Open Research Case Studies: Faculty of Biological Sciences

Table of Content

Open Research Case Studies – School of Biology with Amanda Bretman

What does open research mean to you?

For me, open research is about making our research practice and output as transparent as possible to different types of external audiences; doing that in a way that encourages the robustness of our research and allows as many people as possible to interact with it.

How does your research use open research practices?

Open research is very common in my field: we've been making datasets available for quite a while. I can remember the first occasion we had a paper accepted, and the editor asked for a data accessibility statement; this was back in 2005. At that time, I didn't understand what does open research mean. I think it is becoming more common you have to have that data available in the paper. It is also becoming more common to share statistical analysis and codes. It is not something we have done yet, but I would like to do it in the future, so our next papers will have both the data and all the markdown files as well. There is quite a lot of sharing of resources within my research community which is on fruit flies, and lots of people share their mutant or selection line.

How do you deposit your work?

I use both ORCID and Symplectic; they pick them up automatically. In the Faculty of Biological Sciences, we can send our accepted manuscript and a few different figure files to a generic email address. It gets uploaded for us, and I think it is helpful for academics to take out that step. It seems like such a small step to do, but if it's something that you maybe do only a couple of times a year, it is quite difficult to remember what all the steps were.

Do other Faculties have this system, or if not, why?

This difference between the Faculties has only come to my attention recently: as part of my role as Dean of Research Quality, I am thinking about the REF and the support we have. I thought we had made sure everyone's outputs were open access. I was speaking to people, and it turned out that's not necessarily always the case, partly because they had somebody who was doing that support of depositing the work of academics who is now moved on. So, there's a bit of a gap there. I think it's just down to whether the Faculty realises that was something really helpful.

"Getting some more consistency across the Faculties is something we need."

In your field, have you found it difficult or easy to identify other people's open research practices?

I think a lot of the things I talked about are very standard in my field, but what I think is quite difficult is going back to datasets to extract the data from them in a way that is useful for others. We found some barriers to that, potentially because people don't have the correct metadata,

which makes the data useful. It is not good enough making the data available if people can't use it.

How does open research inform your teaching practices?

If the students are involved in, for example, their final-year projects are aligned with the research in my lab, occasionally they will be on papers we publish, so they are involved in that. I don't think we train students in open research at all. I suggested in our School meeting that we could bring it into research projects, and I think it would help us because sometimes I get datasets from students that have been done in my lab, and I am not sure how to use them.

I don't know whether we talk to them enough about the process of doing research: about how that research is funded and gets into the public domain. We've talked to them about publishing, and that touches on open access. Given where the money comes from, we must make those outputs available. I think we could probably do more in terms of teaching open research.

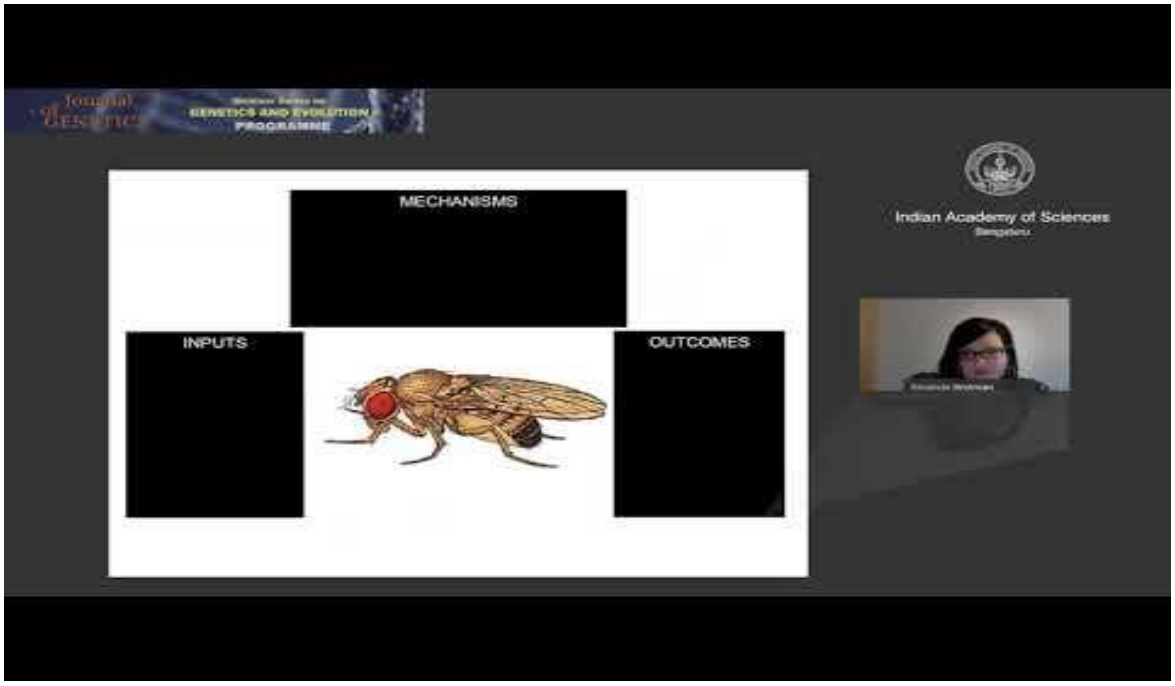
"The Open Research Case Studies project is very important because I think many people perhaps misunderstand what open research is. I mentioned it in one of our School meetings, and I got some blank faces. It's not because people don't do it but because they don't understand what it is and how they are already doing parts of it."

Do you use preprints?

I occasionally use preprints. I think preprints are useful when you are applying for a grant, and you want to show the reviewers that this is a manuscript that is coming through. I don't know how useful preprints are in getting feedback, which you get when you turn it into a peer-

reviewed and published manuscript. I am not sure where I feel about preprints. I have never read a preprint without having a good reason to do so.

No fly is an island: How Drosophila respond to socio-sexual environments



In what ways do you deposit the preprint of your work?

It would be bioRxiv.org or EcoEvoRxiv, so discipline-specific repositories.

Do you have concerns about preprints?

There is a reason why, for all its faults, we rely on peer review before publishing as the gold standard of evidence. It concerns me if people are using a non-peer-reviewed preprint as evidence because it hasn't gone through that process yet.

External to academia, preprints are not well-understood, which can cause confusion. Even internally, I think it can cause confusion when people use the preprint rather than the final published version because they can change a fair amount. If they are not linked properly, it can cause some problems.

Has your attitude toward open access changed in the recent years?

Yes, I think my attitude has changed: I've become more worried about the business model. Ultimately, it's a great idea that everybody should be able to access that knowledge: that is what we want to be able to do. However, some of those open access fees are disgraceful, and

it bothers me that it causes biases in what gets published and where. We are a rich country that can afford some of those fees, but that's not true in other countries: even within the country or the same lab, some differently funded work may or may not be able to get published somewhere that has gold open access as standard. So, the unfortunate side effect of publishing open access does worry that it may cause publishing biases because of who can afford those fees.

In your opinion, what is the current state of open access in your field?

Open access started with the drive from the funders and journals, like PLOS and eLife, who were making open access more common. A lot of the research in my lab has not been funded by UKRI, which makes it slightly more difficult to publish fully open access. That's why I think having the repository we can use to archive the accepted manuscripts has been brilliant. Of course, we always share a paper if asked directly.

Open access is becoming more common, but it depends on how the work has been funded and whether or not people can afford the full open-access fee. So, I think the pictures are still quite mixed.

Are there any negative attitudes toward open access you have come across?

Apart from the costs, I don't think I've heard anything negative about it. If you have published something a bit controversial, people may be contacted more, because it is open access, but I haven't heard of any negative impact.

What does data mean in your field?

The data is often spreadsheets of numbers from various things. So, archiving data means just an Excel spreadsheet. It can also be sequencing data, for example, genomic data. Sometimes we have images as data, but that's rare in my field, so data is mainly spreadsheets of numbers.

How do you manage your data?

We have OneDrive now, and I think that's much more secure, so nobody is wandering around with flash drives with data on them, which is what used to happen. We can also share that much easier between us. Our data is not particularly controversial; we never use human data, so security is not a high priority.

Have you used or published open data?

We haven't published anything as open data yet, but I am currently involved in a project where we are doing a really broad meta-analysis. We haven't got to the point where we're extracting data yet, but it's coming down the line.

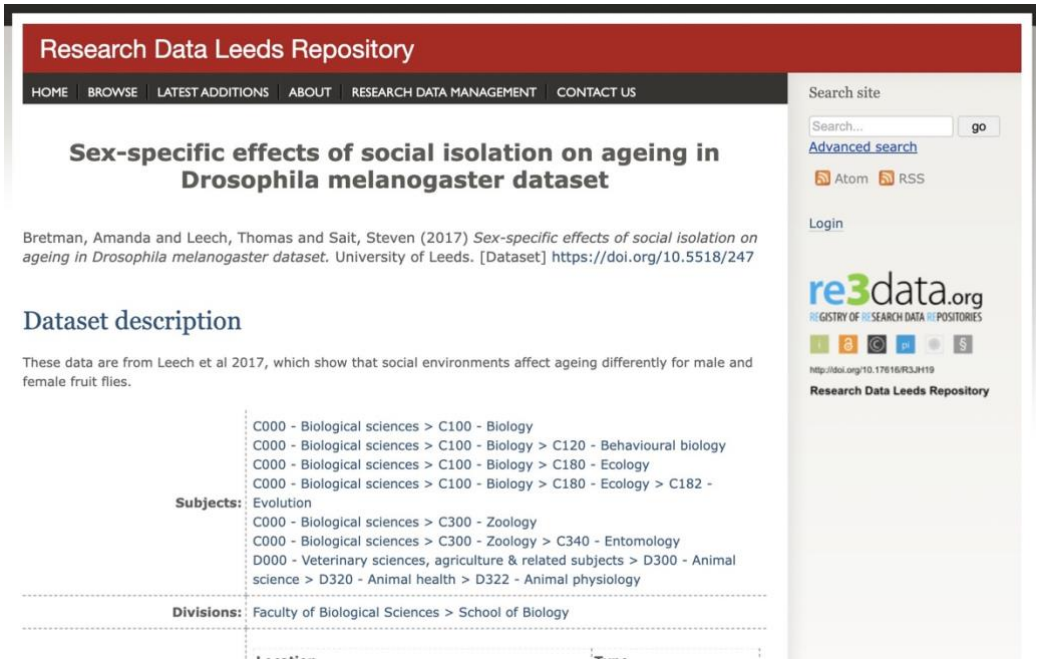
Have you deposited dataset with Leeds's data repository or elsewhere?

Yes, and that has been useful. Some journals we publish in specify the repository, that might be Dryad or the genetic sequencing data repository. There are specific repositories that everybody uses, however, when it isn't specified, I've been using the Leeds repository, which has been straightforward. I have had lots of support from the Library, so I am a big fan of it.

Have you written a data management plan?

For grant applications, yes. I used both the funder's guidance, as well I got help from research data management team. If you are asking them for money, you have to use the guidance of your funder, particularly from UKRI.

Research Data Leeds Repository: Sex-specific effects of social isolation on ageing in Drosophila melanogaster dataset



What kinds of open software do you use?

Our data is maintained in Excel. We do most of our analysis, occasionally an SPSS, mostly in R these days. There are various bioinformatic programmes that we use for sequence analysis too. Most of the tools we use are open-sourced tools.

What does reproducibility mean to you?

I think it is key, and I think people trying to sort of replicate in different labs the same experiment is important so that we understand: is this a general pattern we are seeing? However, it is difficult to both get money to do that because a lot of the time people are looking for novel and transformational ideas, which, if you want to look at the reproducibility of something, is more difficult to sell. It's also difficult when you come to publish as well, because, again, people are always looking for novelty.

How did you become the Dean of Research Quality?

I have always been interested in research development beyond my research. I've been in Leeds for nearly ten years. I looked for roles that allowed me to be involved in further development. So I was the FBS's faculty academic champion for postdocs for a few years. Then, I became our School's Director of Research and Innovation, (DoRI) and within that, I was also involved more widely in the REF preparations last time, which is how I first interacted with Nick Plant. At the time, he was the Dean of Research and Impact, and when he became Deputy Vice-Chancellor, they split his previous role and made it bigger. I've applied to the role, and Cat Davis became the Dean of Research Culture, and I am the Dean of Research Quality.

What do you do as the Dean of Research Quality?

It is quite broad because I am looking across the whole University at the processes, environment, and development we do that support research and help people to think strategically about their research, from individual researchers to Faculties to the University level. So, that brings me into contact with many different people, practices, and groups. REF comes under my remit, so having an eye on how we assess our research internally; how we are going to prepare for the next assessment is a big thing that I'm thinking about at the moment.

What are the University's priorities in Research Quality?

Aligning with our values, we want to do research that is of the highest international quality and has an impact in academia and beyond. It needs to be research that is robust and challenge-led, but that doesn't necessarily mean applied research. Making an impact can mean identifying the key questions in someone's area and then addressing those I think people can think that it is what it means but identifying the key questions in someone's area and then addressing those. Trying to do that in an interdisciplinary way is important and doing it within a healthy culture. This is why Cat Davis (Dean of Research Culture) and I work closely together.

How does your role relate to open research?

I think it's a key part of research quality and likely to be, for example, a larger part of REF next time to show how we support people to do open research and how that encourages robust research.

Half of the time, I am a Dean, and half the time, I am an academic in the Faculty of Biological Sciences. I have been going to School meetings and encouraging people to think about what open research means for them, to be as open as they can be but as close as necessary. I think it's a big movement within modern research, and it's just going to get bigger.

"The people we are training are going to be the next generation's researchers: they absolutely need to get on board with this and understand it and essentially be better than us."

Open Research Case Studies – School of Biology, Citizen Science with Chris Hassall

How do you use open research practices?

My interpretation of open research practices means that we code as much as possible, so our analytical pipelines are mostly developed using scripting languages with good commenting. We also make those scripts with underlying data available for publication. My PhD students have a better track record of using those kinds of open research practices. I use a lot of data published under various open licenses, mostly Creative Commons through biodiversity repositories. We use a fair number of these kinds of datasets.

Have you received any feedback on these practices?

Yes, we had feedback about a particular paper from 2018, which had a GitHub repository with open code and data. We got positive feedback from some people, but a tiny minority of the people who would have read the paper.

How do you deposit your work?

I use the White Rose Repository for publications, and I share the research data and code on Kudos, which is an online tool for pulling together resources about publications on a single webpage.

Do you teach open research practices?

I am very much trying to push for a data science approach that enables students to do open and reproducible science. I teach a second-year statistics module, which is compulsory for biology students. In the module, I talk about reproducibility and the importance of open code and open research. I also teach my PhD students to write code in R to archive their code and data, so that they or anyone else can reproduce their analysis.

We are also undergoing a program of curriculum development under Curriculum Redefined at Leeds. Here, we bring new ways of teaching data science to our undergraduate students. The core part of that is helping them to understand data management, transparency, and reproducibility associated with their work, moving towards scientific programming languages like R and Python that enable visualisations of open scientific work workflows and incorporate data and code into one document.

Do you use preprints?

I use preprints sometimes, but not always. I went through a phase of using them quite a lot, so I probably have about five or six papers that I was submitting to a preprint server before submission to a journal. I think it just never really became part of my general workflow. Now, I tend to use preprint servers when I want to get a paper visible, and it needs to be visible by a grant submission deadline so that I can make it available to reviewers. I use preprint servers strategically. I have used PeerJ preprints before and bioRxiv. They're more discipline-specific repositories.

Generally, people are more positive than negative, I haven't come across much negativity. My personal experiences have been quite neutral because I don't think that people really read them; they are not a part of the standard workflow.

Has your attitude towards open access changed in the recent years?

I have become very sceptical of the growth in open-access journals because of their funding model. I don't think that everything needs to be published in the gold open access. I think that the green open-access model works very well for the democratisation of knowledge. The publishing industry is taking advantage of researchers because of the mandates from funders to make everything open-access.

"I have gone from thinking about open access as this wonderful way to share research with the world to being very cynical about the way that the publishing industry is using it to make to increase their profits. There is a healthy dose of cynicism among the biological community when we look at the open access fees for some journals. "

Hassall Lab website

What does data mean in your field?

It is primary research data collected from the field or lab mostly with the associated analytical scripts to process those data. There are also secondary data that I do not have control of but use, which are mostly citizen science and biodiversity monitoring data available from elsewhere.

How do you manage your data?

I try to keep it machine readable in CSV files with metadata sheets and linking them through analytical scripts as clearly as possible to show how it relates to outputs from papers. In a collaborative project, whoever is leading the work is managing the data.

Do you use open data?

Yes, in the sense that I use freely available biodiversity data. I have used data from supplementary information sections of papers – that is where my field would host data.

What are the stances of major funders on citizen science in your field?

I don't think they have a stance on citizen science. I think there have been a number of funded grants through, for example, NERC (Natural Environment Research Council) that use citizen science data in creative ways. They would be in favour of citizen science as appropriate in applications, depending upon the question.

"From the university perspective, it would be nice to have a bit more support with open research. I tend to use third party tools because I am more familiar with them, but maybe some of my colleagues would be more willing to engage if there was some more step by step training."

Have you conducted research in which citizen scientists were completely involved in the project?

Usually, I design a data collection portal or methodology and then I recruit citizen scientists to contribute data in some way. They would be involved in data collection, and I would share the results after with those who have voiced an interest. I would not do any formal follow-up with individuals. My approaches tend to involve citizen scientists as data collectors and generators.

What benefits has citizen science had in your research?

There are two main benefits: biodiversity data from citizen scientists tends to be collected at a spatial and temporal scale that is simply impossible for academic researchers on their own. That scale is necessary when you are trying to analyse large ecological processes or patterns. The second benefit is the volume: I do work on the evolution of mimicry - that is animals resembling one another - and what I need is to understand how people perceive similarities. The participants act as a cognitive processing tool to analyse patterns and the sheer volume of people who are interested and engage with the projects is a major bonus for data collection.

How do you grapple with managing the different quality of citizen science data?

My citizen science data is collected in a relatively rigorous and clean way, so the format would be relatively clear and articulated through metadata. Almost all publications deriving from citizen science data have to include a caveat about data quality. The problem could be with potential errors associated with the measurements, for instance.

Do you provide training for citizen scientists?

Not extensive training, but I have done two quite large science projects which have evolved people completing computer-based tasks to compare two insects. At the start, there is a check to make sure that the participants have understood the interface and what they are supposed to do. For example, I would show them a couple of different shapes like a blue circle and a blue triangle, then a yellow circle and a blue triangle and ask them to rate how similar they are as a training. This allows me to check if they have understood the task.

Citizen Science: engaging the public in the research process

Do you see an issue in the power imbalance in citizen science and academic research?

I have never heard anybody talk about this apart from academics. I do think there is an issue with power imbalance, but the kind of work that I do tends to be entirely voluntary: people can decide to engage or not engage in citizen science. The result is often that more privileged communities who have free time and access to technology tend to be the ones who contribute, so there is a bias in the demographics of citizen science communities. I have run projects that have involved reimbursing participants, where there are cash incentives, which can themselves create complex ethical issues around engagement.

For the most part, my citizen science work is a very light touch, and I am either using secondary data from people who have been out in the field and spotted a bird and reported it to a local reporting scheme, in which case I was not involved in the data collection. Alternatively, I am asking people to perform a very simple task in which they get to experience something interesting, and I get much more positive feedback from people. I think the positives that people gain from it are much more visible than any problems in citizen science.

Open Research Case Studies – School of Molecular and Cellular Biology with Queen Saikia

Project description:

Queen's PhD research project is about understanding trafficking of a receptor membrane protein VEGFR1, that impacts blood vessel formation using Human umbilical vein endothelial cells (HUVECs) as a model.

What does open research mean to you?

Open research is a practice through which people make their work accessible to the rest of the public, especially to those working on similar grounds. It could be through publications in open-access journals or through websites/social media. Conferences, seminars, and journal clubs are other great ways of sharing information with different audiences.

How does your PhD project use open research practices?

I am at a stage where I am attending conferences and symposia to present the work I have done so far, which is not yet published. Within the University premises, I am in touch with different lab groups with whom we share protocols and talk about the issues/outcomes we face daily.

Have you found it difficult or easy to identify other people's open research practices?

I think open research practices vary across different fields and research groups, and it's hard to identify because there's no specific line drawn to say whether something falls under open research or not. However, the concept is becoming very popular and people are coming up with great ideas on how it could be practised across different disciplines. I would like to mention the ReproducibliTea Network with which I am involved as a co-organiser, and we are engaged in promoting open research practices by holding frequent discussions remotely.

Do you use preprints?

Yes, quite often. It has been useful, especially in being acquainted with the latest techniques or findings before they reach a journal. However, sometimes it's tricky as we cannot blindly follow it, unless peer-reviewed.

What repositories do you use to read preprints?

bioRxiv.org and ASAPbio are my go-to repositories as I am from Life Sciences background. An advantage is, for example, you know what other researchers or a competitor lab group in your field are working on, or what advanced techniques they are using. However, the disadvantage is that you cannot fully trust what they produce because it is not reviewed by editors, so it has not reached a point where it is fully validated.

"There are some risks: we cannot blindly follow what preprints say, but you get an idea of what others are doing in your field."

Has your attitude toward open access changed in the recent years?

"Open access" is a very friendly term, but I didn't realise before that one has to pay a big price for it. And when I say "big price", I mean it literally because we are all aware of the fact that high-impact journals demand high open-access charges to make one's research publicly available. If not, a reader has to subscribe to those journals to access their research. I think it's partly unfair because funding and resources vary from lab to lab across the world. Some labs are engaged in doing valuable research, but might not be well funded, hence that shouldn't deprive them of the open-access benefits they could have otherwise.

As a PGR student, are you happy with the training provided on open research?

I am aware of the initiatives taken by the Library Team at the University of Leeds that are pushing us forward to understand open research in more detail, and I am happy about it.

What does data mean in your field?

Data means an answer to our research questions or proposed hypothesis in general, which could be a YES/NO or somewhere in between (as they say "Biology is a science of exceptions"). In my field, data is mostly generated in the form of numbers, images or experimental videos.

How do you manage your data?

I first collect the raw data after an experiment has been done, and then use different software to quantify them. That follows up with a few repetitions until we are certain that the data is reproducible.

Do you use open-source software?

Yes. For GraphPad, you need a license, but for image processing software, we have a university license.

Leeds ReproducibiliTea - Introducing Qualitative Secondary Analysis and the Timescapes Archive

Have you written a data management plan?

Yes, a data management plan is mandatory in my Faculty, and I had to produce it when I was in the first year of my PhD.

What does reproducibility mean to you?

It means the confidence with which we can say that the answers generated qualitatively or quantitatively stands true, which means they are reproducible.

How did you become interested in advocating open research?

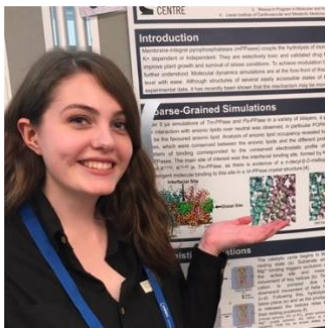
Over time I have realised that certain loopholes exist in and around our work culture. One day I came across an event advert organised by the ReproducibiliTea team, and the theme to some extent resonated with my thoughts. I emailed the Team Lead and expressed my views, and some interesting conversations happened over a video call. That's when I realised that advocating good practices is important in addition to doing good science.

Have you participated in other public engagement activities?

I have been previously involved in some outreach activities as part of an initiative of the Physiological Society. Currently, I am working on a brief project on promoting research assessment practices around the University premises.

Open Research Case Studies – School of Molecular and Cellular Biology with Alex Holmes

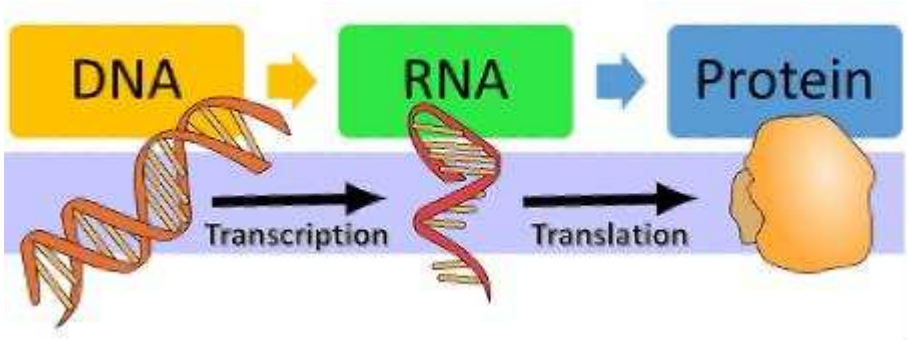
Investigating the Structures and Mechanisms of Ion Translocating Membrane Proteins project description:



Alex Holmes uses computer and laboratory-based approaches to understand how the tools of cells, known as "proteins" work. The proteins she currently works on are PIEZO1, which senses force on membranes and converts it into a cellular response, and membrane-integral pyrophosphatases, which are responsible for the infectivity of protozoan parasites such as malaria.

What does open research mean to you?

Open research means making research available so that others can read and understand what you have done, and that people can contribute their ideas and participate in research.



How does your PhD project use open research practices?

Since my PhD project is funded by the BBSRC (Biotechnology and Biological Sciences Research Council), any research I publish has to be open access. I have worked with Nick Sheppard (Open Research Advisor) to make sure that all my data, even the pre-processed raw data can be made available online, which is linked to each of my publications. I also rely a lot on a protein data bank, a large resource of protein structures and the underlying data, which is freely accessible and open.

How do you deposit your work?

I have an ORCID account. I am probably not the best at keeping up with it, but I make sure that my publications are linked to it. I use Symplectic to track my public engagement activities so that the Public Engagement Team can like put them forward to REF.

In your field, have you found it difficult or easy to identify other people's open research practices?

It depends on who you work with; in my research group, we talk about open research quite a lot because we do have to go through the process of uploading our files, making sure that the right format, and writing README files so people understand what they are. We are a computational group, but we work with a lot more experimentalists whose open research practices I understand a lot less.

They might have raw microscopy images that are much harder to upload online or fully describe the analysis if, for example, it is a 24-hour-long video of mice running around in cages. There is also quite a difference in attitudes: some of the principal investigators make all data open and available, while others like to keep their data only to themselves and close it off from the public until an article has been published.

Do you use preprints?

Yes, I have deposited preprints and I read preprints. I use foremost discipline-based repositories, like bioRxiv.org and there is a Twitter account that tweets out the biophysics and bioRxiv papers, which I look at. I also use the White Rose repository.

Do you have any concerns about preprints?

A little bit since they have not been peer-reviewed. It is useful to have another set of eyes look at your research in terms of integrity, getting the right conclusion, and improving the paper. In my field, there is a research group in America that does very similar work to ours, but a lot of their work is in preprints, so it is hard to gauge how useful it is.

If you are not clued up on these little nuances, you wouldn't realise that. There is quite a lot of conflict between their work that also attempts to disprove our work.

Has your attitude toward open access changed recently?

Previously I didn't know enough about the politics of publishing. Now that I know more, I think open access is the way forward. When you are in the university, you are accessing everything through university subscriptions, and you don't realise how much money is being charged by publishers.

Especially with COVID, having the research rapidly available was essential, and it makes sense to have everyone be able to access the data. There is that conflict about how much some publishers will charge for open-access publication, and that is not very equitable. I am definitely for open access, but I can see the challenges in it.

Are there any negative perceptions or attitudes toward open access that you came across?

Some people are really against it. We had a conversation in a journal club where the prestige element of hard-to-access articles came up. For some, it seems more prestigious to have articles published that are quite hard to access, and the data is even more valuable because people are having to pay to get to it, which I strongly disagree with.

As a PGR student, would you like to see more training in open research?

I would love to know more about what the RDM (research data management) team does since we interact with them; for example, they upload our data. I would like to know more about their role within the whole university, how it fits with careers, and how the university interacts with the team. I think it would be interesting to see what other researchers are saying about open research since it would be very different from my opinion in my department.

What does data mean in your field?

Data is very quantitative, it is the numbers, statistics, and raw data, which get processed into usable data, or in my case, the raw data is large files with coordinates in them. When it becomes analysed data, that is when we share it.

How do you manage your data?

Since I work with very large data sets and I have to use the supercomputer to generate them, I manage it by downloading my raw data directly into an archive and then extracting the relevant files I need for analysis, and then backing up those analysis files later. I structure my data into large-scale projects and then the small data sets within each of those like a pyramid that builds down into different analysis directories; it is a very hierarchical data structure. It takes a while to find individual files, but it is the most useful way for me to manage all my data.

How did science communication become important to you?

Getting into the PhD, I wanted to have as many different experiences as possible to work out what I want to do after my PhD, and I ended up enjoying science communication. I have worked with the University of Leeds's research festival *Be Curious*. Since my work is computational and I look at the screen quite a lot, it is nice to do things with real people and explain to them my research. **This fits into my idea of open research: making research accessible.** It is one thing having a paper, which is not written in an accessible language, for example, my parents would not understand my publications. You need to talk people through the research. On the other hand, receiving feedback and research ideas from the public and hearing what they are interested in has changed my perspective on what I am doing.

What kind of public engagement activities do you do?

I have done hands-on activities at science festivals where you design games to explain research outcomes. I have also done science shows where you stand on stage and talk about science. In the one I am currently doing; we have a scientific debate where my colleague explains how good big things are, such as Jupiter, and I explain how good small things are, for example, proteins inside your body. In 2020, I did a few training courses for A-level students who were thinking about going to university and how to get them interested in science or how to talk about science in personal statements.

Do you think there should be more training on public engagement and science communication?

I think more work needs to be done to highlight public engagement and train people in good practice. It has helped me to see open research as two-way communication between the public and the researchers. Public engagement is something that so many people want, even funding bodies or job recruitment. However, the importance of public engagement is never highlighted to you until you start looking for jobs and you see that I should be able to communicate with different audiences. I think it should be prioritised on an institutional level too.

How important have social media platforms been to you in communicating your research?

I am on Twitter, and I find it useful to find papers and preprints, but also to see what other scientists are doing or saying about, for example, open access. It has given me a wider

understanding of the landscape that I would not have gotten otherwise. I also run the account for my research group: it has a very different aim than my personal account, which is to promote myself and my research. I know some researchers also use Instagram and TikTok for communication or to promote their research.

"I would like to give the general public appreciation of biology beyond just DNA and show that there is a lot is going on in your body, a lot of processes that complicated, and why that matters for medicine, health, technology, and environments. I would like to shine a light on something that is often overlooked."