Research Data Services at ETH-Bibliothek

Journal Article

Author(s):
Petrus, Ana (ID); Töwe, Matthias (ID)

Research Data Services at ETH-Bibliothek

Ana Sesartic[1]

ETH-Bibliothek, ETH Zürich, Switzerland

Matthias Töwe

ETH-Bibliothek, ETH Zürich, Switzerland

Abstract

The management of research data throughout its life-cycle is both a key prerequisite for effective data sharing and efficient long-term preservation of data. This article summarizes the data services and the overall approach to data management as currently practiced at ETH-Bibliothek, the main library of ETH Zürich, the largest technical university in Switzerland. The services offered by service providers within ETH Zürich cover the entirety of the data life-cycle. The library provides support regarding conceptual questions, offers training and services concerning data publication and long-term preservation. As research data management continues to play a steadily more prominent part in both the requirements of researchers and funders as well as curricula and good scientific practice, ETH-Bibliothek is establishing close collaborations with researchers, in order to promote a mutual learning process and tackle new challenges.

Keywords

Data life-cycle, data management plan, libraries, preservation, research data, research data management

Introduction

The growing volume of data produced in research has created new challenges for its management and curation to ensure continuity, transparency and account- ability. Timely and effective management of research data throughout its life-cycle ensures its long-term value and prevents data from falling into digital obsolescence (Corti et al., 2014; Goodman et al., 2014).

Proper data management is a key prerequisite for effective data sharing within a specific scientific com- munity and for data publication beyond any particular target group. This, in turn, increases the visibility of scholarly work and is likely to increase citation rates (Piwowar and Vision, 2013: 25; Piwowar et al., 2007). Managing research data throughout its life-cycle is not only a key prerequisite for effective data sharing but also for efficient long-term preservation because the latter must rely on technical, administrative and rights metadata, as well as sufficient context information being available to make sure that data remains usable and understandable in the long run.

---

[1] Corresponding author:
Ana Sesartic, ETH Zürich, ETH-Bibliothek, Rämistrasse 101, CH-8092 Switzerland.
E-mail: ana.sesartic@library.ethz.ch

Depending on their respective institutional setting, libraries can contribute to research data management in different ways and likewise, expectations from their patrons can vary widely, e.g. by scientific discipline. Therefore, the following report should be understood as a case study rather than a general recommendation.

Libraries are never the only service providers in a university and, ideally, a range of providers caters for researchers' and students' needs. When it comes to data management in particular, IT services will obviously be a strong player on the technical side whereas research offices must take an interest in how researchers comply with internal and external requirements. In such a landscape, it is important to note that libraries should focus on their strengths such as metadata management, content curation, and support and training of their customers, in this case the researchers. Also, the services offered are never carved in stone and should be adapted to the current needs of science.

In the following, we summarize the data services and approach to data management as currently practised at ETH-Bibliothek (ETH Zürich, 2016c), the main library of ETH Zürich (ETH Zürich, 2016b), the largest technical university in Switzerland.
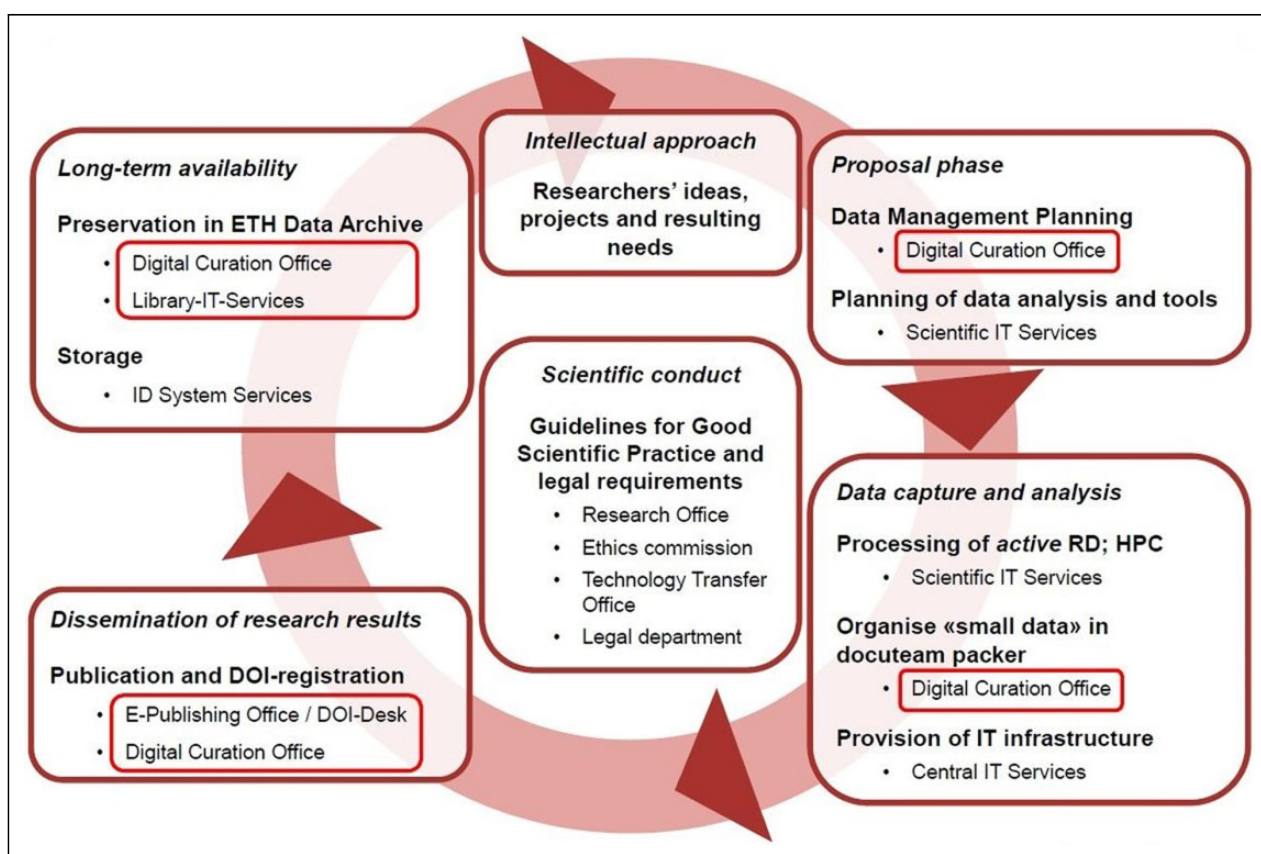

Figure 1. Actors and their tasks at ETH Zürich along phases of the data life-cycle. Units in red boxes belong to ETH Library.

The overall concept

The role of the ETH-Bibliothek within ETH Zürich is to support researchers from early on in the data life- cycle (see Figure 1), starting with general consulting regarding compliance, support

through the development of data management plans, the offer of data archiving and publication services, to the creation of DOIs for easier citation and re-use of data. Important phases of the life-cycle mainly in its active research phase (see Figure 1) require competencies beyond what the library can offer. Other actors within the university support these phases and a clarification of the division of tasks between those service providers is needed. At ETH Zürich, a good level of mutual understanding, for example of the activities in storage and preservation, was achieved between library staff and the storage section of the IT services as part of a common small-scale project starting in 2006. Regular meetings have continued ever since. Nevertheless, it can be challenging to communicate the division of labour transparently to customers. Ideally, there should be only a single point of contact for customers to turn to. This has not been established so far. How- ever, it should not matter who customers contact as long as they are redirected appropriately. This, again, requires a reasonable understanding of other units' services, which obviously evolve over time.

With the Digital Curation Office (ETH-Bibliothek, 2016a), the ETH-Bibliothek offers a point of contact for technical and conceptual questions regarding long-term preservation and management of research data. It also offers help and support to the researchers of ETH Zürich in managing and publishing their data, as well as following the requirements as stated in the *Guidelines for Research Integrity* of ETH Zürich (ETH Zürich, 2011: 31). Regarding intellectual property and research ethics issues, the ETH-Bibliothek closely collaborates with the Technology Transfer Office and the Office of Research, respectively.

Figure 1 illustrates the coverage of roughly defined stages of the research data life-cycle by actors and services at ETH Zürich with their respective tasks. Note that Guidelines for Good Scientific Practice and further legal requirements (centre) apply throughout the cycle, but should obviously be considered from early on.

For a first-time or occasional customer, the multitude of actors can obviously be confusing and users should not be left alone with it. From our experience, customers appreciate being able to get in touch with a contact person they can talk to about their needs. In a first instance, it is not even expected that this person can provide an immediate solution, but it is important that someone takes care of the issue and provides guidance as to where to turn. This means that all actors must be aware of tasks not within their own portfolio as well, and we are aware that this remains a challenge and requires an ongoing learning process. As a detail to underline the importance of personal contacts: as one of the first teams within the ETH-Library, the Digital Curation Office put portrait images of its staff on its website to lower the barrier of writing to an otherwise anonymous email box.

Data management training

Research data management (RDM) has gained increasing attention over the last years, due to growing aware- ness of the value contained in research data and of the risks of losing such data over time. Apart from the need to manage data over the course of a project, there is a need to retain and curate data which one plans to work with in the future. It might be possible and sometimes advisable to repeat an experimental measurement, but in other cases it might either be prohibitively expensive or otherwise ethically unacceptable to do so. For unique observational data, a repetition is not possible at all and accordingly, communities relying on

such data have long been aware of the challenges.

While these issues are mainly related to the efficiency and effectiveness of research and its funding, RDM also addresses the accountability of science. One principle of the scientific process is the requirement to be able to justify results by providing underlying data where necessary. There are significant reputational risks involved for individual researchers, principal investigators and institutions who fail to comply with good scientific practice of which RDM is just one part. To address these questions a workshop within the ETH Critical Thinking Annual Programme (ETH Zürich, 2016e) was developed. The overall aim of this programme is to strengthen critical thinking and a responsible approach to taking actions beyond disciplinary competences. The workshop introduced some services and tools for RDM, as well as encouraging the participants to share both their experiences and the methods and tools they use, during the interactive parts of the workshop. The goal of the workshop within the Critical Thinking programme was to increase the critical thinking skills of undergraduate and graduate students, as well as scientists, regarding RDM. The philosophy behind this approach is the conviction that researchers themselves must be empowered to make informed decisions on their data, as they are the experts with the most intimate knowledge of their own data. The workshop was focused on activating teaching methods, engaging the participants in group work and discussions.

The majority of the participants at this event were working on their doctoral studies with few Master's students and post-docs present. They showed very varied needs and levels of knowledge, but they were well aware of the problems regarding RDM before-hand and were looking for solutions. This is why we also offer tailored training courses in RDM for groups and departments. These can range from so-called Tools & Tricks mini lectures, short 15-minutes inputs over coffee for lunch break, to fully-fledged one-day training workshops. It is important to note that certain departments and institutes already offer similar or overlapping internal training, which is why communication and coordination are key. As of now, there is no dedicated course on RDM within the ETH curriculum; however, some departments offer methodological courses including research ethics and scientific writing which might cover some aspects. With increasing concerns about data management issues, it is to be expected that the topic will figure more prominently in curricula in the future.

Data management plan checklist

Today, the availability of well-managed data is part of good scientific practice and ensures the reproducibility of research results, a key requirement at the core of the research process. Many funding organizations prescribe the use of data management plans and insist on open access publication of the research results they funded.

In some parts of EU's research programme Horizon 2020, DMPs will, for example, be evaluated as part of the impact of a proposal and in the reporting during the course of a funded project (European Com- mission, 2013: 6). But even if a funding body does not explicitly demand data management, following professional curation and preservation concepts has numerous advantages (DLCM, 2016):

- It greatly facilitates the reuse of research data;

- As a result, this increases the impact of research results;

- It saves precious research funds and ultimately natural and human resources by avoiding unnecessary duplication of work

As effective and efficient data management becomes more and more challenging for both researchers and information specialists, the question arose how they can best be reached and supported on a national level regarding best innovative practices in digital preservation so as to succeed in this ambitious enterprise. This led to the creation of the Swiss Data Life-Cycle Management (DLCM) project (Blumer and Burgi, 2015: 16), facilitated by *swissuniversities* (swissuniversities, 2016) and involving collaboration between eight Swiss higher education institutions (EPF Lausanne, ETH Zürich, Universities of Basel, Geneva (lead), Lausanne, Zürich, Geneva School of Business Administration at the Western Switzerland University of Applied Sciences and Arts, and SWITCH, the national IT service provider for higher education institutions).

The Data Management Plan (DMP) Checklist (ETH-Bibliothek and EPFL Library, 2016) is among the first tangible deliverables of the DLCM project. It is meant to be an essential tool aiding researchers in the management of their data, thus preparing them for later publication and preservation, as needed. By giving clear guidelines, it should facilitate this task for researchers and eventually save them time and effort. The list has been customized for Switzerland based on pre-existing national and international policies. It covers general planning and the phases of the data life-cycle, from data collection and creation to data sharing and long-term management. Special sections cover documentation and metadata, file formats, storage, ethical and intellectual property issues. Ideally, the list should be used by researchers to critically assess their data management and to gather information they might need to create a data management plan. It can also serve as a starting point for further face-to-face discussions of data management issues within research groups and with support staff if required. The list is static with no further functionality and it will be observed whether a more interactive solution will be required later.

The checklist was created in close collaboration between the Digital Curation Office at ETH- Bibliothek and the Research Data Team at EPFL Library. It is currently available through the ETH-Bibliothek and EPFL Library websites (ETH- Bibliothek and EPFL Library, 2016) and will soon be disseminated on a national portal on DLCM which will be launched in the next few months. The portal itself will touch on many aspects, aggregating and providing further information about data organization, training (in person and online) of end- users, and consulting regarding best practices among many others.

Active data management

Data management is understood as a comprehensive task throughout the data life-cycle. It therefore needs to comprise the handling of research data while the actual research is carried out. We call this active data management to signify that data at this stage is usually not static, but keeps being analysed and worked upon as part of the research. At this stage of the life-cycle, subject-specific tools are employed in data processing which may be implemented and run by a specialized support unit or by research groups themselves.

At ETH Zürich, the section Scientific IT Services of the central IT Services provides this kind of sup- port. Research groups from life sciences are known to be among the most intensive users of their services. Among them figures the data management software platform openBIS (ETH Zürich, 2016d), which has also been extended with components to serve as an electronic laboratory notebook (ELN) and/or as a laboratory information management system (LIMS). Obviously, this platform and other tools with similar aims must already capture a lot of information, which is relevant for current research. Part of this information might be gathered automatically, while more will be required to be entered by researchers, with quality depending on their willingness to comply and therefore on the ease of the process. Most of this input and possibly additional documentation will be needed in order to use and make sense of the research data at a much later stage and with the active data platform no longer being available.

While such systems themselves are not meant as publishing or preservation tools for research data, they can very well serve as sources for these pro- cesses. Ideally, researchers working in such a system should be able to decide which part of the content must be preserved or can be published to trigger an export, e.g. to a long-term preservation solution. Such a process has not yet been implemented, but it is envisaged that an interface from openBIS to the ETH Data Archive will be developed in the current DLCM project.

It should be noted that the exchange between the different systems must not be understood merely as a technical transfer. Data at this stage also crosses a boundary between the research world in a narrow sense and a curation domain (see Figure 2). Such a transition is prone to suffer from misunderstandings between the parties involved due to a lack of common standards or even a uniform vocabulary.
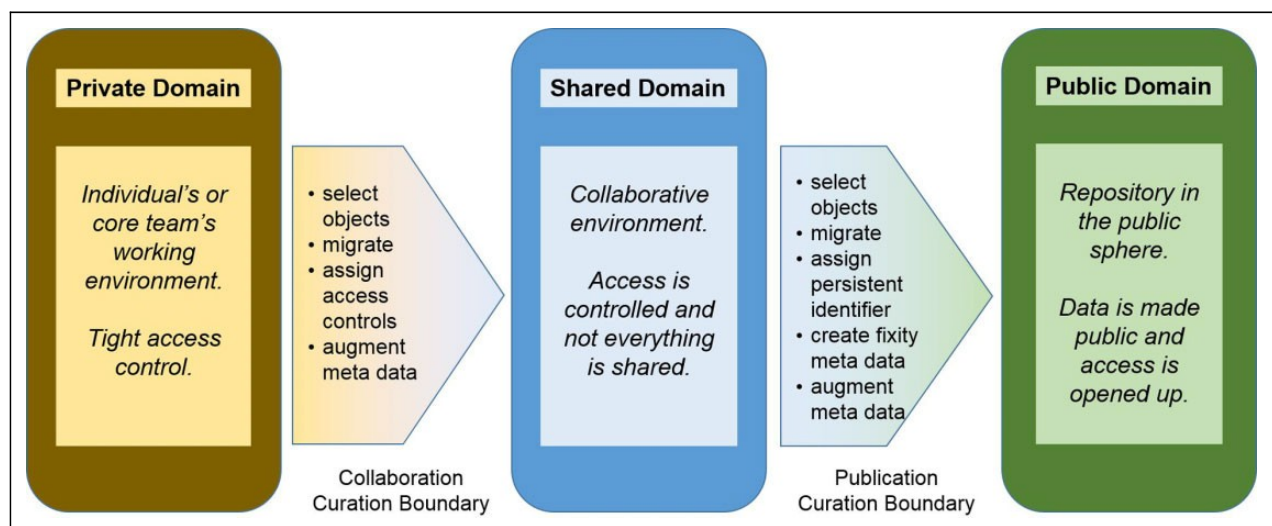

Figure 2. Transitions between curation domains along the research data life-cycle. Adapted from (Treloar, 2012).

This already highlights the importance of close collaboration not only of researchers and service providers, but also among different providers with complementary competencies. An essential part of the work to achieve this has been a constructive exchange over several years between ETH Library and the Central IT Services. Even before the Digital Curation Team existed,

staff from various teams in the library (e.g. University Archives of ETH Zürich, Library IT Services, Consortium of Swiss Academic Libraries) engaged in early pilot projects mainly with the Central IT Services team in charge of storage management. These activities from 2006 onwards served to understand the required functions, to get an idea of the available competencies and not the least to build trust within an, albeit loose, network of players in the field of data management within the university. This helped to achieve a common understanding of the tasks and to raise awareness, e.g. for the need to define more precisely what each stakeholder under-stands when talking about 'archiving'.

Two rather surprising outcomes arose from the exchange with a retiring professor on how to transfer his well-managed archive to the library. Firstly, one of his post-docs involved in the operation could be hired to work with ETH Library's Digital Curation Team, and secondly, the discussion of the principles underlying this research group's archive recently led to a publication highlighting the general concepts behind their approach which had proven useful for almost three decades of research practice (Sesartić et al., 2016).

Publication and preservation of data

From the start of its activities around research data, ETH-Bibliothek saw a central role in support of the processes for publication and preservation of such data. This was in line with expectations from both researchers and other service providers in the university. At the same time, it was decided to address preservation issues with a view on all kinds of scientific and cultural heritage unique to ETH Zürich. This includes research data from ETH Zürich staff, administrative records or personal bequests to ETH Zürich University Archives (ETH-Bibliothek, 2016c) and digital born and digitized content of ETH- Bibliothek, in particular doctoral theses which must be deposited in digital form and master files from several large scale digitization projects.

Since 2014, ETH-Bibliothek has offered the ETH Data Archive as a productive service. It is based on the commercial long-term preservation system Rosetta (Ex-Libris Group, 2016b). The application itself is maintained by the library's own IT services, while both virtual servers and different types of storage at two different sites of the university are provided by ETH Zürich's central IT services.

ETH researchers can deposit content into the ETH Data Archive and define the appropriate access rights. This can be done manually via a web client, e.g. for supplementary material belonging to an article. Depositors are quite free to define on which level they want to or can provide metadata: an archival package or intellectual entity may contain just a single file with its individual description or a ZIP- or TAR-container with thousands of files included under only one metadata set. The latter is sometimes made use of when a collection of files belonging to a concluded thesis or another publication needs to be retained for a defined period of time (10 years minimum). In this case, it is often assumed that the thesis itself is the most comprehensive documentation of the data package and should suffice for the professor to answer any inquiries. This might not be an ideal arrangement, but it represents considerable progress compared to some previously existing group archives held exclusively on CD-ROM and DVD. When the focus is on safeguarding data for a limited period of time only, no strict requirements on the longevity of file formats are enforced. If requested by researchers,

support is provided with identifying suitable formats.

In other cases, metadata is routinely gathered for research data on the level of individual folders and files in research groups who have already seen a need to operate a managed archive. They may use the open source editor and viewer docuteam packer (docuteam, 2016; ETH-Bibliothek, 2016e) locally to organize files in a defined structure and add metadata as required. To support this potentially laborious task, certain metadata can be pre-defined or inherited from the top-most level. A researcher can then decide when to submit the whole or part of the structure she or he created to the ETH Data Archive. For data in docuteam packer, DOIs (Digital Object Identifiers) can already be reserved which will be registered to become active after submission to the ETH Data Archive. The advantage of this method is that the DOI can already be used in a manuscript even before the data has actually been deposited. DOIs are registered with the consortium DataCite (2016). Likewise, access rights and a retention period can be defined which will be enforced after submission.

Submission itself is handled by the tool docuteam feeder, which processes the Submission Information Package (SIP) put out by docuteam packer into the according Archival Information Packages (AIP) to be submitted to the ETH Data Archive. Docuteam packer will also be used for submissions to ETH Zürich University Archives with a modified configuration and a more interactive workflow between depositors and University Archives staff.

Obviously, there are limits to which kind of sub-missions can be comfortably handled via a web user interface and via docuteam packer. Automated pro-cesses with submission applications relying on existing sources for metadata and content are currently in use only for library content from the institutional repository ETH E-Collection, for master files from the digitization of rare books from ETH-Bibliothek in e-rara (ETH-Bibliothek, 2016b), and for digitized material from ETH Zürich's Archives of Contemporary History (ETH Zürich, 2016a). In the future, such interfaces should also be created with existing sources for research data, such as the platform openBIS mentioned above. Apart from fully automated processes, there might still be others requiring a trigger from the data producer to give them more control of which part of their content is archived when. For example, producers might want to collect, (re-) structure and describe their data before finally deciding to submit an archival package. In other cases, professors want to review datasets from their groups before submission and start the transfer themselves.

Deposit in the ETH Data Archive is not coupled with an immediate publication of the content. While ETH Zürich encourages open access also to research data, it is currently up to the data producer to decide which data they want to make accessible as long as they are observing existing requirements, e.g. from funders. They may opt for an embargo period of e.g. two years or for limited access within the IP-range of ETH Zürich only. Even very restrictive access rights for defined persons only are currently accepted.

Metadata of published content are published in the ETH Knowledge Portal (ETH Zürich, 2016c), in the Primo Central Index (Ex-Libris Group, 2016a) and in the Data Citation Index (Thomson Reuters, 2016).

Currently, the workflow for the deposit of research data is largely separated from the one for

publications, e.g. articles to be published via the green road of open access. In the future, a new platform will pull several workflows for publications, bibliographic records and research data together into one service.

For published research data in particular, it is a reasonable expectation that it should remain available in the long term similar to what is expected from formal publications. Whether this will actually be possible in the future depends to a large degree on the file formats being employed. In the heterogeneous environment of a university with numerous contradictory pressures and constraints on researchers, it would not be a realistic approach to admit only a limited number of well-documented open formats to a data archive. Rather, the Digital Curation Office at ETH-Bibliothek offers some guidance on preferred formats and recommends a few of them depending on the expected retention period (ETH-Bibliothek, 2016d). This may limit the chances of actual preservation measures in the future to the extent where only bitstream preservation remains as viable. This is made transparent to researchers submitting data and usually they are fully aware of this serious limitation. However, those depositing data 'just in case' do not consider this as a major problem because they expect a need to invest effort into using such data in the future, anyway. Their perspective then is to post- pone this effort to the point where it is actually needed rather than making an upfront investment which might be in vain, given that only a very small part of their data might ever be re-used.

On the other hand, data producers who regularly share and exchange data with colleagues in their own community have usually overcome the barrier of proprietary formats and often rely on open community standards. However, this might only apply to the core of research data, while accompanying material may contain less suitable formats. Given the high level of awareness of these users, they might want to re-consider those formats, as well.

A particular format issue concerns the vast number of research data files in plain text formats, but with a large variety of sometimes misleading file extensions. While text files with documented encoding are actually very suitable for preservation purposes, it can be challenging to identify them in the first place and in many cases, the identification will not be a technical one, but will rely on information from the data producers.

Taking the sections above into account, it is obvious that communication with researchers forms an essential part of providing research data services for publication and preservation. It is in the best cases rewarding for customers and staff alike, but nevertheless time consuming when time might already be pressing, for example when a manuscript is about to be submitted and supplementary material needs to be deposited on time. Obviously, this kind of communication is much facilitated if appropriate skills are available on both sides. It is therefore very helpful to have staff with a scientific background in the Digital Curation Office, although it is obvious that they cannot cover all fields in depth.

Conclusions

With the growing digitization of science and society, a curation gap between research practice and curation needs opens up. However, if there is collaboration and communication between IT services, libraries and researchers, the discrepancy between research practice and research

content preservation can be minimized and the curation gap closed. In order to do so, university libraries and data centres must continue to support and educate researchers, which also requires a thorough understanding of researchers' work practices and the challenges they meet. The heterogeneity of their needs limits the possibility to generalize services – or the other way round: in some cases, it may only be possible to serve needs close to the smallest common denominator between various interests. This is the reason why libraries should not aim at serving all communities equally themselves, but rather also keep an eye on subject specific solutions, which are created by third parties to address specific needs of one discipline. A combined and well-integrated landscape of institutional, networked and subject-specific approaches might then cover most needs over time.

While libraries need to build on their strengths to become an active part of the overall landscape for RDM in a university, they must also consider that the services offered are not set in stone but have to be adapted and developed continuously. As scientific practice evolves rapidly, a constant learning and innovation process is needed to keep up with changing requirements. Libraries – and other service providers – need to open up and reach out further towards researchers and collaborate more closely with the researchers, in order to establish a mutual learning process on the part of both libraries and researchers. Requirements of researchers will not only evolve through technical and scientific developments in their field of research, but it can also be expected that RDM will play a more prominent part in curricula and in good scientific practice in the years to come. Constant efforts in information and training on RDM should help to further implement it as an essential task of researchers with each new generation of, for example, doctoral students. A top-down commitment from universities can certainly support this, provided it is appropriately translated into activities, which really reach researchers at their workplace.

References

Blumer E and Burgi P-Y (2015) Data Life-Cycle Management Project : SUC P2 2015-2018. *Revue électronique suisse de science de l'information* 16: 1–17. Available at: http://archive-ouverte.unige.ch/unige:79346 (accessed 5 October 2016).

Corti L, Van den Eynden V, Bishop L et al. (eds) (2014) *Managing and Sharing Research Data*.

London: SAGE. Available at: http://ukdataservice.ac.uk/manage-data/handbook/ (accessed 7 October 2016).

DataCite (2016) *DataCite*. Available at: http://www.datacite.org (accessed 6 May 2016).

DLCM (2016) *Data Management Checklist*. Available at: http://www.dlcm.ch/ressources/data-management-checklist (accessed 22 September 2016).

docuteam (2016) *docuteam packer*. Available at: http://www.docuteam.ch/en/products/it-for-archives/software (accessed 10 May 2016).

ETH Zürich (2011) *Guidelines for Research Integrity*. Zürich: ETH Zürich. Available at: www.vpf.ethz.ch/services/researchethics/Broschure.pdf (accessed 7 October 2016). ETH Zürich (2016a) *ETH Archives for Contemporary History*. Available at: http://www.afz.ethz.ch (accessed 22 September 2016).

ETH Zürich (2016b) *ETH Zürich*. Available at: http://www.ethz.ch/en.html (accessed 11 May 2016).

ETH Zürich (2016c) *ETH-Bibliothek*. Available at: http://www.library.ethz.ch/en/ (accessed 11 May 2016).

ETH Zürich (2016d) *OpenBIS*. Available at: http://openbis-eln-lims.ethz.ch (accessed 2 September 2016).

ETH Zürich (2016e) *Critical Thinking Initiative*. Available at: https://www.ethz.ch/services/en/teaching/critical-thinking-initiative/ct-annual-programme.html (accessed 22 September 2016).

ETH-Bibliothek (2016a) *Digital Curation Office*. Avail- able at: http://www.library.ethz.ch/Digital-Curation (accessed 11 May 2016).

ETH-Bibliothek (2016b) *E-Rara*. Available at: http://www.e-rara.ch/?lang=en (accessed 6 May 2016).

ETH-Bibliothek (2016c) *ETH Zürich University Archives*. Available at: http://www.library.ethz.ch/en/Resources/Archival-holdings-documentations/ETH-Zurich-University-Archives (accessed 11 May 2016).

ETH-Bibliothek (2016d) *File Formats for Archiving*. Available at: http://www.library.ethz.ch/en/Media/Files/File-formats-for-archiving (accessed 11 May 2016).

ETH-Bibliothek (2016e) *Intended Purpose of Docuteam Packer*. Available at: http://www.library.ethz.ch/en/Media/Files/Intended-purpose-of-docuteam-packer (accessed 10 May 2016).

ETH-Bibliothek and EPFL Library (2016) *Data Management Checklist*. Available at:

http://bit.ly/rdmchecklist (accessed 22 September 2016).

European Commission (2013) *Guidelines on Data Management in Horizon 2020*. Available at:
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (accessed 22 September 2016).

Ex-Libris Group (2016a) *Primo Central Index*. Available at:
http://www.exlibrisgroup.com/category/PrimoCentral (accessed 6 May 2016).

Ex-Libris Group (2016b) *Rosetta*. Available at: http://knowledge.exlibrisgroup.com/Rosetta (accessed 10 May 2016).

Goodman A, Pepe A, Blocker AW et al. (2014) Ten simple rules for the care and feeding of scientific data. *PLoS Computational Biology*. 10(4): e1003542.

Piwowar HA and Vision TJ (2013) Data reuse and the open data citation advantage. *PeerJ* 1: e175 Available at: https://doi.org/10.7717/peerj.175 (accessed 5 October 2016).

Piwowar HA, Day RS and Fridsma DB (2007) Sharing detailed research data is associated with increased citation rate. *PLoS ONE* 2(3). Available at:
http://dx.doi.org/10.1371/journal.pone.0000308 (accessed 5 October 2016).

Sesartić A, Fischlin A and Töwe M (2016) Towards narrowing the curation gap: Theoretical considerations and lessons learned from decades of practice. *ISPRS Inter- national Journal of Geo-Information* 5(6): 91. DOI: 10. 3390/ijgi5060091.

swissuniversities (2016) *Rectors' Conference of Swiss Higher Education Institutions*. Available at: http://www.swissuniversities.ch/en/ (accessed 12 May 2016).

Thomson Reuters (2016) Data Citation Index. *Web of Science*. Available at:
http://wokinfo.com/products_tools/multidisciplinary/dci/ (accessed 6 May 2016).

Treloar A (2012) *Private Research, Shared Research, Publication, and the Boundary Transitions*. Available at: http://andrew.treloar.net/research/diagrams/data_curation_continuum.pdf (accessed 22 September 2016).

Author biographies

Ana Sesartic is an environmental scientist with a doctoral degree in Atmospheric Physics. She brings hands-on experience in working with and managing big data sets, and is now bridging the gap between scientists and librarians at the Digital Curation Office. Currently she is working on the Data Life-Cycle Management (DLCM) project.

Matthias Töwe is a chemist with a doctoral degree in Experimental Physics and a trained scientific librarian. Since 2003 he has been working at ETH-Bibliothek in different roles, first for the Consortium of Swiss Academic Libraries and later for the Swiss electronic library: e-lib.ch. Since late 2010 he has been head of the Digital Curation Office of ETH Zürich at ETH-Bibliothek.