Foundation (DFG), up to 90% of the digital generated research data and results are still getting lost (Winkler-Nees 2011, p. 5) or "disappear in the drawer" (Kramer 2014) shortly after completion of research projects and are therefore not available for further use and reuse.

Method

As a structured way to gain information, we decided to follow the six stages process recommended for survey research (Müller et al. 2014). In addition, the survey is based on the relevant articles published in the handbook "Methods of Library and Information Science" by Umlauf, Fühles-Ubach & Seadle (Umlauf et al. 2013). The six steps are briefly explained below.

The Internet survey, the online questionnaire as well as the detailed report are published on the DCH-Website (http://dch.phil-fak.uni-koeln.de/umfrage-2016.html; see also Kronenwett 2017).

Research goals and constructs

The goal of the survey is to contribute to the conceptual development of the Data Center for the Humanities (DCH), which was founded as a central infrastructure service institution by the faculty dealing with humanities research data in 2013. In practice, the enquiry aims to characterize the present situation and to obtain information on the demands in the sectors RDM and consultation services offered by the DCH in cooperation with the University and City Library of Cologne (USB), one of the local partners of the DCH. Another goal was the comparability with other surveys conducted in the field.

Population and sampling

Because RDM should be handled in a way specific to each discipline (Sahle et al. 2013), the survey targeted only researchers at the Faculty of Arts and Humanities of the University of Cologne - one of the largest humanities faculties in Europe. The survey's population is limited to the academic staff of the Faculty of Arts and Humanities of the University of Cologne. In particular, the survey focused on researchers who are responsible for data-driven research projects.

Questionnaire design and biases

Firstly, with regards to content, conceptual and methodical design of the survey, Internet surveys and online questionnaires on research data at national and international scientific institutions and

research institutes were analyzed so far available (the website "forschungsdaten.org" offers an overview regarding national and international surveys on research data). Secondly, the questionnaire design was tailored and adapted to suit the unique circumstances which can be found at the Faculty of Arts and Humanities (DCH 2016, CCeH 2016). Finally, the results of a series of expert interviews carried out by the DCH with researchers at the Faculty were taken into account (Blumtritt 2016, p. 16). The questionnaire addresses five issues, namely (1) research data, (2) use of data archives, (3) support for research data, (4) discipline and position, (5) interest.

Review and survey pretesting

After a review of the questionnaire with stakeholders, such as the dean's office, the library and the data protection officer, a test link was sent to 20 potential subjects. This included representatives of all subject groups of the Faculty of Arts and Humanities (Faculty of Arts and Humanities 2016) as well as external experts (mostly sociologists and colleagues with RDM-background). After several feedback loops, the questionnaire was further modified and optimized.

Implementation and launch

The questionnaire was compiled using Kronenwett & Adolphs online survey tool (Kronenwett & Adolphs 2017). It was put online from 2016-05-30 to 2016-06-12 (2 weeks). Depending on individual answers the questionnaire contained up to 24 questions.

Data analysis and reporting

The questionnaire was completely answered by 136 participants (out of 191 persons who started the survey) which is 71.20% completion rate. The following selection of data analysis and reporting takes into account only these participants (n=136).

Results

Our objective in the compilation of the questionnaire was to answer the following questions:

1) What research data are available?
2) What is the need for research data?
3) What support do the members of the Faculty of Arts and Humanities want from the DCH?

Regarding the first question, sustainability and data volume were important to us. As far as sustainability is concerned, the majority of respondents are storing their research data on their local computers: 70% work computer, 70% private computer, multiple re-

sponses were possible (see fig. 1). Only 14% are storing their data in a data archive, a number that is also reflected in other questions like how many participants can imagine their data being stored in a data archive.
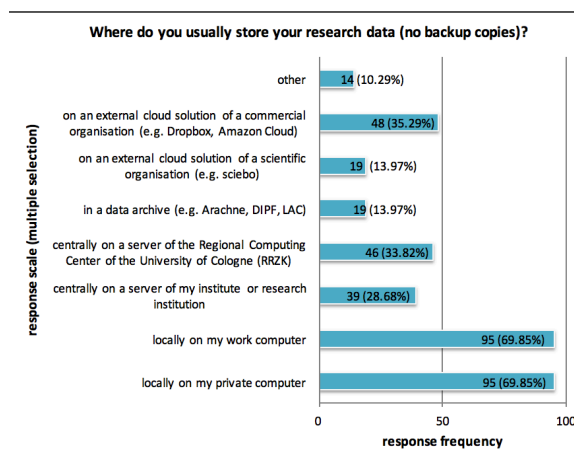

Figure 1. Storage places (n=136)

Regarding sustainability standards the given answers are fatal results since structured access and retrievability of research data are only ensured in professional data archives. Cloud solutions are also quite popular (35% use by commercial vendors and 14% of scientific vendors) because they ensure overall data access and data share. But regarding aspects like data plausibility, traceability or even long-term preservation they are totally unsuitable.

This result could be explained by the fact that the participants do not reflect their approach to sustainability and traceability. The vast majority of the respondent's self-assessment regarding their own skills in RDM is rated to be average or even less (71%) (see figure 2).
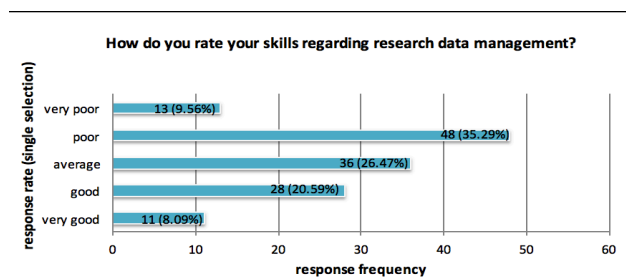

Figure 2. Self-assessment RDM-skills (n=136)

Sustainability is seen as a problem. 66% of the participants state that data could be lost when there is no one to be responsible for the website. 60% fear problems with data conversion. There is some sensibility towards the issues of finding the data (45%) and documentation of the data (41%). Figure 3 shows all answers concerning problems the participants see with preserving research data. Interesting is that both privacy and data theft are the concerns

2

voiced the least frequent (11% each), despite the fact that we encounter these concerns frequently in our consulting practice. But this could be an anomaly due to us asking from the user perspective rather than the data giving perspective which is more typical to our consulting.
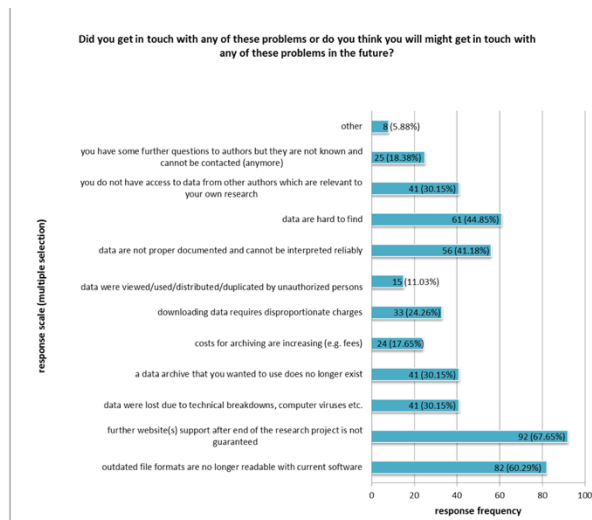

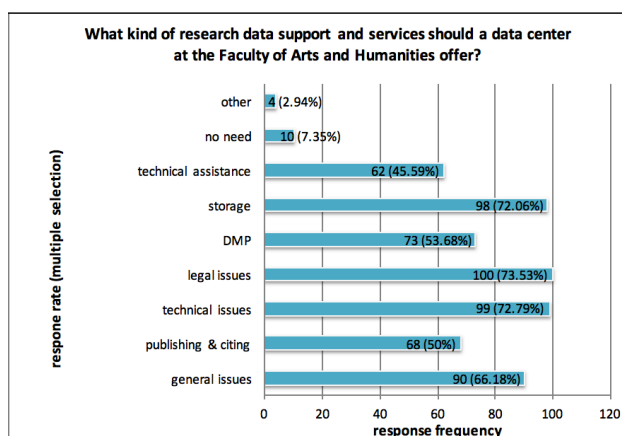Figure 3. Past and future problems (n=136)

Figure 4. Support and services needed (n=136)

In an effort to improve our services towards the faculty, we also asked which services should be provided by the DCH. Here legal and technical issues featured prominently (74% and 73% respectively; also cf. Fig. 4). Requests for storage (72%) and consultation (66%) on general issues are also in high demand.

Conclusion and outlook

As a result of the survey, we propose the following recommendations for action for the University of Cologne on the one hand and for the DCH on the other hand. Together with the local library (USB), the DCH now offers legal counseling on the subject of research data with a specialized lawyer. We are cur-

rently planning a project for improving the sustainability of living software systems, since the survey showed that this is an eminent problem in our faculty. Based on the projected storage space from the survey, we have negotiated with the computing center to provide that space centrally for all the members of our faculty.

In our talk, we will give more details on the study and its results and will also compare it to other surveys conducted internationally. We feel that surveys of this nature are an important tool to shape strategic decisions made in institutions concerned with research data.

	English	German
CCeH	Cologne Center for eHumanities	-
DCH	Data Center for the Humanities	Kölner Datenzentrum für die Geisteswissenschaften
DFG	German Research Foundation	Deutsche Forschungsgemeinschaft
DIPF	German Institute for International Educational Research	Deutsches Institut für Internationale Pädagogische Forschung
GWK	Joint Science Conference of Germany	Gemeinsame Wissenschaftskonferenz
HRK	German Rectors' Conference	Hochschulrektorenkonferenz
LAC	Language Archive Cologne	-
RfII	German Council for Scientific Information Infrastructures	Rat für Informationsinfrastrukturen
RRZK	Regional Computing Center of the University of Cologne	Regionales Rechenzentrum der Universität zu Köln

Figure 5. List of abbreviations used.

Bibliography

Kronenwett & Adolphs (n.d.), http://www.kronenwett-adolphs.com/en (accessed on 2017-03-26)

Averkamp, S., Gu, X., Rogers, B. (2014). Data management at the University of Iowa: A university libraries report on campus research data needs. University of Iowa, http://ir.uiowa.edu/lib_pubs/153/ (accessed on 2016-10-31)

Bauer, B., Ferus, A., Gorraiz, J., Gumpenberger, C., Gründhammer, V., Maly, N., Mühlegger, J. M., Preza, J. L., Sánchez Solís, B., Schmidt, N., Steineder, C. (2015). Researchers and their data. Results of an Austrian survey - report 2015. e-infrastructures austria, Vienna, http://e-infrastructures.at/das-projekt/deliverables (accessed on 2017-03-12)

Blumtritt, J. (2016). Consulting Workflow for Humanities Research Data, Talk at Humanities Research Data Conference (FORGE 2016): Beyond Data - Sustainability for Research Application and Software, 2016-09-15, University of Hamburg, https://www.gwiss.uni-hamburg.de/gwin/ueber-uns/forge2016/praesentationen/f1606-blumtritt-mathiak.pdf (accessed on 2017-03-12)

Cologne Center for eHumanities (Ed.) (2016). Digital Humanities. Structures - Teaching- Research, Cologne, http://cceh.uni-koeln.de/broschure-digital-humanities-2016/ (accessed on 2016-10-12)

Data Center for the Humanities (n.d.) http://dch.phil-fak.uni-koeln.de (accessed on 2017-03-26)

Data Center for the Humanities (2016). Research Data Survey 2016, http://dch.phil-fak.uni-koeln.de/umfrage-2016.html (accessed on 2017-03-26)

Faculty of Arts and Humanities (Ed.) (2016). Research 2016/17, Cologne

Faculty of Arts and Humanities, University of Cologne (n.d.) Subject Groups Overview, http://phil-fak.uni-koeln.de/9785.html (accessed on 2017-03-26)

Forschungsdaten.org, (n.d.) Overview of Surveys on Research Data at Scientific Institutions, http://www.forschungsdaten.org/index.php/Umfragen_zum_Umgang_mit_Forschungsdaten_an_wissenschaftlichen_Institutionen (accessed on 2017-03-26)

Hochschulrektorenkonferenz (2014). Management von Forschungsdaten als strategische Aufgabe der Hochschulleitungen. Empfehlung der 16. HRK-Mitgliederversammlung am 13. Mai 2014 in Frankfurt am Main, https://www.hrk.de/uploads/tx_szconvention/HRK_Empfehlung_Forschungsdaten_13052014_01.pdf (accessed on 2017-03-12)

Hochschulrektorenkonferenz (2015). Wie Hochschulleitungen die Entwicklung des Forschungsdatenmanagements steuern können. Orientierungspfade, Handlungsoptionen, Szenarien. Empfehlung der 19. Mitgliederversammlung der HRK am 10. November 2015 in Kiel, https://www.hrk.de/uploads/tx_szconvention/Empfehlung_Forschungsdatenmanagement__final_Stand_11.11.2015.pdf (accessed on 2016-08-12)

Kramer, B. (2014). Datenflut an Unis. Forscher müssen teilen lernen, in: Spiegel Online, 26.02.2014, http://www.spiegel.de/unispiegel/jobundberuf/umgang-mit-daten-der-glaeserne-forscher-a-954958.html (accessed on 2016-10-31)

Kronenwett, S. (2017). Forschungsdaten an der Philosophischen Fakultät der Universität zu Köln (Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft, Bd. 78) Köln

Kuipers, T., van der Hoeven, J. (2009). Insight into digital preservation of research output in Europe. Survey report. PARSE insight, European Commission, Brussels, http://parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf (accessed on 2017-03-12)

Müller, H., Sedley, A., Ferrall-Nunge, E. (2014). Survey research in HCI, in: J. S. Olson, W. A. Kellogg (Ed.): Ways of Knowing in HCI, New York, p. 229-266

Rat für Informationsinfrastrukturen (2016). Leistung aus Vielfalt [Performance by diversity]. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland, Göttingen, http://www.rfii.de/?wpdmdl=1998 (accessed on 2017-03-12)

Ray, J. M. (Ed.) (2014). Research Data Management: Practical Strategies for Information Professionals, West Lafayette/USA

Sahle, P., Kronenwett, S. (2013). Jenseits der Daten: Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner 'Data Center for the Humanities', in: LIBREAS. Library Ideas, 23, p. 76-96, http://edoc.hu-berlin.de/libreas/23/sahle-patrick-1/PDF/sahle.pdf (accessed on 2017-03-12)

Schöpfel, J. Prost, H. (2016). Research data management in social sciences and humanities: A survey at the University of Lille (France), in: LIBREAS. Library Ideas, 29, p. 98-112, http://nbn-resolving.de/urn:nbn:de:kobv:11-100238193 (accessed on 2017-03-12)

Winkler-Nees, S. (2011). Vorwort, in: S. Büttner, H.-C. Hobohm, L. Müller (Ed.): Handbuch Forschungsdatenmanagement, Bad Honnef, p. 5-6, http://www.forschungsdatenmanagement.de (accessed on 2016-08-30)