Universität Zürich[UZH]

ZB Zentralbibliothek Zürich

A second life for old research data: the challenge of post-research data publication and how academic libraries could support it

MAS Thesis in Library and Information Sciences

Universität Zürich / Zentralbibliothek Zürich

Submitted by:

Aline Frank (Universitätsbibliothek Bern)

Supervised by:

Prof. Dr. René Schneider (Haute école de gestion de Genève)

Dr. Dirk Verdicchio (Universitätsbibliothek Bern)

Bern, 8th March 2019

Universität Zürich[UZH]

ZB Zentralbibliothek Zürich

Declaration of originality

Selbständigkeitserklärung

Studiengang in Bibliotheks- und Informationswissenschaft 2017-2019

Der/die Verfasser/in der Arbeit mit dem Titel

A second life for old research data: the challenge of post-research data publication and how academic libraries could support it

erklärt hiermit ausdrücklich,

- die vorliegende Arbeit selbständig, ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt zu haben;
- die aus fremden Quellen direkt oder indirekt, wörtlich oder sinngemäss übernommenen Texte, Illustrationen und/oder Gedanken ausnahmslos als solche kenntlich gemacht zu haben und sie gemäss wissenschaftlichen Richtlinien korrekt nachgewiesen zu haben.

Der/die Verfasser/in hat das Merkblatt „Plagiate“ zur Kenntnis genommen und ist mit einer allfälligen Überprüfung der Arbeit auf Plagiate unter Einsatz entsprechender Software einverstanden. Er/sie ist sich bewusst, dass Arbeiten, welche die Grundsätze der Selbständigkeitserklärung verletzen – insbesondere solche, die Zitate oder Paraphrasen ohne Herkunftsangaben enthalten –, als Plagiat betrachtet werden und die entsprechenden rechtlichen und disziplinarischen Konsequenzen nach sich ziehen können (gemäss §§ 7ff. der Disziplinarordnung der Universität Zürich).

Frank, Aline

Bern, 8. März 2019

Ort, Datum

Unterschrift

Abstract

The interest in Open Science is growing within the scientific community. Nevertheless, many researchers still keep unpublished data stored on internal institutional servers or personal devices, hardly accessible to the scientific community. Scientists now increasingly intend to publish such data "post-research", i.e., sometime after the respective research projects were terminated. Academic libraries could step in here and provide support for the publication of old research data. This MAS thesis aimed at 1) evaluating if post-research data publication represents an issue to scientists at the University of Bern, 2) detecting challenges that may appear when publishing old research data in practice, 3) deriving guidelines on post-research data publication for researchers, and 4) proposing library expertise, services, and infrastructure to assist this kind of data publication.

A non-representative survey was conducted among biologists and geologists at the University of Bern. More than two thirds of the 21 respondents were aware of unpublished data in their research environment. Most of them considered these data still worth publishing, especially to make them findable, accessible, and reusable. They had a positive attitude towards several potential support options. A self-experiment was performed to experience the process of publishing old long tail research data in practice. Major challenges included the lack of information and guidance, in particular concerning data publication options and legal questions, and the underestimation of effort and time needed for preparing old data and metadata. Guidelines were proposed to provide advice to researchers. They comprise eight chronological steps that highlight key tasks and players during post-research data publication. In addition, they indicate at what stages support by academic libraries and external funders may be requested. The proposed library expertise, services, and infrastructure indicate that academic libraries could specifically assist post-research data publication by providing basic informative support, such as the presented guidelines. Furthermore, libraries could generally foster the publication of research data by providing thorough technical assistance and innovative IT infrastructure for data management. Basic services will require little effort for implementation, whereas advanced and expert support will involve major investments to hire specialized library staff and establish new institutional publication services.

Although post-research data publication will likely remain a sidetrack of current data publication efforts, it represents an issue that academic libraries – in collaboration with scientists and partner institutions – should approach to foster Open Science. The results of this thesis will be relevant to libraries that have the opportunities to further expand and strengthen their research data management portfolios.

Table of contents

Keywords

academic library, data management, data publication, library services, Open Access, Open Research Data, Open Science, post-research data publication, research data, research support

Tables

Figures

1. Introduction

Movements towards open research data

Within the last years, Open Science – "the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society" (FOSTER, 2019a) – has evolved to a global trend in scholarship. First efforts towards Open Science already took place in the 1990ies in the United States, triggered by initiatives such as the US Policy Statements on Data Management for Global Change Research (GCRIO, 1991) or the US Human Brain Project (Huerta, Koslow & Leshner, 1993). These early approaches aimed at facilitating full and free access to high-quality data for research reuse. Open Science postulates that "research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods" (FOSTER, 2019b). Hence, open access to research data, the basis of science, represents one prerequisite to Open Science (Pampel & Dallmeier-Tiessen, 2014). Only if research data are openly available, they will represent a source for data validation and reuse, thereby fostering innovation, transparency, and trustworthiness in science. Open research data facilitate new discoveries, reduce the amount of duplicate work, enhance researchers' impact, and help them make new collaborations with other researchers (Simons & Richardson, 2013). For these purposes, open research data should adhere to the FAIR Data Principles, i.e., be findable, accessible, interoperable, and reusable (Collins et al., 2018; Wilkinson et al., 2016). As defined by Pampel and Dallmeier-Tiessen (2014), the term *open research data* differs from *open data*, "which is mainly used in the context of Open Government initiatives and neglects the special requirements of science".

Research data publication

Publication is the keyword and path that renders research data open, discoverable, and reusable. Research data are considered to be published when they are persistently available, documented, citable, and potentially validated in a certain way (Kratz & Strasser, 2014). The term *research data publication* overlaps with *research data sharing*, and *open research data* – three terms that are used as synonyms in this thesis. In addition, the term *data publication* is used here for the complete process of making data openly available, including all steps from data preparation to final upload to a repository or release of a data paper. At first, each dataset is thoroughly reviewed and accompanied by metadata that appropriately describe its content, purpose, origin, format, size, and other features (Schiermeier, 2018). Then, three options exist to share research data (Kratz & Strasser, 2014; Pampel & Dallmeier-Tiessen, 2014):

Research data and metadata can be uploaded to a disciplinary or general online repository that meets the FAIR principles. Each dataset is thereby tagged by a Digital Object Identifier (DOI). The data may also be stored in an online database or platform that is freely accessible, such as GitHub. Nevertheless, all metadata should be made available in a data repository in order to

render the data discoverable, both by humans and machines. Once available and well documented in a repository, the major purposes to sharing research data are fulfilled – they can easily be found, accessed, and cited.

Research data can also be published as supplement to a research article, forming a so-called "enriched publication". In this case, the dataset usually represents the basis of the correspondent article (but may represent only a subset of all data collected), and is either deposited directly on the journal's platform or in an external data repository. The interest in supplemental material is decreasing, since repositories are considered to be better suited for providing long-term storage and access to the data (Kratz & Strasser, 2014).

Furthermore, research data can be published in the form of a "data paper". A data paper is defined as "scholarly publication of a searchable metadata document describing a particular online accessible dataset, or a group of datasets, published in accordance to the standard academic practices" (Chavan & Penev, 2011). As such, a data paper contains all information about the dataset – e.g., details on collection methods, potential for reuse, availability, licenses – but no analyses or conclusions (Kratz & Strasser, 2014). It follows a short, tightly structured format, with the data itself being mostly stored in a separate repository. Compared to data deposition in a repository, a data paper offers certain additional benefits (Chavan & Penev, 2011; GBIF, 2019; Schneider & Prongué, 2015). Most data papers undergo peer review, which helps to assure a dataset's technical and scientific quality. The publication and citation model of data papers closely resemble those of regular research articles and, thus, are well known to scientists. Furthermore, a data paper counts as additional, scholarly approved publication to the record of the data creator and, therefore, represents a reward for his/her extra effort invested into data publication. These benefits render data papers promising tools for boosting Open Science (Chavan & Penev, 2011; Rushby, 2015).

Current state and trends in data publication

Now that science funders (e.g., SNSF, 2018; swissuniversities & SNSF, 2017), and publishers (e.g., ESA, 2018) increasingly require researchers to make their data publicly available during the publication process, the interest and effort in data management and publication is continuously growing within the scientific community. Illustrated in an idealized, simplified way, the Open Science movement will drive current and new research projects to increasingly publish their data over time – until a high percentage of open research data will be available (Figure 1, blue solid line). This percentage will not reach 100%, since certain datasets will never be completely open, such as sensitive data from medical trials. The number of data repositories has recently grown to a wide variety, with more than 2000 repositories by now listed on re3data.org, "the most comprehensive source of reference for research data infrastructures globally" (re3data, 2018). Also, the number of data papers published and data journals operational worldwide has grown remarkably between 2000 and 2013 (Candela et al., 2015), and presumably continued to increase during the last five years from 2014 to 2019.

Consequently, one can assume that this trend towards Open Research Data will continue. At the moment, however, we are probably only at the beginning of the curve's incline (Figure 1, red dashed line) – and at the start of what Stephanie Simms, a research-data specialist from the California Digital Library, called "a profound shift in research culture" (Schiermeier, 2018).

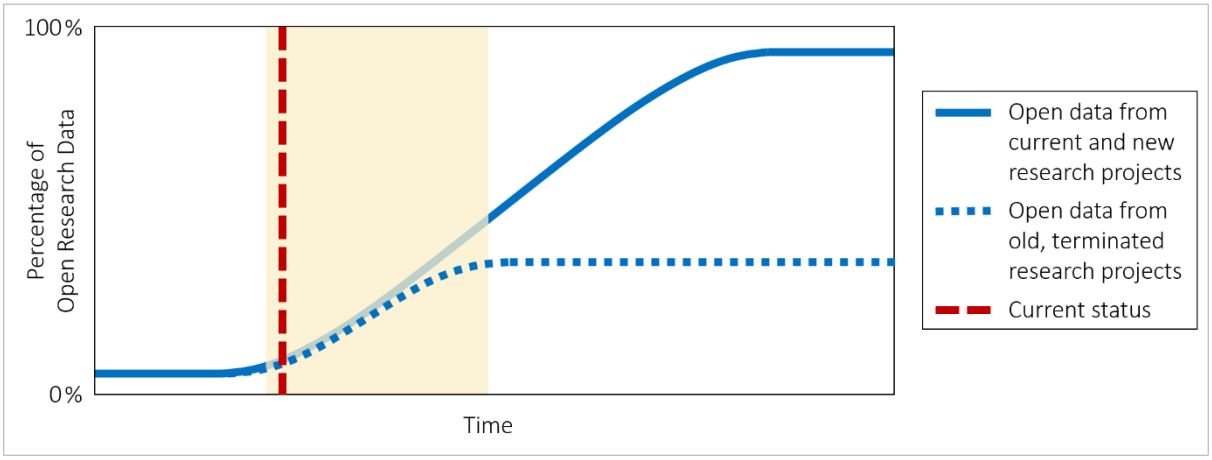


Figure 1: Simplified and ideal representation for the development of open research data across time. The two curves refer to data from current and new research projects, and data from old, terminated research projects. The yellow area illustrates the period during which a focus on post-research data publication support would be meaningful.

Researchers are increasingly supported during research data management and publication. Many research institutions and academic libraries offer guidance by specialized staff when it comes to managing, curating, and archiving data (reviewed in the book of Kellam & Thompson, 2016). In addition, researchers can rely on a broad variety of data repositories and data journals worldwide (Candela et al., 2015; Chavan & Penev, 2011; re3data, 2018). Recommendations for research data management, in particular data management plans, are available from various institutions (e.g., Science Europe, 2018). All researchers who start a publicly funded new research project today will come across certain requirements and guidelines on how to manage and publish their data from scratch. These measures help assure that all steps of the research data's life cycle are fulfilled – from planning research to sharing and (re)analyzing data (e.g., DataONE, 2019; DCC, 2008).

The special case of old research data

When it comes to old data from terminated research projects, however, the situation is different. These data are not in the focus of institutions or research funders. At the time when such data were created (and that period may be as short as 5 years back from now), often no data management requirements existed, such as the adherence to a data management plan. The lack of personal experience in data publishing, insufficient personal and monetary resources for data curation, unclear legal frameworks, or marginal scientific importance (small datasets, failed experiments, etc.) may be additional reasons for old research data remaining

unpublished. In addition, fast job rotations in science, especially at PhD and postdoc positions, impede proper research data management. Consequently, many old research datasets remain stored on internal institutional servers or personal devices. Monastersky (2013) reported on the example of an animal biomechanics research group at Johns Hopkins in Baltimore, USA, who aimed at reanalyzing 7-year-old data that were stored on a personal hard drive. It took them a several months to make sense of the data, in particular because multiple versions existed and metadata were of poor quality. Such old and privately stored research data are not or only hardly discoverable and accessible to the scientific community. Especially small datasets in the "long tail" of research data are not well maintained in many cases (Ferguson et al., 2014; Palmer et al., 2007). Consequently, old long tail datasets and associated knowledge often remain within research groups, or even disappear once the accountable persons leave the institution.

Compared to an ideal data life cycle, such as the one proposed by the UK Data Service (2019), unpublished data from old research projects only partly fulfill the proposed data management steps (Figure 2). An old research project was likely planned thoroughly, the data was collected, processed, and analyzed (steps 1–3). Then, however, the life cycle of old, unpublished research data was interrupted (red barrier). Since such data is not published, it cannot be shared, nor well preserved, nor reused (steps 4–6). This illustrates that old unpublished research data has to be considered as partly beyond the ideal data management processes that underlie current data management support services of academic libraries (e.g., DataONE, 2019; Van den Eynden et al., 2011).

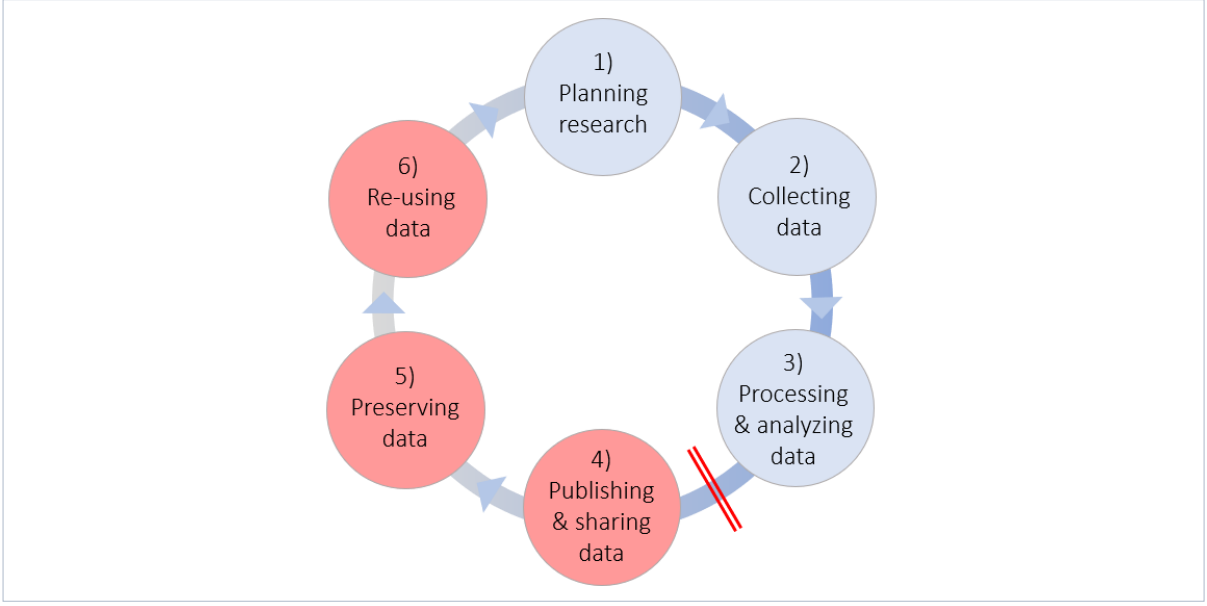


Figure 2: Research data life cycle from the UK Data Service (2019), adapted for the case of old unpublished research data. Completed data management steps for unpublished research data are presented in blue, uncompleted steps in red. A red barrier indicates where the life cycle of old unpublished research data is interrupted.

Growing desire for post-research data publication

Increasingly, also people responsible for such hidden research datasets are being sensitized to Open Science and becoming aware of the importance to openly sharing their data. Indeed, it seems logical that the Open Science movement not only encompasses current and new research projects, but also older, terminated research projects. Thereby, unpublished yet still valuable research data could be brought back into the circular data life cycle (Figure 2).

The publication of data after the termination of a research project is defined here as "post-research data publication". "Post-research" means some time – months to years – after the respective research article(s) had been published and after the activities of the research project had been terminated. Ideally, also the amount of open research data from old, terminated research projects will increase over time (Figure 1, blue dotted line). However, the percentage of old data being published post-research is likely to stagnate at a much lower level compared to the data from current and new research projects, for example because many datasets will disappear from the focus of interest after some time.

These considerations about post-research data publication were provoked by my own data management experiences from my PhD thesis in forest ecology, a large genecological study[1] (Frank, 2016). For this project, I had collected data on tree seedling growth and phenology, and on environmental variables describing the seedlings' origins. I used these data to compare three major tree species in Switzerland – Norway spruce, silver fir, and European beech – with respect to their risk of being poorly adapted to climate change. Together with my research colleagues, I published three papers based on these data (Frank et al., 2017a; 2017b; 2017c). In addition, two follow-up projects use(d) the data for their analyses (e.g., Frank et al., 2019).

Obviously, the data are valuable, not only because they cost a lot of time and money to collect, but also because such datasets generally are rare. Nevertheless, I did not publish the data, mainly because during my research project I was never asked – and certainly not forced – to do so. In 2016/2017, when I finished my PhD thesis and published the three papers, neither my funder nor my publishers required data publication. Furthermore, during the final stage of my PhD thesis I was pressed for time, which did not allow for additional data publication efforts (at least I thought so). Now, three years after the completion of my PhD thesis, I asked myself if and how data publication would still be possible. I believe that this is a question not only relevant to myself, but an issue that many other researchers would find relevant in regards to their former projects.

[1] Genecology denotes the study of genetic variation in relation to the environment (Aitken, 2004).

Post-research data publication: a field of activity for academic libraries?

In order to make old and hidden research datasets visible and FAIR (FORCE11, 2016), academic libraries could start providing support to facilitate post-research data management and publication. So far, many libraries offer a variety of services supporting the management of current research data (Pampel et al., 2010), but no guidelines, personal assistance, or other services were designed and offered specifically for the management of old research data. Yet, such services would help to further increase the amount of research data openly accessible (and reusable), and at the same time enhance the researchers' and institution's publication output and reputation. As Hofelich Mohr et al. (2014) stated, libraries can "offer specialized skills that are not present elsewhere on campus and are crucial to data management […]". Indeed, libraries "have a significant amount of trust capital" and "are particularly well positioned to fill the data management gap" (Monastersky, 2013). Support in post-research data publication might represent exactly such a data management gap. The provision of suitable library expertise, services, and infrastructure[2] are particularly relevant now and during the next couple of years (Figure 1, yellow area), which renders the topic a certain urgency.

Objectives of this thesis

This case study aims at establishing the basics needed for academic libraries to develop potential new services to assist post-research data publication. In particular, the thesis addresses four questions:

1. Is post-research data publication an issue for scientists? The hypothesis of this study is that considerable amounts of unpublished research data exist in the sciences and, thus, post-research data publication is of interest to many researchers.
2. What are the challenges of post-research data publication? The publication of research data imposes several challenges to researchers anyway, but is likely to be even more difficult if the datasets are old, i.e., belonging to terminated research projects.
3. What guidelines for researchers could facilitate post-research data publication? Existing recommendations on data management and publication focus mainly on current and new research data, but not on old data from terminated projects.
4. What expertise, services, and infrastructure could academic libraries offer to support post-research data publication? Libraries have already positioned themselves as experts and service providers in the field of data management. They might further expand their data service portfolios towards post-research data publication.

[2] The term *infrastructure* is used in this thesis for referring to a library's IT equipment, such as networks, databases, online platforms, or repositories, but not to its buildings or facilities.

2. Methods

A **survey** was conducted among researchers from the University of Bern to evaluate the relevance of post-research data publication in the sciences. The survey was created using the webservice SurveyMonkey (2018). It consisted of ten questions addressing the existence of unpublished data, the expected value of these data for post-research publication, the reasons for still publishing these data, and the need for support during the process of post-research data publication. In addition, the survey asked for the scientists' current positions, their interest in the survey's outcome, and their disposition to answer additional questions. Personal comments could be posted at the end of the survey. The questions were based on personal experiences and literature, e.g., the data service tasks listed by Mizzy and Hayslett (2014). One researcher tested the final questionnaire (see Appendix 1) to guarantee that it was working well in practice. An invitation message and a web link to access the survey were sent out on November 15, 2018, to 167 scientists from the Department of Biology and the Institute of Geological Sciences at the University of Bern. The recipients were mainly postdocs and researchers in higher scientific positions. Most likely, they had some experience in data management. The number of responses was stimulated by sending out one reminder email on December 3 and by personally contacting several researchers. The responses were manually extracted from SurveyMonkey and transferred to Excel for data visualization.

A **post-research data publication self-experiment** was conducted to explore the challenges of publishing old research data. For this purpose, typical unpublished long tail research data from the PhD thesis of Frank (2016), conducted in the field of forest ecology, was used. The intention was to share the data on a publicly available repository and to write a data paper about it. The procedure and challenges of this data publication experiment were documented.

Guidelines for researchers were derived from the data publication self-experiment. They addressed key steps along the process of post-research data publication where guidance appeared to be important. Each step was complemented by information about the protagonists of the respective task, i.e., whether the step involves mainly the researcher themselves, their (former) research colleagues, the university library, or external funders. The guidelines were kept short and simple to best appeal to researchers.

Suggestions for library expertise, services, and infrastructure were deduced from the survey, data publication self-experiment, and guidelines (sections 3.1 to 3.3). Several support options were proposed to assist post-research data publication along the publication process. The suggestions were directed towards academic libraries, proposing new ideas on how they may complement existing or planned services in research data management. The support options were kept at a general level due to the restricted length of this thesis.

3. Results and Discussion

3.1 Survey among scientists at the University of Bern

Respondents of the survey

The survey among biologists and geologist from the University of Bern generated 21 responses until December 21[3]. The respondents of the survey were mostly postdocs (10 mentions), followed by several principle investigators and senior researchers (4 mentions each), and professors (3 mentions; Figure 3). The remaining persons were master or PhD student, project leader, Oberassistent/in, lecturer, and/or group leader. Note that each respondent could declare several positions. No "other position" was mentioned.

This pattern of research positions can be explained by the fact that primarily researchers from postdoc level upwards were invited to participate in the survey. The idea of approaching these scientists was that they presumably had managed research data on their own, i.e., had enough personal experience to answer the survey. In general, postdocs as well as principle investigators, senior researchers, and professors regularly manage data. Therefore, they form target groups for research data publication support (see 3.4).

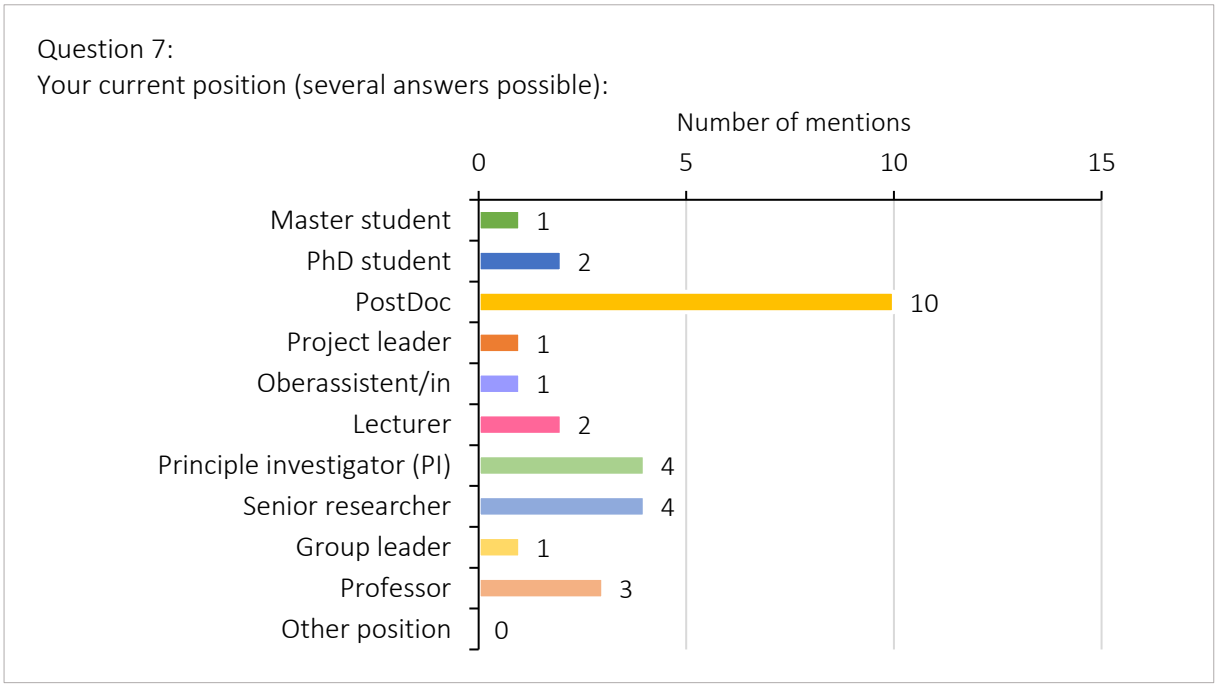


Figure 3: Summary of 21 answers given to Question 7 addressing the respondents' current position.

[3] Data available at https://drive.google.com/open?id=1utYj5GmzRQ5oE521KOkdGezpKCMUFaJ3

Unpublished research data in biology and geology at the University of Bern

Two thirds of all respondents knew of research data that they had created themselves and used for writing scholarly publications, but that they had not published as dataset(s) so far (Figure 4, A). Even 76% knew of unpublished research data within their research environment, i.e., their group, institute, or project (Figure 4, B). In contrast, 33% and 10% of all respondents said that they or their group did not create any unpublished research data (Figure 4, A and B).

This result supports the initial hypothesis that unpublished research data exist in the sciences. Although knowledge about the importance of Open Research Data is widespread by now (see also the responses given in Figure 6), the publication of research data seems not to have taken place routinely within the last years, leading to several old, unpublished datasets in the fields of biology and geology at the University of Bern.

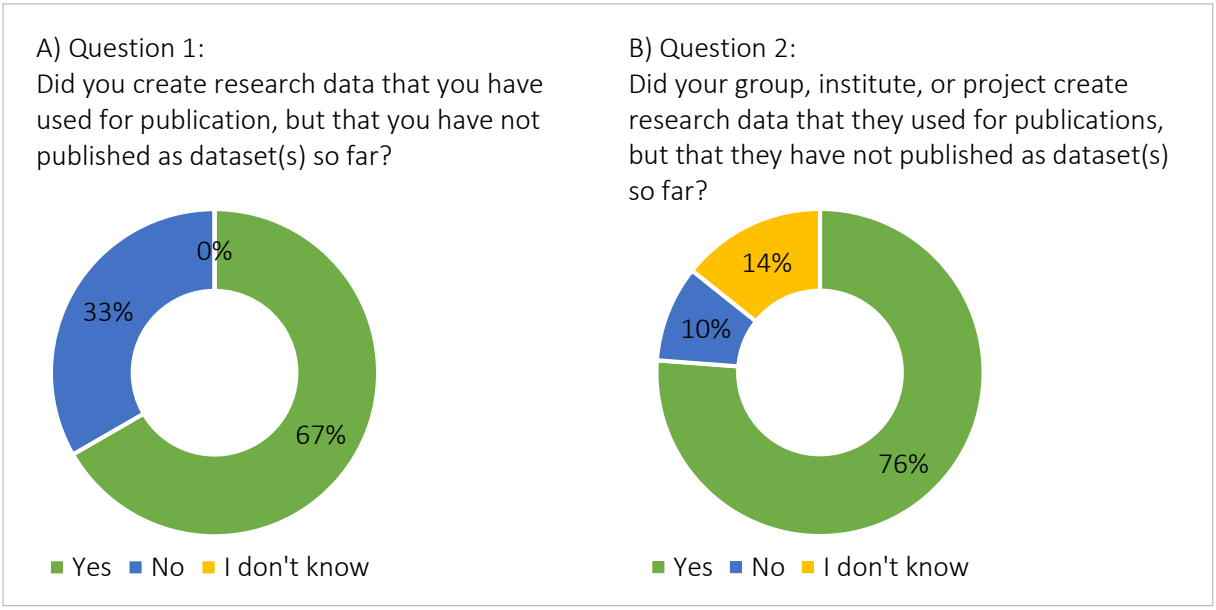


Figure 4: Summary of 21 responses given to Questions 1 and 2 of the survey addressing the existence of unpublished research data created by the respondent (A) and unpublished research data created by the respondents' research group, institute, or project (B).

Value of old research data for publishing

The majority of all respondents considered their own (11 responses) and/or their groups' (12 responses) unpublished data still worth publishing (Figure 5). Only two respondents did not consider their own or their groups' data valuable for post-research publication, and three did not know.

These results suggest that the majority of old and unpublished research data mentioned in the survey are not useless remains of former research activities. Instead, they represent valuable goods that scientists still consider worth for publication. Hence, these unpublished old data merit being addressed by initiatives that aim at fostering open research data.

Figure 5: Summary of 20 responses given to Question 3 addressing the value of unpublished data for post-research publication.

Reasons for post-research data publication

Almost all respondents (19 out of 20) considered the findability, accessibility, and reusability of research data the major reason for post-research data publication (Figure 6). In addition, seven respondents mentioned that it is important to make the data citable. Only four researchers considered the scholarly incentives of a data paper a relevant reason for post-research data publication. Additional reasons for post-research data publication were given by three researchers. They said that data publication was general knowledge and important to enable data reuse by other researchers. In addition, one person mentioned that the publication of research data should focus on data that support or refute a scientific hypothesis, rather than any research data.

The respondents of this survey agreed on the findability, accessibility, and reusability of research data as main reason for publishing old data. Therefore, the general principles of FAIR data management (Wilkinson et al., 2016) seem already well known among these scientists. Interestingly, the prospect of getting additional scientific reward through the publication of a data paper did not seem to greatly motivate the respondents to publish their old research data. Probably, they consider the citability of research data sufficient for acknowledging data publication, or they are not yet familiar with the publication format "data paper". Interesting was also the input of one researcher that not any data should be published, yet any data relevant for retracing scientific findings. This consideration does certainly help to narrow down the amount of data that merit publication, but may also restrain Open Science efforts by preventing certain data from publication and removing them from the pool of discoverable and reusable data.

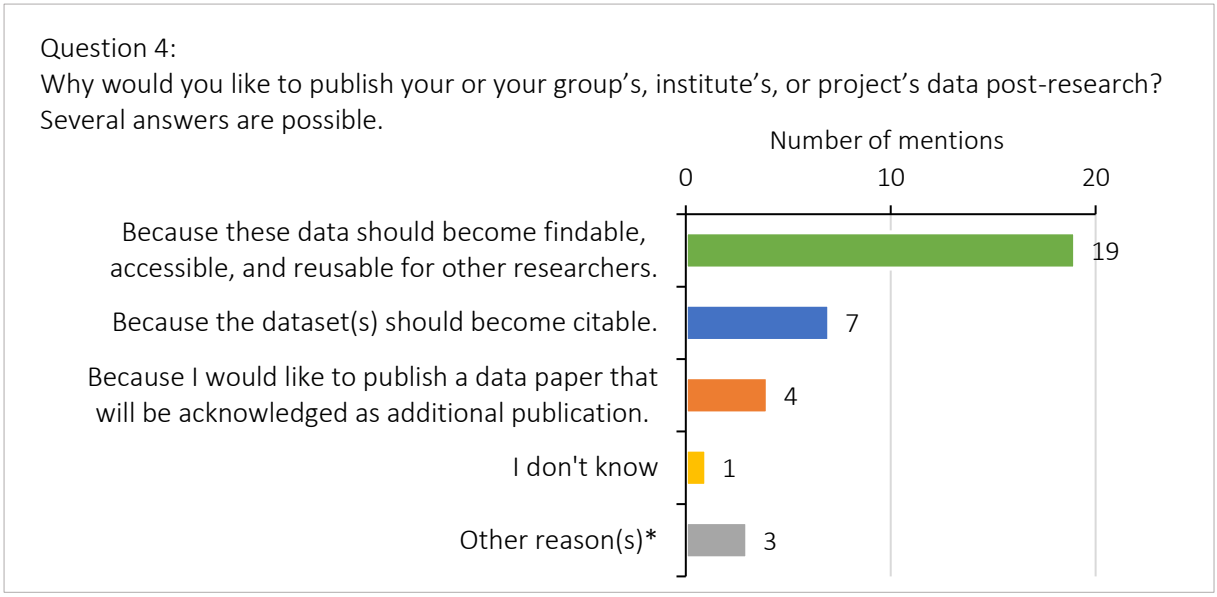


Figure 6: Summary of 20 responses given to Question 4 addressing the reasons for post-research data publication. *Other reasons for post-research data publication: making data available to the public is general knowledge; data publication allows others to build-up on already conducted work; published data should support or refute scientific hypotheses.

Estimation of support for post-research data publication

Approximately two thirds of all respondents said that they would appreciate to get support for post-research data publication (19% "Yes" and 48% "Probably yes"), whereas one fifth does not (14%) or probably not (5%) appreciate getting such support (Figure 7).

This result shows appreciation for support in post-research data publication among several biologists and geologists at the University of Bern. The most frequent answer was "Probably yes", which might signify that the respondents were generally positive with the idea to get help – but that they were not completely sure whether they really would need it, or what the support might comprise. Otherwise, a higher percentage of clear "Yes" might have resulted as answer to Question 5. Nevertheless, the positive attitude towards potential support services in post-research data publication shows that libraries might indeed step in here (see 3.4).

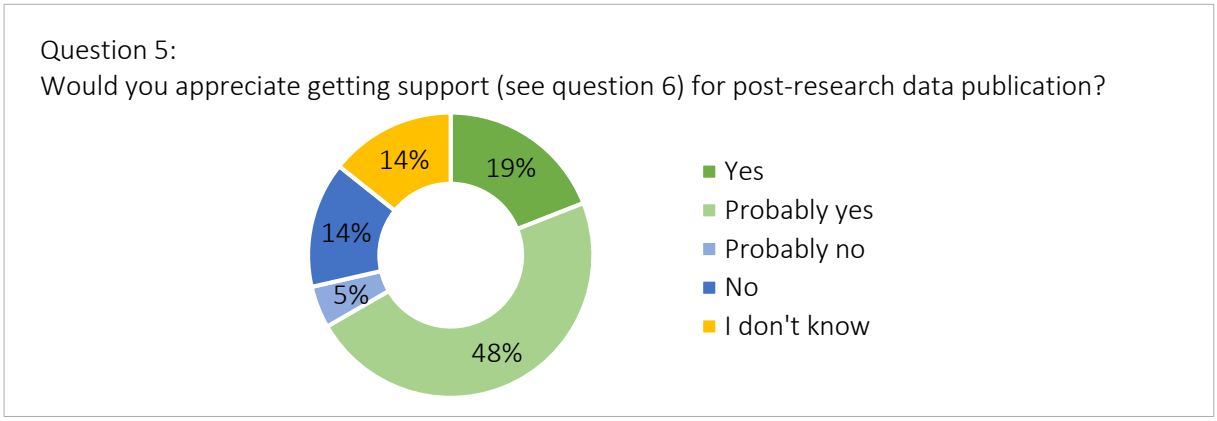


Figure 7: Summary of 21 responses given to Question 5 addressing the estimation of support.

Useful support during post-research data publication

Five types of potential support services in post-research data publication were chosen eight to ten times (Figure 8). These included advice about data publication options, guidelines, staff support for data preparation, staff support for metadata preparation, and information about legal aspects. In contrast, assistance during the data publication process was only mentioned four times. Two additional comments addressed the lack of time and money for data preparation, and the lack of a free database for data publication.

Based on these results, services designed to support post-research data publication should be diverse. Five options seemed to be equally useful and attractive to the respondents of Question 6. There was no single service that was required exclusively. Given that each research project and dataset is diverse, and that each scientist and research group works differently, it seems logical that also support services have to be diverse. Consequently, a variety of library services was outlined in part 3.4, designed to address all stages of the post-research data publication process.

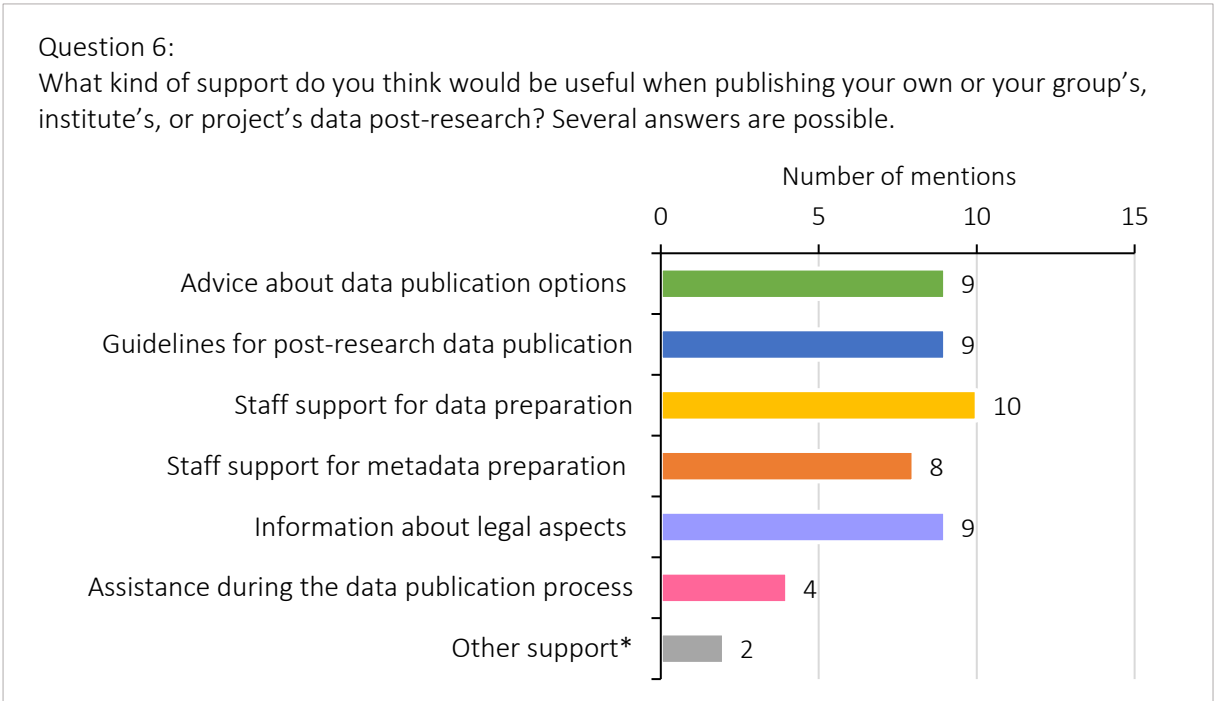


Figure 8: Summary of 20 responses given to Question 6 addressing the type of potential support for post-research data publication. *Two mentions within the category "Other support": need for time and money to prepare old datasets, need for a database where publishing data is free.

Limitations of this survey

The total number of 21 respondents render this survey a non-representative study. In addition, the results cannot be directly extrapolated to other scientific disciplines or institutions. Nevertheless, the survey's results allow us to get a first insight into the current situation of unpublished research data in biology and geology at the University of Bern. I assume that a similar picture would appear in other subjects and other universities. A follow-up study among researchers from different subjects and universities could be conducted to verify the results of this small survey in more depth.

The respondents likely consisted of researchers interested in post-research data publication; otherwise they may not have participated in the survey. These persons are more likely to be aware of old research data, which may have led to a higher number of positive answers to Questions 1 to 3 (Figures 4 and 5) – representing a potential bias in the survey's results towards an overestimation of existing and publishable old datasets.

Designed as introductory, but not major, part of this MAS thesis, the survey was kept short. A potential follow-up survey should be designed more broadly, not only to verify and precise the results of the present survey, but also to cover further aspects of post-research data publication. Additional questions might address the amount of time that researchers are willing to spend on post-research data publication, and the type of old data researchers consider worth for publication.

3.2 Post-research data publication self-experiment

Procedure

This post-research data publication experiment was started in May 2018. Since then, several steps have been accomplished, although the publication project was not terminated by the time of completing this thesis in March 2019. The progress was recorded in Table 1. Given the fact that this self-experiment represents a case study, the results have to be considered as such, i.e., as subjective insight into post-research data publication in the field of ecology.

Table 1: Data publication protocol for the procedure between May 2018 and March 2019.

Month & Year	Workflow	Effort	Time Used
Steps accomplished so far			
May 2018	Collect ideas, get familiar with the topic, conduct preliminary trial	Reading, browsing through own data, writing MAS "Leistungsnachweis 4"	8 days
June/July 2018	Approach former research colleagues, discuss potential post-research data publication	Writing emails	1–2 hours
Jan. 2019	Evaluate data repositories and data journals	Searching the internet	1 day
Jan. 2019	Decide on the amount of data, the data repository and data journal	Creating data overview, writing emails, meeting with former research colleagues	2 days
Jan. 2019	Get introduced to the data repository EnviDat (explained on page 16)	Meeting with repository manager	1 hour
Jan. 2019	Check requirements of former research funder and data providers	Writing emails, studying contract	3 hours
Jan. 2019	Check copyright aspects of data publication	Writing emails, studying publication license agreements	4 hours
Jan. 2019	Prepare data paper draft: Title, Authors, Abstract, Keywords, Introduction, Metadata (partly), and Acknowledgements	Writing, getting feedback from former research colleagues, polishing text	2 days
Jan./Feb. 2019	Prepare data and metadata for upload to data repository	Gathering and reviewing data files, adjusting files, completing and translating metadata	Several hours, in progress
Feb. 2019	Create dataset record in repository, write overall dataset description, discuss licensing	Establishing EnviDat profile, adding new dataset, describing project and data	4 hours
Next steps to approach			
	Complete data and metadata preparation, choose license		
	Upload (meta)data to repository		
	Complete data paper with metadata and submit paper		
	Publish data paper		

Since having left research two years ago, I kept the **idea** of still publishing data from my former PhD project (Frank, 2016). I used the "Leistungsnachweis" of MAS BIW module 4 in May 2018 as a preliminary trial for these data publication plans. The resulting essay represented a first step towards the publication of my PhD dataset (recently summarized in a blogpost; Frank, 2019). In summer 2018, I approached my former research colleagues with the idea of publishing our old research data. By email, we discussed the relevance and feasibility of doing so, and decided to try it. We already decided that once we take the effort of data revision and publication anyway, we would also try to publish a data paper with the aim of receiving scholarly perception and recognition.

For about half a year, my data publication self-experiment came to a halt, due to other projects and obligations. I resumed the proceedings at the beginning of January 2019 and went on by evaluating data repositories and, in particular, data journals suitable for the publication of data from the field of forest ecology. Together with my former research colleagues, we decided on the amount of data to publish, the data repository to share our data on, and the data journal to publish our future data paper in:

We defined the **amount of data** as the smallest number of data subsets that is needed to retrace and comprehend the genecological analyses conducted during my PhD studies (Figure 9). It should be possible for other researchers to use the data, e.g., for developing new genecological models, or for teaching purposes. We divided the data into four subsets, A–D, that cover all aspects of a genecological study, i.e., seedling phenotypes, maternal trees, populations and seed source locations, and test sites.

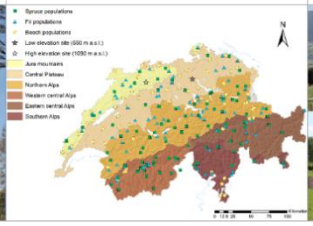
The genecological project ADAPT studied: - Quantitative genetics - Phenotypic plasticity - Adaptation / maladaptation to climate change		Species: Norway spruce (*Picea abies* [L.] Karst.) Sivler fir (*Abies alba* Mill.) European beech (*Fagus sylvatica* L.)	
A) Seedling data	B) Mother tree data	C) Seed source data	D) Test site data
phenotypes of 17,000 seedlings - *Growth* - *Phenology*	1-3 mother trees per population *Tree descriptors*	260 tree populations from across Switzerland - *Past climate* - *Soil*	2 test sites (common gardens) - *Temperature* - *Soil moisture*
			

Figure 9: Overview of data scheduled for post-research data publication self-experiment.

The choice of a **data repository** was not difficult given that well-maintained lists of repositories were available for consultation online (e.g., re3data, 2018). We chose EnviDat[4], the institutional data repository of the Swiss Federal Institute for Forest, Snow and Landscape Research WSL, the institute at which I had conducted my PhD thesis. This repository seemed to perfectly suit our needs, because it represents an institutional repository with a focus on environmental data and offers in-house assistance. I arranged a meeting with the repository manager and was introduced to the technical requirements and how to get started on EnviDat.

The choice of a suitable **data journal** was far more difficult compared to the choice of the data repository. No complete, updated, and well-curated list of peer-reviewed data journals in general, and in the field of ecology in particular, could be found on the internet. Consequently, gathering all information needed to find a suitable peer-reviewed data journal required several hours of investigation. First, I evaluated the journals I already knew from previous publications (Table 2, first three rows). Only one of these, *Ecology*, does in fact publish data papers. Second, I evaluated five journals from the field of ecology that publish data papers among other publication types (Table 2, rows 4–8). Finally, I also checked three interdisciplinary, pure data journals (Table 2, rows 9–11). For each journal, I evaluated if it was open access, published data papers, provided freely available abstracts online including links to the (meta)data, offered data paper templates, and collected article processing costs for data papers. We were looking for a subject-specific open access data journal with low APC – there was no money left from our former research project – and that provided templates for data papers. The resulting short, non-exhaustive list of available data journals for ecological data showed that such an ideal data journal presumably does not exist (Table 2). The subject-specific journals were not open access or required high APC for the publication of data papers. At second sight and after inquiries at the publishers' editorial offices it became clear, however, that although most subject-specific journals were subscription-based, their data paper abstracts and access links to data and metadata are available online to anyone. This does not represent genuine Open Access – the complete data paper including detailed descriptions may remain behind the paywall (e.g., in *Annals of Forest Science*) – but may be considered as "Virtual Open Access"[5]. Data paper templates were only provided by the interdisciplinary, pure data journals. The final choice was a compromise: We decided for *Ecology*, due to its subject-specificity, the "virtually open" availability of abstract and data online despite paywall, and because we had published the first of my PhD project's papers in this journal. At the same time, we accepted the APC of $250, lack of a data paper template, and absence of a practical, up-to-date metadata scheme.

[4] EnviDat: The Environmental Data Portal of the Swiss Federal Institute for Forest, Snow and Landscape Research WSL. https://www.wsl.ch/en/about-wsl/programmes-and-initiatives/envidat.html

[5] But careful: Although such "Virtual Open Access" enables anyone to read data paper key messages online and access the data, it does not meet the OA-requirements for projects funded by the Swiss National Science Foundation (SNSF). In addition, the SNSF does not sponsor OA publications in hybrid journals. (Tobias Philipp, SNSF, email from March 5, 2019)

Table 2: Overview of data journals considered for the publication of data from the field of ecology. OA: open access; DPs: data papers; APC: article processing charges; PPEES: Perspectives in Plant Ecology, Evolution and Systematics; BMC: BioMed Central. "s" refers to subject-specific journals, "i" to interdisciplinary journals.

Journal name	OA	Publication of DPs	Open abstract incl. link to (meta)data	Template for DPs	APC for DPs	Publisher	Focus
Ecology	No	Yes, among other publication types	Yes[a]	No	$250	Wiley	s
PPEES	No	No, but collaborates with Data in Brief	--	--	--	Elsevier	s
Global Change Biology	No	No	--	--	--	Wiley	s
Annals of Forest Science	No	Yes, among other publ. types	Yes[b]	Yes	No[b]	Springer	s
Ecological Research	No	Yes, among other publ. types	Yes[c]	No	No[c]	Wiley (from Vol. 34)	s
Global Ecology & Biogeography	No	Yes, among other publ. types	Yes[d]	No	No[d]	Wiley	s
BMC Ecology	Yes	(Yes), publication of database articles	Yes	No	$2,170	Springer Nature	s
BMC Plant Biology	Yes	(Yes), publication of database articles	Yes	No	$2,145	Springer Nature	s
Data in Brief	Yes	Yes, only data publications	Yes	Yes	$500	Elsevier	i
Scientific Data	Yes	Yes, DPs called "Data Descriptors"	Yes	Yes	€1,390	Springer Nature	i
Data	Yes	Yes, DPs called "Data Descriptors"	Yes	Yes	CHF1,000 after 6/30/19	MDPI	i

[a] Information provided by Anne Marie, esajournals@esa.org, on February 15, 2019.
[b] Information provided by Isabelle Fabrissin, Managing Editor, annforsci@inra.fr, on February 19, 2019.
[c] Information provided by Shoko Nakamura, Editorial Coordinator, ecores2@mail.esj.ne.jp, on February 20, 2019.
[d] Information provided by Ruth, geboffice@wiley.com, on February 18, 2019. Access to data paper abstract and (meta)data without subscription tested on February 23, 2019.

In addition, it was necessary to evaluate if there were **legal requirements** to data publication by our former research funder, the research program "Forests and Climate Change"[6]. Consulting the contract that existed between my PhD project ADAPT and this research program revealed no conflicts in data ownership and no restrictions to data publication. In addition, I verified that there were no conflicts in data ownership between our project and the providers

[6] Between 2009 and 2018, the research program "Forests and Climate Change" funded research projects to find out how Swiss forests will react to climate change. The program was a joint effort of the Federal Office for the Environment (FOEN) and WSL. https://www.wsl.ch/en/forest/forests-and-climate-change/research-programme-forests-and-climate-change.html

of soil and modelled climate data. Furthermore, I contacted a legal specialist[7] and asked him if there were copyright issues connected with the publication of my data. It became evident that the data – mostly numerical and categorical values – were not subject to copyright and, thus, could be published without restrictions. Indeed, the copyright transfer agreement between me as author of my research papers and the publishers Wiley and Elsevier did not contain any restrictions to publishing the data. Also, given that the data was neither sensitive nor confidential, no anonymization of data was needed in this case.

Now that all preconditions were clear – the amount of data defined, repository and data journal chosen, and legal matters clarified – I started to prepare the **data paper** following the data paper instructions of *Ecology* (ESA, 2018b). A data paper for *Ecology* has to be submitted according to the following structure: Title, Authors, Abstract, Keywords, Metadata, Acknowledgements, and Literature Cited. The final data paper, published in print and online, will contain only the parts Title to Keywords. Actual metadata and data files will be attached as supplementary material online. The ESA data paper instructions recommend to consult recent data papers published in *Ecology* for further orientation. These mostly contain an Introduction section in addition to the sections mentioned above (e.g., Bello *et al.*, 2017). By the end of January 2019, I completed all sections of the future data paper except the Metadata, which remained work in progress (see Appendix 2 for the data paper draft).

Finally, I started to prepare the **data and associated metadata** for upload to the repository and for integrating the metadata into the data paper draft. I followed the data paper instructions of *Ecology*, in short:

Data

- Logical and consistent formatting.
- Conversion of any proprietary data format to a plain text format.
- Submission of software as source code and as compiled (executable) code.
- Compression of multiple files as self-extracting .ZIP or .RAR archives.

Metadata

- Description of the content, context, quality, and structure of the data.
- Submission as single .DOC or .DOCX file.
- Adherence of metadata content to the standards from Michener et al. (1997).
- Adherence of metadata text to the instruction for ESA print journals.
- Inclusion of a minimum of statistical analyses, if needed.

[7] Dr. iur. Cyrus Beck, MAS, Stv. Leiter IK, Digitale Dienste & Entwicklung, Zentralbibliothek Zürich

The progress was documented in Table 3, structured according to the data subsets given in Figure 9. The table contains information about the steps needed to prepare the data – from original files (names, formats, data and metadata condition) to adjusted files (names, formats) – and includes an approximation of (meta)data preparation time. About half of the data subsets could be prepared for publication before March 2019.

Table 3: Insight into the data preparation process. Data subsets correspond to Figure 9. Columns 2–5 contain information about the original data files, i.e., names, formats, and whether data and metadata were already in publishable condition. Columns 6 and 7 contain information about new data file names and formats. In addition, approximate preparation time is given in column 8. "…" indicates that the (meta)data could not be prepared before the completion of this thesis in March 2019.

Data subset	Original data file names	Original formats	Original data OK?	Original metadata OK?	New data file names	New format	Prep. time (h)
A) Seedling data							
Spruce	Growth curve traits SPRUCE 09.12.2014	txt	yes	almost	adapt_growth_spruce_ 2013	csv and xlsx	1
Fir	Growth curve traits FIR 09.12.2014	txt	yes	almost	adapt_growth_fir_2013	csv and xlsx	1
	…	…	…	…	…	…	…
B) Mother tree data							
	Mutterbaumdaten gesamt	xlsx	no	no	adapt_mother_trees	csv and xlsx	2
C) Seed source data							
General data	Populationsdaten_ gesamt	xlsx	no	no	adapt_populations	csv and xlsx	2
Soil	Populationsdaten_ Bodenprofile	xlsx	no	no	adapt_seed_sources_ soil_pits	csv and xlsx	2.5
	Populationsdaten_ Bodenchemie	xlsx	no	no	adapt_seed_sources_ soil_chemistry	csv and xlsx	2
	Dokumentation_Boden -ansprache 2011_FiTaBu_final_ korrAF	pdf	yes, except naming	no	adapt_documentation_ soil_analyses_german	same as original	0.25
Past climate	adapt_hist_day	folder with txt	yes	no	adapt_clim_1931_1960 _day	same as original	1
	adapt_hist_swb_mon_ sv0_140123	folder with txt	yes	no	adapt_clim_1931_1960 _mon	same as original	0.5
	Klimadaten-Waldmodellierung_21_ W+K	pdf	yes, except naming	no	W+K_documentation_ climate_ modelling_german	same as original	0.25
D) Test site data							
	…	…	…	…	…	…	…

It became evident that the effort needed to prepare the data and metadata was much higher than expected. Most data files and metadata sheets were written in German, and not in English. At least column headers and metadata descriptions had to be translated into English in order to make the data understandable and reusable worldwide. In addition, it was sometimes difficult to determine the status in which the data should be published, i.e., as very first raw data after data collection or as corrected raw data together with the corresponding data preparation R-scripts. These considerations were time-consuming because it was important to make the right, scientifically sound decisions.

In order to be prepared for the data upload to EnviDat, I created a **dataset record on the repository**, which will later on function as container for the data (Figure 10). This record contains a general description of the research project and its dataset, including authors and **data sharing license**. I first considered to choose the Open Data Commons (ODC) Open Database License (ODbL; Open Data Commons, 2019), an open license for data resources, which was recommended by the EnviDat data manager. However, ODB licenses "apply only to sui generis database rights and any copyright in the database structure, they do not apply to the individual contents of the database" (Creative Commons, 2013). Consequently, the database contents may (or even have to?) be assigned individual licenses, such as Creative Commons (CC) licenses (Creative Commons, 2019). Alternatively, I could choose one of the open CC licenses, CC BY or CC BY-SA, for the whole dataset. Although the dataset record was uploaded to EnviDat, its properties, inclusive licenses, can still be changed. At the time of completing this thesis, the discussion on which license to choose was ongoing.

Figure 10: Screenshot of the EnviDat record created for the publication of the genecological dataset from the former research project ADAPT.

Altogether, I spent more than two weeks on this data publication self-experiment (Table 1) – and there are several more tasks to be completed before the data publication will be terminated. The next steps will be to make a decision on the licensing, complete the data and metadata preparation, upload all files and relevant information to the repository, and complete the data paper draft with all metadata needed. Finally, the data paper will be handed in to *Ecology* for peer review and (ideally) publication.

Challenges encountered

In retrospect, my attempt to publish old data from my former PhD project was rather naïve and uninformed. At the start, I knew little about the process of data publication in general and the challenges of post-research data publication in particular, even after the first trial in May 2018. Consequently, I spent a lot of time gathering information needed to fulfill the different steps of the data publication process. The definition of the final amount of data for publication, the evaluation of a suitable data journal, and the clarification of legal questions were particularly time-consuming. For example, I would have wished to have a curated list of suitable data journals at hand, or a legal counsel available to thoroughly answer my questions on data sharing licenses. Moreover, my approach to post-research data publication was, in retrospect, not systematic enough. From the start, I already planned to write a data paper and started to do so early, even before the data were prepared and available on the repository. This shortcut turned out to be unfavorable, because writing a data paper without exactly knowing the dataset's details is difficult.

My level of motivation for the data publication experiment changed frequently. From the start, I felt very much committed to share my data post-research, and I still do, mainly because I had once collected these data myself. Nevertheless, I frequently doubted whether this publication project would come to a successful end. In such moments, this MAS thesis itself was an incentive to continue.

Finally, time management was a challenge. I underestimated the time and effort for preparing the data and corresponding metadata. My initial dataset was not of bad quality, but most files were written in German, which required time for translation. In addition, certain information concerning the dataset content had not been noted and, thus, was forgotten over the years, which made it more difficult and laborious to compile the metadata. Furthermore, it was hard for me not to lose track of data publication, as this project was not part of my principal job. I worked on this self-experiment partly during my 20 % study time, and partly during my spare time. In addition, there were no deadlines, no schedule, no pressure. As a result, the publication proceeded more slowly that I had expected.

Positive experiences

It was enriching to collaborate again with my former research teammates. I experienced a lot of good will for this project. In addition, the support provided by the EnviDat repository manager, the friendly provision of legal advice by the expert from ZB Zurich, and discussions among colleagues at UB Bern were encouraging. Finally, this self-experiment allowed me to get a deep insight into the processes of data publication, which enhanced my personal expertise in scholarly publication. I concluded that investing time and effort in data publication does not only contribute to Open Science, but also to researchers' personal expertise.

Lessons learnt

Presumably, this data publication self-experiment proceeded very realistically, following a wavy rather than a straight way. This was not always comfortable to experience for me as data publisher. However, these very uncomfortable experiences are useful to build on, because they clearly showed where support during post-research data publication would be helpful:

First, information and expert support are particularly relevant to become familiar with the data publication process, to decide on publication options and legal matters, and to master technical aspects of data and metadata preparation and repository upload. Relevant sources of information should be well known to researchers and easily accessible. For example, guidelines covering the most relevant steps of post-research data publication would be useful. Furthermore, expert staff should be ready to provide advice and one-on-one support. In particular, thorough legal support is important, because most scientists are non-experts in legal matters and may feel hesitant when it comes to publishing data.

Secondly, incentives for researchers and their personal commitment to Open Science are essential preconditions for the publication of old scientific data; no funder, institution, or publisher specifically asks researchers to share their old data, contrary to data from new research projects. Incentives have to be strong to motivate researchers throughout the complete data publication process and to compensate for their effort.

Finally, only sufficient time (and money) allows scientists to realize a post-research data publication project outside regular work, as the costs involved are often not covered by the funders of terminated research projects (see page 24). A realistic idea of the effort needed to prepare and publish old datasets should be gained early during the publication process. Indeed, a dataset's deficiencies, if discovered too late, can significantly delay the data publication process – and may have a demotivating effect, too.

These findings were incorporated in the results of sections 3.3 and 3.4, in which guidelines and library services for supporting post-research data publication were proposed.

3.3 Guidelines on post-research data publication for researchers

Content of the guidelines

The guidelines derived from the data publication self-experiment (chapter 3.2) encompass eight key steps that researchers likely pass when approaching post-research data publication (Figure 11). Each step provides advice on what to do, and comprises questions and keywords as thought-provoking prompts.

Step 1 encourages researchers to think about the idea of post-research data publication before getting started. The reader is encouraged to learn about their peers' opinion and potential previous experiences. Thereby, a researcher will find out if their plan is supported by (former) colleagues and what kind of support they might expect from them.

Step 2 introduces academic libraries as key players in the post-research data publication process. Approaching the research institution's library at an early stage will help answer a researcher's questions and provide relevant information needed to successfully publish their data. The researcher will get to know their library as an information and service provider. The library, in turn, will be able to promote its support services (see also section 3.4) and, ideally, provide advice in post-research data publication from the start.

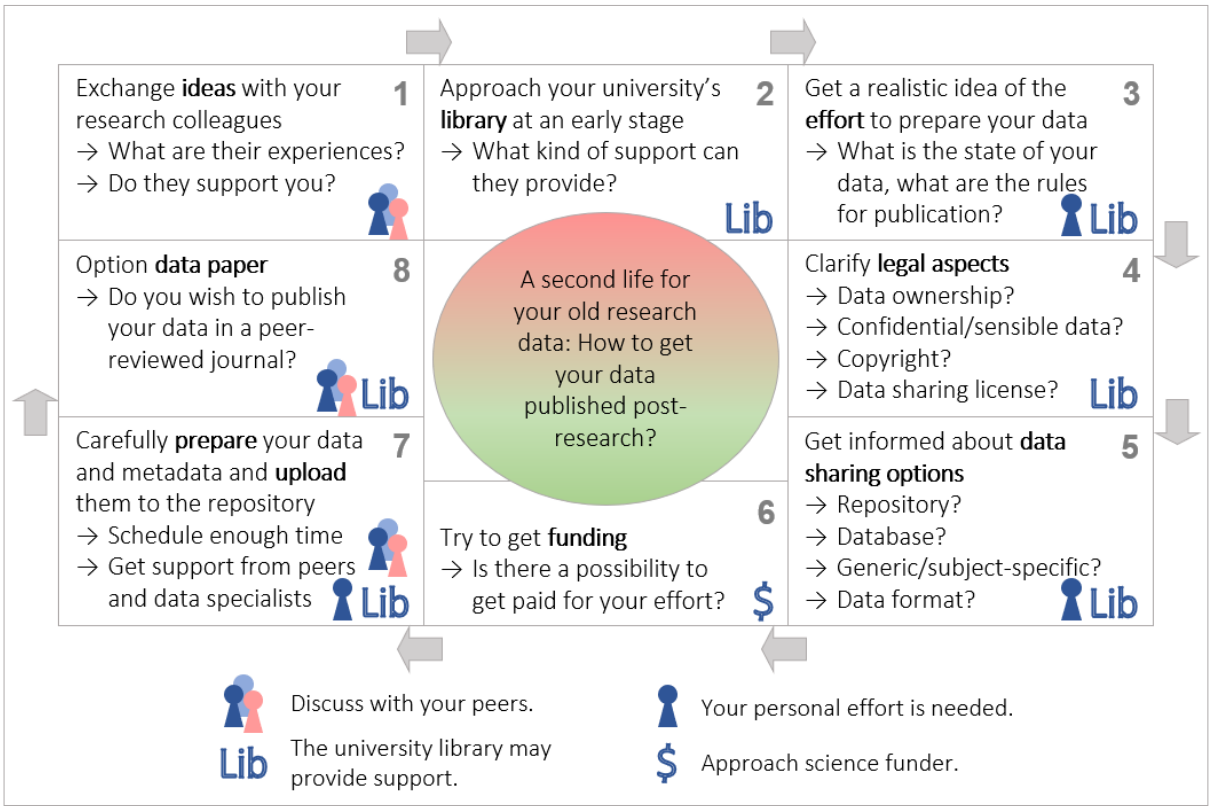


Figure 11: Suggestion of guidelines for researchers covering the process of post-research data publication in eight steps.

Step 3 invites researchers to thoroughly review the data they intend to publish and clarify what effort it will need to prepare both data and metadata for publication. Information about a dataset's weaknesses, such as poor documentation or non-English metadata, might be forgotten after some time. It is therefore crucial to first evaluate the data's current condition (format, documentation, metadata, etc.) and to learn about the rules and standards for publication. Here, both researchers and academic libraries are involved.

Step 4 includes the clarification of legal matters. Data ownership, copyright, confidentiality, sensible data, and licensing are issues that require legal considerations. These should be addressed early during the data publication process. Since only few researchers are legal experts themselves, academic libraries might provide critical advice, for example by a copyright librarian or legal counsel (Hofelich Mohr et al., 2014; see also 3.4).

Step 5 tackles the decision regarding which data repository and data formats to choose. Again, an academic library might provide advice here. This step focusses only on the path of data repository publication, or the combination of database and repository publication. In contrast, a so-called "enriched" publication where data is added as supplement to an article is likely not possible for old research data, assuming that the article(s) based on the dataset of interest had already been published. The option of a data paper is only raised in step 8, because also data papers mostly require the data to be published in a repository first.

Step 6 addresses the challenge of external funding. Researchers might want to publish data post-research, but they (and their institutes) may not have time and money to do so (which may also be the case with current research data; Tenopir et al., 2011). The publication of old research data is likely more time-consuming compared to current data, because essential knowledge about handling the data and the data's specifications may have been lost between the time of data collection and the time of post-research data publication. Consequently, researchers would appreciate being paid for the time they invest in publishing old research data (Figure 8). In addition, costs may arise from data archiving or data paper publication. It is now possible to get funding for the data publication of new research projects (e.g., SNSF, 2019), but the possibilities to get funding for data publication activities of terminated research projects are very limited[8]. Therefore, step 6 of the guidelines may be unpromising in many cases. Still, it will be important to keep it in the guidelines for raising awareness of this issue.

Step 7 includes the actual realization of data publication. This step involves the laborious preparation of the dataset and corresponding metadata. It is completed when the data and metadata are uploaded and openly available in the repository that was chosen in step 5. Time is a key factor during this step, but also appropriate scientific and technical support that might be provided by research colleagues and/or the university library. If the primary goal of post-

[8] It is possible, for example, to get funding by the SNSF for the publication of old research data within the context of an ongoing project (e.g., the continuation of a former project), but only if both projects are related with regards to contents (information by email from Cornelia Sommer, SNSF, on March 7, 2019).

research data publication is to make the data openly available and time is short, then the publication process may be terminated at step 7. However, this requires that the metadata comprise a thorough description of the data including contact details for further inquiries.

Step 8 addresses the potential supplemental step of publishing a data paper. A data paper can provide additional benefits to the data creator (see Introduction) and, therefore, represents an attractive additional step of post-research data publication – even if not standing out as primary publication reason in the survey (Figure 6).

Advantages of the guidelines

First, the presented guidelines have the potential to render researchers generally aware of the possibility to publish old research data. The guidelines are kept short, which makes them suitable for being used as a means of communication, for example in the form of a flyer, website, or similar. Such a flyer might attract researchers' attention and serve as starting point for discussions between scientists and research data specialists, in this case an academic library's staff. In order to enable communication, contact details of the library's data management team could be provided together with the guidelines. Ideally, the guidelines will be completed by further information and detailed descriptions of each data publication step, for example on the library's webpage.

Second, the guidelines potentially provide useful guidance to anyone interested in post-research data publication. In particular, the guidelines are suited to shortly answer the following questions:

- What is a meaningful way to proceed if someone plans to publish old research data?
- What are the tasks that have to be completed for post-research data publication?
- Who is involved in each step?

The guidelines propose how researchers may proceed in a chronological order, but can be altered as required. For example, the question of funding (step 6) may be clarified earlier during the process (see also Figure 12). In addition, each step does not only recommend what task to complete, but also names the major protagonists involved, i.e., the researcher themselves, their colleagues, the university library, or a science funder.

Finally, the guidelines promote academic libraries as fundamental players in the field of post-research data publication specifically, and in the field of research data curation generally. Several steps proposed in the guidelines indicate where a library might get active (Figure 11, steps marked with "Lib"). Researchers should perceive academic libraries as competent information and service providers in the field of research data management (Schiermeier 2018). Potential library roles in supporting post-research data publication are further discussed in chapter 3.4.

Comparison to existing guidelines for research data management

To my knowledge, these are the first guidelines designed to assist specifically the publication of old research data. Existing guidelines focus on the process of research data management in general, encompassing all steps of the research data life cycle (DataONE, 2019; ETH Library, 2016; Van den Eynden et al., 2011). Consequently, existing guidelines and best practices are generally broader. In addition, they also address early steps within the data life cycle, i.e., research planning, data collection, and data analysis (Figure 2), that altogether no longer apply to old research datasets; these steps were already completed earlier in terminated research projects. Furthermore, certain steps of post-research data publication might require more time and effort compared to current research projects, in particular data and metadata preparation. Therefore, time and effort were explicitly included in the guidelines as key factors during post-research data publication (Figure 11, steps 3 and 7), which seems not to be the case in general research data management recommendations.

Potential and challenges of the guidelines' practicability

I consider the presented guidelines to fit well into already designed and scheduled research data management support services of academic libraries. They represent a desirable supplement to existing support options. Guidelines for supporting post-research data publication were amongst the options that were most often named by biologists and geologist at the University of Bern (Figure 8). Designed as information material, the guidelines may be part of a library's basic data publication support (discussed in 3.4). Short and to the point, they will make it easier for users to assimilate the relevant points – and prevent them from being scared off by too much information. The implementation will first require some effort for data librarians to revise and adjust the guidelines to the library's actual services. Then, the guidelines will incur little direct costs, for example for printing and distributing flyers.

Nevertheless, the guidelines still represent an ideal and theoretical representation of post-research data publication. The following issues might be critical for their implementation:

- Post-research data publication is based on researchers' voluntary commitment. No one who decides to publish data after the termination of a research project is driven to do so by institutions or funders, which is a fundamental difference compared to current research projects that increasingly espouse strict open access standards. Consequently, the interest in considering and following the presented guidelines might be lower for post-research data publicists. Still, most biologists and geologists consulted in the survey (chapter 3.1) mentioned that they would appreciate getting support, for example guidelines (Figures 7 and 8).

- Academic libraries will have to get prepared for and be able to provide services as indicated in the guidelines. Consequently, library services for post-research data management should be implemented first, before the guidelines can be put into practice. As discussed in chapter 3.4, several basic informative services will likely not require much more than what academic libraries in Switzerland already do in supporting research data management (e.g., ETH Library, 2017; Universität Bern, 2019b).
- Indirect costs will appear through the involvement of expert staff, i.e., research data specialists and legal specialists, if not yet present in an academic library's data management team (see also 3.4).
- Funding of post-research data publication (Figure 11, step 6) will remain difficult to achieve for researchers as long as research funders do not have this kind of data publication efforts on their agenda. Nevertheless, funders might be interested in seeing old research data published, in particular if they had already supported the corresponding projects earlier on. Hence, it might be worth approaching science funders, such as the Swiss National Science Foundation SNSF or Horizon 2020, and make them aware of the issue.
- Implementing the guidelines among *former* researchers will be difficult. Anyone who has left science has also left the area of influence of their former university library. In contrast, current researchers represent a main target group of academic libraries. This facilitates the distribution of information between library and researchers, and legitimates potential effort and money that an academic library spends on potential post-research data publication services. Consequently, the guidelines will likely have to address *present* researchers, even if the issue also concerns former researchers.

Finally, these guidelines are based on my personal, subjective experiences (see 3.2). This means that they are not validated by now. As a next step, I suggest to discuss the guidelines with experts, both researchers and data management specialists at academic libraries, in order to evaluate their applicability in practice.

3.4 Suggestions for library expertise, services, and infrastructure

Introduction to data publication support model

The support of research data management and open access publication has evolved to a substantial part of academic libraries' service portfolio (Kellam & Thompson, 2016; Pampel et al., 2010). By now, the major university libraries in Switzerland have established specific units, called for example "Digital Curation Team" (ETH), "Research Data Team" (EPFL), or "Open Science Team" (Bern). Their existing services could now be complemented by the suggestions for post-research data publication support outlined below.

Based on the results of the first three parts of this thesis, a model for library expertise, services, and infrastructure was established (Figure 12). This model shows where academic libraries could step in to support the publication of old, yet still valuable research data. The 15 proposed support options are presented in chronological order. They follow the process of post-research data publication – from hidden (red) to open (light green) and peer-reviewed open old research data (dark green) – and are directly linked to the steps that researchers encompass during post-research data publication. The suggestions were divided into options that specifically refer to the publication of old research data (solid grey rectangle on top), and options that generally address the publication of research data, i.e., from both terminated and current projects (dashed grey rectangle below). The support options are presented in colored boxes referring to their proposed support level, i.e., basic (yellow), advanced (orange), and expert (blue). These levels were introduced to account for differing magnitudes of resources needed for implementation (money, personnel, time) that exist among the proposed support options.

Basic support

Basic support refers primarily to the distribution of information concerning data publication (Figure 12, yellow boxes). Such basic support services were proposed for all major steps of publishing data in order to supply researchers with information throughout the complete process. Researchers should know from the start about existing options and requirements for (post-research) data publication, useful guidelines, and existing library services and infrastructure (options no. 1–3). Information and advice should also be provided concerning legal matters (no. 4), data repositories and journals (no. 5 and 11), (meta)data preparation (no. 6), and upload to repositories (no. 8).

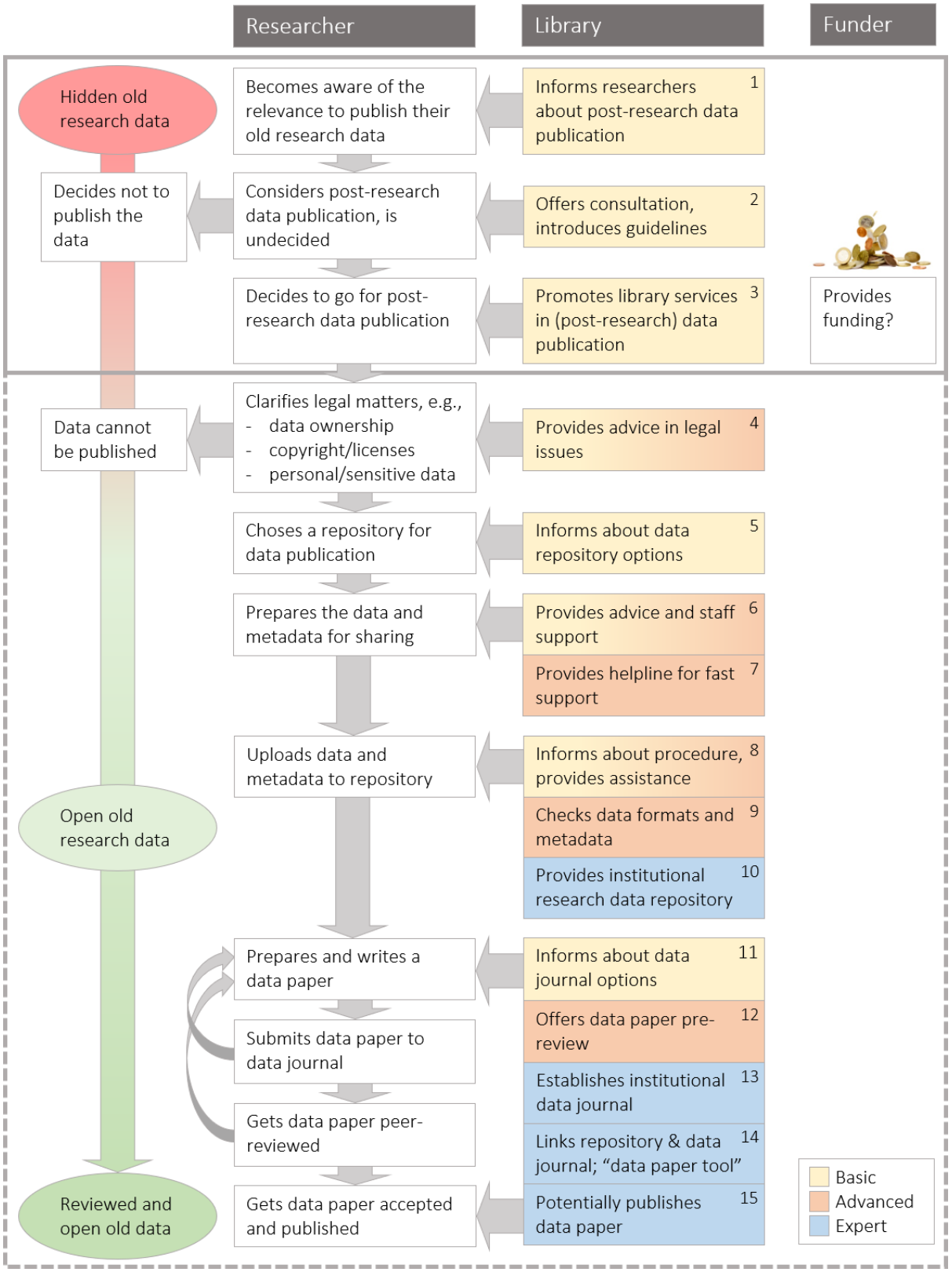


Figure 12: Model for library expertise, services, and infrastructure along the process of data publication from hidden (red) to open and reviewed old research data (green). The support options are classified according to three service levels: basic (yellow), advanced (orange), and expert (blue). The solid grey rectangle refers to post-research data publication in particular; the dashed grey rectangle refers to data publication in general.

The costs for providing basic support are expected to be low. Given that many academic libraries already offer basic services in research data management, such as data skills training and consultancy, or management and sharing plans (Whyte, 2014), basic support can be provided without much additional effort. In most cases, current library staff will be able to resume the additional function of informing researchers about various aspects of (post-research) data publication straight away – or may be trained for this role. Staff needed to fulfill basic support includes information and data specialists (referred to here as "data librarians"; Mizzy & Hayslett, 2014) and subject librarians, with both job categories already being present in many academic libraries. Data librarians might form the core team responsible for research data support, whereas subject librarians might form the library's "outposts" at the border between library and institutes. A similar model is currently implemented at the University Library of Bern (personal communication by Anna Keller, UB Bern).

Advanced support

This second support level refers to the close **assistance** of researchers by library staff during the research data publication process (Figure 12, orange boxes). Assistance at this level will be provided by email, phone, and one-on-one consultations. It includes the provision of advice in legal questions (no. 4), the support of (meta)data preparation (no. 6), and the assistance during (meta)data upload to the repository (no. 8). One specific idea presented here is the establishment of a "helpline", i.e., a specific call number or internet chat application with service hours lasting far into the evening, which would allow researchers to get fast remote support when they encounter problems (no. 7). Critical steps are, for example, metadata preparation or data upload to the repository. An additional service will be the verification of data formats and metadata at the time of data upload to the repository (no. 9). This step to oversee the quality of (meta)data input will be especially crucial for libraries that run their own data repositories. Such a check could also be provided for external repositories that may not validate metadata quality in detail. In addition, library staff could offer the revision of data paper drafts on demand, called here "data paper pre-review" (no. 12). This service would allow researchers to get thorough feedback on their data paper draft before handing it in to a journal for peer review.

In terms of resources, this second level of support requires more specialized library personnel. Therefore, advanced support will be more expensive for academic libraries to implement compared to basic support, since they likely have to hire additional staff. The corresponding job profiles include librarians for general advice in data management, law experts for in-depth support in legal matters, and IT specialists for support concerning data repositories, data upload, and IT-related technical matters. In fact, new roles and professional profiles in data curation at academic libraries are currently being discussed and developed, such as "data librarian", "data steward", or "data archivist" (Lyon, 2014). Pampel et al. (2010; after Corrall, 2008), for example, characterized "data librarians" as experts working at the intersection of

context, content, and IT-related research support. There is, however, no consensus on the exact definition of a "data librarian" (Mizzy & Hayslett, 2014). Instead, each library will have to define its particular data librarian – and legal counsel, IT specialist, and other – profiles.

Expert support

Finally, expert support involves the provision of infrastructure that facilitates the publication of research data (Figure 12, blue boxes). First, this includes the establishment of an institutional research data repository (no. 10), as already planned or implemented by several academic libraries in Switzerland (e.g., ETH Library, 2017; Universität Bern, 2019b; WSL, 2019). An institutional repository provides long-term storage and access to research data of various kinds. Thereby, data curation remains under the control of the institution itself, long-term operation is more likely to be assured than in non-institutional repositories, and support for users can be provided in-house.

A second potential expert support option is the establishment of a pure, institutional data journal (Figure 12, no. 13 and 15). Compared to mixed journals, pure data journals are not numerous today (Candela et al., 2015). Given that the interest and effort in research data publication will likely increase in the next years, more pure data journals might be needed in the future. Taking the role of a data journal may be new to academic libraries, although many of them have already launched online publishing programs and are familiar with scholarly publishing (ACRL, 2016; Hswe, 2015). The Library Publishing Coalition (LPC), for example, represents an association of academic libraries engaged in scholarly publishing, which "extends the impact and sustainability of library publishing and open scholarship by providing a professional forum for developing best practices and shared expertise" (Educopia Institute, 2019). An example for a single academic library publishing initiative in Switzerland is Bern Open Publishing (BOP), a publication platform for reviewed open access journals and periodicals run by the University Library Bern (Universität Bern, 2019a). Similarly, the foundation of an institutional data journal could be accomplished by a single library or by a network of libraries, potentially at the national level. The latter seems more promising, because establishing and operating a new journal in the long term will likely require investment from more than one institution (see below). The idea of a "Swiss Open Access platform for data papers" has already been proposed by Schneider and Prongué (2015), who identified the "need of a consortial activity in Switzerland towards the creation of a data paper platform for the long tail of research data". In addition, the formation of a "Swiss Data Journal" has been promoted (Zwahlen & Schneider, 2016). Both ideas, however, have not been put into practice so far.

The establishment of a data journal requires major investments in both IT infrastructure and personnel. Information and IT specialists will be needed as well as editors and an advisory panel that guide the policies, standards, and editorial scope of the data journal, and an editorial board that oversees the peer review process (e.g., Springer Nature Publishing, 2019). For the review of data papers, it is useful to distinguish between technical and scientific review

(Callaghan et al., 2012). Technical review – the revision of metadata completeness and data file formats – can be accomplished by specialized library staff and may be allocated to the institutional data repository. Scientific review, however, requires domain expertise to survey the data's collection methods and reuse potential, and will require the collaboration and commitment of researchers. Importantly, the functioning of a data journal involves the participation of scientists in different roles (Zwahlen & Schneider, 2016) – as data paper authors, reviewers, and members of the editorial and advisory boards. This has to be considered carefully, because researchers are already overloaded with reviewing articles and may not be receptive to participate in a new (data) journal (Diederich, 2013).

The third proposed expert support option (Figure 12, no. 14) comprises the creation of an interactive "data paper tool" that links the institutional data repository (no. 10) with the data journal (no. 13). The advantage of such a tool is simple: as soon as a dataset and its description are available on the repository, it allows to automatically extract metadata from the repository and to compile these to a data paper – all at the touch of a button. For example, the Global Biodiversity Information Facility (GBIF) has, together with Pensoft Publishers, developed a data paper tool as part of their Integrated Publishing Toolkit (GBIF, 2019). They advertise their data paper tool as follows: "At a click of a button you can download the metadata as an RTF-formatted manuscript ready for editing and submission for peer review". This idea appears convincing, given that the compilation of a data paper from scratch is time-intensive (see 3.2). The approach of a data paper tool could even be expanded to a web-based platform where "the entire process from collaborative writing, commenting, and editing, over submission and peer review, to publication and post-publication revising is handled" (Pensoft, 2019).

Not included in the library support model – but listed on the chart anyway – is the question of external funding of post-research data publication (Figure 12, top right). Presented at the step where researchers decide to publish their old data (or rather not), this indicates that funding could essentially contribute to the success of post-research data publication (see also the need for money expressed in Figure 8). The challenge of achieving external funding is discussed in section 3.3.

What is new? Comparison to existing data management services of academic libraries

A good deal of the proposed library expertise, services, and infrastructure is already included in existing or planned data management services of (intermediate to larger) academic libraries, inclusive the establishment of institutional research data repositories (e.g., ETH Library, 2017; Universität Bern, 2019b; WSL, 2019). Nevertheless, all three suggested support levels comprise fresh, innovative ideas that may further promote academic libraries as key players in the field of research data publication:

The specific promotion of post-research data publication as opportunity to make old research data discoverable and reusable (Figure 12, no. 1–3) is new and, to my knowledge, not yet covered by existing academic library expertise and services in Switzerland. The guidelines on post-research data publication for researchers (Figure 12, no. 2; chapter 3.3) represent a completely new service within the basic support level. Therefore, the informative support options at the basic level concerning specifically the publication of old research data represent fresh ideas that could be tested and possibly implemented in the near future – especially because the consulted biologists and geologists at the University Bern expressed appreciation and interest for such services (see chapter 3.1). Due its challenges, the publication of data post-research will likely remain a sidetrack of data publication efforts. Nevertheless, potential support services should be initiated now, simply because the interest in old datasets usually declines with time.

At the advanced support level, the focus on providing more and in-depth legal advice may be rather new to academic libraries (Figure 12, no. 4). Law experts are often present at a university library, for example in the position of a subject librarian, in particular if the library comprises a law section. However, this person is not necessarily part of the university library's research data management support team and, thus, usually not in charge of answering researchers' questions on legal aspects of data publication. An official "law counsel" or "copyright librarian" might take on this role instead (Hofelich Mohr et al., 2014). Further new advanced library services include the provision of instant support through a helpline or internet chat application, and the pre-review of data papers (no. 7 and 12). Similar services already exist, such as the Online Chat provided by the library catalogue swissbib Basel Bern (Universität Basel, 2019), or the data management plan review service by the UB Bern Open Science Team (Universität Bern, 2019b). Yet, no similar support options exist for the purposes described here, i.e., instant-support and paper revision in the context of research data publication.

At the expert support level, the establishment of an institutional data journal and the creation of an interactive data paper tool (Figure 12, no. 13–15) represent new, visionary, and far-reaching thoughts. It may be too early, too expensive, or too complex to approach these ideas right now. In addition, these services may only pay off if several institutions collaborate – in the sense of a national data publication platform (Schneider & Prongué, 2015; Zwahlen & Schneider, 2016). These ideas will have to be discussed on a strategic level first, before their implementation will come to the fore.

Practicability of the proposed library support options

Most of the library expertise, services, and infrastructure for post-research data publication presented here do not specifically refer to old research data only, but are generally applicable to the process of scholarly data publication. Therefore, the results of this thesis are suitable not only to complement existing library services towards the integration of specific services for post-research data publication, but also to support data publication in general. The proposed

options may serve as starting point for expanding existing and planned data publication services in academic libraries – and may be of broad interest in the field of academic data publishing. However, similar to the guidelines presented in section 3.3, the proposed library expertise, services, and infrastructure will have to be discussed and validated together with experts from academic libraries, research, and – in the case of an institutional data journal – also with expert scholarly publishers.

The implementation of the suggested support options very much depends on academic libraries' budgets designated to research data curation activities. Pampel and Dallmeier-Tiessen (2014) stated that "a professionalization of the Research Data Management, which supports scientists in the sharing of their data, is necessary to ensure the permanent accessibility […]. In this context, priority must be given to the structuring and networking of the research data repositories and their long-term financing". It appears that library as well as research budgets are generally tight and may at first not leave much freedom to develop new services (Monastersky, 2013). Nevertheless, provided that the Open Science movement has reached top priority at institutional and political levels by now, the prospects to achieve additional funding in the future seem promising. This is illustrated, for example, by the current trend of new institutional data repositories being initiated at academic institutions and libraries (e.g., BORIS Research Data; Universität Bern, 2019b).

Another critical aspect will be the collaboration with the scientific community. As outlined above, researchers will be the key players in the process of (post-research) data publication (Figure 12). It will be only their interest in publishing reach data, their need to get support, and their willingness to collaborate with their library that will drive the expansion of data management services in academic libraries. Consequently, the success of Open Science initiatives very much depends on the contribution of scholarly societies (Pampel & Dallmeier-Tiessen, 2014). If this precondition is not fulfilled, academic libraries will have a much harder job in supporting research data management, and may lack the justification of doing so. In the case of the idea to promote an institutional data journal or even a "data paper tool" as shown above, the question remains whether such services would attract enough interest by researchers and meet real needs (compare, for example, Figures 6 and 8) – and not just represent innovative ideas from the perspective of libraries as service providers. Thus, it will be essential for academic libraries to closely connect to their institution's researchers, to make themselves visible as competent data management service providers, but also to carefully evaluate the need for support services. This will only be possible if academic libraries interact with researchers on an equal footing. Consequently, the relevance of subject librarians – and, more generally, scientists with research experience working in academic libraries – as bridge-builders between libraries and the scientific community will increase and be a key premise to foster Open Science.

4. Conclusions

The publication of data "post-research", i.e., after the termination of a research project, represents an issue for scientists. Several biologists and geologists at the University of Bern are aware of unpublished research data and consider part of this data still worth sharing. Their interest in the publication of old data and suitable support services represent a potential new opportunity for Open Science initiatives.

In practice, the publication of data post-research requires substantial amounts of information, personal commitment, and resources (in particular time). All three preconditions have to be fulfilled for successful post-research data publication (Figure 13).

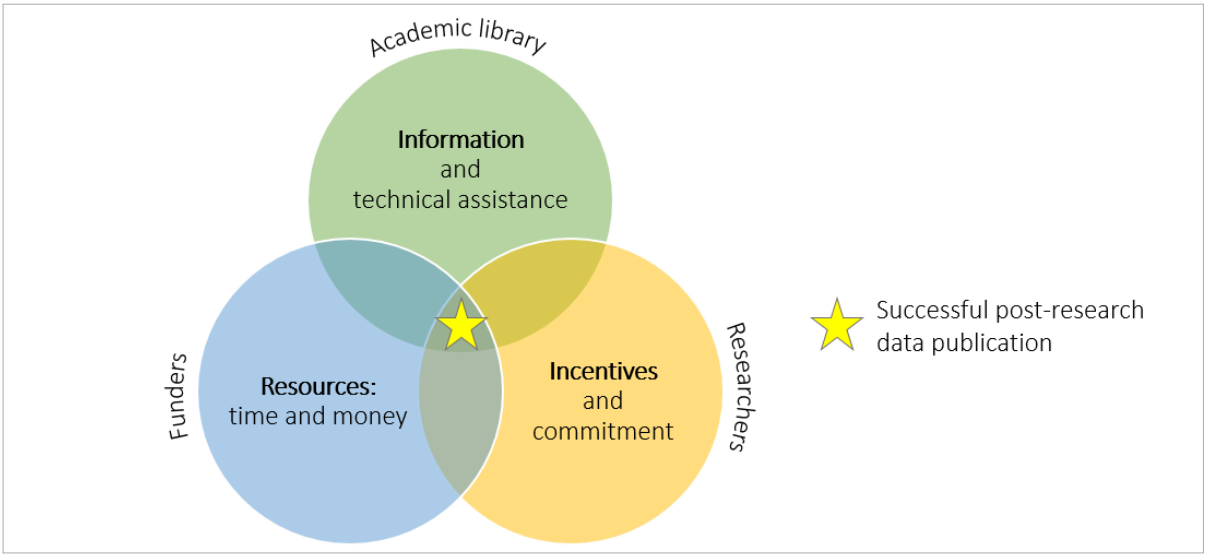


Figure 13: Preconditions for successful post-research data publication.

Guidelines for researchers should address all major steps of post-research data publication in a short and informative manner. The proposed guidelines – ideally complemented by expert staff support – represent a new tool for academic libraries to promote and assist the publication of old research data.

Academic libraries could specifically contribute to information supply and technical assistance (Figure 13, green), besides highlighting incentives for publishing old data when dealing with researchers (yellow). Basic services, such as guidelines, will be simple to integrate into existing or planned data management portfolios. In addition, libraries may further invest in advanced and expert support to foster data publication in general, involving specialized library staff and tailored IT infrastructure.

In summary, a diversification and expansion of library services is recommended to promote the sharing of both old and new research data. Innovative data publication support may start with providing information, but may lead as far as turning libraries into data journal publishers. The extent of these measures will likely depend on scientists' demands and university policies.

Acknowledgements

This MAS thesis was supported by several persons. I wish to thank René Schneider and Dirk Verdicchio for their supervision, my former research colleagues at WSL, Christoph Sperisen, Peter Brang, and Caroline Heiri, for their support in preparing the data paper, Cyrus Beck for legal advice, Ionut Iosifescu Enescu for assistance with the repository EnviDat, Gero Schreier, Anna Keller, Jennifer Morger, and Jonas Stillhard for stimulating discussions, Denis Maier for citation style adjustments, and my husband Martin Salzmann for his encouragement throughout the writing process.

Literature

ACRL (2016). Scholarly communication toolkit: library publishing programs. Retrieved on February 22, 2019, from https://acrl.libguides.com/scholcomm/toolkit/librarypublishing

Aitken, S. N. (2004). Genetics and genetic resources: genecology and adaptation of forest trees. In J. Burley (Ed.), *Encyclopedia of Forest Sciences* (pp. 197–204). Oxford, UK: Elsevier.

Bello, C., Galetti, M., Montan, D., Pizo, M. A., Mariguela, T. C., Culot, L., … Jordano, P. (2017). Atlantic frugivory: a plant–frugivore interaction data set for the Atlantic Forest. *Ecology*, *98*(6), 1729–1729. http://doi.org/10.1002/ecy.1818

Callaghan, S., Donegan, S., Pepler, S., Thorley, M., Cunningham, N., Kirsch, P., … Wright, D. (2012). Making data a first class scientific output: data citation and publication by NERC's environmental data centres. *International Journal of Digital Curation*, *7*(1), 107–113. http://doi.org/10.2218/ijdc.v7i1.218

Candela, L., Castelli, D., Manghi, P. & Tani, A. (2015). Data journals: a survey. *Journal of the Association for Information Science and Technology*, *66*(9), 1747–1762. http://doi.org/10.1002/asi.23358

Chavan, V. & Penev, L. (2011). The data paper: a mechanism to incentivize data publishing in biodiversity science. *BMC Bioinformatics*, *12*(Suppl 15), S2. http://doi.org/10.1186/1471-2105-12-S15-S2

Collins, S., Genova, F., Harrower, N., Hodson, S., Jones, S., Laaksonen, L., … Wittenburg, P. (2018). Turning FAIR data into reality: interim report from the European Commission Expert Group on FAIR data. http://doi.org/10.5281/zenodo.1285272

Corrall, S. (2008). Research data management: professional education and training perspectives (presentation). *Research Data Management Forum*. Manchester, UK. http://www.dcc.ac.uk/sites/default/files/documents/RDMF/RDMF2/07 Corrall.pdf

Creative Commons (2013). Data. Retrieved on March 3, 2019, from https://wiki.creativecommons.org/wiki/Data

Creative Commons (2019). Share your work. Retrieved on February 23, 2019, from https://creativecommons.org/share-your-work

DataONE (2019). Best Practices. Retrieved on February 15, 2019, from https://www.dataone.org/best-practices

DCC (2008). DCC Curation Lifecycle Model. Retrieved on March 3, 2019, from http://www.dcc.ac.uk/resources/curation-lifecycle-model

Diederich, F. (2013). Are we refereeing ourselves to death? The peer-review system at its limit. *Angewandte Chemie International Edition*, *52*(52), 13828–13829. http://doi.org/10.1002/anie.201308804

Educopia Institute (2019). Library Publishing Coalition. Retrieved on February 22, 2019, from https://librarypublishing.org

ESA (2019a). Ecological applications: data policy. Retrieved on February 14, 2019, from https://esajournals.onlinelibrary.wiley.com/hub/journal/19395582/resources/data-policy-eap

ESA (2019b). Ecology: data paper instructions. Retrieved on February 14, 2019, from https://esajournals.onlinelibrary.wiley.com/hub/journal/19399170/resources/data_paper_inst_ecy

ETH Library (2016). Data management checklist. Retrieved on February 17, 2019, from https://documentation.library.ethz.ch/pages/viewpage.action?pageId=6685357&preview=/6685357/6685350/Data Management Checklist.pdf

ETH Library (2017). ETH Zurich Research Collection: repository for publications and research data. Retrieved on March 3, 2019, from https://www.research-collection.ethz.ch

Ferguson, A. R., Nielson, J. L., Cragin, M. H., Bandrowski, A. E. & Martone, M. E. (2014). Big data from small data: data-sharing in the “long tail” of neuroscience. *Nature Neuroscience*, *17*(11), 1442–1447. http://doi.org/10.1038/nn.3838

FORCE11 (2016). FAIR Data Principles. Retrieved on February 14, 2019, from https://www.force11.org/group/fairgroup/fairprinciples

FOSTER (2019a). Open Science. Retrieved on February 5, 2019, from https://www.fosteropenscience.eu/taxonomy/term/7

FOSTER (2019b). Open Science Definition. Retrieved on February 5, 2019, from https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition

Frank, A. (2019). A second life for old research data: challenges of post-research data publication. In *Universitätsbibliothek Bern: Bern Open Science Blog*. Retrieved from https://biblio.unibe.ch/web-apps/wkblog/2019/03/second-life-data

Frank, A. D. (2016). Genecology of Norway spruce , silver fir , and European beech in Switzerland: Are current populations adapted to future climates? ETH Dissertation Nr. 23644. Zürich: ETH. https://doi.org/10.3929/ethz-a-010878283

Frank, A., Heiri, C. & Kupferschmid, A. D. (2019). Growth and quality of *Fagus sylvatica* saplings depend on seed source, site, and browsing intensity. *Ecosphere*, *10*(1), e02580. http://doi.org/10.1002/ecs2.2580

Frank, A., Howe, G. T., Sperisen, C., Brang, P., St.Clair, J. B., Schmatz, D. R. & Heiri, C. (2017a). Risk of genetic maladaptation due to climate change in three major European tree species. *Global Change Biology*, *23*(12), 5358–5371. http://doi.org/10.1111/gcb.13802

Frank, A., Pluess, A. R., Howe, G. T., Sperisen, C. & Heiri, C. (2017b). Quantitative genetic differentiation and phenotypic plasticity of European beech in a heterogeneous landscape: indications for past climate adaptation. *Perspectives in Plant Ecology, Evolution and Systematics*, *26*, 1–13. http://doi.org/10.1016/j.ppees.2017.02.001

Frank, A., Sperisen, C., Howe, G. T., Brang, P., Walthert, L., St.Clair, J. B. & Heiri, C. (2017c). Distinct genecological patterns in seedlings of Norway spruce and silver fir from a mountainous landscape. *Ecology*, *98*(1), 211–227. http://doi.org/10.1002/ecy.1632

GBIF (2019). Data papers: getting scholarly recognition for your datasets. Retrieved on February 14, 2019, from https://www.gbif.org/data-papers

GCRIO (1991). United States policy statements on data management for global change research. Retrieved on May 4, 2018, from http://www.gcrio.org/USGCRP/DataPolicy.html

Hofelich Mohr, A., Johnston, L. R. & Lindsay, T. A. (2014). The data management village: collaboration among research support providers in the large academic environment. In L. M. Kellam & K. Thompson (Eds.), *Databrarianship: the academic data librarian in theory and practice* (pp. 51–66). Chicago, IL: ACRL.

Hswe, P. (2015). Peering outward: data curation services in academic libraries and scientific data publishing. In M. Bonn & M. Furlough (Eds.), *Getting the word out: academic libraries as scholarly publishers* (pp. 221–248). Chicago, IL: ACRL.

Huerta, M. F., Koslow, S. H. & Leshner, A. I. (1993). The human brain project: an international resource. *Trends in Neurosciences*, *16*(11), 436–438. http://doi.org/10.1016/0166-2236(93)90069-X

Kellam, L. M. & Thompson, K. (Eds.). (2016). *Databrarianship: the academic data librarian in theory and practice*. Chicago, IL: ACRL.

Kratz, J. & Strasser, C. (2014). Data publication consensus and controversies [version 3; referees: 3 approved]. *F1000Research*, *3*, 94. http://doi.org/10.12688/f1000research.3979.3

Lyon, L. (2014). Learning about research data in the lab at the Pitt iSchool. Retrieved on February 15, 2019, from https://libraryconnect.elsevier.com/articles/learning-about-research-data-lab-pitt-ischool

Michener, W. K., Brunt, J. W., Helly, J. J., Kirchner, T. B. & Stafford, S. G. (1997). Nongeospatial metadata for the ecolgical sciences. *Ecological Applications*, *7*(1), 330–342. http://doi.org/10.1890/1051-0761(1997)007[0330:NMFTES]2.0.CO;2

Mizzy, D. & Hayslett, M. (2014). Data librarianship: a day in the life-science edition. In L. M. Kellam & K. Thompson (Eds.), *Databrarianship: the academic data librarian in theory and practice* (pp. 355–371). Chicago, IL: ACRL.

Monastersky, R. (2013). Publishing frontiers: the library reboot. *Nature*, *495*(7442), 430–432. http://doi.org/10.1038/495430a

Open Data Commons (2019). Open Database License (ODbL) v1.0. Retrieved on February 11, 2019, from https://opendatacommons.org/licenses/odbl/index.html

Palmer, C. L., Cragin, M. H., Heidorn, P. B. & Smith, L. C. (2007). Data curation for the long tail of science: the case of environmental sciences. *Third International Digital Curation Conference*. Washington, DC.

Pampel, H., Bertelmann, R. & Hobohm, H.-C. (2010). "Data Librarianship" – Rollen, Aufgaben, Kompetenzen. *RatSWD Working Paper*, *144*. Berlin: Rat für Sozial- und Wirtschaftsdaten. http://hdl.handle.net/10419/43622

Pampel, H. & Dallmeier-Tiessen, S. (2014). Open research data: from vision to practice. In *Opening Science* (pp. 213–224). Cham: Springer International Publishing. http://doi.org/10.1007/978-3-319-00026-8_14

Pensoft (2019). Arpha Writing Tool. Retrieved on February 22, 2019, from https://arpha.pensoft.net

re3data (2018). re3data.org – registry of research data repositories. Retrieved on May 14, 2018, from https://www.re3data.org

Rushby, N. (2015). Editorial: data papers. *British Journal of Educational Technology*, *46*(5), 899–903. http://doi.org/10.1111/bjet.12337

Schiermeier, Q. (2018). Data management made simple. *Nature*, 555, 403–405. http://doi.org/10.1038/d41586-018-03071-1

Schneider, R. & Prongué, N. (2015). Data papers as a catalyst for Open Access and the long tail of research data (poster). https://zenodo.org/record/30761#.XIJ9BelYY-J

Science Europe (2018). Practical guide to the international alignment of research data management. Retrieved on March 7, 2019, from https://www.scienceeurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf

Simons, N. & Richardson, J. (2013). Research data: the new gold. In *New content in digital repositories: the changing research landscape* (pp. 115–146). Oxford, UK: Chandos Publishing.

SNSF (2018). Open Research Data. Retrieved on May 4, 2018, from http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx

SNSF (2019). General implementation regulations for the Funding Regulations. 2.13 Material costs: Costs for granting access to research data (Open Research Data). Retrieved on February 17, 2019, from http://www.snf.ch/en/funding/documents-downloads/Pages/regulations-general-implementation-regulations.aspx#ar_a_2_13

Springer Nature Publishing (2019). Scientific Data: editors, advisory panel & editorial board. Retrieved on February 22, 2019, from https://www.nature.com/sdata/about/editorial-board

SurveyMonkey (2018). Get the answers you need. Retrieved on November 10, 2018, from https://www.surveymonkey.com

swissuniversities & SNSF (2017). Nationale Open-Access-Strategie für die Schweiz. https://www.swissuniversities.ch/fileadmin/swissuniversities/Dokumente/Hochschulpolitik/Open_Access/Open_Access__strategy_final_DE.pdf

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., … Frame, M. (2011). Data sharing by scientists: practices and perceptions. *PLoS ONE*, *6*(6), 1–21. http://doi.org/10.1371/journal.pone.0021101

UK Data Service (2019). Research data lifecycle. Retrieved on from https://www.ukdataservice.ac.uk/manage-data/lifecycle

Universität Basel (2019). swissbib Basel Bern. Retrieved on March 3, 2019, from https://baselbern.swissbib.ch

Universität Bern (2019a). Universitätsbibliothek: Bern Open Publishing. Retrieved from https://www.unibe.ch/universitaet/dienstleistungen/universitaetsbibliothek/service/open_science/bern_open_publishing

Universität Bern (2019b). Universitätsbibliothek: Research Data Management. Retrieved from https://www.unibe.ch/university/services/university_library/services/open_science/research_data_management

Van den Eynden, V., Corti, L., Woollard, M., Bishop, L. & Horton, L. (2011). *Managing and sharing data: best practice for researchers* (3rd ed.). Colchester, UK: UK Data Archive. https://data-archive.ac.uk/media/2894/managingsharing.pdf

Whyte, A. (2014). From RDM strategy to action – a glass half full! Retrieved on March 3, 2019, from http://www.dcc.ac.uk/blog/rdm-strategy-action-glass-half-full

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. http://doi.org/10.1038/sdata.2016.18

WSL (2019). EnviDat: das Umweltdatenportal. Retrieved on February 22, 2019, from https://www.wsl.ch/de/ueber-die-wsl/programme-und-initiativen/envidat.html

Zwahlen, S. & Schneider, R. (2016). Swiss Data Journal: étude de conceptualisation d' un modèle de partage des données de recherche basé sur le concept de Data Paper dans le contexte helvétique. Haute École de Gestion de Genève (HEG-GE). https://doc.rero.ch/record/278093/files/TDB_Zwahlen_Stephane.pdf

Appendix 1: Survey questionnaire

A second life for old research data

Is "post-research" data publication an issue for scientists? Survey for MAS thesis

Dear researchers

Data management and data publication have become hot topics in research (e.g., Ferguson et al. 2014, SNSF 2018). Most interest and effort focusses on the management of current research data from new or ongoing projects. **However, what happens to older research data that have been relevant in the recent past for research, but that have not been published as dataset(s)?** You may let these data stored on your devices and servers – or you may decide to publish them "post-research" in order to make them findable, accessible, and reusable for other researchers. I use the term "post-research data publication" for referring to data publication after the active research process, i.e., after the papers were published, or after a project was terminated.

Do you (or your group, institute etc.) own old research data that are still worth publishing? Do you wish or plan to actually publish data post-research? Moreover, would you appreciate to be supported during that process, i.e., get guidelines, information, or even staff support for post-research data publication? **Your opinion matters and might help to develop new library services.**

I am currently conducting a survey among scientists at the University of Bern for my MAS thesis in library and information sciences at the University of Zurich. **My goal is to find out if post-research data publication is an issue for you.** This survey will be evaluated anonymously. Thank you for participating!

References:

Ferguson, A. R., J. L. Nielson, M. H. Cragin, A. E. Bandrowski, and M. E. Martone. 2014. Big data from small data: Data-sharing in the "long tail" of neuroscience. Nature Neuroscience 17:1442–1447.

SNSF. 2018. Open Research Data. http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx.

1. Did **you** create research data that you have used for publications, but that you have not published as dataset(s) so far?

- ◯ Yes
- ◯ No
- ◯ I don't know

2. Did **your group, institute, or project** create research data that they used for publications, but that they have not published as dataset(s) so far?

- ◯ Yes
- ◯ No
- ◯ I don't know

3. If you know of old, unpublished research data (in that case, you have answered "yes" in question 1. or 2. above):
Do you think these research data are still worth publishing?

- ☐ Yes, at least some of my own old research data are still worth publishing.
- ☐ Yes, at least some of my group's, institute's, or project's old research are still worth publishing.
- ☐ No
- ☐ I don't know

4. Why would you like to publish your or your group's, institute's, or project's data post-research? Several answers are possible.

- ☐ Because these data should become findable, accessible, and reusable for other researchers.
- ☐ Because the dataset(s) should become citable.
- ☐ Because I would like to publish a data paper that will be acknowledged as additional publication.
- ☐ I don't know.
- ☐ Other reason(s):

5. Would you appreciate to get support (see question 6) for post-research data publication?

- ◯ Yes
- ◯ Probably yes
- ◯ Probably no
- ◯ No
- ◯ I don't know

6. What kind of support do you think would be useful when publishing your own or your group's, institute's, or project's data post-research? Several answers are possible.

- [ ] Advice about data publication options (repositories, data journals, data papers)
- [ ] Guidelines for post-research data publication
- [ ] Staff support for data preparation
- [ ] Staff support for metadata preparation (i.e., preparation of dataset descriptions)
- [ ] Information about legal aspects (data ownership, embargos, licenses, etc.)
- [ ] Assistance during the data publication process
- [ ] Other support:

7. Your current position (several answers possible):

- [ ] Master student
- [ ] PhD student
- [ ] PostDoc
- [ ] Project leader
- [ ] Oberassistent/in
- [ ] Lecturer
- [ ] Principle investigator (PI)
- [ ] Senior researcher
- [ ] Group leader
- [ ] Professor
- [ ] Other position:

8. Would you like to get information about the results of this survey?

- ( ) No, thank you.
- ( ) Yes. Please contact me by this email address:

9. May I contact you in case that I have questions concerning your answers?

- ( ) No.
- ( ) Yes. You may contact me by the email address given above (question 8).
- ( ) Yes. You may contact me by this email address (if not given above):

10. Thank you for your contribution. Would you like to add any comments?

Appendix 2: Data paper draft

Genecological data of Norway spruce, silver fir, and European beech in Switzerland

Aline Frank[1], Christoph Sperisen[2], Peter Brang[2], Caroline Heiri[3]

[1]University Bern, University Library, Sidlerstrasse 5, CH-3012 Bern, Switzerland; aline.frank@ub.unibe.ch

[2]Swiss Federal Institute for Forest, Snow and Landscape Research WSL, CH-8903 Birmensdorf, Switzerland; christoph.sperisen@wsl.ch; peter.brang@wsl.ch

[3]Kanton Bern, Waldabteilung Bern Mittelland, Molkereistrasse 25, CH-3052 Zollikofen; caroline.heiri@vol.be.ch

---

Internal note:

red text = parts to be completed

Abstract

Genecology – the study of genetic variation in relation to the environment – is useful to investigate adaptation, phenotypic plasticity, and genetic maladaptation of trees. Understanding genecological patterns in forest trees is critical for gene conservation, predicting the effects of climate change, and successful reforestation. We conducted a genecological study for Norway spruce (*Picea abies*), silver fir (*Abies alba*), and European beech (*Fagus sylvatica*), the three most abundant tree species in Switzerland, to judge whether current populations are adapted to future climatic conditions. To this end, we measured growth and phenology of 17,000 seedlings of all three species. The seedlings originated from 77 to 92 populations per species, sampled across their natural ranges in Switzerland from 1–3 mother trees per population, and were grown at two common garden test sites. This dataset consists of the seedling phenotype data for all three species measured at one (Norway spruce and silver fir) or two test sites (European beech), the descriptions of sampled mother trees, populations and seed source environments (soil and past climate), and test site conditions (temperature, precipitation and soil moisture). The dataset represents one of the largest tree seedling common garden datasets in Europe. It forms a unique basis for assessing the three tree species' quantitative genetic variation, phenotypic plasticity, and adaptive significance of environmental drivers. We found that population differences and associations with climate variables were strongest in Norway spruce, intermediate in European beech, and weakest in silver fir. Genecological models for all three species provided evidence of natural selection and adaptation to local climates, but not to soils and other site characteristics. The models were useful for assessing the species' risk of genetic maladaptation to climate change. The dataset improved our understanding for the genecological patterns and climate change vulnerability of the three tree species in Switzerland.

Keywords

Abies alba; adaptive genetic variation; common garden; *Fagus sylvatica*; genecology; growth; *Picea abies*; phenology; quantitative genetics; seedling; Switzerland

Metadata

Introduction

Climate change affects forest ecosystems in various ways (Lindner et al. 2010). The high speed of current climatic changes is expected to cause serious adaptation lag in many tree species (Rehfeldt et al. 2002), resulting in reduced fitness and changes in forest composition, structure and health. Genecology is the study of genetic variation in relation to the environment (Aitken 2004). It is a research method suited to investigate adaptation, phenotypic plasticity, and genetic maladaptation of species to environmental change, such as current climate warming. It can be used to identify sensitive tree species and populations, and to project the extent of maladaptation to future climates (St. Clair and Howe 2007). Such information is important for guiding forest management, but is still lacking for many tree species.

Typically, genecological studies involve collecting seeds from natural populations, cultivating the progeny under uniform environmental conditions, and measuring potentially adaptive traits, such as growth, survival, and bud phenology (St. Clair et al. 2005, Bussotti et al. 2015). These measurements can be used to derive quantitative genetic estimates for within- and among-population genetic variation and to develop genecological models that relate population differentiation to environmental variables (e.g., St. Clair et al 2005). These models provide the basis to evaluate the species' risk of maladaptation due to climate change (St. Clair and Howe 2007).

In the project ADAPT (2009–2016) we studied the genecology of Norway spruce (*Picea abies*), silver fir (*Abies alba*), and European beech (*Fagus sylvatica*) in Switzerland for evaluating to

what extent current populations are adapted to future climatic conditions (WSL 2016, Frank et al. 2017a). To this end, we established an extensive common garden experiment at two field sites with seedlings originating from 77 to 92 populations distributed across the natural range of the three tree species in Switzerland (Figure 1). We recorded traits of growth and phenological traits, i.e., key adaptive traits of young trees (Bussotti et al. 2015), on [number] seedlings (# Norway spruce, # silver fir, and # European beech) during two consecutive years. We derived quantitative genetic estimates and developed genecological models that associate population variation with seed source environments. Finally, we estimated the relative risk of maladaptation to current and future climates for key phenotypic traits using the climate projections of three regional climate models in Switzerland.

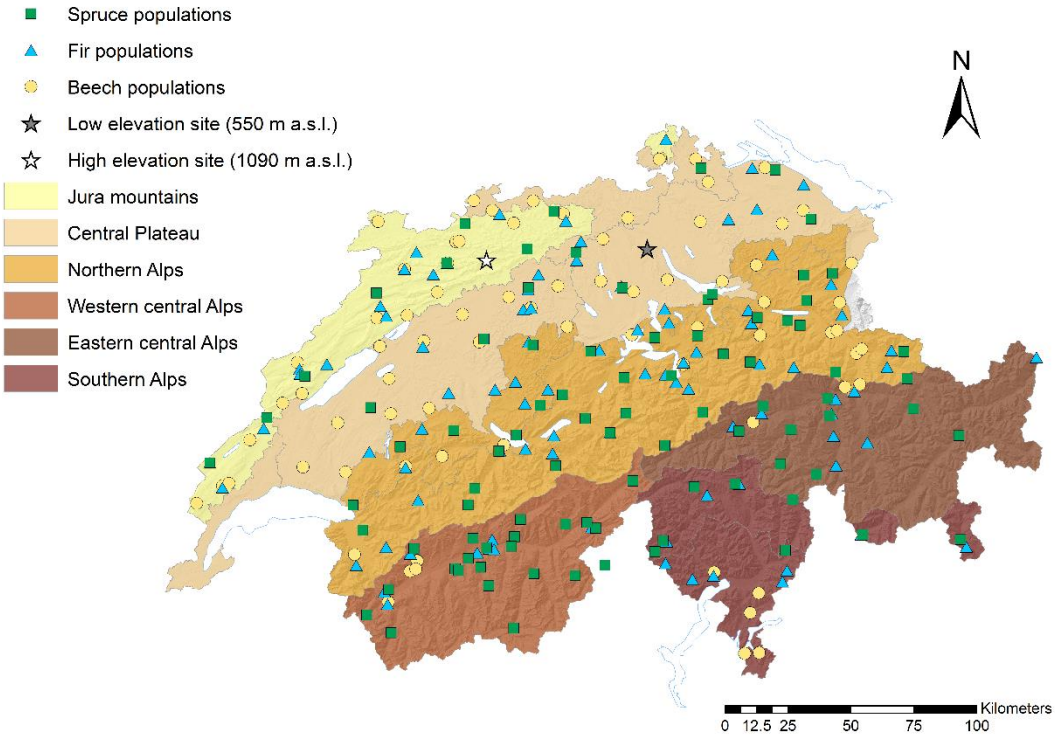


Figure 1: Distribution of the 92 Norway spruce (*Picea abies*), 90 silver fir (*Abies alba*), and 77 European beech (*Fagus sylvatica*) populations sampled across the six main biogeographic regions of Switzerland. Stars indicate test site locations.

The dataset presented here represents the basis for the analyses conducted within the project ADAPT. It includes A) the seedling phenotype data for all three species measured at one (Norway spruce and silver fir) or two test sites (European beech), B) the descriptions of sampled mother trees, C) the descriptions of populations inclusive seed source environments (soil and past climate), and D) test site conditions (temperature and soil moisture) (Figure 2).

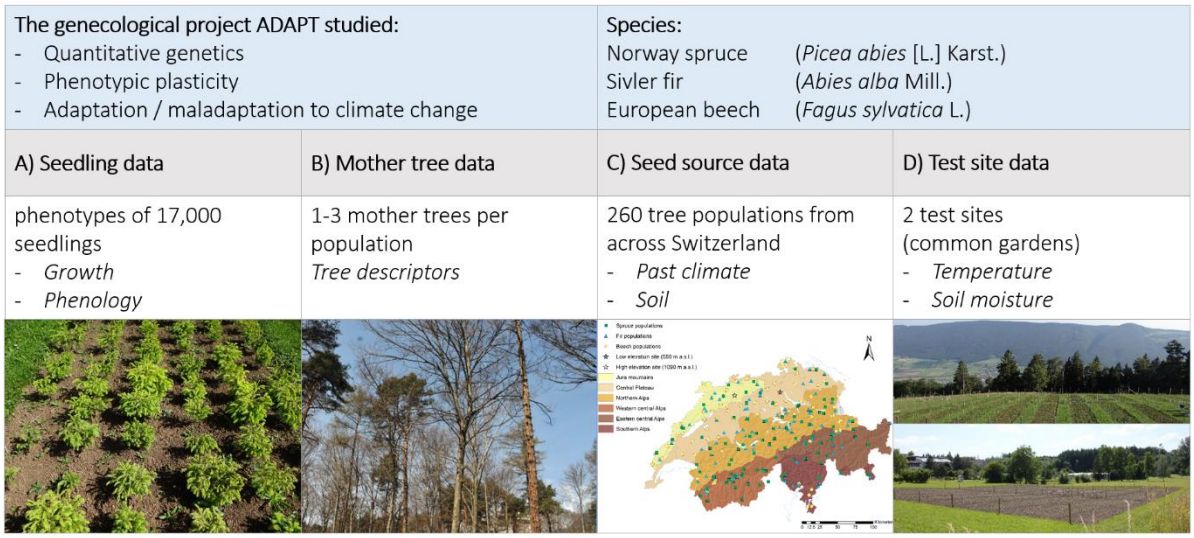


Figure 2: Overview of data produced and studied within the project ADAPT. The dataset consists of four subsets containing data on A) the seedling phenotypes measured during the common garden experiment, B) the seedlings' mother trees, C) the seed sources, i.e., environments of sampled tree population across Switzerland, and D) the test sites.

I. Dataset descriptors

A. Dataset identity

Genecological data of Norway spruce, silver fir, and European beech in Switzerland

B. Dataset identification code

DOI

C. Dataset description

1. Principle investigators

- Aline Frank, University Library Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland, aline.frank@ub.unibe.ch
- Christoph Sperisen, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, CH-8903 Birmensdorf, Switzerland, christoph.sperisen@wsl.ch
- Peter Brang, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, CH-8903 Birmensdorf, Switzerland, peter.brang@wsl.ch
- Caroline Heiri, Kanton Bern, Waldabteilung Bern Mittelland, Molkereistrasse 25, CH-3052 Zollikofen, wald.mittelland@vol.be.ch

2. Abstract

Same as above.

D. Keywords

Same as above.

II. Research origin descriptors

A. Overall project description

1. Project title

German: Adaptive genetische Variation von Fichte, Tanne und Buche (ADAPT)

English: Adaptive genetic variation of Norway spruce, silver fir, and European beech (ADAPT)

2. Originators

See principle investigators as listed above (I C. 1.)

3. Period of study

2009–2016

4. Objectives

The objectives of the project ADAPT were to

- identify adaptive traits and associated selective forces for Norway spruce, silver fir, and European beech populations in Switzerland
- compare the adaptive strategies of these three species
- infer their potential for climate change adaptation
- examine the extent of phenotypic plasticity in potentially adaptive traits
- estimate the amount of genetic maladaptation of current populations of spruce, fir, and beech in Switzerland to climate change
- identify the species and regions most vulnerable to future maladaptation

5. Abstract of overall project, adapted from Frank (2016)

Climate change impacts on forest ecosystems are of great societal concern. The high speed of current climatic changes is expected to cause serious adaptation lag in many tree species, resulting in reduced fitness and changes in forest composition, structure and health. Genecology, the study of genetic variation in relation to the environment, can help to identify sensitive tree species and populations, and to project the extent of maladaptation to future climates. Such information is valuable to guide forest management strategies for preparing forests to climate change, but is lacking for many tree species.

In the project ADAPT, the genecology of Norway spruce (*Picea abies*), silver fir (*Abies alba*), and European beech (*Fagus sylvatica*) – the three most abundant tree species in Switzerland – was investigated and used to judge whether current populations are adapted to future climatic conditions. To this end, an extensive common garden experiment with two field test sites ("Birmensdorf" and "Matzendorf") was established with seedlings originating from 77 to 92 populations distributed across the natural range of the three tree species in Switzerland. Traits of growth and phenology were recorded during two consecutive years. Quantitative genetic

estimates were derived, and genecological models were developed that associate population variation with seed source environments. Relative risk of maladaptation to current and future climates was estimated for key phenotypic traits using the climate projections of three regional climate models.

Overall, this project improved our understanding of the genecological patterns and climate change vulnerability of Norway spruce, silver fir, and European beech in Switzerland. The findings of this study are valuable for adjusting management strategies to promote climate change adaptation of our major forest trees, and might also be relevant for other disciplines, such as landscape genomics and vegetation modeling.

6. Sources of funding

- Swiss Federal Institute for Forest, Snow and Landscape Research WSL
- Swiss Federal Office for the Environment FOEN

B. Specific subproject description

1. Site description

The lower elevation test site, called “Birmensdorf”, was located in the Central Plateau, close to the Swiss Federal Institute for Forest, Snow and Landscape Research WSL. The higher elevation test site, “Matzendorf”, was situated in the Jura Mountains (Figure 1). The differences in test site conditions were reported by Frank et al. (2017b) and shown in Table 1. Both sites were formerly as pastures for sheep (Birmensdorf) and cattle (Matzendorf).

Table 1: Environmental conditions at the two ADAPT test sites. Location, topography and soil characteristics (a), and temperature and water availability during the 2014 measurement period, during which both sites were analyzed (b). Table from (Frank et al. 2017b).

a)

Variable	Unit	Low elevation site "Birmensdorf"	High elevation site "Matzendorf"
		Value or type	Value or type
Latitude	° ' ''	47°21′44″	47°19′35″
Longitude	° ' ''	8°27′22″	7°36′42″
Elevation	m a.s.l.	550	1090
Aspect		west	southeast
Slope	%	5	22
Soil type		Gley	Rendzina
Rooting depth	cm	ca. 45–70	ca. 40
pH*		7.2	6.9
Sand*	%	31.9	7.5
Silt*	%	34.1	36.4
Clay*	%	34.1	56.2
Fine earth density*	kg/dm^3	0.8	0.7

b)

Variable†	Unit	Low elevation site "Birmensdorf"					High site elevation "Matzendorf"				
		Mean	SD	Min.	Max.	Sum	Mean	SD	Min.	Max.	Sum
Mean air temperature in spring	°C	10.0	5.2	-2.1	27.2		7.1	4.7	-3.4	20.8	
Mean air temperature in summer	°C	17.1	4.4	7.2	34.7		14.1	4.1	5.2	29.0	
Mean soil temperature in summer*	°C	19.4	2.0	13.8	25.7		16.7	2.3	11.0	23.4	
Summer precipitation sum	mm					385					471
Summer soil water potential*	kPa	-14.7	19.2	-189.2	-5.0		-19.1	37.8	-391.1	-4.8	

*Physical and chemical soil characteristics, soil temperature and soil water potential refer to the top soil layer (0–15 cm).

†Spring refers to March–May; summer refers to June–August.

2. Experimental design

Design characteristics:

We used seedlings grown from seed that was collected from 1–3 mother trees of 260 populations from across Switzerland. For the field experiment, the seedlings were planted at both test sites at 30 cm × 40 cm spacing in 16 blocks. Within blocks, each family was represented by one seedling, whereas each pooled seedlot was represented by three seedlings. All seedlings were randomized within blocks without regard to population origin. For families that had fewer than 16 seedlings in the nursery, we set a threshold of at least 12 seedlings for being included in the field experiment.

Data collection period:

Seedling growth and phenology data were collected in Matzendorf during 2013, 2014, and 2015 and in Birmensdorf during 2014. Mother tree data were collected during 2009 and 2010. Seed source data were collected between 2009 and 2012. Test site data were recorded between 2013 and 2015.

3. Research methods

Genecological seedling common garden experiment with two test sites, similar to, e.g., the experiment performed by St. Clair et al. (2005). The detailed procedures were described by Frank et al. (2017c, 2017b) and in the field reports published on the project webpage (WSL 2016).

4. Project personnel

Same as listed above (I C. 1.). For a more detailed list of people that contributed to the project ADAPT, see Acknowledgements in Frank (2016).

III. Dataset status and accessibility

A. Status

1. Latest update: …
2. Latest archive date: …
3. Metadata status: …
4. Data verification: …

B. Accessibility

1. Storage location & medium: The data is available at EnviDat, DOI, https://www.envidat.ch/dataset/adapt-genecology-data
2. Contact person: Aline Frank, University Library Bern, Sidlerstrasse 5, CH 3012 Bern, Switzerland, aline.frank@ub.unibe.ch and aline.frank@alumni.ethz.ch
3. Copyright restrictions: Open Data Commons Open Database License (ODbL); to be defined
4. Proprietary restrictions: Please cite this data paper when data are used for publications or teaching.
5. Costs: none

IV. Data structural descriptors

A. Data set file

To be completed

B. Variable information

To be completed

V. Supplemental descriptors

References

Aitken, S. 2004. Genecology and adaptation of forest trees. *In:* Burley, J. Encyclopedia of Forest Sciences:197–204. Elsevier, Oxford, UK.

Bussotti, F., M. Pollastrini, V. Holland, and W. Brüggemann. 2015. Functional traits and adaptive capacity of European forests to climate change. Environmental and Experimental Botany 111:91–113.

Frank, A.D. 2016. Genecology of Norway spruce , silver fir , and European beech in Switzerland : Are current populations adapted to future climates? ETH Dissertation Nr. 23644. ETH, Zurich, Switzerland.

Frank, A., G.T. Howe, C. Sperisen, P. Brang, J.B. St.Clair, D.R. Schmatz, and C. Heiri. 2017a. Risk of genetic maladaptation due to climate change in three major European tree species. Global Change Biology 23:5358–5371.

Frank, A., A.R. Pluess, G.T. Howe, C. Sperisen, and C. Heiri. 2017b. Quantitative genetic differentiation and phenotypic plasticity of European beech in a heterogeneous landscape: indications for past climate adaptation. Perspectives in Plant Ecology, Evolution and Systematics 26:1–13.

Frank, A., C. Sperisen, G.T. Howe, P. Brang, L. Walthert, J.B. St.Clair, and C. Heiri. 2017c. Distinct genecological patterns in seedlings of Norway spruce and silver fir from a mountainous landscape. Ecology 98:211–227.

Lindner, M., M. Maroschek, S. Netherer, A. Kremer, A. Barbati, J. Garcia-Gonzalo, R. Seidl, S. Delzon, P. Corona, M. Kolström, M.J. Lexer, and M. Marchetti. 2010. Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. Forest Ecology and Management 259:698–709.

Rehfeldt, G. E., N.M. Tchebakova, Y.I. Parfenova, W.R. Wykoff, N.A. Kuzmina, and L.I. Milyutin. 2002. Intraspecific responses to climate in *Pinus sylvestris*. Global Change Biology 8:912–929.

St.Clair, J.B., and G.T. Howe. 2007. Genetic maladaptation of coastal Douglas-fir seedlings to future climates. Global Change Biology 13:1441–1454.

St.Clair, J.B., N.L. Mandel, and K.W. Vance-Borland. 2005. Genecology of Douglas fir in Western Oregon and Washington. Annals of Botany 96:1199–1214.

WSL 2016. Projekt ADAPT: Adaptive genetische Variation von Fichte, Tanne und Buche. https://www.wsl.ch/de/projekte/projekt-adapt.html. Accessed: February 28, 2019.