SWISS NATIONAL SCIENCE FOUNDATION

www.snsf.ch
Wildhainweg 3, P.O. Box 8232, CH-3001 Berne

Call for proposals

Open Research Data: Landscape and cost analysis of data repositories currently used by the Swiss research community and required in future

1. Background and purpose

Mandated by the Federal Government, the Swiss National Science Foundation (SNSF) supports basic science in all academic disciplines, from history to medicine and the engineering sciences. The SNSF values Open Research Data (ORD) as a fundamental contribution to the impact, transparency and reproducibility of scientific research. Accordingly, the SNSF has formulated its Open Research Data Policy statement and defined the criteria, which are expected to be met by all its funded researchers. One of these is to deposit data and metadata onto trusted, publicly accessible repositories in formats that anyone can find, access and reuse without restrictions. The infrastructure landscape in this regard is complex. Repositories can differ widely in their strategic focus, in particular with regard to target groups and services. They either serve as archives for data sets from different research disciplines (e.g. Zenodo, Dryad, Harvard Dataverse), or focus on specific disciplines (e.g. Archaeology Data Service, Cambridge Crystallographic Data Centre, Proteomics Identifications Data-base). Depending on the focal topic or breadth of the data archive, different services are offered pertaining to data management (curation) and data analysis. As far as funding bodies are concerned, this can include the most diverse national and multi-national partners (governments, development organizations, institutions, foundations, etc.).

The purpose of the landscape and cost analysis is to gain a better understanding of the data sharing activities of the Swiss research community and to prepare a multi-layered overview of national and international data archives as well as their features, functioning and financial flows. The results will serve as a basis for swissuniversities, in the frame of its national open access strategy, to develop a concept that analyzes the current need for repositories and their types and describes different funding scenarios. At the same time it will serve the future development of the SNSF ORD policy and related infrastructure funding strategies of both the SNSF and swissuniversities. See also Appendix for further background information and resources.

2. Objective

The objective of this study is to draw conclusions regarding the status quo of data repositories and their cost structures, and to propose concepts for supporting the establishment, repurposing or expansion of data repositories from all areas of analysis. These proposals can include national as well as international promotional measures and should identify the considerations pertaining to archives that focus on specific scientific domains as well as non-specific archives.

The study should address the following points:

- Definition of the types of repositories/archives/registries for research data
- Data collection and survey methods
- Definition of outcomes
- Sample size consideration and statistical power
- Strategic analysis of the repositories (features, functioning) in order to identify the need as well as the advantages and disadvantages of specific solutions
- Financial flow analysis, cost calculations or estimates for the different data archiving solutions
- Coordination with other stakeholders (eg Swiss Data Science Center, SWITCH)

Ethical and data protection issues must be considered and a draft of the ethics submission to the Ethics Committee of the Canton of Bern included if required.

3. Questions to be addressed by the landscape and cost analysis

Mapping of data archives

Nowadays thousands of data repositories and data archives exist. A full landscape analysis would be very time intensive. It therefore appears more productive to approach the mapping study from the perspective of Swiss researchers, i.e. to limit the study to data archives that are used by these researchers. The mapping study should include

- the assignment of data archives to specific countries and scientific domains,
- the comparison of the services offered and of the levels of data quality and accessibility in accordance with the FAIR principles, and
- a structured survey in order to have an overview of the data archives (and their allocation to scientific domains) used by Swiss researchers.

In order to avoid redundancies, this survey should also consider data and findings that are already available from other surveys and analyses. All levels of analysis should be widely supported by data so that it will be possible to apply different classification systems to the data archives and also to identify any gaps. This requires the most complete capture possible of all research data archives used by or known to Swiss researchers, or planned to be installed by Swiss Universities and the ETH domain, and how they are attributed to scientific domains. As regards the services and the degree of compliance with the FAIR principles, information can be obtained from re3data.org, the data archives themselves, analyses provided by the SNF (see 4. Scope and synergies) and a list of repositories with data quality seals that is kept by Science Europe, i.e. can be provided by the SNSF.

Costs of data archives

Depending on the initial conditions and the services on offer, the establishment, operation and use of data archives can have very different financial dimensions, requirements and funding bodies. It is important to collect this information systematically in order to prepare cost estimates for future data archive solutions supported by the Confederation. Considering the complexity of the data archiving landscape, we assume that a cost analysis of all archives used by Swiss researchers will not offer significant added value. It is therefore all the more important that such analyses should be based on all data archives rooted in Switzerland (whether FAIR or not) or planned to be established by the Swiss Universities and the ETH domain in near future, and supported either on a national basis or by an institution, as well as on international data archives that are often used by Swiss researchers and combine these in cost models. As many of the analysed data archives are presumably still being developed and do not yet offer a full range of services or do not yet follow the FAIR principles, cost analyses will also have to take account special issues such as the technical level of expandability, an estimate of the costs required to achieve compliance with FAIR, and an estimate of the operating costs once a full range of services can be offered.

These analyses are tied to the above mentioned structured survey and specific inquiries directed to the operators of data archives and other stakeholders (SDSC, SWITCH, possibly others) will need to be made.

Requirements of Swiss researchers

Mapping and cost analyses will provide a picture of what is on offer. For future investments in data repositories to be effective, however, the available resources have to be compared with the requirements of Swiss researchers. It will presumably be costly and time-consuming to collect all requirements by means of a survey and to use the answers to draw up a structured comparison. Interviews with users of data archives that are frequently mentioned or are specially promoted in Switzerland as well as with researchers from domains for which there are hardly any or no data archives available could be a possible solution as regards the analyses. The persons selected for interviews should be representative in number and membership of the different scientific domains and research institutions in Switzerland and their size, and the interviewees and interviewers should not have conflicts of interests.

4. Scope and synergies

The tendering groups are invited to consider and comment on potential synergies with existing databases and completed studies.

The SNSF holds data on DMPs submitted as part of grant proposals since October 2017. Based on these, a summary of the repositories that are known to its applicants and the extent to which these are used is in preparation. The summary will focus on aspects such as the listing of the identified data archives, the frequency with which a specific archive is mentioned, the number of different data archives that are planned to be used in each project, and the attribution of the data archives to broad research domains and their compatibility with the FAIR principles. It will be based exclusively on the DMPs submitted to the SNSF, and as such will offer only a reduced input for the mapping study. The SNF data will also only offer rough indicators for the allocation of the data archives to scientific domains.

5. Qualifications expected

Tendering groups should have documented expertise and experience in the conduct and analysis of surveys and semi-structured interviews, technical and user knowledge on different types of data infrastructures (archives/repositories/registries), and a good overview on current international open research data trends and developments, especially on the European Open Science Cloud initiative and FAIR principles. Active participation in current and past working groups on related topics is an asset.

6. Deliverables and timing

The SNSF will review the tenders in collaboration with swissuniversities and issue the mandate in early July 2018. Monthly meetings and close collaboration with SNSF/swissuniversities, as well as information on progress is mandatory for the whole duration of the project.

- A draft study design, for discussion with the SNSF and swissuniversities (mid July 2018)
- A final study design, data collection instruments and methods, defined milestones, established advisory board (July 2018).
- A rough estimate of the investments planned by the Swiss universities and the ETH domain for the years 2021-2024 (end August 2018)
- A progress report (mid-September 2018)
- A draft report (end October 2018)
- Final report (end November 2018
- Consultation meetings with other stakeholders in Switzerland

Compensation is on a time and material basis. The total cost of the concept should not exceed CHF 65'000, incl. VAT. The first instalment of CHF 35'000 starts when the mandate is issued. The final instalment is paid once the project report has been approved and the final, detailed invoice is issued.

7. Tender

Interested parties are requested to indicate by 18 June 2018 whether they wish to submit a tender (div3@snf.ch)

Tenders must be submitted in writing by 2 July 2018 to div3@snf.ch. They should include:

- Approaches and methods, including sampling, survey, semi-structured interviews and data collection approaches.
- Practicalities of study execution and data collection to be considered, including comments on expected response rates.
- Information and resources required from SNSF and swissuniversities
- Conditions of cooperation between project partners: respective roles in study design, communication, reporting etc.
- A timeline, including milestones and budget.
- Project team, references, project portfolio.

Decisions will be communicated by 10 July 2018.

Applicants should not have any conflicts of interests with the SNSF or swissuniversities.

This invitation for tenders will be re-issued if proposals are insufficient in number or content.

8. Selection Criteria

Criteria for selection are the concept and feasibility of the study and prior experience with comparable topics and projects, as well as cost efficiency.

Contacts:

Aysim Yilmaz (until 21 June 2018)
Head of Division
Biology and Medicine
aysim.yilmaz@snf.ch
Tel: +41 (0)31 308 23 30

Martin von Arx (from 22 June 2018)
Scientific Officer
Biology and Medicine
martin.vonarx@snf.ch
Tel: +41 (0)31 308 21 89

For swissuniversities:
Gabi Schneider
Co-Head of Program Coordination P-5
gabi.schneider@swissuniversities.ch
Tel: +41 31 335 07 83

Appendix:

- Background information and Resources

Appendix: Background Information and Resources

The following data, sources and contacts are available for the study:

1. The SNSF

- SNSF's Open Research Data working group
- SNSF's member in the working group "Research Data" of Science Europe
- Analysis of Data Management Plans submitted by applicants to the SNSF (since October 2017)
- Research Database P3 (link)

2. Program "Scientific Information" of swissuniversities

- Contact to project leaders of ongoing and completed projects (link; for further information contact Head of program coordination Gabi Schneider)

3. Other resources

- Swiss National Strategy on Open Access (link)
- API Documentation of re3data.org (link)
- OECD study "Business Models for Sustainable Research Data Repositories" (link)
- OECD study (Strengthening the effectiveness and sustainability of international research infrastructures" (link)
- Report vof Knowledge Exchange "The Evolving Landscape of Federated Research Data Infrastructures" (link)