

Keep Calm and Fill In Your DMP: Lessons Learnt from a Swiss DMP Template Initiative

Lorenza Salvatori
EPFL Library

Ana Sesartic
ETH Zurich

Nathalie Lambeng
EPFL Library

Eliane Blumer
EPFL Library

Abstract

Aligning with other funders such as Horizon 2020, the Swiss National Science Foundation (SNSF) requires researchers who apply for project funding to provide a Data Management Plan (DMP) as an integral part of their research proposal. In an attempt to assist and guide researchers filling out this document, and to provide a service as efficient as possible, the libraries of the Ecole Polytechnique Fédérale de Lausanne (EPFL) and ETH Zurich took the lead to elaborate on a DMP template with content suggestions and recommendations. In this practice paper, we will describe the collaborative effort between the two Swiss federal institutes of technology, namely EPFL and ETH Zurich, as well as some partners of the national Data Life Cycle Management (DLCM) project, which resulted in a very helpful document as reported by our researchers.

Received 19 January 2018 ~ *Accepted* 20 February 2018

Correspondence should be addressed to Eliane Blumer, Station 20, Rolex Learning Center, 1015 Lausanne. Email: eliane.blumer@epfl.ch

An earlier version of this paper was presented at the 13th International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution Licence, version 4.0. For details please see <https://creativecommons.org/licenses/by/4.0/>



Background

Proper data management is part of good scientific practice and a key prerequisite for effective data sharing throughout the scientific community. In order to facilitate and organize data management, the best practice is to create a Data Management Plan (DMP) – a written document that describes the way a research team manages its data, how research data is produced during a research project, how it is collected, and what to do with it during and after the research process. It defines every step of the data lifecycle, starting from data creation, to transformation, analysis, storage, sharing and reuse.

Following the principle that research data and research findings funded by public money should be freely accessible to all, the Swiss National Science Foundation (SNSF) recently aligned with US, UK and European Commission policies, and announced that researchers applying for SNSF project funding have to supply a DMP as part of the submission process.

From that perspective, SNSF expects funded researchers to:

1. store the research data they have worked on and produced during the course of their research work;
2. share this data with other researchers, unless they are bound by legal, or ethical clauses, as copyright and confidentiality;
3. deposit their data and metadata into existing public repositories in formats that anyone can find, access and reuse without restriction.¹

According to the SNSF, this DMP has to be understandable, suit the researcher's project, and meet the standards set by the researcher's community. The SNSF DMP form, together with detailed guidelines, became available in May 2017. It comprises four sections, (1) data collection and documentation, (2) ethics, legal and security issues, (3) data storage and preservation, and (4) data sharing and reuse, and requires researchers to reflect upon their research data management during the whole project lifespan (and beyond).²

Approach and Methodology

Considering the tight timing imposed by the Swiss National Science Foundation and the high estimated number of applications, the EPFL and ETH Libraries analysed the possible resulting workload compared to the available human resources and tried to anticipate possible common issues in DMP writing in order to be ready to provide a good service to their communities.

¹ SNSF policy on open research data: http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx#SNSF%20policy%20on%20Open%20Research%20Data

² SNSF Data Management Plan – Content of the mySNF form: http://www.snf.ch/SiteCollectionDocuments/DMP_content_mySNF-form_en.pdf

In addition to the tight schedule, a second important constraint had to be taken into account in designing this focused support. Differently from other funders, the SNSF requested researchers to use a specific DMP form that is integrated in the web-based grant proposal submission form.

It is not possible to submit a document completed outside of the SNSF's system. The solution conceived by the EPFL Library, and proposed to the ETH Library, has been to prepare a 'ready-to-use' document for the researchers, based on the SNSF DMP template, giving as much advice as possible through the document itself. The aim of such a solution was to make the users autonomous in filling their DMP, and for the library to be able to focus on other personalized support activities.

Both EPFL and ETH Libraries propose a wide range of research data support services for researchers who have to deal with data management.

One of the core support activities at EPFL and ETH Zurich – started in 2015 – is the guidance to complete Data Management Plans delivered through personalized sessions (see also Sesartic and Dieudé, 2017). This approach helps in better assessing the needs of the researcher, providing a tailored support. However, such an approach requires a considerable effort in terms of time and human resources for the Library. For each session, one or two members of the Research Data Library team and – in the case of EPFL – a liaison librarian³ are involved to provide either information on existing services and tools for RDM to fill the DMP, or to give constructive feedback on DMP drafts.

To address these questions at ETH Zurich,⁴ the ETH Library Digital Curation Office developed RDM training sessions in the form of a basic one-and-a-half hour training and an extended half-day workshop. The aims of the trainings are to raise awareness of existing requirements and of benefits of proper RDM, to introduce some services and tools for RDM as well as to encourage participants to share both their experiences and the methods and tools they use during the interactive parts of the workshop. Researchers must be empowered to make informed decisions on their data, as they are the experts with the most intimate knowledge. Activating teaching methods, which engaged the participants in group work and discussions, facilitated direct exchange between the peers as well as with the trainers. The trainings are free of charge and open to everyone, with a focus on members of the ETH Zurich. Most participants taking part in the workshop and the short one-and-a-half hour training about RDM are doctoral students, nevertheless some postdocs, senior scientists and technical staff also attended the training.

They generally showed various needs and levels of knowledge, but all were aware of the problems surrounding RDM and were happy to learn about possible solutions. The training also proved to be an excellent marketing instrument, as nearly every training led to invitations from research groups for more tailored trainings, after one or several group members participated in the courses.

In order to better cater to the various groups, the ETH Library in general offers tailored training sessions for groups and departments. These can range from 15-minute short mini lectures over coffee or lunch breaks, to full-fledged one-day training workshops. As some departments and institutes already offer similar internal training, and communication and coordination with them is key.

³ The team members have various backgrounds, either in data management or in information science with concentration on data management. A liaison librarian is responsible for a specific discipline.

⁴ See also Sesartic and Töwe, (2016) for further information on research data services at the ETH Zurich.

Concept, Design, Partners

To be ready in due time, the EPFL Library started the preparation work right after the publication of the SNSF guidelines in May 2017.

First, the decision to use the SNSF form as a base for preparing the support document was taken. Starting from the proposed template, the aim was to adapt it and propose tailored recommendations for the ETH domain and a few disciplinary examples for each section along with the funder's guidelines (see the following section for a detailed description).

An important step in the design process has been to identify the internal and possibly other external partners to develop this document.

Internally, it was crucial not only to define the right actors to contribute to the document, but also to engage the EPFL senior management in this process. For this reason, a working group has been established and led by the Library, including actors from different units of EPFL as well as the Vice President for Research and the Vice President for Information Systems. The EPFL Presidency was also represented by the Open Science scientific advisor to the President.

During the summer, the concerned actors⁵ worked on the document and more specifically:

- the EPFL Library led the whole writing process and coordinated the different involved units;
- the Research Office, as responsible for Research Ethics at EPFL, provided information and text for the ethics section. It also contributed with useful insights into the SNSF submission process that helped in estimating the possible workload related to support services for researchers;
- the Technology Transfer Office was involved in the preparation of the EPFL Data Publication Decision tree (one of the support tools added to the template), thanks to its expertise on licenses and intellectual property;
- the Vice Presidency for Information Systems was responsible for the storage section of the template.

In terms of external partners, in addition to the partnership with the ETH Library, the first draft of the document was then sent for comments and feedback also to all the Swiss institutions participating (as the EPFL and the ETH Zurich) in the national project Data Life Cycle Management⁶.

The Template and its Writing Process

A core document based on the SNSF DMP template was proposed to the researchers to support them in the preparation of their own DMP. This document combined the (binding) funder's guidelines and the recommendations specific to EPFL/ETH Zurich. For each section, the two libraries prepared a help text tailored to the disciplinary context of the Swiss Federal Institutes of Technology. In addition to that, a selection of

⁵ The EPFL Library, the Research Office and the Technology Transfer Office are part of the Vice Presidency for Research.

⁶ Data Life Cycle Management project: <http://dlcm.ch/>

relevant examples from existing DMPs (duly anonymised, if needed) has been added to better guide in the DMP writing process.

For some sections, an adaptation of the recommendations to one specific institution was needed (for example in the case of the storage solutions or the ethical process).

Generally speaking, the aim of this adapted template was to help researchers in quickly understanding what the funder expected from them, reduce the anxiety often linked to the preparation of such a document⁷, and get something easily reusable, as some sentences (and even examples) could be readapted and used for a specific project.

At EPFL, it was also decided to add two complementary practical tools to the template, to give a more complete guidance to the community:

- a data publication decision tree⁸, to orient researchers in the process of making their data publicly available and choosing an appropriate license
- a recommended file formats document⁹, to give advice on the most suitable file formats for sharing and preservation.

At ETH Zurich, the document was supplemented by the already existing ETH file formats recommendation document¹⁰, customized to ETH purposes and in use since several years.

To prepare the document, the EPFL and ETH Libraries put in place a collaborative writing process, feeding both each section and providing feedback one to each other. The whole process took about two months and the template was ready by the end of August 2017.

An open and critical question was how to propose the document, considering that no online tool (such as the DMPonline) exists for the Swiss academic community. The EPFL Library decided to make an editable version of the template (Word document) available from the EPFL research data website¹¹, so that the researcher can work directly on it and once satisfied, just copy and paste text into the corresponding section of the SNSF grant proposal submission form. The document is likewise provided by the ETH Library as well.

Workshops

In order to enhance the usage of the SNSF DMP template, the EPFL Library organised eight workshops in the weeks prior to the first SNSF grant application deadline following the DMP requirement. The goal of the workshops was to provide guidance about Data Management to a broader audience and to promote the SNSF DMP companion document. For this purpose, a 90-minutes format during lunch hours seemed most appropriate. To increase audience homogeneity, EPFL Library suggested specific workshop dates for each Faculty. The presentations were the same for all faculties, but

⁷ During the three years of tailored DMP support service, it was clear that for researchers it is usually scarier to start from an empty DMP template, compared to filling it out with appropriate information and an explanation provided.

⁸ Data publication decision tree: <https://infoscience.epfl.ch/record/230281>

⁹ Recommended file formats: https://researchdata.epfl.ch/wp-content/uploads/2018/05/Recommended_DataFormats_-2018_03_05_Final.pdf

¹⁰ ETH Zurich file formats recommendation document: <https://documentation.library.ethz.ch/display/DD/File+formats+for+archiving>

¹¹ Research data at EPFL: <https://researchdata.epfl.ch/>

the answers given to specific questions were useful to more participants because they could more easily relate.

After a short reminder about Data Management Plans, the specific requirements of SNSF were presented, including the FAIR principles. Trainers also explained how to deposit and open up data, presenting various repositories. A live demo was used to present the SNSF grant submission interface and how it is linked with the DMP companion. At the end, participants were able to work with DMP use cases and to see how to use the DMP companion when writing a DMP for the SNSF. Throughout the workshops, participants were asking the trainers specific questions as well as sharing their own experience with data. For more personalised guidance, they were advised to get an appointment with the EPFL Library Research Data Team. EPFL Library is planning other similar workshops before the next the SNSF grant application deadline.

The ETH Library, together with the Scientific IT Services of ETH Zurich, organized a specific information event prior to the SNSF grant application deadline, focusing on DMP plan requirements and services offered by the two teams to help the researchers. Several hundred researchers attended the event and the presentation slides were shared online.¹²

Following the event, there was a short one-and-a-half hour comprehensive training and a half day workshop, the latter again in collaboration between the ETH Library and Scientific IT services. These trainings were open to anyone within the ETH Domain. In addition, trainings for specific institutes were held as well upon invitation.

To complement the training, an ETH-wide mail was sent out pointing the researchers to the guidance document compiled by EPFL and ETH Zurich, aiming to aid them in filling out the SNSF plan.

Following this dissemination of information, the Digital Curation Office at the ETH Library, which is the main point of contact for questions concerning DMPs, saw an enormous increase in questions and requests for consulting from the researchers. Thanks to the combined effort of all involved and increased working hours, all requests were resulting in a high level of satisfaction and very positive feedback from the researchers. However, the high workload also showed the need to increase the supporting personnel in the future in order to meet the rising needs from the ETH community.

Benefits and Limitations of the Initiative

The service has been going on now for some time, so it is the right moment to make an assessment. In the next few lines the benefits and limitations will be addressed.

First of all, one can say that for both Schools the DMP template is a highly used and also appreciated service, as the regular positive feedback and the increasing number of demands from researchers shows.

Since March 2017, namely for the first submission phase of SNSF with a mandatory DMP, a total number of 68 participants have participated in the eight DMP-workshops provided by the EPFL Research Data Library Team. 31% of those participants came from four different faculties (Basic Sciences, Life Sciences, Architecture and Engineering) and asked for a review of their DMP (21 DMPs in total), covering two funding pillars (Prima, Ambizione).

¹² Data Management Plans for SNSF Grants: <https://de.slideshare.net/ETH-Bibliothek/data-management-plans-for-snsf-grants>

As for ETH Zurich, 434 participants attended various training and education events regarding DMPs and research data management since March 2017. The gender of participants is nearly equally balanced, which is surprising considering the gender imbalance of only little more than 30% female students¹³ at ETH Zurich, suggesting that female researchers seem to be more interested in data management. Approximately three quarters of the participants belong to the ETH Domain, while the last quarter of participants came from non-ETH higher education institutions (mainly University of Zurich). The department of environmental systems sciences provides the largest number of participants with 26%, followed by the department of biosystems science and engineering with 23%, the department of information technology and electrical engineering with 13% and the department of architecture with 12% of participants.

Generally, DMP support consists of the Library preparing a feedback in advance by proof-reading the DMP and then inviting the concerned person to the Library and to discuss possible formulations of answers or the use of examples matching their research approach. The nearly 30-pages strong DMP-template with examples serves as a basis for discussion. This template is of big use for the team, as it reduces the workload by empowering researchers to advance further in the process by themselves and by normalizing the review process.

Second, the DMP-reviewing service also permitted the EPFL and ETH Libraries to become the centres of competence for most of the questions around data management planning and therefore act as a first point of contact for the entire EPFL and ETH Zurich community, respectively. In this context, the DMP template and service have also been asked for and reused by other institutions in the ETH Domain, such as EAWAG¹⁴, PSI¹⁵ as well as IDIAP¹⁶.

Finally, yet importantly, this first success may be recurrent in time, as the SNFS is continuing with the DMP-initiative and the next submission period for funding applications with DMP is already scheduled for the beginning of April 2018. Therefore, the ETH and EPFL Libraries continue to adapt their service and to stay up-to-date with the new demands.

However, it is very difficult to find out the exact number of researchers benefitting from the service. The template, for instance, is shared freely by email and on the website, which also makes it difficult to track how many people the template has actually reached in total and what its real impact was. There remains a challenge to answer questions about whether the document is too long, or whether the examples are useful or not and how they should be adapted. For instance, when a researcher copy-pastes an entire example out of the template, is it a question of time or a question of the length of the document? In the future, the EPFL and ETH Libraries will try to find solutions to be able to answer those questions.

¹³ ETH Zurich student facts and figures: <https://www.ethz.ch/services/en/finance-and-controlling/facts-and-figures/students.html>

¹⁴ EAWAG: <http://www.eawag.ch/en/>

¹⁵ PSI: <https://www.psi.ch/>

¹⁶ IDIAP: <http://www.idiap.ch/en>

Next Steps

As already mentioned, funders continue with data management policies in 2018 and the SNSF is not an exception. That's why the EPFL and ETH Libraries will also continue to adapt and further develop their DMP-services. In this context, specific trainings for selected labs are planned at the EPFL and new webinars at the ETH Zurich.

There is also an ongoing reflection about an interactive version of the template, which will be continued in the upcoming months and which could help to tackle a part of the question about the impact. The idea would be to offer a more discipline-oriented service, as the researchers are now generally aware of the main basics of data management and increasingly ask specific questions from their disciplines. The EPFL and ETH Library Data Management Teams will also continue to evolve their educational offer regarding data management.

Acknowledgements

We would like to thank warmly both respective research data library teams for their very useful input and precious rereading work for this article. At ETH Zurich, we would like to thank Matthias Töwe and at EPFL, Karine Delvert, Fantin Reichler, Mathilde Panes, Jan Krause, Aude Dieudé and Raphaël Rey.

References

- Sesartic, A. & Dieudé, A. (2017). Research data management training and support services at both ETH Zurich and EPFL Lausanne. Conference Proceedings: E-Science-Tage 2017: Forschungsdaten managen, Heidelberg, Germany, 16-17 March 2017. doi:10.3929/ethz-b-000224456
- Sesartic, A. & Töwe, M. (2016). Research data services at ETH-Bibliothek. IFLA Journal 42(4), 284-291. doi:10.1177/0340035216674971