PAPER

Setting standards – how information centres can help to foster open science

Keywords: research data management, open data, quality criteria, usability, qualitative data

# Introduction

As research aims at being more open and transparent to raise quality standards and prove validity of results (Fecher and Friesike 2014), regional, European-wide and discipline-specific infrastructures are developed and funded to support the goals of open science. Moreover, researchers themselves change their practices and apply tools and services to manage and share their research (Kramer and Bosman 2016). With regard to sharing and archiving research data, professional research data centres like the research data education alliance[[1]](#footnote-1) (Bambey et al. 2018) aim at supporting scientists in their data management and offer search platforms to make data searchable, findable and accessible. Besides established data centres, there are numbers of online and free to use services that allow researchers to share their data. Online services have the advantage to offer a quick and direct service researchers can use immediately without any restriction.

Data centres are managed by data scientists that curate incoming data, including processes like quality and data privacy proof as well as metadata and license control. Thus, processes usually take a bit longer, but at the same time the data archived fulfils high-quality standards and researchers seeking reusable data can rely on proper resources. With regard to the different options researchers have for storing their data, our objectives in the following study were to compare the online repository Zenodo[[2]](#footnote-2) with the research data education centre and to apply quality criteria standards of this centre on open online data sets. We aimed at doing a qualitative study with a small data set sample (29) that was analysed manually and will give hints on the quality of online data sets. We focus on the field of educational research, a highly heterogeneous field with different disciplines involved, like e.g. pedagogics, psychology, sociology. Those fields have different methodological approaches and thus produce diver research data with quantitative and qualitative characteristics. Standards of sharing educational research data are developed by professional data centres, which as well curate research data and do quality checks. The question arises, whether shared online educational data has the same or similar qualities and if online repositories allow to set the same quality criteria at all.

The research questions are:

* What are relevant criteria for educational research data?
* Which quality criteria does educational research data shared online fulfil?

We want to contribute to a better understanding of researcher practices with regard to uploading open data and gives recommendations on how data centres could help facilitate data sharing and data re-using practises.

# Theoretical framework

Open science practices are not yet a default. Studies show that researchers have diverse understandings of the concept of openness (Levin and Leonelli 2017). Among other issues, they are concerned about data ethics and rights (Ünal et al. 2019). Moreover, researchers are influenced by their communities and external factors like research policies (Kim and Stanton 2016; Linek et al. 2017; Kim and Nah 2018). A concern often stated is that researchers lack explicit guidelines and best practices to efficiently and quickly manage their data for further reuse. Professional data curation centres and libraries support researchers with data management plans, guidelines and search platforms, new programmes are established to prepare institutes for their new role and satisfy researcher needs (Read et al. 2019; Koltay 2019; Si et al. 2019). The facilities provide support and are responsible for long-term research data curation. As such, data scientists developed quality standards applicable for any data to be professionally stored (Rat für Sozial- und Wirtschaftsdaten 2016). However, not any research data can be stored in those centres – for example data from other disciplines or small studies. Researchers practicing open science switch to other open repositories. The website re3data[[3]](#footnote-3) gives an overview of research data repositories and helps scientists find appropriate services according to their needs and discipline. They do not benefit from the centre’s support and quality check. However, the developed quality standards might be applicable for online data sets and researchers might be aware of good standards. The following study aimed at proving these assumptions and used a heuristic approach to develop and analyse quality criteria for educational research data sets.

# Methodology

Quality criteria for research data sets were developed based on standards by the research centre for educational research (guidelines and interview with staff), additional open science principles like those set by the FAIR initiative[[4]](#footnote-4) and as well own criteria that we consider relevant for proving research data appropriateness for a researcher.

We choose Zenodo[[5]](#footnote-5), an OpenAIRE project funded by the EU, which is currently one of the most well-known and interdisciplinary repository for storing any research-related data like publications, white papers, instruments and generated data.

# Results & Discussion

The study approach was in June 2019 and the search at Zenodo resulted in 49 uploaded research data sets found with the keyword “education”. We checked the data sets manually according to their topical scope and excluded 20 studies that did not fit the educational field, i.e. the studies did not investigate education as research topic, or that included too less information to judge the scope. This resulted in 29 data sets which we included in the analysis.

Open licences at Zenodo are default, we only had one restricted data set. The most striking aspect regarding information on data sets is its presentation. Research data centres apply structured and controlled vocabulary on their archived data sets. Information on methods and target group sample are shown in specific metadata fields. This information is collected by data scientists and in the best case managed together with the researchers before their study starts. Zenodo offers metadata fields, which researchers do fill out, but only a few fields allow specific information relevant to describe research data. We will discuss relevant fields in the full paper as well publish our research data.

# Conclusion

We analysed open online data sets from educational research and the appropriateness of quality criteria from a research data centre. Many criteria are highly relevant for finding good research data and reusing it. Most of the criteria a research centre applies are adaptable for open online repositories. Criteria like transcription rules properly set are hard to adapt as they need manually proof. At the data centre, data scientists, with advice from researchers, evaluate this quality criterion. We recommend data centres to make their standards transparent for researchers and to offer standard protocols to describe research data in open repositories. A first attempt is done with projects on domain research protocols[[6]](#footnote-6) that we will introduce in our paper (Science Europe 2018). Here, data centres and libraries might need to work on common efforts as research data management will face future challenges when researchers start fully adopting open science principles and funders require data sharing without exemption. As Timmermann says: “Research data management is a broad and complex field with many actors involved. It needs collective efforts by all actors to work towards aligned policies that foster Open Science” (Timmermann 2019).

# References

Bambey, Doris, Alexia Meyermann, Maike Porzelt, and Marc Rittberger. 2018. “Bereitstellung und Nachnutzung qualitativer Daten in der Bildungsforschung.” In *Bereitstellung und Nachnutzung qualitativer Daten in der Bildungsforschung*, edited by Doris Bambey, Louise Corti, Michael Diepenbroek, Wolfgang Dunkel, Heidemarie Hanekop, Betina Hollstein, Sabine Imeri et al., 59–68. https://www.ratswd.de/dl/RatSWD\_WP\_267.pdf.

Fecher, Benedikt, and Sascha Friesike. 2014. “Open Science: One Term, Five Schools of Thought.” In *Opening science*, edited by Sönke Bartling and Sascha Friesike, 17–47. New York: Springer.

Kim, Youngseek, and Seungahn Nah. 2018. “Internet researchers’ data sharing behaviors.” *Online Information Review* 42 (1): 124–42. doi:10.1108/OIR-10-2016-0313.

Kim, Youngseek, and Jeffrey M. Stanton. 2016. “Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis.” *Journal of the Association for Information Science and Technology* 67 (4): 776–99. doi:10.1002/asi.23424.

Koltay, Tibor. 2019. “Accepted and Emerging Roles of Academic Libraries in Supporting Research 2.0.” *The Journal of Academic Librarianship* 45 (2): 75–80. doi:10.1016/j.acalib.2019.01.001.

Kramer, Bianca, and Jeroen Bosman. 2016. “Innovations in scholarly communication - global survey on research tool usage.” *F1000Research* 5: 692.

Levin, Nadine, and Sabina Leonelli. 2017. “How Does One "Open" Science? Questions of Value in Biological Research.” *Science, Technology & Human Values* 42 (2): 280–305. doi:10.1177/0162243916672071.

Linek, Stephanie B., Benedikt Fecher, Sascha Friesike, and Marcel Hebing. 2017. “Data sharing as social dilemma: Influence of the researcher's personality.” *PLoS ONE* 12 (8): e0183216. doi:10.1371/journal.pone.0183216.

Rat für Sozial- und Wirtschaftsdaten. 2016. “Forschungsdatenmanagement in den Sozial-, Verhaltens- und Wirtschaftswissenschaften.” doi:10.17620/02671.7.

Read, Kevin B., Jessica Koos, Rebekah S. Miller, Cathryn F. Miller, Gesina A. Phillips, Laurel Scheinfeld, and Alisa Surkis. 2019. “A model for initiating research data management services at academic libraries.” *JOURNAL OF THE MEDICAL LIBRARY ASSOCIATION* 107 (3): 432–41. doi:10.5195/jmla.2019.545.

Science Europe. 2018. “Presenting a Framework for Discipline-specific Research Data Management.” https://www.scienceeurope.org/media/nsxdyvqn/se\_guidance\_document\_rdmps.pdf.

Si, Li, Yueliang Zeng, Sicheng Guo, and Xiaozhe Zhuang. 2019. “Investigation and analysis of research support services in academic libraries.” *ELECTRONIC LIBRARY* 37 (2): 281–301. doi:10.1108/EL-06-2018-0125.

Timmermann, Marie. 2019. “A collective challenge: Open Science from the perspective of Science Europe.” *Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare* 72 (2). doi:10.31263/voebm.v72i2.2831.

Ünal, Yurdagul, Gobinda Chowdhury, Serap Kurbanoglu, Joumana Boustany, and Geoff Walton. 2019. “Research data management and data sharing behaviour of university researchers.” *Information Research: An International Electronic Journal* 24 (1).

1. <https://www.forschungsdaten-bildung.de> [↑](#footnote-ref-1)
2. <https://zenodo.org/> [↑](#footnote-ref-2)
3. <https://www.re3data.org> [↑](#footnote-ref-3)
4. <https://www.go-fair.org/go-fair-initiative> [↑](#footnote-ref-4)
5. <https://zenodo.org> [↑](#footnote-ref-5)
6. <https://www.dipf.de/en/research/current-projects/domain-data-protocols-for-empirical-educational-research> [↑](#footnote-ref-6)