Open Research Data: The current and future use of repositories by the Swiss research community

Markus von der Heyde[1]

[1]vdH-IT, Weimar, Germany, info@vdh-it.de

1. INTRODUCTION

Research is based on many different sources, including historical artefacts, simulations, empirical research data, concepts, and primary literature. However, every discipline of science[1] produces results and makes them accessible through publication. Often, background information is shared within the discipline for the sake of projects or collaborations.

One way to share the results of scholarly production is to upload publications, in conjunction with the research data underlying them, into an Open Data Repository. The Swiss National Science Foundation (SNSF) and swissuniversities are encouraging this and plan to support the research community with appropriate funding. In light of the SNSF's decision to make access to data from funded projects mandatory as of 2017, this study mandated by the SNSF examined the sharing and reuse behaviour of researchers in the Swiss community in 2018.

2. SURVEYS

The landscape survey across the complete Swiss research community collected information from 2,384 scientists about their data sharing practices and data reuse via an online questionnaire. The range of the questions was very broadly designed and questions from earlier international studies were used for comparability. Additionally, a second questionnaire addressed international repositories in order to learn about their perspectives and plans for future development. This repository survey added the perspective of 208 international repositories in terms of their genesis, provided services and use, cost and finance structure, and self-assessment of the degree of FAIR principle implementation (Wilkinson et al., 2016). The results were analyzed using statistical methods and can be regarded as representative.

The research was carried out with the intention of applying standard procedures as often as possible. However, crucial terms like "data" and "sharing" are known to have different meanings throughout the scientific community. The project therefore adopted with care a definition which is both close to those of the contracting authorities (SNSF, swissuniversities) and used in previous work in the area. We therefore defined data in this context using the NIH definition of 'Final Research Data' (National Institutes of Health (NIH), 2003). Also, the word "open" is very general and can be seen from many different perspectives. The SNSF expects that data generated by funded projects are publicly accessible in digital databases provided there are no legal, ethical, copyright or other issues[2]. The survey's use of "open" focused on aspects of free access and accessibility within the scientific community. The access often has to be managed by means of an authorisation infrastructure, which in itself is not the focus of this research. For all terms concerning Open Science, we refer to the Foster taxonomy of open science (Knoth & Pontika, 2015).

---

[1] We use the terms "science" and "research" as synonyms. Social sciences, humanities, life sciences, natural sciences, engineering and all potential other fields of research are considered to be equally relevant for this project.

[2] See http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx#.

3. RESULTS

Generally, the motivation and concerns for sharing data and reuse in the Swiss community are not different from other scientific communities. Differences in sharing and reuse behaviour are found according to the disciplines of the researchers, which were assessed using the bepress taxonomy (Warner, 2018). Different methods used by the researchers did not result in different sharing behaviour, but in where the data was shared. While the sharing is done equally in general repositories and smaller disciplinary repositories, of which a great number exist, the researchers prefer to use disciplinary repositories if they want to reuse data.

Overall, research is shared by ¾ of all scientists; however, only about a third of the Swiss research community share data in repositories. The main reason for not sharing was researchers' plans to publish their results first. Also, many participants claimed to have a different concept of data; while we tried to define terms carefully, apparently there is a need for more discipline-specific information and discussion on the topic.

The Swiss research community uses and therefore relies on international repositories extensively: Institutions from Switzerland are involved in only 25% of the repositories mentioned. The other 75% are represented within the EU or internationally. Switzerland provides institutional or financial support for 13% of all repositories mentioned. However, future requirements for services from the Swiss community are not yet met by the international repositories' plans. In particular, services for the support of legal issues and security in general have to be more in focus, as this has the highest overall demand.

Extensive supplementary material for the landscape and repository surveys are going to be provided within the SNSF community[3] on Zenodo in conjunction with data papers which describe the survey design and basic evaluation in detail.

4. REFERENCES

Knoth, P., & Pontika, N. (2015, September 2). Open Science Taxonomy. https://doi.org/10.6084/m9.figshare.1508606.v3

National Institutes of Health (NIH) (Ed.). (2003, March 5). NIH Data Sharing Policy and Implementation Guidance. Retrieved from https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

Warner, A. (2018). *Disciplines: Digital Commons Three-Tiered Taxonomy of Academic Disciplines* (p. 28). bepress. Retrieved from https://www.bepress.com/reference_guide_dc/disciplines/

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016, March 15). The FAIR Guiding Principles for scientific data management and stewardship [Comments and Opinion]. https://doi.org/10.1038/sdata.2016.18

5. AUTHORS' BIOGRAPHY

Dr. Markus von der Heyde received his PhD in Computer Sciences from the University of Bielefeld, Germany for his work at the Max Planck Institute for Biological Cybernetics Tübingen in 2000. His approach to adopt biological principles into distributed computer applications in order to enhance stability and robustness was applied in DFG and EU funded research projects. Between 2003 and 2011 he served as ICT director of Bauhaus University in Weimar and focused on topics such as information security, service management, strategy and governance. Since 2011 he is CEO of vdH-IT and management consultant specializing in IT governance and digital transformation in higher education. In cooperation with various partners he has conducted the German CIO studies since 2014. Recently he conducted the Open Data Repository Landscape Analysis for the Swiss National Science Foundation (SNSF).
See more details on https://www.researchgate.net/profile/Markus_Von_Der_Heyde3.

---

[3] See https://zenodo.org/communities/snsf/.