Beyond Open Access - Tools and Methods for Open Research

Jenni Hyppölä
CSC IT center for Science
Keilaranta 14
FI-02100 Espoo
+358 50 381 8520
jenni.hyppola@csc.fi

Jessica Parland-von Essen
CSC IT center for Science
Keilaranta 14
FI-02100 Espoo
+358 50 501 2543
parland@csc.fi

Esa-Pekka Keskitalo
National Library of Finland
P.O.Box 15
FI-00140 University of Helsinki
+358 2941 23296
esa-pekka.keskitalo@helsinki.fi

ABSTRACT

This is an extended abstract of the workshop "Beyond Open Access - The Changing Culture of Producing and Disseminating Scientific Knowledge", organized by the Open Science and Research Initiative in the Academic Mindtrek Conference, Tampere, on September 24th 2015. The workshop included a short introduction to existing services, questions of sustainability and case study presentations followed by a workshop to map user needs for platforms and services for open scientific cooperation.

Categories and Subject Descriptors

H.m MISCELLANEOUS

General Terms

Documentation

Keywords

open research, open science, services for researchers, research tools

1. Introduction

Open Science began as Open Access for the traditional journal, but the digital revolution spurs a change much bigger than that. Opening up the research process into what we have called ultra open science changes the way in which scientific information is created, processed, managed and shared. Micropublications, version control, persistent identifiers, ontologies and linked data are requirements that become obvious as science turns more digital. Still most of our information systems, publications, and archiving methods reflect an older paradigm of the printed matter, which creates barriers for doing open science in an efficient and dynamic way.

AcademicMindTrek'15, September 22-24, 2015, Tampere, Finland
ACM 978-1-4503-3948-3/15/09.
http://dx.doi.org/10.1145/2818187.2818280

The Ministry of Education and Culture of Finland promotes research information availability and open science through the Open Science and Research Initiative (ATT), which is set out for the years 2014–2017. On 25th November 2014, the Ministry released The Open Science and Research Roadmap 2014–2017. The target is that Finland becomes one of the leading countries in openness of science and research by the year 2017, and to ensure that the possibilities of open science will be widely utilized at our society. In addition to this, the initiative aims at promoting trustworthiness of science and research, supporting the culture of open science within the research community, and increasing the societal and social impressiveness of research and science. The vision 2017 is that open research leads to surprising discoveries and creative insights. [1, 2]. This means a situation in which research data and materials move freely throughout society; from one researcher or research team to another, between disciplines, to innovative businesses, and to decision-makers and citizens. Information flow is facilitated by clear policies and best practices, and by providing services to safeguard the availability of scientific and research results." The Initiative has produced some services to support open data and a handbook for researchers. It is important to have open and good cooperation with researchers in order to meet their needs.

The goal of this workshop was to identify 1) existing tools and services meeting different needs during the research process, and 2) gaps in existing service infrastructure. These could be platforms, services, or tools that researchers require in order to share and crowd source their work.

The workshop was composed of two 1,5 hours sessions. The first one consisted of three presentations: one presenting services that Open Science and Research Initiative provides for researchers, and two practical case examples of open science methods. This session was live streamed and recorded. The presentations can be viewed on: http://livestream.com/ITstriimIT/beyond-open-access.

In the second session, we organized a roundtable discussion on existing and required open science services and tools. The workshop created valuable information for the Open Science and Research Initiative that coordinates and delivers services for open science in Finland. Due to limited space, we concentrate in this second session, i.e. the roundtable discussion in this extended abstract. All materials produced in the workshop will be shared openly. No scientific paper will be produced.

2. Open Science Tools – Reality And Requests

We had planned to organize discussions in small groups, and a roundtable discussion following the group work in the workshop. However, due to a limited number of participants (n<10), we organized a roundtable discussion and brainstorming around services as one group. The discussion was led by a researcher, and a facilitator. The discussion was based on three questions:

- What open science tools do you use in your research?
- What kind of open science tools would you like to use in your research?
- What kind of features do you expect from the tools you use or would like to use, and what kind of features would be inutile for you?

The questions were planned to be discussed separately but in the course of discussion, they often got mixed. This was the case especially for the second and third questions. Questions and results will be discussed in the following sections. We will list tools and services that were named in the discussion. Most of them were used by some of the participants, or they were somehow familiar with them. Taking into account the small number of participants of the workshop, the following is only a snapshot of tools and services offered for researchers. We are neither aiming at a comprehensive list of all the available tools and services, nor claiming our sample would be statistically representative. On the other hand, the participants came from different disciplines and universities, and among them were pioneers of ultra open science. Thus, we believe in gaining valuable information for the Open Science and Research Initiative's service development as planned.

2.1 Open Science Tools Used In Research

The first question, "What open science tools do you use in your research", appeared to be loosely formulated as a definition of an open science tool varies. In a narrow sense, an open science tool was defined in the discussion as a tool that enables opening up a research process. In a broad sense, an open science tool may be anything that helps open research in some way (thus, it may sometimes be used also in "normal" or "closed" research). These tools include e.g. a large number of software. For instance, *Skype* (www.skype.com) and *Slack* (www.slack.com) were named as tools of communication between team members, *Word* and *LaTeX* were mentioned as text processing tools, and *AtlasTI* and *SPSS* as data processing and analysing tools. *Rmarkdown* (http://rmarkdown.rstudio.com) and *iPython* (http://ipython.org/notebook.html) can be used to create different types of documents by using computing languages (R and Python, respectively).

Some of the aforementioned applications provide some open science features and options for collaborative working method. However, we limited our definition of open science tools, software and applications, to those that fully support open collaborative working method, or open dissemination of research results and outputs. Based on the above, we distinguished three categories of open science tools: those supporting 1) research, 2) dissemination, and 3) archiving or permanent storage. However, some services may contain properties of more than one category, and thus, they are partly overlapping. Some tools are also compatible, and can be used together e.g. via extensions.

The first category, tools supporting research, include a number of software and services that enable researchers to work in a collaborative manner. In this category, we identified three subcategories, namely 1) platforms, 2) reference tools, 3) sharing services.

The platform subcategory includes first blog platforms, such as *Blogspot* (www.blogspot.com) and *Wordpress* (www.wordpress.com), which offer researchers a way to share their ideas or research in blogs. The NMRlipids research blog (www.nmrlipids.blogspot.com) is an example of ultra open science, using Blogpsot platform. All research on the blog is openly shared and anyone can participate by offering insights or technical help on blog comments. However, blog platforms are not designed for conducting collaborative research and thus, they lack some properties we will discuss later. Another type of platform is *Github* (www.github.com) that enables sharing files (including code) and collaborative working on code, i.e. social coding. In the discussion, Github was estimated to be the most used and widespread, and it is strengthening its position. It offers collaboration, code review, and code management for open source and private projects, and it is free of charge for public projects [3]. *Authorea* (www.authorea.com) was mentioned in the discussion as a platform option, but none of the participant had personal experiences of it. SatureApp (www.satureapp.com) was the only tool mentioned for collaborative qualitative research. We also counted *Google Drive/Google Docs*, and *Overleaf* (www.overleaf.com) among platforms that enable collaborative working method, such as simultaneous editing of text or other types of documents.

The second subcategory: reference tools, consist of software or services designed for collecting and organizing references. Some of them offer social aspects, such as creating groups or libraries that can be shared with (a limited number of) collaborators. In the workshop, *Mendeley* (www.mendeley.com), *Zotero* (www.zotero.com) and *EndNote* (www.endnote.com) were mentioned. The payable version of EndNote is offered by some universities for their researchers.

In the third subcategory, namely sharing services, we classified services that are mainly focusing on organizing and sharing files, not necessarily modifying them within the service. Examples of this kind of tools include *Dropbox* (www.dropbox.com) and *Pearltrees* (www.pearltrees.com).

Dropbox and Pearltrees can also be used for dissemination to a wider public, and thus, they can be included in the second main category of services, namely dissemination. There is a great number of tools and services for dissemination, Twitter (www.twitter.com), *Slideshare* (www.slideshare.com) and *ResearchGate* (www.reserachgate.com), to name a few examples of different types of services. These are aimed at sharing results, such as articles and presentations, whereas data sharing services were classified under the next category. ResearchGate can also be seen as a social network type of service.

The third category, archiving or permanent storage services, includes organizational, national, and international tools and services that enable storing, opening and sharing research results and data. For instance, *Zenodo* (www.zenodo.org), run by CERN and developed under the EU FP7 project OpenAIREplus, *DataDryad* (www.datadryad.org) (payable), *AILA* (https://services.fsd.uta.fi/index), *IDA* (access via https://sui.csc.fi/web/guest), and *AVAA* (http://avaa.tdata.fi/) offer possibilities of storing and sharing research data. (AVAA and

IDA are offered by the Open Science and Research Initiative, and provided by CSC). However, in the workshop it came apparent that none of these provide a universal solution for all requirements. IDA is reserved mainly for researchers and projects hosted by Finnish academic institutes, or funded by the Academia of Finland. Zenodo was seen as an easy to use solution, open for all and provided by a recognized organization. Zenodo also provides DOIs as permanent identifiers. However, it only suits as the final destination for the data, as data cannot be removed and it is always public, at least after an embargo period. Zenodo also has a size limit for files, in general, they only accept files up to 2GB.

Yet another important resource for researchers was noted, namely open "ask and answer" discussion boards on the internet, such as *Stack Overflow* (http://stackoverflow.com/) where anyone can ask a question about different matters, or propose solutions to questions asked by others.

2.2 Requested Open Science Tools And Services

The second question was "What kind of open science tools would you like to use in your research?" The discussion revealed gaps in existing offer on open science tools. At the same occasion, we discussed also the third question: "What kind of features do you expect from the tools you use or would like to use, and what kind of features would be inutile for you? "

The questions were tackled in relation with a metaphor of research process as a cycle, with the following phases: Hypothesis, Data Collection, Processing, Storing Data and Results, Long-Term Preservation, Publication and Distribution, and Reuse. This is a model created and used by the Open Science and Research Initiative. It aims to be universal in the sense that it recognizes phases, which usually take place during a research project. However, as there are remarkable differences between disciplines and sometimes even within a discipline, this approach has its limits. Despite the limits, we considered this model as a good starting point for discussion.

A missing tool identified by researchers was a specific project management tool designed for open science with collaborative working methods. In open research process, collaboration may be based partly or totally on voluntary efforts and thus, the "leading researcher" has no formal authority over the participants. The smaller the required effort is, the easier it is to engage people. Thus, it may be beneficial to divide the research in very small tasks. A project management tool or service should support microwork or microtasking (i.e. small tasks of a large project are completed by many people, and tied together online). The service should also take into account the characteristics of scientific work, providing tools and support in every phase, starting from the project idea and funding application preparation.

Funding agencies impose requirements that must be taken into account. The research plan approved by a funding agency must be respected, as a funding contract is strictly binding. Requesting changes in the plan is complicated and expensive, yet impossible, and modifications in the plan may weaken one's credibility. This leads to imprecise project plans or projects aiming at granted results. Therefore, funding agencies should allow some flexibility and ease adapting the project plan if required. It might thus be valuable if a tool or service could improve their involvement and knowledge about the research and it outcomes by means other than formal reporting.

It was also noted that in open science, publication and dissemination are essential parts of the process from the very beginning. Open collaboration and peer review are also present all the time, not only when judging results presented in an academic article.

From the open science perspective, the concept of a scientific paper as the main forum of discussion and verification of results is outdated. Peer review providing corrections is a crucial part of the scientific method. In many cases, data and code cannot be widely reviewed and corrected afterwards but the quality control must be present throughout the research process. Thus, there is some analogy to lean startup philosophy. One must learn fast, adapt work, and adjust the parameters accordingly to ensure the research proceeds in an efficient and scientifically sound manner.

Tools for social annotation, tagging possibilities, and attributing permanent identifiers, e.g. DOIs, to unconventional references, e.g. blog posts or comments, were also identified important and required features that are missing at the moment. Social annotation is needed for linking general discussion to a certain discussion, early publishing, and when working jointly on the text. This might be complicated, on the other hand, some parts may exist, but as far we know, these cannot be found in a one single service. Permanent identifiers are required as references must be permanent, and thus, normal urls cannot be used as such. On the other hand, when using permanent identifiers, the text must remain unchanged[1]. Use of permanent identifiers rises an issue of organizations providing them, as they require a formal process. On the other end of annotating, hashtags are widely used in social media with no formalities or additional work. Permanent identifiers are, however, a key point is the long term accessibility and usability of the document in question.

3. Conclusions

A great variety of tools and services for open science research already exists, but as for our understanding, there is no ultimate service responding to all needs researchers have. It may be disputed whether such a service could be created as the needs vary from discipline to another, as well as on the individual or group level. For instance, research groups or collaborations vary in their size and location. Researchers, or participants of an open science research project may never meet each other personally, but all communication depends on digital services. Some researchers utilize large amounts of data in their research and require a lot of data storage capacity, perhaps also computing power. For some others, the storage of data is not an issue, but they must take legal and ethical concerns into account. This applies e.g. to medical, human, and social studies. The challenges and issues the researchers face vary from one to another. The services offered to researchers must adapt and address only the needs that are relevant to the researcher in question.

Open science requires tools in every phase of the research process to ease communication and dissemination. Accessibility is a crucial part of openness. "Open" is not open unless it is known, accessible and reusable by everyone. All these aspects must be considered when planning and creating tools and services. Usability and user experience are also important. Easy access

[1] For instance W3C standards: every draft has a permanent identifier.

demands understandable terms of use and technical solutions that are compatible with host organizations' policies in terms of e.g. legal and computing security matters. Price matters, too. Some universities offer certain software for their researchers, or they can be purchased, whereas others have more limited options or restricted policies. Time is an issue, and therefore services should be easy to use and adopt. In short, if a researcher wants to start using certain tool or service, neither negotiations nor user training should be required. This imposes challenges for services, from user experience to the service's business model.

It is also important to keep in mind that tools and services are only means to help researchers. Changing research culture to be open and turning all research to open science requires partly radical changes in research and publication culture, in people's attitudes, and their ways of working. For instance, opening data and culture of reusing data should be more heavily emphasized. Challenging one's own attitudes and habits is often the hardest part, but it is required to change a culture.

4. ACKNOWLEDGMENTS

The authors thank the Open science and research initiative in Finland for the opportunity to prepare and conduct this workshop under the project framework. We also want to thank Infocrea Oy and CloudStreet Oy for making live streaming and recording of the workshop possible.

5. REFERENCES

[1] http://openscience.fi/about

[2] The Ministry of Education and Culture's Open Science and Research Initiative 2014–2017 2014. The Open Science and Research Roadmap. Juvenes Print – Suomen Yliopistopaino Oy.

[3] www.github.com. Accessed September 25th, 2015.