Could Open Research benefit Cambridge University researchers?

Report from a discussion with researchers on 8 June 2016, Department of Engineering

Written by Lauren Cadwallader, Joanna Jasiewicz and Marta Teperek

Since its inception in January 2015, the Office of Scholarly Communication has largely focused on providing services to researchers around open access for publications and research data management and sharing. This was born out of necessity to comply with funder requirements. However, should the University be doing more than merely ensuring compliance with funders' policies? Should it focus more on the bigger picture: what are the benefits of open access? And perhaps we should start thinking about "Open Research" as a whole and making sure that we encourage transparency and sharing across the whole research process. Open research does include open access and data sharing, but most importantly, also open access to theses, to book chapters, to presentations, protocols, reports and all other types of research in all flavours that it comes.

The rationale behind organising a discussion with researchers on the 8th June was to learn more about the aspects of the research process that they found frustrating and which could be improved if the research process was more open. More importantly, however, we asked researchers if open research could solve (some) of these problems and what the potential solutions might be. This blog posts summarises the solutions put forward by the researchers and invites the broader Cambridge community to discussion on whether the adoption of some of these solutions could benefit Cambridge academics and the research community as a whole.

We believe that at least some problems discussed at the event could be addressed if the researchers and the University worked more actively towards promoting open research.

Working together with researchers – democratic approach to problem-solving

To get an initial idea of what are the expectations of the research community in Cambridge, we created a simple registration page and anyone who wished to register for the event was asked three questions: **what frustrates you** about the research process as it is now? **Could you propose a solution** that could solve that problem? Would you be willing to speak about your ideas publicly?

Interestingly, around fifty people registered to take part in the discussion and almost all of them contributed very thought-provoking problems and appealing solutions to solve them:

Problems with research as it is
- Unfair reward mechanisms: the only thing that counts in academia is publication in a high impact factor journal
- Lack of article-level impact measurements and evaluation
- Rewarding of "the fast and loud research" while the slow and careful is valued less
- The quantity of produced research matters more than its quality
- Problems with re-usability of data and pressure on publishing only positive results
- No incentive for researchers to engage with open research - focus primarily on compliance
- Exploitative publishers

Solutions

- Suggestion for recruitment decisions:
  - Commitment to Open Research should be an essential requirement in job descriptions
  - Requirement for the reproducibility of research as an employment criteria instead of focusing on the h-index
- Development of training and outreach programme:
  - Department-specific training on open research
  - Awareness raising about the benefits of data sharing
  - Promoting ethical publishers
  - Raise awareness about pre-print services
- Suggestions for policy development:
  - University might adopt an institutional licence on scholarly publishing to protect the rights of authors
  - Three key elements needed to be made available: the raw data supporting the publication, the source code and the methods
  - Non-positive research results need to be shared
  - Formalised data management assessment to help to improve data management practice
- Role for publishers:
  - Employ in-house data experts and implement mechanisms to measure the quality of submitted data
- Role for funders:
  - Grant schemes aimed at reproducing existing research
  - Grant renewal schemes for researchers who focus on doing reproducible research
  - Mandate data sharing in a similar way as open access to publications

To our surprise, **half of the people expressed their will to speak publicly about their ideas**. We selected four people to speak about their frustrations and four to speak about possible solutions.

Frustrations

Dr Lauren Cadwallader from the Office of Scholarly Communication facilitated the event and set the scene for discussion. We first invited our four selected speakers to explain to the audience what they found frustrating about the research process.

Flawed metrics

Dr Jenny Molloy
Coordinator, Synthetic Biology Strategic Research Initiative and Open Plant
Director, ContentMine

Jenny has been an advocate for Open Research for a while and she coordinates several projects into the 'Open'. She mentioned that it is difficult to get most researchers engaged with the Open Research ideas. Instead, they worry more about compliance with journal policies and the impact factor of the journal they wish to publish in. However, t**he current metrics of impact do not reflect the real importance of the research described**. What is needed is not a journal-level impact measurement, but article-level impact measurements. But it is difficult to solve this systemic problem: **any new journal which wishes to introduce new metrics system has no journal-level**

impact factor to start with, and therefore researchers do not want to publish in it thus it is doomed for failure.

Jenny also mentioned that more training effort is needed to ensure that researchers are equipped with skills to allow them to properly use the internet to disseminate their research, as well as with skills allowing them to effectively manage their research data. **It requires effort and skills to make research open, re-usable and discoverable** by others, and if researchers are to sign up for the Open Research agenda, they need to be equipped with these skills.

Academia feeding money to exploitative publishers

Dr Corina Logan
Leverhulme Early Career Research Fellow, Department of Zoology

The talk from Corina further elaborated on the publishing frustrations introduced by Jenny. Corina raised the issue of exploitative publishers. At the moment, the cost of publishing is around £3,777 on average per article. The big four publishers (Elsevier, Wiley, Springer and Informa) have a typical profit margin of 37% per article. **It means that on average £1,397 is leaving academia each time an article is published with these publishers. Articles which were donated to the publishers for free by the academics, and reviewed by other academics, also free of charge**.

Cost/benefits analysis when publishing with an exploitative versus ethical publisher
Corina then undertook a compelling comparative analysis of cost/benefits when publishing with an exploitative versus an ethical publisher. When publishing with an exploitative publisher, the researcher does not have to pay anything, but it costs academia £3,777 as a whole per publication. The goal of an exploitative publisher is to maximise profits, which go directly to shareholders. The end product, the article, is available to subscribers only, which is indirect discrimination (not everyone will be able to afford to read it). If a researcher decides to publish with an ethical publisher, the cost of publishing is £150 - £2,042 (paid by the researcher or the funders of research) and since the goal of an ethical publisher is to share research, profits stay mostly within academia. Additionally, articles are available open access to everyone, which in turn brings more readers, higher citation rates and potentially more prestige to the authors.

Corina's presentation spurred a lot of interest and many supportive voices. Another frustration added by one of the researchers in the audience was that **most researchers are completely unaware of who are the exploitative and ethical publishers and the differences between them**. Researchers typically do not directly pay the money to the exploitative publisher and are therefore not interested in looking at the bigger picture of sustainability of scholarly publishing.

Reproducibility crisis

Dr Avazeh Ghanbarian
Research Associate at the Department of Genetics

Avazeh started her talk stating that having 9 years of experience as a bioinformatician, she is frustrated about discrepancies between published papers and their corresponding published datasets. Frequent lack of metadata and detailed methods and also occasional unavailability of raw data, makes it impossible to reuse the published data. She quoted the results from a recent study on

research reproducibility [published in *Nature*](). According to this article, 52% of 1,500 surveyed researchers said there was a significant crisis of research reproducibility and 90% of researchers said they were concerned about the low level of research reproducibility.

The benefits of sharing data
Subsequently, Avazeh reminded the audience about the importance of data sharing. Some people do it because the journals ask for data and because data are necessary to validate the findings described in the paper. But most importantly, research data needs to be shared to advance science: to ensure reproducibility (providing the researchers with the possibility to compare and critically assess their results) and re-usability (the same dataset can be used to answer more than one question, and data creators can earn extra citations thanks to this).

Most data is not re-usable
Avazeh concluded that most of the data she has been working with did not satisfy the criteria of reproducibility, nor re-usability. Most of the time there was not enough metadata available to successfully use the data. This leads to frustrations. **Unless the data sharing policies are improved, the missing miraculous method to reach the results and conclusion reported from the little data shared would prohibit any effort to reuse or reproduce.**

Publishers don't respect authors' rights

[Dr Ross Mounce]()
Postdoc at the Department of Plant Sciences

Ross contributed yet another voice of frustration about exploitative publishers. He was invited by a guest editor for the Association for Information Science and Technology to write an article about [altmetrics and open access](). He got assurance from the editors that the article and indeed the whole special section would be open access - he wouldn't have agreed to submit there if it wasn't open access. It was initially freely available from the original society publisher website, but after some time Wiley made the article available behind a paywall on their website for $45 per copy (inc tax). It took Ross a long time to argue with Wiley to remove the paywall from his article, and the rest of the special section is *still* paywalled. Ross was frustrated about the ways in which Wiley did not respect his rights as an author.

Universities should protect researchers from exploitative publishers
He suggested that **it is outrageous that publishers can take away all copyright from the authors** and that perhaps universities should protect their researchers from being exploited by publisher copyright transfer agreements. And this is part of general frustration about the lack of copyright and legal issues awareness among researchers. Ross suggested that perhaps Cambridge could adopt an institutional licence on scholarly publishing (similar to that used by [Harvard]() and [Princeton]()) which could protect the rights of the authors.

Other frustrations

Discipline-specific norms need to be respected
These initial points started a long discussion with the audience about related research frustrations. Two important contributions were made by researchers from Arts and Humanities. Firstly, it frustrated them that most of the time **only the fast and loud research got rewarded, and the slow and careful seemed to be valued less**. Secondly, a concern was raised that discipline-specific norms need to be taken into account. Researchers from Arts and Humanities publish mostly monographs,

book chapters and books, and making them accessible open access presents different challenges than making a journal article openly available.

Quality vs quantity
Researchers from the sciences also supported the idea that the slow and careful research needed to be rewarded. Someone mentioned that the volume of produced research is higher and higher in terms of quantity and science seem to have entered an 'era of quantity'. This raises a concern that **the quantity matters more than the quality** of research. Researchers are under pressure to publish and they often report what they want to see, and not what the data really shows. This approach leads to reproducibility crisis and lack of trust among researchers. So perhaps mandating all research data to be open would not be helpful, as this could lead to even more research of a poor quality being released into the public domain.

After a brief discussion the audience agreed that, **as a minimum, everything that directly supports publications needed to be shared**, and that good quality non-positive research data (produced using appropriate scientific methodology) needed to be published as well in order to move scientific knowledge forward.

Impact Factor is all that matters
The discussion around data sharing quickly came back to the frustrations with journal impact factor being the only metric taken into account when appointing new researchers. It was noted that as a result, early career researchers had no motivation **to share their data and to publish their work in open access journals, which usually have lower (if any) impact factor**.

Solutions

Dr Avazeh Ghanbarian

Research Associate at the Department of Genetics

The minimum requirements for making shared data useful
Avazeh [proposed solutions](#) to make research data more re-usable. She suggested that in order to make research reproducible, at least in bioinformatics, **three key elements needed to be made available: the raw data, the source code and the methods**. Raw data is necessary as it allows users to check if the data is of a good quality overall, while publishing code is important to re-run the analysis and methods need to be detailed enough to allow other researchers to understand all the processes involved in data processing. Avazeh referred to an excellent case study [example of Daniel MacArthur](#) who has described how to reproduce all the figures in his paper and has shared the supporting code as well.

Publishers could invest money in improving science
As a solution to lack of proper data sharing, Avazeh proposed that publishers could play an important role in improving the process. She proposed a couple of simple mechanisms that could be implemented to improve the quality of data shared:

- Employment of **in-house data experts**: bioinfomaticians or data scientists, who could judge whether supporting data is of a good enough quality
- Ensure that there is **at least one bioinfomatician/data scientist on the reviewing panel** for a paper

- Ask for the **data to be deposited in a public, discipline-specific repository**, which would ensure quality control of the data and adherence to data standards.
- Ask for the **detailed methods** to be made available as well
- Reward publishing the **source code**

Responsibilities of funders and research institutions
Avazeh also stated that funders and research institutions have important roles to play in ensuring that good research data was shared and that it was of sufficiently high quality.

Funders should implement mechanisms to measure whether data is of a high enough quality. Perhaps one of these mechanisms could rely on funding not only novel research (as it seems to be at the moment), but also **rewarding people who want to reproduce existing research**. Additionally, reproducible research should be rewarded. One possibility could be **grant renewal schemes for those researchers who focus on doing reproducible research**. And finally, Avazeh suggested that funders should mandate data sharing in a similar way as they mandate open access to publications.

Avazeh continued that rewarding reproducible research should also be a responsibility of institutions. They should not only educate the next generation of researchers on how to do reproducible research, but also **embed reproducibility of research as an employment criteria** to be used instead of the h-index. This is especially important since problematic research papers often get cited for being wrong; criteria like h-index are blind in assessing this.

Dr Alasdair Russell
Head of Pre-clinical Genome Editing Lab at Cambridge Cancer Centre

Making data available internally and externally
Alasdair explained that he was leading a large team which produces two types of data: 'embryonic' data and 'mature' data. He stressed that it was relatively easily to share the 'mature' data. On the other hand, the 'embryonic' data, which is immature and preliminary, should not be publicly shared (to ensure the focus on quality in research, and not on the quantity). However, 'embryonic' data needs to be searchable at least within the limited collaborative space to ensure there is no redundancy and people do not waste time repeating the same experiments. He stressed that having all the data available digitally is mandatory for this to succeed.

Use of pre-print services
Alasdair also talked about the value of sharing pre-print papers on platforms like arXiv and bioRxiv. These services allow researchers to share their manuscripts before they become available from the publishers websites. In physics, maths and computational sciences it is common to upload manuscripts pre-publication (and even before submitting the manuscript to a journal) to arXiv in order to get feedback from the community and get the chance to improve the manuscript.

bioRxiv, the life sciences equivalent of arXiv, started relatively recently. Alasdair mentioned that **he was initially worried that uploading manuscripts into bioRxiv might jeopardise his career as a young researcher**. However, he then saw a pre-print manuscript describing research similar to his published on bioRxiv. He was shocked seeing how the community helped to change that manuscript and to improve it. He has since shared a lot of his manuscripts on bioRxiv and this has never hurt him. To the contrary, Alasdair suggested that **using pre-print services promotes one's research**: it allows the author to get the work into the community very early, get feedback, talk to people. And

people will value good quality research and the value and recognition among peers will come back to the author and pay back eventually. Additionally, someone from the audience suggested that publishing work in pre-print services provides a time-stamp for researchers and help to ensure that ideas will not be scooped by anyone – researchers are free to share their research whenever they wish and as fast they wish.

Should researchers be nudged to use pre-print services?
It was mentioned that perhaps researchers could submit their manuscripts to journals after they have been posted on pre-print services. The Rockefeller University Press already enables this and this has already been proposed among biologists. Someone has suggested that perhaps the University of Cambridge should mandate the use of pre-print services. However, the views were mixed as the members of the audience thought that such behaviour should be definitely encouraged, but not mandated.

Ralitsa Madsen

First year PhD student at the Institute of Metabolic Science

Data management inspections
Ralitsa Madsen spoke about tangible solutions to problems with poor quality data. **Research data need to be properly documented and maintained to ensure research integrity and research continuity**. Many researchers have heard or experienced first hand horror stories of having to follow up on somebody else's project, where it was not possible to make any sense of the research data due to lack of documentation and processes. This leads to lot of time wasted in every research group. So perhaps a formalised structure of data management could help to improve data management practice. This could help to ensure that as a minimum, every researchers has a lab book to document the procedures.

The idea seemed to be appealing to researchers in Arts and Humanities as well. Someone mentioned that good data management is very difficult especially for researchers doing fieldwork. During fieldwork it is challenging to document adequate metadata and to record all the notes scribbled while in the field.

Funders now require data management plans as part of grant applications; however they do not follow up on these plans, so researchers have no incentives to adhere to them and see them as a tickbox exercise.

Dr Corina Logan

Leverhulme Early Career Research Fellow, Department of Zoology

Corina proposed two different types of solutions: solutions which can be implemented here and now by every researcher, and solutions which can be adapted institution-wide to better reward transparent research.

Everyone can contribute to Open Research
Corina inspired the audience by **empowering everyone to take the lead in encouraging Open Research**. The simplest way to start is to submit articles to journals which are fully Open Access. This

should be accompanied by making one's reviews openly available. Corina explained that having this commitment in mind improved the quality of the reviews of papers where she was able to publish the review history. All publications should be accompanied by supporting research data and researchers should ensure that they evaluate individual research papers and that their judgement is not biased by the impact factor of the journal.

Additionally, Corina suggested that every researcher could contribute to the community move to Open Research. This could be as simple as approaching scientific societies within their disciplines and suggesting them to evaluate alternative business models and change their journals to fully open access.

Open Research should be an essential requirement in job descriptions
Corina then suggested that perhaps **the University of Cambridge could include a commitment to Open Research as an essential requirement in job descriptions**. Applicants could be asked at the recruitment stage how they achieve the goals of Open Research. Corina mentioned that the LMU University in Munich had recently included such a statement in a job description for a professor of social psychology.

Other solutions

Researchers then exchanged several other ideas related to how the research process could be improved to benefit researchers. There was a suggestion of **boycotting exploitative publishers** and a discussion about the need for **training in open research** skills embedded within departments to provide **discipline-specific advice**.

Help from the Office of Scholarly Communication
It was suggested that potentially **the Office of Scholarly Communication could mandate data statements in papers** and perform some simple checks:

- Is there a data statement in the publication?
- If there is a statement – is there a link to data?
- Does the link work?

This is certainly an interesting idea that could be further explored.

Guidance versus policy
Lot of discussion during the event related to the role the University should have in supporting or promoting Open Research. Some people advocated the University creating a new policy, arguing that there is no point of having guidance if there is no requirement to put the guidance into practice. Others supported the idea of the University having a more modest guiding role, suggesting that researchers already have enough mandates to comply with and that problems cannot be solved by policies – what is needed is a cultural change instead. It is clear that **discipline-specific training and guidance around how to practise open research and what benefits open research has is desired by Cambridge researchers**.

Next steps

We are extremely grateful to everyone who came to the event and shared their frustrations and ideas on how to solve some problems. Joanna who kept noting all the ideas on post it notes was extremely busy, which indicates how creative the participants were in just 90 minutes. It was a very productive meeting, we wish to thank all the participants for their time and effort.

The ideas generated during the discussion were captured on post it notes.

Outlooks for the future

Some researchers said informally after the meeting that they **did not have hopes that the University would address the above discussed problems**. At the Office of Scholarly Communication, we are optimistic as we have seen during the event the level of frustration about the research process as it is at the moment and an appetite for change and **good will**. We think that **by acting collaboratively and supporting good ideas we can achieve a lot**.

As an inspiration, McGill University's Montreal Neurological Institute and Hospital (the Neuro) in Canada have recently adopted a policy on Open Research: over the next five years all results, publications and data will be free to access by everyone.

Immediate steps

We will brief the Open Access Project Board on the results of this discussion at its October meeting. We will conduct a follow up questionnaire on all the ideas you have shared with us to learn more about the frustrations and solutions related to the research process within the wider Cambridge community.

If you would like to host similar discussions directly in your departments/institutes, please get in touch with us at info@osc.cam.ac.uk – we would be delighted to come over and hear from researchers in your discipline.

In the meantime, if you have any additional ideas that you wish to contribute, please send them to us at info@osc.cam.ac.uk. We will be delighted to hear from you.

Everyone who is interested in being informed about the progress here is encouraged to sign up for a mailing distribution list here.