Research Data Management training and support services at both ETH Zurich and EPF Lausanne

Lessons Learned, Best Practices and the Way Forward

Ana Sesartic[1], Aude Dieudé[2]

1 ETH Library, ETH Zurich, ana.sesartic@library.ethz.ch
2 Ecole Polytechnique Fédérale de Lausanne Library, EPFL, aude.dieude@epfl.ch

Abstract. The management of research data throughout its life-cycle ensures its long-term value and preservation, as well as being a key prerequisite for effective data sharing. Many funding bodies mandate the creation of data management plans and open access publication of the research results they funded. In order to concentrate the efforts of different universities, libraries, and IT-services, the project "Research Data Life-Cycle Management: From Pilot Implementations to National Services (Data Life-Cycle Management, DLCM[1])", was launched by eight Swiss Higher Education Institutions on behalf of the former Swiss University Conference (SUC) as part of the programme SUC P-2 "DLCM", which is chiefly designed to improve the handling of scientific information in Switzerland.

With the start of DLCM in 2015 the Libraries of the two Swiss federal institutes of technology, EPF Lausanne and ETH Zurich, rose to the occasion and strengthened their respective initiatives and efforts by creating personalized data management plan (DMP) services and training sessions on research data management. The services and training sessions were set up in close collaboration with the scientific IT departments and the Research Office at EPFL Library and ETH Library, thus enabling the libraries to offer expertise tailored to researchers' needs, and to cover the entirety of the data life-cycle. The closer collaboration with researchers during training sessions and consultations contributed to the establishment of stronger trust relationships between scientists and information professionals at the two Swiss federal institutes of technology. A mutual learning process was sparked, allowing both sides to share best practices and tackle new challenges together.

By building on existing experiences, resources, and tools available within Swiss higher education institutions, the DLCM project focuses on the common goals: to harmonize and strengthen research data management practices across actors (researchers and information professionals), disciplines (sciences, social sciences, and humanities), and institutions (universities, universities of applied sciences, and universities' service providers) in a sustainable way. While universities are obviously competing for funding, students, and scholarly excellence, there is a need to use available resources efficiently and also to share best practices acknowledging the high level of staff mobility between institutions. With this background, we aim to share our insights, experiences, and evolving best practices regarding research data management training, resources, and support services. To this end, our paper focuses on four major aspects: our new data management training, our personalized research data management support services, our lessons learned and how we envision the way forward.

Keywords. Research data management, training, services, best practice, Switzerland

---

1   https://www.dlcm.ch/

Introduction

The rising production of research data has created new challenges for its management. In order to ensure its continuity, transparency and accountability, the timely and effective management of research data throughout its life-cycle is essential (Louise Corti et al. 2014; Goodman et al. 2014). Proper data management is also a key prerequisite for effective data sharing and publication, which, in turn, increase the visibility of scholarly work and are likely to increase citation rates (H A Piwowar and Vision 2013; Heather A. Piwowar, Day, and Fridsma 2007). Managing research data is furthermore a sine qua non condition for efficient long-term preservation because the latter must rely on sufficient metadata and contextual information being available to make sure that data remains reproducible, reusable and understandable in the long run.

As effective and efficient data management becomes more and more challenging for both researchers and information specialists across institutions, the question arose as to how they can best be reached and supported on a national level. The project "Research Data LifeCycle Management: From Pilot Implementations to National Services (DLCM)", aims to concentrate the efforts of 8 Swiss universities (EPFL, ETH Zurich, Geneva School of Business Administration/University of Applied Sciences and Arts Western Switzerland, University of Basel, University of Geneva, University of Lausanne, University of Zurich), represented by their libraries and IT-services including the existing national service provider for HEI (SWITCH). It was initiated on behalf of the former Swiss University Conference (SUC) as part of the programme SUC P-2 "DLCM", which is chiefly designed to improve the handling of scientific information across the country.

With the start of DLCM in 2015 the EPFL and ETH Libraries reinforced their respective efforts, which had already begun in 2012 with the creation of personalized Data Management Plan (DMP) services and training sessions on research data management, as many funding bodies have mandated the creation of data management plans and open access publication of the research results found. The services and training sessions were set up in close collaboration with the scientific IT departments and Research Office at EPFL Library and ETH Library, thus enabling the libraries to offer expertise tailored to researchers' needs, and to cover the entirety of the data life-cycle. Soon after, the EPFL Library and ETH Library joined their efforts by collaborating on a Data Management Checklist in 2015 and establishing training and consulting services on their own in 2016.

The closer interaction with researchers during tailored training sessions and consultations contributed to the establishment of stronger trust relationships between scientists and information professionals at the two Swiss federal institutes of technology. In the following, we are sharing our insights, experiences, evolving best practices and lessons learned regarding research data management training, resources and support services.

Research data management training

In step with the growing awareness of the value of research data and the risks of losing such data over time, the need for research data management (RDM) has gained increased attention over the last years. RDM combines both the need to manage data over the course of a project, as well as the curation and preservation of data for future work and reference.

To address these questions at the ETH Zurich (see also Sesartic and Töwe 2016 for further information on research data services at the ETH Zurich), the ETH Library Digital Curation Office developed RDM training sessions in the form of a basic 1.5 hour training and an extended half-day workshop. The aims of the trainings are to raise awareness of existing requirements and of benefits to be gained from 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 offered free of charge and are open to everyone, with a focus on members of the ETH. Most participants taking part in the workshop and the short 1.5 hour training about RDM are doctoral students, with some postdocs, senior scientists and technical staff present. They generally showed varying needs and levels of knowledge, but all were aware of the problems surrounding RDM and were happy to learn about possible solutions. The trainings also proved to be excellent marketing instruments, as nearly every training led to invitations from research groups for more tailored trainings, after one or several group members visited our courses.

In order to better cater to the varying groups, the ETH Library in general offers tailored training courses 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, communication and coordination with them is key

The ETH Library is planning to establish a dedicated course on RDM and related topics within the ETH curriculum in the future, but to do so will take further time and planning.

To answer the evolving and growing needs of its researchers regarding research data management, the EPFL library set up in January 2015 a steering committee composed of the heads of the IT department, the Research Office and the library to find the best ways to answer them. One month later, on February 2015, the EPFL library received the official green light to start offering in close collaboration with the IT department and the Research Office departments a personalized support service regarding research data management and the preparation of a data management plan. The first six months proved the relevance of such a service, and as a result, the service developed itself and some personalized trainings were also offered to pursue these efforts even further. Therefore in the fall of 2015 two annual trainings on how to optimize research data management were given free of charge by two collaborators of the library, Aude Dieudé[2] and Jan Krause[3], to the entire EPFL community as part of the official staff training service. Due to the success of these two initial workshops in 2015, four training sessions[4] are now offered in both English and French each year and the participants are quite eclectic. They may include principal investigators (PI) and senior researchers, postdoctoral and doctoral students, but also IT specialists, information specialists, project managers, librarians and information specialists. In addition, personalized trainings have been offered on demand to EPFL PIs to train their lab team members in a harmonious way and the feedback we have received has been both encouraging and positive.

Presently in 2017, the EPFL research data management support service team is composed of data librarians, data managers, liaison librarians, and IT project managers. This diversity of back-

---

2   https://people.epfl.ch/aude.dieude?lang=en
3   https:/people.epfl.ch/jan.krause?lang=en%20
4   http://sfp.epfl.ch/files/content/sites/sfp/files/users/132694/public/Descriptifs/Optimiser_gestion_donnees_re cherche.pdf

ground and speciality allows to offer a full picture to the researchers of the different sets of skills relevant for fully considering the data lifecycle management: from its collection and creation, to its description and preservation, and finally to its sharing and reproducibility.

Throughout our training, the team makes sure to fully understand the context, field and specific needs and questions of the participants. Each training is unique and requires dedicated resources and personalized attention; however, here are a series of questions that are regularly asked:

1. What are the best practices for research data management for my lab and my team?
2. How can I ensure that everyone in the lab or all the partnering institutions in this research project is following the same methodologies to save time, energy and money for all of us in the long term?
3. Can you explain to us what a data management plan is, why it is beneficial for us and how to create and maintain one in an efficient, practical and cost-effective way?
4. How much will it cost to store my data during the next ten, fifteen or twenty years?
5. What are the options to safely store, exchange and organize my sensitive data?
6. Do you know of any specific tools that would be best to use in this particular case?
7. Would it be possible to organize a personalized training for my lab in the future?

In providing free expertise and consulting services to researchers, our team is building a solid basis for mutual respect, trust and exchange of best practices. The very fact of having dedicated persons willing to explore with them the best options in their particular scenario while sharing their expertise, networks and knowhow proved to be very beneficial not only for the researchers and their lab, but also for the research institutions.

Knowing how to create a data management plan and how to efficiently manage their data has become a sine qua non condition for receiving research funding from prestigious funding agencies such as the European Commission with its Horizon 2020 program[5] since January 1, 2017, and also from the Swiss National Science Foundation[6] (SNSF) starting in October 2017. These new requirements have the advantage of introducing Swiss researchers gradually but surely to the sensitive questions of reproducibility and open science more generally. In anticipating this transition and responding effectively to these new requirements, the EPFL library paved the way in training a new generation of researchers across disciplines, generations and methodologies to focus on scientific excellence, personalized needs and quality-oriented services to optimize research data management, reproducibility, and open science. In 2016, a specific training for doctoral students was created and developed at EPFL by the library to reach out new generations as early as possible and directly. As a result of all these joined efforts, several institutions across Switzerland and abroad requested and invited the EPFL library to offer personalized training for its participants. Consequently, three trainings focusing on RDM, DMP and data visualization and tools have already been given since 2015 at the Haute Ecole de Gestion (HEG) located in Geneva. In France, two trainings were offered during the Open Access week in 2015 and at the Ecole Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB) in 2016. Recently, the University of Basel invited the EPFL library to give in 2017 a two-day training workshop on how to optimize research data management. The workshop was fully booked within one day and the waiting list is already full, proving how much research data management trainings are both relevant

---

5   https://ec.europa.eu/digital-single-market/en/news/communication-european-cloud-initiative-building-competitive-data-and-knowledge-economy-europe%20
6   http://www.snf.ch/en/researchinFocus/newsroom/Pages/news-170306-towards-open-research-data.aspx

and valuable for everyone. The use of a questionnaire composed of specific questions ahead of the training has proved to be highly important to fine-tune and personalize the workshop according to the participants' needs. In addition a detailed feedback form has been created to fully assess the quality and value of such a workshop for the participants. Both of these tools gave us the chance to rethink creatively and offer a variety of tailored trainings based on the pool of participants, which has been highly appreciated and valued. In a nutshell, providing research data management training at both EPFL and ETHZ paved the way to the creation of new content to offer innovative and personalized solutions to our institutions. Far from being a fashion trend or another administrative burden, RDM trainings are becoming a sine qua non condition to put into practice best practices and excellence in academic research on a daily basis.

Data management plan checklist

Well-managed data is both part of good scientific practice and a key requirement of many funding organizations. A data management plan is required for all projects participating in the extended Open Research Data pilot of EU's research programme Horizon 2020 (EC - European Commission. European Research Area, n.d.). But even if a funding body does not demand data management, following its principles has numerous advantages: it helps make data findable, accessible, interoperable and reusable, thus adhering to the FAIR principles[7].

To aid the Swiss community and the researchers at EPFL and ETH Zurich create DMPs, the EPFL Library and ETH Library created in close collaboration a Data Management Checklist[8], which is one of first tangible deliverables of the DLCM project. Based on pre-existing national and international policies, the list has been customized for Switzerland and covers both general planning and all the phases throughout the data life-cycle. Special sections cover documentation and metadata, file formats, storage, ethical and intellectual property issues.

The list is currently available through the ETH Library and EPFL Library websites, as well as disseminated via the DLCM portal. Further transformation into an online tool using DMP Online is planned within the DLCM project to make the experience even more interactive. The list can both serve as a starting point for face-to-face discussions of data management issues within research groups and with support staff, as well as a concrete starting point for researchers to individually assess their data management and gather information they need for the creation of a data management plan.

Experience proves that this personalized checklist for Switzerland provides an excellent way to get the conversation started and to demystify the researchers' habits, methodologies, and expectations regarding how they perceive the role of the library offering practical and useful resources. As such it is a perfect and gentle ice breaker to concretely unveil what a data management plan is about and how such questions help to raise delicate and sensitive aspects of the data life cycle management, which can be easily overlooked or underestimated. In a nutshell, the researchers can realize that a data management plan represents only the tip of the iceberg. The heart and soul of this new requirement is to engage the researchers to take some distance from their previous habits to think thoroughly about the best practices, resources, and tools that can be used to optimize research data management for the long-term. Using both this checklist with the DMP template[9] of-

---

7   http://www.datafairport.org/
8   https://www.dlcm.ch/ressources/dmp-checklist/
9   https://www.dlcm.ch/ressources/dmp-template/

fers a complementary approach, which demonstrates that both RDM and DMP are not rocket science, but rather ways to guide and support researchers, project managers, IT specialists, data librarians and information specialists towards optimized, personalized and long-term oriented research data management.

Lessons Learned

– Generally positive feedback from researchers. They are very grateful and thankful for such offers and demand more.
– Generally, the researchers are aware of the problems, but do not know how to solve them and do not even know that solutions exist within the university.
– Mind the cultural differences: teaching RDM in Switzerland isn't the same as teaching RDM in another country, as we learned from colleagues.
– Leading by example is important. Get the professor/head of institute into the boat and the doctoral candidates etc. will follow suit regarding RDM.
– Personalized approach is rewarding: from DMP support service to tailored training to word of mouth reputation and credibility.
– RDM training leads to enhanced collaboration with scientists and better visibility for the libraries, as well as improving their image.
– Constructive collaborations are key (win-win approach): among colleagues from the library, among different sets of services within the institutions (including the library, IT, Research Office, and legal experts to name a few) and between institutions nationally and internationally.
– Centralized and harmonized communication, sensitizing actions and quality support service are key to build momentum and trust while changing the image of the library and of librarians.

The Way Forward

Generally, the Swiss data life-cycle management landscape seems to be on the right track regarding RDM services and trainings. To continue improving and expanding our services, the following course of action has been planned:
– Creation of an interactive Swiss DMP tool inspired by already existing tools such as the DMP OPIDoR (Inist-CNRS): https://dmp.opidor.fr/ which is based on https://dmptool.org/
– RDM training offered earning credit points for students to validate their efforts and engagement at their institution, planned at University of Basel and ETH Zurich. Strong wish to put into practice credited training at EPFL in the near future as well.
– Renewal of financial support and funding for the DLCM project for the 2018-2020 period to continue the efforts and harmonization of research data management across Switzerland and between diverse sets of partners (including to name a few Swiss federal institutes of technology, universities, and foundations).
– Continue to change the mentality of the scientific community in close collaboration and communication with the key stakeholders on a national, European and international level: thanks to the new President of EPFL (Martin Vetterli), swissuniversities (the successor of

    the Swiss University Conference), and funding agencies, to name a few, to strengthen research data management, open access, and open science.

– Necessity to increase and strengthen the current team and human resources available in the future to make this change of culture and change of mindset a reality. New skills and a diverse team of experts are necessary in light of new requirements from H2020 and SNSF to continue guiding and training researchers and their team. A robust team is crucial to promote research data management in a personalized way, and to focus on building stronger bridges and connections with researchers based on trust and quality services.

– The library is becoming a one-stop-shop, where open science[10],[11] is one of the core services offered to its institution, putting into light one of the many components and elements following the research data lifecycle management.

Conclusion

In conclusion, we can state the following points and our recommendation are based on our mutual collaboration and national context. Therefore, it goes without saying that they have to be adapted with respect to the respective academic institution, professional environment, national policies in place, and funding resources. Our experience at EPFL and ETH Zurich confirms the relevance of a concrete and need-based approach for researchers. To meet this goal a proactive approach and strategic research data management services are crucial to satisfy new funding requirements. A positive way to emulate a cultural shift in academia is through strong supportive and well-established ambassadors, who can serve as "data champions" within the institution (Higman, Teperek, and Kingsley 2017). Within the academic setting and between institutions, creativity, flexibility and constructive and solid collaborations are key. The ability to think outside of the box and to move beyond academic competitions while building consensus provides a strong basis for a fruitful collaboration among well-known, and yet, usually competitive institutions. In reversing this logic, it becomes possible to see the difference among institutions as complementary assets, which nourish and inspire one another constantly. In this sense, the complementary style and approach between EPFL and ETH Zurich has been particularly valuable for paving the way for further collaborations. As an example, our successful collaboration went from working closely together to create and refine the DMP checklist, to offering personalized RDM training, to participating regularly at international conferences (IDCC[12] and iPRES[13] in 2016, E-Science-Tage[14] in 2017), and to co-write several academic articles (e.g. Burgi, Blumer, and Makhlouf-Shabou 2017). This kind of fruitful collaboration is not the exception to the rule and can be implemented in Switzerland and in other countries worldwide. What it takes is the open mindedness, curiosity and ability to share, engage and collaborate with additional colleagues, teams and projects for the benefit of both parties. It is not the exception, but the norm and will continue to become the case even more in the future in order to save time, energy and money. As an extension of this collaboration, we can for instance think of the recent academic initiative and new strong collaboration

---

10   http://library.epfl.ch/open-science-workshops

11   http://www.library.ethz.ch/Ueber-uns/Veranstaltungen/Think-check-submit-Making-informed-decisions-in-open-access-publishing

12   http://www.dcc.ac.uk/events/idcc16/programme-presentations

13   http://www.ipres2016.ch/frontend/organizers/media/iPRES2016/_PDF/IPR16.Proceedings_4_Web_Broschuere_Link.pdf

14   https://e-science-tage.de

between EPFL and ETH Zurich with the creation of the newly founded Swiss Data Science Center (SDSC) in 2017: https://datascience.ch/

Acknowledgments

References

Burgi, Pierre-Yves, Eliane Blumer, and Basma Makhlouf-Shabou. 2017. "Research Data Management in Switzerland." *IFLA Journal*, January, 34003521667823. doi:10.1177/034003 5216678238.

Corti, Louise, Veerle Van den Eynden, Libby Bishop, and Matthew Woollard. 2014. "*Managing and Sharing Research Data: A Guide to Good Practice*". London, UK: SAGE Publications Ltd. doi:https://uk.sagepub.com/en-gb/eur/managing-and-sharing-research-data/book240297.

EC - European Commission. European Research Area. n.d. 2016. "Open Research in Horizon 2020." http://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf#view= fit &pagemode=none.

Goodman, Alyssa, Alberto Pepe, Alexander W. Blocker, Christine L. Borgman, Kyle Cranmer, Merce Crosas, and Rosanne Di Stefano. 2014. "Ten Simple Rules for the Care and Feeding of Scientific Data." *PLoS Comput. Biol.* 10 (4). e1003542. doi:10.1371/journal.pcbi.100 3542.

Higman, Rosie, Marta Teperek, and Danny Kingsley. 2017. "Creating a Community of Data Champions." *bioRxiv*. http://biorxiv.org/content/early/2017/02/20/104661.abstract.

Piwowar, H A, and T J Vision. 2013. "Data Reuse and the Open Data Citation Advantage." *PerrJ Comput. Sci.* doi:10.7717/peerj.175.

Piwowar, Heather A., Roger S. Day, and Douglas B. Fridsma. 2007. "Sharing Detailed Research Data Is Associated with Increased Citation Rate." *PLoS ONE* 2 (3). doi:10.1371/journal.pone.0000308.

Sesartic, Ana, and Matthias Töwe. 2016. "Research Data Services at ETH-Bibliothek." *IFLA Journal* 42 (4): 284–91. doi:10.1177/0340035216674971.