

datorium: Sharing Platform for Social Science Data

Deposit and Publish Research Data for Better Visibility

Andias Wira Alam, Stefan Müller, Natascha Schumann

GESIS – Leibniz Institute for the Social Sciences
Unter Sachsenhausen 6–8, 50667 Köln, Germany
{andias.wira-alam, stefan.mueller, natascha.schumann}@gesis.org

Abstract

A great number of research articles in scientific publications – particularly in, but not limited to, the social science disciplines – present and analyze results based on data collected from empirical research. However, publishing the research data at the same time requires effort. To name but a few, scientific publishers do not strictly request the supporting materials, e.g. research data, to be published together with the articles. Therefore, despite the importance of publishing the data, many researchers preserve and provide their data in various and less reliable ways. In this paper, we present datorium, a sharing platform for social science research data. Researchers can deposit and publish their data for better visibility, make the data as persistent, unique citable objects as well as enable reuse of the data.

Keywords: Open access, Research data, Sharing platform

In: F. Pehar/C. Schlögl/C. Wolff (Eds.). *Re:inventing Information Science in the Networked Society. Proceedings of the 14th International Symposium on Information Science (ISI 2015)*, Zadar, Croatia, 19th–21st May 2015. Glückstadt: Verlag Werner Hülsbusch, pp. 244–249.

1 Introduction

“Give credit where credit is due”. Scientific publications consider citations as credible sources and an important factor to increase the impact and visibility of scholarly work. Scientific collaboration among researchers comprises not only in writing articles but also collecting data which in return can be cited as well (Callaghan 2014). At the same time, the cost of collecting data can be greatly reduced by reusing existing data. These challenges and advantages relate to the question of how to support the visibility and accessibility of research data, even for small projects carried out by self-funded researchers (Linne 2013).

To answer this question, GESIS – Leibniz Institute for the Social Sciences developed and released a service for sharing social science research data called datorium¹ (Zenk-Möltgen 2014). In short, datorium is a free of charge online platform for easy upload of research data in a secure environment. Figure 1 depicts the process of depositing research data in multiple steps. Researchers can autonomously describe and upload their data online (step 1). It will be stored and reviewed by the GESIS Data Archive (step2) and published with the assignment of a digital object identifier (DOI) (step 3). After publishing the data researchers can search and download data on datorium (step4).



Figure 1. Depositing research data: From *upload* to *download* in datorium

¹ <https://datorium.gesis.org/>

2 Data publication and its challenges

Data publication describes the process of making research data available to a general public of researchers and involved stakeholders. It integrates certain steps like choosing an adequate platform, tool or vehicle for doing it as well as mechanisms and norms established by the scientific community like honouring the publication as a valuable publication in the first place. These steps are described elsewhere and will not be discussed here in an exhaustive manner (Knowledge Exchange 2014). However, defining data publication as a process and not as a singular event reveals the complexity and dimensionality of this phenomenon.

Problems and challenges, needs and advantages may arise on all levels and steps of data publication. It requires a scientific environment that acknowledges data publications as an adequate contribution to scholarship. This “cultural” trait values, among others, integrity, replication and attribution of research. As trivial it may sound it is not: While publication and sharing of research in the form of articles and other written papers that address theoretical and empirical problems is well established in all fields and disciplines of science, the same holds not true for data – ironically, the very foundation of all empirical discovery. This is especially true for smaller research projects in which resources for data management are sparse. Data is often “left alone” after the project is completed and its researchers then are applying and leaving for new projects. To give an exaggerated summary, there often seems to be no time, no funds and no incentives for data publication and since it stays uncommon there also seems to be no appreciation for doing it.

On the other hand, this is changing. More and more funders and scientific journals require that the data underlying the published research is deposited to a suitable data repository. In addition, after the implementation of identifiers like digital object identifiers (DOI) and their assignment to research data an increasing amount of researchers accept data as citable contributions to their scientific work.

To facilitate this development with the backing and experience of an established and reliable data archive GESIS – Leibniz Institute for the Social Sciences implemented datorium. Its mission is to provide an easy tool for researchers to publish their data in a straightforward way. Besides promoting

the abovementioned culture of publishing data it offers the also mentioned platform and vehicle for doing it.

3 System design and implementation

Apart from the technical implementation, the metadata schema is a key factor. We treat research data as publication; therefore the description of the data is very important (Friedrich 2014). However, it is rather not feasible to specify up to the deepest level of the data. The format of the data varies from type to type according to the research disciplines. As described in e.g. Castro et al. (2013), using domain-specific elements in order to fully describe scientific experiments is not avoidable. We therefore oblige the researchers to provide some mandatory information such as title and name of the researchers, and allow researchers to provide more, e.g. abstract or topic classification for better discovery. The metadata schema is based on Dublin Core and some extended elements taken from GESIS – Data Catalogue DBK². The detail of the metadata schema can be found in (Alam 2014).

In the implementation, datorium provides two languages in its interface, namely German and English, to encourage researchers worldwide to deposit their data. Data depositors are required to register in the system and provide their postal address once for validation through our reviewers. Moreover, depositors are able to determine the conditions of data reuse; for instance by applying availability of free or restricted access and appropriate licenses directing the use of the data. The system assigns unique persistent identifiers (DOIs) to the data and the data can be retrieved based on the metadata description, not only in datorium but also in other services, such as dalra³ or DataCite Metadata Store⁴. Furthermore, we also established, since it is one of the datorium's aims, the dissemination of the metadata in other open access repositories, e.g. OpenAIRE⁵, to gather more potential users of the data.

2 <https://dbk.gesis.org/>

3 <http://www.da-ra.de/>

4 <http://search.datacite.org/>

5 <https://www.openaire.eu/>

4 Conclusion and future work

We have shown a user-friendly platform for the autonomous documentation and publication of research data. This platform is targeted not only at social scientists, but also at e.g. information scientists to deposit and publish the data collected from experiments without major effort. It establishes better visibility of the data and makes them as persistent, unique citable objects as well as enables reuse.

As future work, we plan to establish the connection between scientific journals (partner publishers) and datorium. An ad-hoc service will be provided, where the authors of the accepted articles in academic journals are requested to deposit and publish the data along with the articles.

Acknowledgements

This project is funded by GESIS – Leibniz Institute for the Social Sciences. We thank our project members, Reiner Mauer, Wolfgang Zenk-Möltgen, Oliver Hopt, Monika Linne, Sigit Nugraha, and Martin Friedrichs, for their useful inputs and help during the implementation. We also thank the DSpace developers for their effort to continually develop DSpace as an open source project. datorium can be accessed online at <https://datorium.gesis.org/>.

References

- Alam, A. (2014). Dublin Core Metadata for Research Data – Lessons Learned in a Real-World Scenario with datorium. *International Conference on Dublin Core and Metadata Applications 2014*. <http://goo.gl/BVrNKQ>.
- Callaghan, S. (2014). Preserving the integrity of the scientific record: data citation and linking. *Learned Publishing*, 27, 15–24. <http://dx.doi.org/10.1087/20140504>.
- Castro, J. A., Ribeiro, C., Rocha da Silva, J. (2013): Designing an Application Profile Using Qualified Dublin Core: A Case Study with Fracture Mechanics Datasets. *International Conference on Dublin Core and Metadata Applications 2013*. <http://goo.gl/e4Q4H8>.
- Friedrich, T., Kempf, A. O. (2014). Making research data findable in digital libraries: A layered model for user-oriented indexing of survey data. *IEEE/ACM*

Joint Conference on Digital Libraries (DL) 2014. <http://dx.doi.org/10.1109/JCDL.2014.6970150>.

Knowledge Exchange (2014). Sowing the seed: Incentives and motivations for sharing research data, a researcher's perspective. <http://goo.gl/UkVvOA>.

Linne, M. (2013). Sustainable data preservation using datorium: facilitating the scientific ideal of data sharing in the social sciences. *iPRES 2013, Portugal*. <http://goo.gl/WvjH8E>.

Zenk-Möltgen, W. (2014). datorium: benefit from data sharing. *IASSIST 2014, Toronto*. <http://goo.gl/UV4Hru>.