# The New Open Research Handbook for UKRNs Training Community of Practice

Joe Corneli and collaborators

**Abstract**

This is my open research handbook. There are many like it, but this one is mine.

@def title = “Intro to Open” @def tags = [“syntax”, “code”]

## Introduction to the Open Research Book

## Contents

## Background

The initial contents of this web page are 15 scripts for videos that were produced at Oxford Brookes to introduce basic topics in Open Research. We followed an outline shared in image we first encountered care of Anton Muszanskyj. (If you know where it came from please let us know.)

@@row @@container @@left @@ @@ We followed the high-level outline breaking the sections down into **open** ~~~ Infrastructure, Methods, Communities, and Knowledge. In some cases we combined some of the topics into one script, but in general we followed the lower-level outline as well. The hope is that this can be used as a starting point for further improvements. ~~~ ~~~

~~~ @@

## Quick tour

Have a look around via the menu at left. You'll notice that most of the talks include some Oxford Brookes specific pointers towards the end. These are flagged in a box as "⚠ Practice Example". Other related practices may be in place at other institutions.

Last modified: {{ fill fd_mtime }}.

# Open Licensing

Hello, and welcome to this short talk about Open Licensing. In this session, we will look at what licences are, and how they are traditionally used to manage rights in "intellectual property". We'll contrast traditional exclusive rights-holding with open licensing, which distributes permissions widely and non-exclusively, while imposing certain requirements on people who use the material. We present some examples showing how this can allow remixing material from different sources in a new form, and offer some guidance on online tools and further reading.

### Why are open licences relevant?

Without being especially technical, open licences are a core part of the infrastructure for open research. They provide the legal foundation for forms of "sharing" which go beyond simply gaining access to material. Open licences give recipients the right to make use of the shared material in some potentially innovative ways.
~~~

### What are licences, anyway?

First, let's set the "open" aspect aside for a moment and talk about what licences do. They are legal agreements which give someone permission to do something which they wouldn't have the right to do otherwise. Think of a hunting, fishing, or driving licence. A patent is not dissimilar in that if granted, it gives the applicant a temporary monopoly on their invention. However, unlike a driving licence or a patent, contemporary copyrights don't need to be applied for: they automatically come into force when a work that qualifies for protection is created. Almost anything that is written down or recorded is subject to copyright, but crucially it is the *creative expression* that is copyrighted. Importantly, ideas can't be copyrighted, and if you want to protect a method you'll have to seek a patent. It's worth mentioning that, typically in the UK, if work is created for an employer then the copyright belongs to the employer — although academic writing is an exception.

### How are traditional licences used?

A licence can be issued by whoever owns a patent or copyright to allow another party to make limited use of the patented invention, or to publish a copyrighted work. Some traditional alternatives to licensing are to transfer copyright to the publisher — possibly in exchange for royalty payments — or to sell a patent.

### What's special about open licences?

Against this backdrop, open licensing does something pretty different. These licences permit anyone to do certain things which would otherwise have been an exclusive right. Depending on the preferences of authors and publishers and the terms of their selected licence, this may include: copying the work without limitation, selling those copies, and creating and sharing modified derivative works.

### Open licences impose restrictions on use

Most open licences impose some additional requirements, typically requiring copies and modified versions to make a clear attribution to whoever owns the original copyright. Licences may allow people to create new derivative works based on the original — perhaps a film adaptation for example, or a remix of contents from multiple sources (I'll come to an example shortly). Some licences require that any derivatives be released under the same terms as the licenced work that they derive from — this is a so-called "viral" clause. Software

licences may impose the requirement that the source code for any derived versions be made available at no cost, alongside the compiled runnable software.

Perhaps the most common place for academics to encounter open licences is in the context of Open Access publishing. While the concept of "Open Access" puts *getting access to material* front and centre, open licences permit people to do more than just download and read. Popular open licences include the Creative Commons Attribution or "CC BY" licence and the CC Attribution-ShareAlike (or "CC By-SA") licence. These licences permit republication and the creation of derivative works, requiring that the original author be acknowledged; CC By-SA's "ShareAlike" clause additionally requires that the person reusing or adapting the material must share their derived work under the terms of the CC By-SA licence (hence "share *alike*").

**Open licences allow remixing**

Where this gets particularly exciting is that if two works use the same or compatible licences, say, Wikipedia (which is published under CC-By-SA) and an open access journal article, then the contents can be combined, either by including excerpts of the research work in Wikipedia (with attribution, in line with the licence requirement, and with a citation a matter of good encyclopaedic practice), or by remixing the selected material in new publications. It is worth mentioning another offering from Creative Commons, the CC Zero (CC0) Copyright Waiver, which is not a licence: instead, it is used to transfer copyright and other associated rights to the public domain. In principle, this can facilitate reuse in settings where attribution would be cumbersome. CC0 is used, for example, by the European Commission for "raw data resulting from instrument readings, bibliographic data and other metadata."'

**You can get help choosing a suitable licence!**

The range of available licensing options speaks to the fact that there is no one-size-fits-all licence. If you're just getting started thinking about open licences, it could be useful to have a look at the Creative Commons's "licence chooser" and the human-readable summaries of their licences.

Before publishing work online with an open licence, it's wise to be sure you have the right to do so.

⚠ **Practice Example:** *The Brookes IP policy is linked below, and, where needed, members of the Knowledge Exchange and Impact team in RIE can assist with IP questions: Email rieenquiries@brookes.ac.uk . For help with open licences in relation to*

*Open Access publishing, you can check out the University website, and the Open Access team at the Library is also there to help you: email them at 'openaccess@brookes.ac.uk'.*

As indicated above, there is no one-size-fits-all licence. Even though CC By is widely used in Open Access publishing, Punctum Books for example, have made an argument (linked below) for using the more restrictive CC By-NC licence on their humanities monograph publications. This licence permits derivatives, but forbids commercial use of these derivatives. For those who want to learn more about these topics, here are some books on open licences and adjacent topics (which are themselves available as open access).

Suber, Peter, *Open Access* (Cambridge, Mass., MIT Press, 2012)

Montgomery, Lucy, Hartley, John, et al., *Open Knowledge Institutions: Reinventing Universities* (Cambridge, Mass. MIT Press, 2021)

https://punctumbooks.pubpub.org/pub/creative-commons-by-nc-licensing-open-access/release/9

https://www.gov.uk/guidance/ownership-of-copyright-works#works-created-for-an-employer

https://www.brookes.ac.uk/library/resources-and-services/for-researchers/open-access

https://drive.google.com/file/d/1M6BEvVhaMxGqktrQNOGsTDhztTcioFwb/view (IP Policy)

## Open Platforms: Toolboxes for Open Research

Hello, and welcome to this short talk about Open Platforms. Platforms bring together technologies, processes, and conventions for their users: *open platforms* introduce aspects of openness to some or all of these dimensions. In this session, we will begin by introducing some familiar platforms from outside the research domain, describe how to recognise more "open" platforms, and then talk about using open platforms to do open research. The talk will highlight how the "openness" of open platforms is particularly useful for eliciting contributions, giving examples of research collaborations that have used open platforms and of a new open platform that is experimenting with fundamental changes to the way research is done.

### To begin: Platforms outside of research

First, let's think about some familiar platforms from outside the research sphere. These days, when we go online, we often access "platforms", whether we notice it or not. For instance, Amazon is home to Amazon Marketplace, Google runs Google Cloud, and Facebook runs the Facebook Platform. These are all collections of integrated hardware and software technologies, business models, and cultural conventions which allow third parties to do things online: to buy and sell products, to run software services, or to access social network data. It is worth noticing that all these platforms are 'open' in some senses of the word: they "provide the hardware and software foundation for others to operate on."

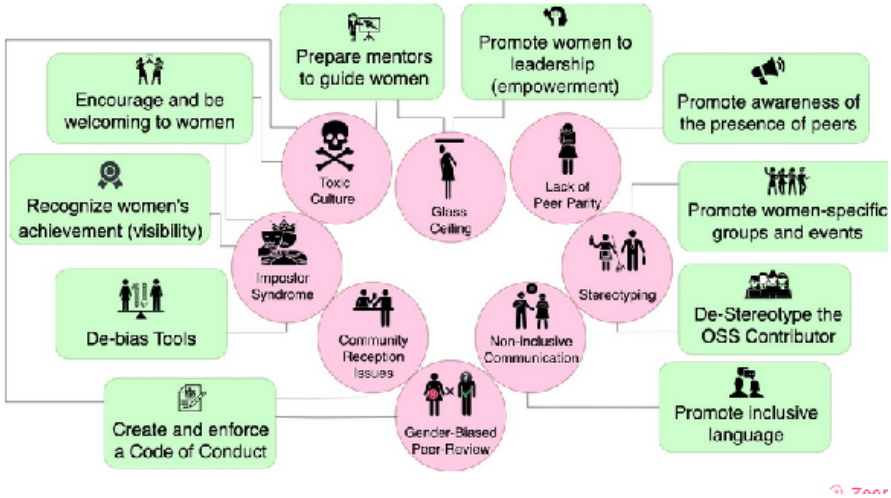
### Further aspects of "openness" in open platforms

Some platforms have additional aspects of openness baked into their design, including open enrollment, open software, open governance, and a commitment to transparent operation more generally. Wikimedia, for example, exemplifies all of these aspects of openness. Not only can anyone with an internet connection edit the sites on the Wikimedia network, anyone with coding skills can get involved with improving the underlying wiki software; and anyone with time on their hands can get involved with operational work, or put themselves forward for leadership roles. This contrasts with the corporate platforms mentioned previously, which are more centrally controlled.

### Open Platforms for open research

The openness of open platforms can make them attractive as sources of information about human behaviour. However, some open platforms are attractive to researchers for the more immediate practical reason that they have support for research as their primary practical focus. As such they are not just potential "research sites" but toolboxes for open research.

It's useful to recall that any aspect of the research lifecycle can be more or less "open". Research can be planned in public or behind closed doors; data can be collected centrally or by a distributed pool of contributors; and so on.




Some platforms only deal with one or a few of these phases. Rather than being a comprehensive survey of available platforms and their features, this talk aims to illustrate how the "openness" of open platforms can be helpful to researchers.

### The openness of open platforms make them particularly useful for eliciting contributions

Some open platforms focus on sharing research results in a way that supports **sharing** and **re-use** (for example, "preprint platforms" achieve this). Some platforms focus on opening up the **process** in a way that helps people replicate a study after it's been carried out (e.g., "Protocols.io is a secure open access platform for developing and sharing reproducible methods."). Using tools like these can certainly change the way research works (for example, some researchers devote considerable time to keeping up with new preprints in their area of study). Platforms that open up other aspects of the research lifecycle can change who directly participates in a given research project. Let's turn to a couple of examples of how open platforms have been used to elicit contributions to research-in-progress.

### Example: OSF support for a large-scale research collaboration

The Open Science Framework (or OSF) describes itself as "a free, open platform to support your research and enable collaboration." Many people use it to share preprints and reports, but it can also be used earlier in the research cycle. One flagship example was a collaboration between some 270 authors on an intensive review article called

"Estimating the reproducibility of psychological science", which was subsequently published in *Science*. OSF was used to host and discuss the components of the paper — in the open — as they were being developed by contributing authors. Referring back to our research lifecycle keywords, the platform features particularly helped with the `Process` phase of work.

### Example: Custom platform design to create a novel research site

Whereas the previous example used an online platform in an instrumental way, some research in computing explicitly focuses on exploring the properties of tools and platforms, themselves. For example, Bill Tomlinson led a 30-author contribution to CHI 2012 which looked at "Massively distributed authorship of academic papers". In this case, the research was mostly qualitative, and based on a reflective exploration of "the experiential aspects of large-scale collaborative research." The platform was assembled by the paper's contributors: co-authors were recruited via existing mailing lists, and a co-authoring space was set up using open source software. This time, tooling features were in place to support the `Collect` and `Analyse` phases, i.e., to gather observational data, and to help surface trends in that data.

### Co-authoring and other forms of collaboration: new directions in open platforms

Both of these examples mentioned above describe contemporary co-authoring workflows involving quite a few co-authors. However, not only is the process of writing research papers evolving, the way we carry out basic research is itself slowly changing. For example, Octopus is "an open access platform where researchers can read, review and publish findings at all stages of the scientific process." It is part of a broader trend of tools for sharing research, such as infrastructures for creating and sharing laboratory notebooks and data science notebooks are helping scientists do research in the open. The basic idea is that sharing ideas and preliminary results can foster discussion and collaboration, avoid redundancy, and accelerate progress (though there is more research needed to check these claims!).

### Creating an open culture

Beyond the technical and resourcing considerations, there are various "soft" design features which should be considered when selecting or developing a research platform. While focusing on workplace issues, the 2020 Wellcome Trust Report on "What Researchers Think About the Culture They Work In" notes that researchers say that their work-

ing culture is best when it is collaborative, inclusive, supportive and creative. You could look for and seek to foster these qualities in collaborative platforms as well.

Even a platform that isn't open in all senses (like Google Workplace) can be used in ways that are more or less open. For example, you can easily change your settings on Google Docs so that "Anyone on the Internet with the link can edit" or "Anyone on the Internet with the link can comment".

Recent research also suggests that just about any form of co-authoring tends to lead to quality improvements and also to more citations, a finding which seems to be true across disciplines. While open platforms can be useful for various research-relevant purposes, as we discussed, some are particularly useful for eliciting contributions, and it is encouraging to know that this can lead to improvements in research quality. While open platforms can be useful for various research-relevant purposes, as we discussed, some are particularly useful for eliciting contributions, and it is encouraging to know that this can lead to improvements in research quality.

Thelwall, M., Kousha, K., Abdoli, M., Stuart, E., Makita, M., Wilson, P., & Levitt, J. (2023). Why are coauthored academic articles more cited: Higher quality or larger audience?. *Journal of the Association for Information Science and Technology*, *74*(7), 791-810.

- https://wellcome.org/sites/default/files/what-researchers-think-about-the-culture-they-work-in.pdf
- https://staff-learning.brookes.ac.uk/totara/dashboard/index.php?id=71
- https://www.brookes.ac.uk/sites/research-support/open-research-at-brookes
- https://www.octopus.ac/
- https://openlabnotebooks.org/

# Interoperability

Hello, and welcome to this short talk about Interoperability. In this session, we will talk about what interoperability is, and how giving it attention can support effective collaboration and re-use. We touch on several points to consider when choosing and or designing research tools, and give examples of the way interoperability has been con-

sidered in two academic fields — with some implications that can be useful in any field.

### What is interoperability?

"Interoperability" is a characteristic of a system that means it can work with other systems.
Many organisations are independent in their day-to-day operations, but also able to share information when needed. For example, a hospital and a clinic in different health authorities have protocols they need to follow when they work together to treat a patient. Another example is Wikipedia. In fact, Wikipedia works well partly because there are multiple Wikipedias: one for each major language. This way, the sites can share a core technical infrastructure, without becoming a Tower of Babel — or requiring everyone to learn English. The workflows interoperate smoothly.

### How does interoperability relate to open research?

In an open research context, "interoperability" means that the way things work in one research project can be taken up in another context without too much friction. As such, the concept of interoperability encompasses both "how you can build on pre-existing work", and "how others can build on yours." For these things to be possible, one set of data and methods needs to be able to interoperate with another. Often, this happens without direct interaction between the two different research teams. While there may be a technical dimension to this (for example, related to sharing or exchanging data in suitable formats), even more fundamentally 'interoperability' is related to what Baroness Onora O'Neill has referred to as *intelligent openness*: "ensur[ing] that others can not only locate but also understand and assess material."

### When is attention to interoperation needed in open research?

You will be able to develop the necessary dimensions of 'interoperability' between your work and others' in dialogue, by asking them what they need you to do for you to be able to work together compatibly. In this way, people who use your research effectively become "collaborators". However, open research can also involve setting things up so that people who you never interact with directly are able to easily re-use your research. This means anticipating interoperability requirements in advance, while being open-minded to the possibility of interactions that you can't anticipate. This has the practical implication that you may want to check back from time to time using tools

like Google Scholar, or with more hands-on follow-up studies, to see how people are (in fact) making use of your research.

### Interoperation makes cross-setting co-working possible

Interoperation can apply at the level of tools and data, or the interpersonal level (or both). Open source software includes aspects of both. For example, a relevant design principle taken over from the UNIX operating system is for individual open source software packages to "do one thing well", and to work as an interoperable part of a software tool chain. To achieve this, the individual projects need to involve a sufficiently large and diverse team. Another adage from the open source world is that "with enough eyes, all bugs are shallow." The main take-away here is that to build vibrant projects, you need to give attention to creating a healthy collaboration culture that is suitable to participants' ways of working.

### Choosing and designing suitable tools

For simple data-focused workflows, a baseline consideration might be 'how easy it is to get your data in and out of the system?'. For more complicated workflows which involve building entirely new systems, research might be needed to look at the design considerations that allow others to use the system for their own purposes. Let's have a look at a couple of dimensions of interoperability, through the lens of different academic disciplines.

### Interoperability can be useful even when it's not universal

Building information modelling (BIM) involves "the generation and management of digital representations of the physical and functional characteristics of buildings and other physical assets." More succinctly, BIM is about building "a shared digital representation founded on open standards for interoperability". BIM allows teams to collaborate and coordinate work, and reduce conflicts and errors. That being said, the standards that apply to European buildings don't necessarily apply in the developing world. The fact that it doesn't work equally in all contexts doesn't take away from the value of BIM in the contexts where it does work.

### Creating interoperability also doesn't mean removing all friction

You might convene a workshop that brings some of the people who use your research together, and that creates opportunities for them

to interact with each other. This might, for example, lead to some interesting debates. In general, if you're able to work with people who think about things in different ways, this can help you make unexpected discoveries (we will discuss the related topic of research diversity in another video)

Whether you are working across geographical and economic regions or across neighbourhoods, interoperability can get complex. Grappling with this complexity may require working across disciplinary boundaries as well.

⚠ **Practice Example:** *Interdisciplinarity happens to be the focus of the various Brookes RIKE networks. One possibility for Brookes researchers in interoperability could be to reach out to a network that you're not currently part of, and explore what new ways of working would need to be in place to establish a collaboration.*

Sennett, R., & Sendra, P. (2020). *Designing disorder: Experiments and disruptions in the city*. Verso Books.

Merton, R. (1936). The Unanticipated Consequences of Purposive Social Action, 1AM. *SOC. REV*, *894*, 903-04.

- https://www.brookes.ac.uk/research/networks
- https://www.pnas.org/doi/10.1073/pnas.1424329112
- https://www.scientificamerican.com/article/graphic-science-some-of-the-best-science-can-slumber-for-years/
- https://www.jstor.org/stable/2084615
- https://gcs.civilservice.gov.uk/publications/in-case-a-behavioural-approach-to-anticipating-unintended-consequences/

## Repositories and Persistent Research Identifiers

Hello, and welcome to this short talk which will look at the related topics of Repositories and Persistent Research Identifiers. We begin by saying what we mean by a repository, and how repositories relate to open research. We'll give some examples of the alternative kinds of repositories you might use, depending on the needs of your project. We will then move on to describe Persistent Research Identifiers — which are codes that make it easy to find and share materials stored in a repository. We'll conclude with a pointer to guidance on how to

prepare information for before deposit, and some tips on working with persistent identifiers.

### What is a repository?

A repository is, basically, a place to put stuff. It has material "deposited" in it, with people able to retrieve it later and generally allow others to access it. In the area covered here, we are generally referring to a digital repository: basically a database with an interface that makes it straightforward to deposit, search for, retrieve, and share materials. Most research projects will have data storage needs: if you're being systematic about how you handle data, you're probably already using a repository.

### How are repositories related to open research?

Ranging from simple data stores to complex digital library systems, repositories are quite widely varied as to contents and protocols. One way in which they can contribute to open research is on the sharing side: making research material viewable to the public, and available under the terms of an open licence and/or open access.

They can also contribute to open research in another way, by being "open" to submission of materials from a variety of people. This may be at the level of a single project with multiple collaborators — or the repository may collect materials from multiple projects.

⚠ **Practice Example:** *For example, the Research And Digital Assets Repository (RADAR) at Brookes is open to submissions from people at Brookes and is open to anyone to browse. RADAR also allows people depositing material to choose how the material they deposit can be reused, for example, under the terms of a Creative Commons licence.*

The people maintaining repositories which are open to contributions from the general public — like the Open Science Framework which hosts a "free, open platform to support your research and enable collaboration" — may enforce some light-weight moderation processes, though these are less intensive than scholarly peer review. This helps researchers share and discuss ideas at an early stage.

### Flow

### Alternatives when selecting a repository

Use of an institutional repository is strongly encouraged for the practical reason that publications must be deposited in an institutional repository no later than three months from the date of acceptance in

order to be eligible for the REF. You may additionally (or, if the material isn't REF-able) choose to upload some of your research materials in some other research repository, perhaps a discipline-specific archive like Arxiv or BioRxiv, or a general-purpose repository like Zenodo or Figshare.

Another example that may be of particular interest for people working in the social sciences is the UK Data Archive. The UK Data Archive provides curation and data handling safeguards, which distinguish it from a self-service data depository. If your concern has more to do with setting up a repository for an active project, one good alternative is Git: although it is typically used for sharing code, there is a lot of documentation available for how to use it to share research materials more broadly. Sites like Github and Gitlab make using Git fairly straightforward.

### Persistent Research Identifiers

As mentioned above, repositories not only support deposit, search, and retrieval, they also support sharing. One well-established way to do this is for the repository to systematically assign a unique identifier to every deposit. This is referred to as a Persistent Research identifier. The most common type of identifier is in use across different repositories: it is a DOI or "Digital Object Identifier". A DOI is a reference code which might appear rather inscrutable on its own. However, doi.org provides a service allowing the code to be used as a hyperlink, which enables people to share, access, and cite the published material.

### Some other persistent research identifiers

The DOI is not the only persistent identifier out there, for example, researchers themselves can have a persistent identifier (ORCiD) as can research institutions (ROR). If you're not familiar with ORCiDs, you may wonder why you might want your own personal digital identifier. Much like DOIs are for documents, ORCiDs make it easy for people to find and share information about you. For example, the UKRN's Open Research Project plans to use ORCiD profile pages to check whether participants are making use of open research practices (like creating preprints) and also plans to use the profiles hosted on orcid.org as a place to post certifications of open research training. Another benefit of the ORCiD system is that in many cases, your profile on orcid.org is kept up to date automatically by computational services running behind the scenes. While the focus for most is on the identifier, the website orcid.org also has a repository of information about researchers. It also has an advanced search function which can show how many people at an institution have ORCiDs.

⚠ **Practice Example:** *At Oxford Brookes University, for example, there are 1866.*

**Repositories with additional features**

Having had a look at repositories of documents and repositories of profiles, it's worth a quick look at another repository, Wikidata, which is a collection of all kinds of facts. It can be queried with the SPARQL language to answer simple statistical and inference questions, like "What are the largest cities in the world that have a female mayor?" and "What airports are located within 100km of Berlin?" Its particular relevance to open research is that Wikidata also allows you to upload your own structured data, which can support downstream uses in ways that a more static "data dump" would not. The broader point to consider when selecting a repository is: what added value does it have? How can people use the material they find there?

**Keep it practical**

**Working with DOIs and persistent identifiers**

As mentioned above, the website doi.org is useful for looking up a paper from its DOIs. Additional tools like doi2bib.org can turn the DOI into a usable citation. Although it's a somewhat self-explanatory example, shortdoi.org can be used to shorten DOIs, which could be useful if you're ever scrambling for column space. If you have set up a working repository on Github, it may be useful to know that you can use Zenodo to obtain a DOI for the repository, making it easier to cite work in progress there. Other identifiers like ORCiDs and RORs may be helpful when submitting material to journals, repositories, or other platforms (for example, if you upload to Octopus, you can flag the organisation that funded your work using their ROR).

- https://www.brookes.ac.uk/library/resources-and-services/forresearchers/publishing-research-data#metadata
- https://www.sciencedirect.com/science/article/pii/S0164121221002144
- https://en.wikipedia.org/wiki/List_of_preprint_repositories
- https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/A_gentle_introduction_to_the_Wikidata_Query_Service
- https://citation.crosscite.org/

# Crowdsourcing and Citizen Science

Hello, and welcome to this short talk about Crowdsourcing and Citizen Science. In this session will discuss what we mean by citizen science or crowdsourcing and when a researcher might want to use these methods. We'll look at a couple of examples, and also talk about how to manage practicalities such as participant motivation and quality control.

### Crowdsourcing and Citizen Science are worth introducing together

The term crowdsourcing was coined in 2006 by Wired magazine author Jeff Howe, as a variant of "outsourcing". Since then the term has taken on a somewhat broader meaning, spanning a variety of ways to attract participants and divide work among them to achieve a cumulative result. This often involves an **open call** for participants. Participants may be invited, for example, to complete small tasks online in exchange for payment, or some other reward. However, the term is also sometimes used in a colloquial way to refer to an informal process of gathering ideas from any group.

In research settings, crowdsourcing is often used to process data in an operational way, and contributors are not the subject of the research. However, in some cases, contributors can be formally enrolled into a research project as research subjects, for example, as survey respondents. Doing this requires careful treatment of matters such as informed consent and confidentiality: some specific guidance provided by the UC Berkeley Human Research Protection Program is linked below. Citizen science projects typically recruit participants to join as active hands-on contributors, and not as "research subjects". Indeed, Alan Irwin — a British sociologist who was one of the first people to use the term citizen science — puts the emphasis on "developing concepts of scientific citizenship[,] which foregrounds the necessity of opening up science and science policy processes to the public". This way of thinking about citizen science goes beyond the hands-on practicalities of gathering and analysing data. Nevertheless, the practicalities are a good place to start.

### These are simple forms of "participatory research"

Both crowdsourcing and citizen science projects typically assume that participants will restrict their participation to carrying out discrete bite-sized tasks — for example, classifying texts or images, or gathering samples. Typically, these small pieces of work don't require significant expertise to complete — if you find yourself designing a project

which needs *lots* of *engaged expert* contributions, you may need to adopt an entirely different research design. Other aspects of participatory research which are beyond the typical scope of crowdsourcing and citizen science projects are the subject of a separate video in this series.

**Benefits: Why might you want to use crowdsourcing and citizen science methods?**

Both crowdsourcing and citizen science open up aspects of a research process to participation from people who might not otherwise contribute to research. This can come with quite a number of potential benefits. For instance, opening up a research process in this way can be inexpensive in comparison to other ways to gather data. Indeed, these methods can also yield a broader range of data than can be readily sourced through other means. Your research may benefit considerably from the unexpected observations and innovative thinking that your participants provide. Taking part in a research project can have benefits to participants as well; and, indeed, there may be further benefits to all parties, and the public at large, building on public engagement with the research process. Let's have a look at an example.

**An early example**

Contemporary crowdsourcing and citizen science are often facilitated by computer technology. However, similar approaches were in use much earlier. One stand-out example was the creation of *The Oxford English Dictionary*: James Murray, a Victorian lexicographer, was the lead editor of the team responsible for the OED. Faced with the enormous task of producing a comprehensive dictionary, he enlisted the help of dozens of amateur philologists as volunteer researchers to provide quotations illustrating the uses of each meaning of each word, and with evidence for the earliest use of each.

**Participant motivation**

Training for participants can be crucial in citizen science projects. Firstly, it may be necessary to ensure quality of contribution, whether that is in primary data collection or other tasks such as analysing data that has already been collected. In the OED example, participants had to understand what kinds of submissions were being sought, what form to put them in, and how to go about sourcing them. Moreover, learning new things could be a big part of what participants feel they have signed up for in the first place. Understanding your

participants' motivation and helping them realise their goals will help your project succeed. In connection with this, some citizen science projects invite participants to take on leadership or coordinating roles, or to help support others' learning.

### A more contemporary example

Let's take a look at a contemporary citizen science project to get a sense of how they work.

*Surfers Against Sewage* is a UK charity focused primarily on marine conservation and related topics like beach water quality. Recently, it ran a citizen science project to monitor water quality in inland waters. Participants would fill a couple of plastic bottles of river water each week, and post them off to a firm ~~called **Pathology Management Services**~~ to test for *e. coli* and other contaminants. Results were aggregated, showing trends and disparities. In order for all of this to work, the participants didn't have to know much about water testing. They were instructed on how to collect good samples and sterilise their equipment. All further technical details were taken care of by the lab. Participants also didn't need significant resources beyond their time: Surfers Against Sewage sent out the necessary kit, and paid for the tests. Monthly Zoom calls gave participants an opportunity to look at the data collected so far, and talk about shared concerns (for example, changing regulations around applying for designated Bathing Water status). The project coordinators helped participants navigate obstacles, ranging from problems with the courier to tensions in local groups. Putting it all together, by the end of the summer, SAS were able to announce the headline-grabbing result: "Citizen science data shows that 60% of the inland bathing sites we monitored didn't meet minimum safety requirements for water users in England." While similar data could in principle have been gathered by one person who travelled to forty different riverbank sites around the UK each week, that would have been an exhausting job — furthermore, the benefits for participants would have been missed, and the final news story would have had a very different feel.

### Risks and risk management strategies

Along with the various benefits touched on earlier, crowdsourcing and citizen science methods come with some risks and limitations. You may decide that the research you're working with can't be dealt with ethically or practically by relying on low-paid or unpaid contributors; for example, confidentiality concerns might make these strategies inappropriate for some research settings. In settings where crowdsourcing or citizen science is suitable, from a pragmatic standpoint,

the project will need to be made robust against low-quality contributions, reporting inaccuracies and participant drop-out. Some of these risks can be sorted out with a good research design. For example, one typical strategy that can help ensure data quality in crowdsourcing projects is to test out would-be participants on a set of screening questions, in order to assess their suitability as contributors, before they start contributing research data. It may also be necessary to spot-check individual responses, and to screen the contributions for statistical outliers or other indications of less-useful contributions (for example, if your participants have disengaged and started to mindlessly click through the questions).

### Citizen science as a route to engagement

On that note, if it employs a solid design, citizen science can work well as a route to "public engagement"; often more so than crowdsourcing. As mentioned earlier, you may choose to include both technical training and more wide reaching activities to help participants develop scientific literacy. If you're thinking about how to incorporate citizen science into your research, or wondering whether another form of engagement would be better for the aims you have in mind, it is worth having a look at the relevant research support pages, and getting in touch with the Public Engagement team (`publicengagement@brookes.ac.uk`).

Mechanical Turk (MTurk) for Online Research, https://cphs.berkeley.edu/mechanicalturk.pdf

SAS citizen science project: https://waterquality.sas.org.uk/england/

Irwin, A. 1995. *Citizen Science: A Study of People, Expertise and Sustainable Development.* London: Routledge.

Bassi, H., Misener, L., & Johnson, A. M. (2020). Crowdsourcing for Research: Perspectives From a Delphi Panel. In SAGE Open (Vol. 10, Issue 4, p. 215824402098075). SAGE Publications. https://doi.org/10.1177/2158244020980751

Moss, A. J., Rosenzweig, C., Robinson, J., Jaffe, S. N., & Litman, L. (2020). Is it Ethical to Use Mechanical Turk for Behavioral Research? Relevant Data from a Representative Survey of MTurk Participants and Wages. Center for Open Science. https://doi.org/10.31234/osf.io/jbc9d

https://www.brookes.ac.uk/sites/research-support/pen/whyengage

# Participatory Research

Hello, and welcome to this bite-sized training on Participatory Research. We'll look at how the research process itself can be opened up to become "participatory", right through from the design stage to the dissemination of results. We'll discuss what this means for participants and how participatory research relates to a growing mandate from research funders around Responsible Research and Innovation.

### Expanding what it means to "participate" in research

"Participation" in research typically means being a source of data, perhaps gathered via interviews or through interaction with a pre-existing experimental setup.Traditionally, research participants are "human subjects" whose interviews, personal details or test responses provide the project's primary data. However, active participation can be invited in just about any — or, potentially, every — phase of research, including research design. This can change the dynamic from *research "about"* a group of human subjects, to *research "with"* active participants, who become co-investigators, collaborators or correspondents, rather than just data sources.

### Opening up the research process

Discourse around open research often considers the **outputs** — *which are made available as* "open access" — or **inputs** — *which are made available as* "open data". Here, instead, we'll talk about opening up the research **process**, inviting people to engage with and potentially shape this process as it develops. It's worth mentioning that there are also ways to open up a research process that don't involve inviting participation — such as publishing your research protocol so that others can reproduce your research — and that topic will be discussed in a separate video. Participatory research tends more in the direction of "robustness" than "reproducibility", and can help to ensure that the research is actually fit for purpose. Funders may want to see plans about how a research project will meaningfully involve stakeholders. This may be referred to as 'co-production', 'co-creation', or 'co-design' — or Responsible Research and Innovation.

### Logic of participation

If this is new to you, the idea of opening up a research process to participation may sound strange. Traditionally, the design, oversight, and administration of a research process have primarily been carried

out by experts, with participants taking on the role of data sources. However, expertise and participation are not actually at odds. In medicine, for instance, there is a well-established and exemplary form of participatory research in the sense that we have been talking about, namely public and patient involvement or PPI. PPI is a research practice wherein people with health conditions, carers and members of the public work together with researchers, and influence what is researched and how. The balance of involvement in PPI can range from 'consultation' to 'patient-led research'. One study found that around 45% of recent papers in the general health-research journal *BMJ Open* contain PPI.

Here's an example that illustrates how this kind of project can work. The Maternal health And Maternal Morbidity in Ireland (MAMMI) study recruited 3047 women as participants, following their initial visit to a maternity hospital and a subsequent telephone interview. The project ran from 2012 to 2017. Participants co-created a set of resources about women's health after motherhood, suggested several follow-up studies, and co-presented and co-published findings. The year after the project, a public contributor panel was convened — running quarterly on Saturday mornings — in which 88 of the study participants took part in discussions about what further related research they would like to see funded.

The principal investigator, Dr Deirdre Daly, reflects that "To be successful, public involvement ought to be planned as a core essential component of the study's design" and that the process must be "cultivated in order to achieve true partnership." These points would apply across domains, though of course different research topics would come with different requirements for participants (such as demography, expertise, time available, and so on). Participation can, accordingly, be structured in different ways, ranging from traditional data-gathering methods (like interviews and surveys) to more-involving forms of engagement (such as activity-based workshops, possibly including training for participants, and so on).

### Benefits of participation

We can probably all agree that "survey fatigue" is real. If we're going to take time to answer someone's questions, we want to know that our contribution will actually impact matters we care about. Simply inviting people to participate in your research often won't be enough to spark their interest, unless you make the benefits for them clear. In the example of the MAMMI study, the women involved had all recently given birth and their participation reflected that direct lived experience; moveover, they were motivated by the opportunity to

have two-way information exchange about topics of concern with the study team. While some benefits may only appear at the end of the project, meaningful participation in a research project can potentially be its own benefit, providing upskilling opportunities and camaraderie, along with all the other things that make work in research interesting and enjoyable.

### Requirements around Responsible Research and Innovation

EPSRC now requires a statement on Responsible Research and Innovation (RRI) to be included in submissions proposals. They offer the acronym "AREA", for "**anticipate, reflect, engage, and act**" as a framework for RRI. Whether we refer to engagement, inclusion, or participation, a key part of being a responsible researcher is to

Open up … visions, impacts and questioning to broader deliberation, dialogue, engagement and debate in an inclusive way.

This process can end up opening up possible futures that the researcher wouldn't have anticipated on their own; this illustrates why opening up the research process can constitute such a vital aspect of "open research".

⚠ **Practice Example:** *There are several well-developed strands of participatory research at Brookes, for example a recent report led by Dr Sarah Quinton includes good practice guidelines on participatory research with older people. While that might not be your area of focus, the Public Engagement team offers sessions that can help you develop your own approach to engaging the public in your research ('publicengagement@brookes.ac.uk'), with a particular eye to impact.*

Stilgoe, J., Owen, R., & Macnaghten, P. (2013). Developing a framework for responsible innovation. In Research Policy (Vol. 42, Issue 9, pp. 1568–1580). Elsevier BV. https://doi.org/10.1016/j.respol.2013.05.008

Nielsen, M. (2020). Reinventing discovery: the new era of networked science. Princeton University Press.

Featherstone, L. (2018). *Divining desire: focus groups and the culture of consultation*. OR books.

Benkler, Y. (2002). Coase's penguin, or, Linux and "The nature of the Firm". *Yale Law Journal*, 369-446.

Gowers, T., & Nielsen, M. (2009). Massively collaborative mathematics. In Nature (Vol. 461, Issue 7266, pp. 879–881). Springer Science and Business Media LLC. https://doi.org/10.1038/461879a

Lang, I., King, A., Jenkins, G., Boddy, K., Khan, Z., & Liabo, K. (2022). How common is patient and public involvement (PPI)?

Cross-sectional analysis of frequency of PPI reporting in health research papers and associations with methods, funding sources and other factors. In BMJ Open (Vol. 12, Issue 5, p. e063356). BMJ. https://doi.org/10.1136/bmjopen-2022-063356

Ten Holter, C. (2022). Participatory design: lessons and directions for responsible research and innovation. In Journal of Responsible Innovation (Vol. 9, Issue 2, pp. 275–290). Informa UK Limited. https://doi.org/10.1080/23299460.2022.2041801

https://www.brookes.ac.uk/getmedia/06ad3597-b6c7-4be5-9b3f-18d71fa343d1/older-people-and-participatory-research-2022.pdf

https://www.ukri.org/who-we-are/epsrc/our-policies-and-standards/framework-for-responsible-innovation/

https://www.youtube.com/watch?v=7r9lYl4CtKI

https://www.brookes.ac.uk/sites/research-support/pen/training-and-support

# Open peer review

Hello, and welcome to this short talk about Open Peer Review. We will begin by describing some “open” alternatives to traditional scholarly peer review practices. Since open peer review isn’t always an option, we mention some other further alternatives in a similar spirit. We touch on the potential advantages and limitations of open peer review, and give some examples of how it is used alongside other open research practices. The talk concludes with some suggestions for how you can try it for yourself.

### What is open peer review?

Operationaly, *open peer review* covers a range of models. The name implies that scholarly peer review processes are not open by default. Let’s begin by reviewing the typical state of affairs. Fundamental to the traditional peer review model is the fact that:

- the identities of at least one of the parties in the review process are not *disclosed*
- the reviews themselves are not publicly available

Accordingly, open peer review reverses one or both of these conditions, and opens up the peer review process through dis-closure. Along the way, additional other aspects of the review process may also be made open to scrutiny.

This depends on how it is implemented: open peer review can involve sharing the identities of reviewers with authors, or publishing reviews alongside an article. This may expand to include ongoing open/online dialogue about a published article. A journal might introduce it across several steps. The *British Medical Journal* (*BMJ*) for example, started its journey towards open peer review by sharing the names of reviewers with authors, and now publishes the full pre-publication history of articles alongside them, including reviewer comments.

### Advantages of open peer review

Advocates of open peer review practices contend that they "enhance the quality, reliability and effectiveness of the peer review process" and simultaneously "enhance the visibility, recognition and reputation of reviewers".

### Limitations of open peer review

Arguments have also been put forward that say that open peer review isn't always the best strategy. Detractors of the practice suggest that reviewers may be less critical in an open review process, because this could expose the critics to criticism themselves if it turns out they are mistaken — or, even worse, they might be subject to reprisals even if they are correct. When open peer review was introduced in *BMJ*, the editor wrote "we hope our small move will contribute to a broader culture change." By and large this change has not yet spread to the high-stakes world of grant funding decisions: applicants may know who was on a given review panel, but have no idea who said what inside the panel.

### Related practices

Open peer review isn't universally adopted, but it can be seen within a context of related scholarly practices, which open up other aspects of scholarly communication. Some scholarly communities publish "commentaries" or "critiques" alongside a "target" or "plenary" article. The journal may publish the authors' responses to those commentaries, effectively transforming a scholarly journal into an open discussion for a particular purpose. (Journals vary as to whether commentary articles are themselves peer reviewed.) Some review communities put a system of rebuttals in place in the review stage: this allows people who receive reviews to respond to the reviewer, although the reviewer remains anonymous.

### Open peer review within open research ecosystems

Open peer review can be used together with other related open practices. For example, Octopus is "a new publishing platform for scholarly research", created with support from JISC. When publishing on Octopus, rather than appearing all at once, the various sections of a research paper are uploaded by the author incrementally, as the work develops. Each of these segments can then be peer reviewed by any other Octopus user. Although there's no guarantee from the people who run Octopus that submitted contents will be reviewed, any contributed reviews are 'open' insofar as both the content and identity of the reviewer are shared publicly. In this connection, there are related practices which do come with the guarantee of a review, as described in the companion video on Open Protocols. And there are further variations: for instance, a so-called "overlay journal" is built by combining articles hosted on a preprint repository with additional structure hosted elsewhere, which collects and may also openly publish peer reviews. (Incidentally, one thing you should not do is publish any anonymous reviews that you've received from a journal without permission: they remain copyrighted material, even though they are anonymous!)

### You can try it for yourself.

A practical next step would be to find a journal (or another venue) in your research area that employs open peer review, and try it out for yourself! openreview.net is a free website that runs a reviewing system that quite a few conferences use to collect and share open reviews.

⚠ **Practice Example:** *At Brookes, you may be able to help with the internal review of grant proposals (as a matter of policy this applies to grants with value above a certain threshold). In this collegial setting, the reviewer's identity and constructive comments are shared with the proposer, but not made fully public. Contact your research office to find out how you can get involved in the grant review process.*

Review of Corneli's Method for using Octopus to evaluate Open Research training [Example of an open peer review on the Octopus platform] https://www.octopus.ac/publications/xtsb-xs05/versions/latest

Pros and cons of open peer review. (1999). In Nature Neuroscience (Vol. 2, Issue 3, pp. 197–198). Springer Science and Business Media LLC. https://doi.org/10.1038/6295

Smith, R. (1999). Opening up BMJ peer review. In BMJ (Vol. 318, Issue 7175, pp. 4–5). BMJ. https://doi.org/10.1136/bmj.318.7175.4

https://www.ukri.org/publications/review-of-peer-review/review-of-peer-review-june-2023/ http://www.nytimes.com/2010/08/24/arts/24peer.html https://www.dariah.eu/2024/02/29/introducing-the-dariah-overlay-journal-an-alternative-and-transparent-publishing-model/

# Open Protocols

Hello and welcome to this short talk on open protocols. We'll begin by reviewing what "research protocols" are, and how sharing them — especially *before* you carry out a research study — is an open research practice that contributes to both transparency and research quality.

### What are open protocols?

A protocol is basically a recipe for carrying out a research project, or, to put that more formally, an "explanation of all the aspects of a **future research** project given in a precise, understandable manner." An *open protocol*, accordingly, is a research methodology that has been shared in such a way that other researchers can easily replicate the exact procedures that it describes — and, at least potentially, arrive at similar results. As such, open protocols are not only linked to the concept of "openness" per se, but to another aspect of good research: "replicability".

### The difference between method and documentary

Crucially, a protocol is not a play-by-play description of *what happened* in a research project, but a methodology that is written down in advance. The protocol will likely talk about the overall shape of the data to be gathered, and may include a data management plan. However, due to variability in contextual factors, simply having a protocol in place does not guarantee that the results of running two studies with the same methods will actually be the same: "*methods replicability*" is necessary but not sufficient for "*results replicability*". A good test of the readiness of a protocol is whether or not it could be followed by someone who wasn't previously familiar with the research project — indeed, when it has reached that level of detail and accuracy, it may well be ready to publish as a stand-alone document. We'll talk more about that shortly.

### Associated practices of pre-registration and registered reports

Assuming you've created a protocol in advance of your study (which is in and of itself a good practice — because it constitutes a logical check on how you expect the project to develop), you then have to decide where to put it. One option is to publish it before you actually carry out the work it describes. *Preregistration* is the practice of specifying your research plan and registering it to a public repository in advance of undertaking your study. *Registered Reports* are a further evolution of this practice. A registered report is a form of journal article in which peer review of the study protocol and the (in-principle) decision to publish the results occur before the study is run (this is called "in principle acceptance" or "IPA"). In this case, the research is submitted in two stages: one part is pre-registered and reviewed before the research is carried out, and the other submitted after the study has completed. Reviewers will check whether the protocol was actually followed. Journals have different approaches as to when the Stage 1 material is published.

### Why do this?

Of course, you're free to write up your methods, implement them, and only publish them after the results are in. One reason not to do things in that order — and also a reason for all formality around Registered Reports — is to avoid any possibility of "HARKing" or **Hypothesizing After the Results are Known**. Another benefit is that sharing your methods early on can allow them to be scrutinised and improved by others. While there's a chance that someone could "scoop" your methods and carry out the data gathering before you have a chance to complete your study, preregistration itself is evidence of the priority of an investigator's hypotheses. (Furthermore, in the case of Registered Reports, publication is guaranteed no matter how many similar papers are published between the two stages.)

### Working with open methods

Not all fields or research projects lend themselves to the full formality of preregistration, however, it's worth mentioning that it may be possible to share your "methods" openly, even when the "data" you gather cannot be shared (for whatever reason). Aspects of your process may be both rigorous and relatively easy for others to adopt — for example, a structured debrief protocol that you might prepare in advance of carrying out field work — even if other aspects of your methodology are deeply inductive or interpretive.

Initial meta-research suggests that preregistration helps to counteract the bias towards publishing primarily "positive" and "statistically significant" findings, i.e., findings that appear to confirm the researcher's hypothesis. Registered reports may provide the antidote to other "Questionable Research Practices". Here's one example. A 2021 paper found that *non-replicable* papers tend to be cited more frequently than *replicable* papers. As the authors point out, registered reports could help reverse this rather discouraging trend: this is because registered reports are reviewed before data collection has taken place, and so they will not be accepted simply because they happen to include some remarkable (but in the end non-replicable) finding.

One place you can find examples is on the "Peer Community In" platform (https://rr.peercommunityin.org/). PCI is an open peer reviewing service that both runs its own multi-disciplinary journal, and also maintains relationships with other journals that are willing to accept PCI-reviewed registered reports. In various examples hosted on the platform, you'll see that "Stage 1" of a Registered Report can include multiple rounds of reviewing. For example, the review of "Voice preferences across contrasting singing and speaking styles" took four rounds, spanned 15 months and "resulted in largely a new manuscript." After getting familiar with the medium, another practical step could be to find a protocol describing a project you're interested in, and carry out a replication study, possibly working together with colleagues or students. Have a look at "Ten simple rules for designing and conducting undergraduate replication projects" for further ideas.

https://web.archive.org/web/20210730224723/https://how-to-open.science/share/open-protocols/

https://rr.peercommunityin.org/articles/rec?id=750

https://rr.peercommunityin.org/about/pci_rr_friendly_journals

Serra-Garcia, M., & Gneezy, U. (2021). Nonreplicable publications are cited more than replicable ones. In Science Advances (Vol. 7, Issue 21). American Association for the Advancement of Science (AAAS). https://doi.org/10.1126/sciadv.abd1705

Scheel, A. M., Schijen, M. R. M. J., & Lakens, D. (2021). An Excess of Positive Results: Comparing the Standard Psychology Literature With Registered Reports. In Advances in Methods and Practices in Psychological Science (Vol. 4, Issue 2, p. 251524592110074). SAGE Publications. https://doi.org/10.1177/25152459211007467

Moreau, D., & Wiebels, K. (2023). Ten simple rules for designing and conducting undergraduate replication projects. In R. Schwartz (Ed.),

PLOS Computational Biology (Vol. 19, Issue 3, p. e1010957). Public Library of Science (PLoS). https://doi.org/10.1371/journal.pcbi.1010957

Henderson, E. L., & Chambers, C. D. (2022). Ten simple rules for writing a Registered Report. In PLOS Computational Biology (Vol. 18, Issue 10, p. e1010571). Public Library of Science (PLoS). https://doi.org/10.1371/journal.pcbi.1010571

https://replicationindex.com/tag/qrp/

# Research Diversity

This short talk introduces 'research diversity', why it matters, and how it relates to the question "who is involved in research?" As we'll see, there is a clear relationship between the diversity of research — in terms of its contributors, participants, methods, output formats, and possible routes to impact — and some of the widely discussed aims of open research. We'll have a look at some evidence about the relationship between Open Access and research diversity, supporting the conclusion that open access alone isn't enough to bring about some of the key aims of the open research agenda. We'll also have a look at some reflections on Open Source Software developer culture for some reflections on what not to do when trying to build a more diverse research culture. Lastly, we'll offer some practical suggestions about how to incorporate more of the benefits of research diversity in your own research work.

### What is research diversity?

The term **Research Diversity** itself spans several different meanings. What these different meanings have in common is that they are all about including a wide variety of perspectives, experiences, and voices in research. One way to achieve this is through the inclusion and representation of diverse participants and members within research teams. "Research diversity" can also refer to the *diversity of research practices*, output formats, and impact types.

### Why does research diversity matter?

Research diversity can be linked to themes such as *academic plurality* — a widely held value within academic communities that embraces the harmonious coexistence of different interests, convictions and lifestyles. These values might be exercised (or even put to the test) within multi- or interdisciplinary research. This might also call

to mind the varied activities and roles that are involved in creating a given piece of research in the first place. All of these considerations relate to another question: **who is involved in research?** As we'll see, the diversity of research topics and output formats is linked to the diversity of researchers themselves.

### Diversity of research participants

Other videos in this series point out that certain kinds of research simply cannot be done without meaningfully involving a suitable population of participants. As in the MAMMI study (mentioned in the "Participatory Research" video), first-time mothers were an eminently appropriate group to involve in a research project on maternal health. Similary, in the CLARISSA project (mentioned in the "Open Communities" video), children working in hazardous occupations are an important group to include in research that aims at systematic changes to child labour. Additional factors of diversity and inclusion need to be considered in such projects for the research to be valid and ethical. For example, the MAMMI study needed to reflect relevant factors of diversity among new mothers to be representative and useful. These factors included ethnicity, educational background and socioeconomic status. Researchers used **purposive sampling** to recruit women from more diverse groups, including those who were single mothers, LGB identifying, or who had experienced homelessness. While involving children working in hazardous occupations — as well as their employers — the CLARISSA project had to take particular care when it came to safeguarding. The overall points to make here are that involving diverse participants can have a range of beneficial impacts, including improving the accuracy of results, reducing inequalities, and avoiding biased results.

### Diversity of research contributors

Although open research practices make barriers between academic research and wider communities more permeable, when considering research diversity, it's useful to consider the demography of the academic research sector itself. The first thing to note is that there are many different factors and markers of diversity, including race, ethnicity, gender, age, and social class, among others. Empirically, there is a strong relationship between the characteristics of researchers and the research topics they study. For example, few people would be surprised to learn that publications about pregnancy are predominantly written by women, or that ethnic minorities tend to write more about racial discrimination. What's more intriguing is the fact that people from underrepresented groups tend to make more innovative

and novel insights, and this is often true for more diverse teams as well. Accordingly, a lack of diversity in research is a potentially significant liability for a country like the UK where women, minorities, and people with working class backgrounds tend to be underrepresented in academia, and particularly so at the higher levels of the academic hierarchy and within elite institutions. One striking set of numbers communicates in stark terms how these kinds of inequalities matter for academics themselves. According to UKRI's equalities monitoring,

"For the principal investigator role type, applications from the White ethnic group had a significantly higher award rate (31%) than principal investigators from both Asian and Black ethnic groups (23% and 21%, respectively)."

In other words, all else equal, your chances of getting a grant may be up to one third less than that of another applicant based on what's widely regarded as a protected personal characteristic. The reality, though, is that "all else" typically isn't equal. Firstly, the make-up of UKRI funding panels doesn't reflect the population of applicants, which may be associated with a difference of opinion about what kind of research is worth carrying out. In the US, bibliometric research revealed that although black and white investigators published the same number of papers during their PhD and postdoc years, papers from black investigators were cited less than half as much, had fewer co-authors, and often appeared in journals with lower impact factor. Preliminary evidence suggests that robust mentoring interventions and enhanced professional network development could help correct these trends, and reduce discrepancies in grant funding rates. However, if these interventions are given at the PhD level, it may be too late for many people who would have had a lot to offer as researchers — in both the US and UK, black students are significantly underrepresented at PhD level.

**Who is involved in open research?**

The inequalities that characterise research trajectories are reflected in the classic sociological concept of cumulative advantage, otherwise known as the Matthew effect (based on the principle "For to every one who has will more be given", or in short, "success breeds success"). Open research doesn't make this issue go away. As Ross-Hellauer et al. write in *Royal Society Open Science*,

"[M]aking processes open will not *per se* drive wide reuse or participation unless also accompanied by the capacity (in terms of knowledge, skills, financial resources, technological readiness and motivation) to do so. These capacities vary considerably across regions, institutions and demographics. Those advantaged by such factors will remain po-

tentially privileged, putting Open Science's agenda of inclusivity at risk of propagating conditions of 'cumulative advantage'.

A more explicitly inclusive approach to building an open research culture would be to foster discussions around what works and what doesn't work, who is included, and who is excluded by default. Let's turn to some of these points now.

### Open Access and research diversity implications

In terms of getting ideas out there, **open-access papers draw more citations from a broader readership**: this hints at key benefits including a diverse readership and diverse possibilities for reuse. However, it's currently the case that OA articles with Article Processing Charges (or APCs) are more likely to be written by people possessed of the male gender, employment at a prestigious institution, association with a STEM discipline, more research funding, and a more advanced career stage. It's the Matthew effect all over again. It should come as no surprise that the Higher Education Policy Institute concludes that Open Access is not enough to foster a vibrant open research ecosystem — they touch on the necessity of developing capabilities, connections, coordination, collaboration, and co-production strategies: themes that have also appeared in this video series.

Building more diverse teams can start with something as simple as working on projects that you wouldn't normally work on, with people you wouldn't normally interact with. When it comes time to publish, mechanisms like the CRediT taxonomy can help make sure that the diverse inputs to a research publication are suitability accounted for. That said, the rise of open research is associated with a broader ongoing culture shift away from an exclusive focus on publications and citation counts, and towards other ways of evaluating the impact of research. In the next REF, the weighting given to academic publications will be reduced from 60% in REF 2021 (down from 65% in REF 2014) to just 50% in REF 2029. There will be a corresponding increase in weighting given to 'people, culture and environment' — i.e., the factors that support the production of high-quality research and impact in the first place. As we've noted, involving diverse participants — both as research subjects and as research collaborators — can increase the robustness and novelty of your research. However, research diversity it's not just about research quality in terms of outputs. Diversity in research is crucial both for reasons of social equity, and because building our capacity to work with difference enhances our own learning.

⚠ **Practice Example:** *At Brookes, the new "Equality, Diversity and Inclusion Strategy 2024-29" [will be paired](https://www.brookes.ac.uk/staff/people/related-*

*strategies/equality-diversity-and-inclusion-strategy) with an annually-developed implementation plan, which will be worth consulting for ways you can get involved.*

Matthew Gandy. (2024) Losing control: REF 2029 and the downgrading of academic outputs
https://rgs-ibg.onlinelibrary.wiley.com/doi/10.1111/tran.12713

Hui, W. (2020). The Humanities in China: History and Challenges. In History of Humanities (Vol. 5, Issue 2, pp. 309–331). University of Chicago Press. https://doi.org/10.1086/710276

*Ecological Thinking: The Politics of Epistemic Location* (Oxford University Press, 2006)

David Rock and Heidi Grant (2016). Why Diverse Teams Are Smarter. Harvard Business Review. https://hbr.org/2016/11/why-diverse-teams-are-smarter

Robin J. Ely and David A. Thomas (2020). Getting Serious About Diversity: Enough Already with the Business Case. Harvard Business Review.
https://hbr.org/2020/11/getting-serious-about-diversity-enough-already-with-the-business-case

Bas Hofstra, Vivek V. Kulkarni, Sebastian Munoz-Najar Galvez, Bryan He, Dan Jurafsky, and Daniel A. McFarland, "The Diversity-Innovation Paradox in Science," *Proceedings of the National Academy of Sciences*, vol. 117, no. 17, 2020, pp. 9284-9291.

Ginther, D. K., Basner, J., Jensen, U., Schnell, J., Kington, R., & Schaffer, W. T. (2018). Publications as predictors of racial and ethnic differences in NIH research awards. In L. G. Koniaris (Ed.), PLOS ONE (Vol. 13, Issue 11, p. e0205929). Public Library of Science (PLoS). https://doi.org/10.1371/journal.pone.0205929

Ben Green (ed.). "Technology Ethics in Action: Critical and Interdisciplinary Perspectives," Special Issue of the Journal of Social Computing. 2021.

Marylin Vantard, Claire Galland, Martina Knoop; Interdisciplinary research: Motivations and challenges for researcher careers. *Quantitative Science Studies* 2023; 4 (3): 711–727. doi: https://doi.org/10.1162/qss_a_00265

# Open Communities: Working together for research impact

Welcome to this short talk on open communities and open research. To explore this topic, we'll draw links between open research, public engagement, and the theme of impact. We'll look at how an open and collaborative way of working can sometimes help bring about rapid positive change. We'll look at how a specific kind of community, called a "community of practice" builds the skills of its members. Then we will turn to several current examples of public-facing research projects. The talk concludes with some practical suggestions for working with communities, and pointers to places where you can get involved.

Whatever your interest in open research may be, it's worth remembering that we do open research — and, indeed, most kinds of research — in communities. Sometimes we might focus on "research communities" or "disciplines" — often, communities of experts. Other times, we might focus on the people in our neighbourhood, or on stakeholder groups who could be distributed internationally. As discussed in the video on "Participatory Research", some open research practices make the boundary around a given research project or team more permeable, inviting relevant people to get involved with the research. In this talk, we will deal more with the way open research percolates out into society at large. Since the open research agenda is largely concerned with ensuring the maximal public benefits from research, this is a crucial topic.

### Defining open communities

It's helpful to explain what we mean by 'open community' in connection with open research. An 'open community' gives everyone within that community the opportunity to be involved in activities related to a shared interest. Many communities in the open research space are open regarding materials and methods, and participation-friendly governance structures, helping people to get involved. A central point in this talk is that similar aspects of openness can help all kinds of communities — not just communities of researchers — get organised around problems that matter to them.

### "The societal impact of open science"

One way to think about the role that open research can play in communities is to have a look at the evidence surrounding its broader societal impact. In a recent review by Cole and colleagues, published in the journal *Royal Society Open Science*, most of the sampled studies that

attest to some form of societal impact due to open research practices focus on the impacts of *citizen science*. Indeed, as we covered in another video, citizen science frequently has impactful aspects "baked in" — for example, in terms of learning impacts for participants. By contrast, in all of the sampled papers, the benefits of Open/FAIR data were only ever discussed in speculative terms! The authors suggest that this is because the impacts of Open Data are not so easy to measure. This reaffirms something we've touched on in other talks in this series. Making data "open" typically is not the same as making that data concretely useful (or even usable). The concept of an 'open community' introduced earlier suggests two further hurdles to overcome. In order for open data to achieve measurable impact through re-use firstly, other people need to find the data interesting in some way. Secondly, they need to have ways to get involved in activities related to that data.

### 'Intelligent openness' and 'public engagement' for impact

This hints at links between "intelligent openness" and 'public engagement'. The National Co-ordinating Centre for Public Engagement defines the latter term as follows:

Public engagement describes the myriad of ways in which the activity and benefits of higher education and research can be shared with the public. — https://www.publicengagement.ac.uk/introducing-public-engagement

NCCPE further highlights that "Engagement is by definition a two-way process, involving interaction and listening, with the goal of generating mutual benefit." Similarly, 'intelligent openness' involves meeting people where they are at. Let's have a look at a straightforward example, showing how open dialogue can turn an existing community into a site of shared inquiry and problem solving.

### "Rural Self Help": An anecdote from Sri Lankan activist A T Ariyaratne

The late Sri Lankan activist A. T. Ariyaratne told a story about how he visited a rural village and asked the villagers what their biggest problem was. This was an example of public engagement at the village scale. Indeed, everyone was welcome to participate in the discussion. It quickly emerged that the biggest problem the villagers faced was repairing their water storage system. They explained that they had already been corresponding with the government for 15 years, trying to get the matter taken care of, and in all that time nothing had happened. Ariyaratne asked them to think together about what

it would take to solve the problem themselves. By talking it over, thinking it through, and getting organised, the villagers were able to identify a more direct solution to the problem. Indeed, they managed to carry out the necessary repairs themselves in one day of hard work! Ariyaratne shows that this story highlights three criteria for effective community problem solving: (i) community members need to have a thought which unites them, (ii) **techniques within their capacity**, and (iii) an organisational structure under their control.

The whole process was 'intelligently open', first in an open dialogue and then in a shared problem solving process using methods that were understandable and effective for everyone involved.

### Communities of practice as places to build capacity

In the scenario just described, once they had thought through the problem, the community could address their water storage problem in one day of hard work. Sometimes, though, the problems that need addressing are not one-off — and the best way to approach them is by gaining new skills. In the terms used by Etienne Wenger-Trayner and William Snyder:

"Communities of practice exist to develop members' capabilities, and to build, and exchange knowledge. People self-select to participate in the community. The community is held together by possession, commitment, and identification with the group, and with its domain of expertise. The community lasts as long as there is sufficient interest."

Let's turn to some examples of open communities.

### Example 1: A Developing Community of Practice around open research training

The first example is explicitly being developed as a Community of Practice. The UK Reproducibility Network is currently running a series of train-the-trainer courses that are intended to help embed open research in institutional practice. People who have attended these courses are expected to go on to deliver training in their home institutions. They will also be invited to join a Community of Practice for Open Research Trainers (https://www.ukrn.org/ukrn-community-of-practice/). The Community of Practice provides a space where peer-to-peer interactions help participants get better at sharing open research practices. The degree to which people are involved can vary: it's perfectly OK to just come along to some talks, but attendees are also invited to get more involved, for example, by sharing presentations on what they are doing locally with the rest of the group. This illustrates the principle of 'legitimate peripheral participation' that is

associated with communities of practice. Metaphorically, as people get more involved, they move from the periphery of the group into the core.

### Example 2: Science Together

Science Together is an initiative here in Oxford, which actively seeks the involvement of researchers and support staff from both Oxford University and Oxford Brookes, as well as local organisations. "Through **facilitated workshops** researchers are able to connect with local communities as a starting point to identify and develop funded, collaborative projects that help these communities to overcome a challenge or seize an opportunity." Incidentally, the "science" label in the name of this project is something of a misnomer. Current projects range from assessing the impact of community hubs and dance, to creative projects including rap and legally-sanctioned graffiti, to co-production of research with young people, on to the co-design of a platform for healthy eating. Science Together is not primarily organised as a community of practice, but rather as a space for using existing research skills in new contexts, and finding routes to impact by working with members of the community who aren't based at one of the city's two universities. The accompanying video on "Research Diversity" elaborates on the potential benefits of this way of working.

### Example 3: "People-driven solutions"

"People-driven Solutions" is the title of a 2024 report by a multi-party research consortium which tackled topics related to Child Labour in South and South-Eastern Asia (or CLARISSA for short). As in Ariyaratne's example, this project describes an outside team coming into existing communities, and working to turn them into a space for inquiry. In this case, their aim was to understand how to alleviate — or better yet entirely avoid — the harms that children working in hazardous occupations can experience. To do this, the research team worked to create an 'open community', inviting children involved in this work as well as other stakeholders — such as their employers — to engage creatively in finding alternatives. The authors stress that in this kind of work, it's not appropriate or even realistically possible to push through a "solution" that is imagined in advance. Instead, there is a continuous re-assessment of inputs, outputs, outcomes and impacts. To do this, they used a high-level "learning loop". A lesson from this way of working is that a rigorous review process can bring a new level of openness and clarity to any ongoing research programme — or indeed any programme of work in which research

plays a part.

### Get good at facilitation

Each of these examples required one or more people to take on the role of **facilitator**. Kim Cameron and David Whetten outlined five "P's" for running effective meetings — these can be adapted naturally as a framework for facilitating other kinds of collaborative activities.

- **Purpose**: What is this going to accomplish?
- **People**: Ensure that the relevant people are invited, with responsibilities relevant to the purpose.
- **Process**: Managing the group dynamic, including ground rules, giving people a chance to get to know each other, and keeping things moving along.
- **Plan**: A suitable space, a suitable agenda, and an appropriate decision-making structure
- The fifth "P", **Perspective**, is used to evaluate whether the meeting was, indeed, effective.

A more general perspective-building technique is the "After Action Review" — as used, for example, by the CLARISSA team. Some people leave this item out and shrink the list to four, "P's" — but if you're working to build and evaluate impact, it's really important. NCCPE makes some further guidance available on "How to evaluate public engagement projects and programmes", and similar considerations apply within open research projects.

Some people refer to a sixth "P", suggesting that the outcome of the group effort will be an actionable **Product**, though in general it may be some other outcome or impact. A key theme throughout this talk has been to keep track of these impacts. You can get further help thinking in through research impact matters from the Knowledge Exchange and Impact team within RIE.

### Get involved

⚠ **Practice Example:** *At Brookes there are various groups which describe themselves as Communities of Practice. There are forums for anyone interested in digital technologies; there is an Educational Leaders Forum specifically for Programme Leaders and Subject Coordinators; there are several RIKE networks, and a new [Local Open and Reproducible Research*

*Network](https://groups.google.com/a/brookes.ac.uk/g/open-research-network). The UKRN Open Research Training Community of Practice is open to Brookes staff who complete train-the-trainer courses. Getting involved in one (or more) of the communities of practice here at Brookes could be a natural step if you're looking to develop your workplace role. Other local communities — such as Science Together — could be good places to practise sharing your research outside of the university setting.*

David A. Whetton and Kim S. Cameron. Developing management skills: SUPPLEMENT C CONDUCTING MEETINGS (pp. 652-659) 8th Ed.

2012 Royal Society report on "Science as an Open Enterprise" remarks:

Research Councils UK white paper "What's in it for me? The benefits of public engagement for researchers"

## Open Educational Resources

Welcome to this short talk introducing Open Educational Resources (or OERs for short). We'll point out some of the defining features of OERs, and have a look at the barriers, incentives, and benefits pertaining to their use. We'll also touch on the topic of "Open Educational Practices", which include effective ways to use OERs. We'll begin by talking about how OERs relate to Open Research.

The most obvious connection between Open Research and Open Educational Resources lies in their shared openness. In fact according to UNESCO's definition, open research materials are themselves open educational resources "by default". Here's how UNESCO defines the term:

"Open Educational Resources (OER) are learning, teaching and research materials in any format and medium that reside in the public domain or are under copyright that have been released under an open license, that permit no-cost access, re-use, re-purpose, adaptation and redistribution by others." — https://www.unesco.org/en/open-educational-resources

That said, "research materials" often only become useful and usable when some background knowledge is in place. OERs can play a role in helping to establish that knowledge, thereby opening up research to engagement that goes beyond "access to materials". It's also the case that educational materials are only one part of a learning experience. Depending on the topic domain, effective use of OERs might be supported by a teacher, peers, or structured on-the-job learning op-

portunities, and so forth. The notion of **Open Educational Practices** has been developed to talk about easily shared methods of structuring learning. We'll come back to that topic after a deeper look at OERs.

### OERs and open reseacrh

Open Educational Resources may be intended for classroom use, or for self-study. They can be written by experts, or "crowdsourced" by people with varied backgrounds. In relation to open research, relevant OERs would range from openly-licensed textbooks that educators can use to help their students develop general-purpose research skills — all the way to highly specialised training that helps people get involved with a specific open research project. To put it quite bluntly, if you're developing open research and are not developing accompanying OERs, you may be missing a trick. That's because without suitable OERs to accompany your open research, people are likely to find it harder to build on your work, or get involved with your project.

### You're looking at an example

This series of videos is a collection of OERs related to open research practices. The videos, scripts, and assets (where possible) are not only available to watch online — but are also available under licensing terms that permit reuse and remixing. This helps illustrate a general point, which is that if you produce OERs you — or whoever owns the copyright — can decide how to licence them. However, if you're building on others' OERs, or if you decide to incorporate third-party materials in your work, you need to respect the copyright and any licence that applies to those materials.

### OER creation workflow

In line with the definition presented earlier, the steps to creating an OER are relatively simple: (1) Develop an educational resource, (2) (a) select a suitable licence, and (b) make sure it's clear to readers how the licencing terms apply (e.g., how to attribute derivative works), and (3) share the OER, for instance in a public repository or web page.

### So what…? Reuse and remixing

Let's briefly give some further attention to the questions "why?" and "so what?". A key benefit of OERs is that anyone who can access the resource has the right to re-use and remix it with any other OERs

which are available under compatible licences. As remarked in guidance provided by OER Commons (which is a large repository of OERs):

"[Remixing] is especially helpful when you would like to adapt the material to learners' needs, localize content to make the material more accessible, or add revised data to keep your resource current and up to date while preserving the original."

In short, OERs can help educators save time, can help everyone save a good amount of money, and increase quality and improve the learning experience in several ways.

### Barriers and incentives for using OER

Notwithstanding these potential benefits, people may hesitate to use or create OERs if they have trouble finding relevant quality-assured open materials, or if they aren't familiar with the way open licensing works. Another barrier to the creation of OERs can be summed up as "poor incentives". Why would educators (or educational institutions) — who are typically in the business of charging money to deliver an educational experience — give away some of the core parts of that experience? One answer is that decisions to publish OERs are often more mission-driven than business-driven. That said, OER "branding" can provide marketing and publicity opportunities, and OERs can fulfil other business-relevant functions; for example, including contributions to OER production in assigned coursework can enhance the student experience, by putting students in touch with a global community. Reusing OER produced by others can also save money and potentially increase quality. MIT's OpenCourseWare and the OU's OpenLearn initiatives are relevant examples of the institutional use of OERs. Notably, both of these projects make course materials available under the terms of the **CC BY NC SA** licence, a licence that forbids the sale of reworked materials. By contrast, Wikiversity currently collects 143 courses under the fully open-source-compliant **CC BY SA** licence. The different licensing options support different use cases.

### Students' perspective

All else equal, students are likely to appreciate it if instructors assign open materials, particularly because of the cost considerations. "Ease of use" is another relevant factor for students, both for formally-assigned course materials and any supplemental self-study materials. For students, quality of learning and recognised qualifications are crucial. On that note, one ongoing challenge for learning with OERs is how to incorporate assessment. Even where assessments are built into OERs, those assessments do not always lead to a certi-

fication; when they exist, such certificates may not be recognised by employers. Complementing OERs with replicable “Open Educational Practices” including peer interaction and “peer assessment” can help bring about robust learning outcomes.

### Co-creating and extending OERs: next steps for this project

In closing, let’s return again to this video series as an example of an open-research-relevant OER. The video series has been developed to provide a quick introduction to a variety of open research practices. Each video has accompanying material that people can look into if they want to take a deeper dive. Our plan is to share the scripts for these videos with the UKRN’s Training Community of Practice, and collectively revise the content into a textbook. Several general purpose OERs on Open Research already exist — such as the PaPOR TRAIL and the Global OER Graduate Network OR Handbook (linked below) — the hope is that a community-driven project can build in further specific dimensions of quality around “growing and embedding open research in institutional practice and culture”. Watch this space: the video description will be updated to include information on how you can get involved in future editions of this material.

Egan S, Tobin M, Palmer B *et al.* Developing an open educational resource for open research: Protocol for the PaPOR TRAIL project [version 1; peer review: 2 approved]. *HRB Open Res* 2020, **3**:84 (https://doi.org/10.12688/hrbopenres.13171.1)

Farrow, R. (ed.), et al. (2023). *The GO-GN Open Research Handbook*. Global OER Graduate Network / Open Education Research Hub. https://go-gn.net/gogn_outputs/open-research-handbook/

## Open Software

Welcome to this short talk on Open Software. It will explain what open software is, and describe who writes open software and why they choose to work that way. We’ll then look in further detail at “open research software”, focusing on practical pointers.

Open Software is software that has been published under a suitable open licence.
It may be referred to more fully as

- Open Source Software,
- Free/Open Source Software,

- or Free/Libre/Open Source software

The extra words are used to call to mind the central implications of open software licensing: in particular, the licence permits people to run the program as they wish, for any purpose; the code is available to study; and the code and software can be distributed, even in modified form.

The most straightforward link between open research and open software comes in the form of "open source software used in research settings". It is also possible to do open research that takes some parts of the open software ecosystem as a data set and a site of interaction. In this sense, it's also possible to do 'open research' about 'open research software'. However, let's first develop a big picture view.

### What is open software?

Perhaps the most central feature of open *source* software is that it "anyone can inspect, modify, and enhance its source code." Approximately seventy-seven percent of all code in existence is open source software. As a point of comparison, that is a considerably greater penetration than Open Access, since only around 48% of all research publications are OA.

### Who writes open source software?

With this point in mind, it makes sense that lots of different kinds of people write open source software. These include:

- Professionals working for computer companies
- Computing professionals working for other companies
- Researchers and research students
- Contributors working for charitable purposes
- Contributors working for their own professional development and hobbyists who do it for fun

In recent times, the top open source contributors were all based at companies – particularly at big tech companies.

### Why do people contribute to open source software?

One motivator is the expectation the community will help develop and identify additional features, supporting innovation. Another is that, even if the software is given away for free, companies can sell complementary services. A third motivator is cost saving. In this connection, a Google blog post mentions:

"It's great if [new employees] can hit the ground running and already know and use the tools we have developed. Familiarity with our software and data makes engineers productive from their first day at work." Another factor that applies to both companies and individual contributors is the norm of good citizenship.

### (Open) Research software

How are these points relevant to researchers? In 2014, the Software Sustainability Institute Survey found that 92% of academics use **research software** across disciplines, and 69% say that their research would not be practical without it. Around three-quarters of papers in the journal Nature mention "software". Releasing research software under an "open licence" can have benefits similar to those mentioned above, i.e., supporting innovation, scaffolding complementary offerings, and saving money. To put it in another way, open licensing can help avoid wasting the resources that enable research to develop.

Access to source code has clear repercussions for a research community in terms of reproducibility. Fortunato and Galassi provide a detailed account of the creation of the GNU Scientific Library, comparing it to the other options available at the time for scientific computing. They point out the value of creating a robust design document, outlining the rationale for creating the software, and making it clear how people could contribute.

Packages for specialised research purposes range from neuroimaging to electron microscopy to agent simulations are often released as open source. Other more general-purpose tools ranging from bibliography managing software to document editing suites are also. Many of the more technical tools used in quantitative research — like R and Python, and the scikit-learn Python package for machine learning — rest on an open source foundation. There are often (but not always) proprietary systems for all of these purposes as well. The benefits of open licensing can be even more apparent when working on a 'niche' software package (or niche research topic) for which commercialisation can be less rewarding in financial terms.

### How to create open research software

The first point to make is if there is an existing open source software package out there that does something close to what you need, then the licencing terms allow you to modify the software to address your need. That could be a good place to get started with open source, often bringing with it further benefits associated with participating in an open community. (This can be a great way to learn.) If you're not and have no intention of becoming a programmer, you can often pay someone to make the changes you have in mind — many popular open source software packages (like the Drupal and Wordpress website building tools) have well established "service" economies built up around them.

That said, if you're developing new open source research software from scratch, the following guidance is useful:

- Choose a suitable licence for your software. Popular licences used in the software ecosystem differ as to whether programs that contain open source software must also, themselves, be licensed as open source software. The latter are referred to as 'copyleft' licenses. Guidance is available on choosealicense.com.

- Make use of a code repository when developing your software (many people use Github, but there are other options)

- Document and describe your software, which will enable others to use it and may help them contribute as well — as we saw briefly with the example of the GNU Scientific Library, mentioned earlier.

- Archive key versions or releases to get a DOI: this will make your software citable. You can use Zenodo alongside Github for this purpose, for example.

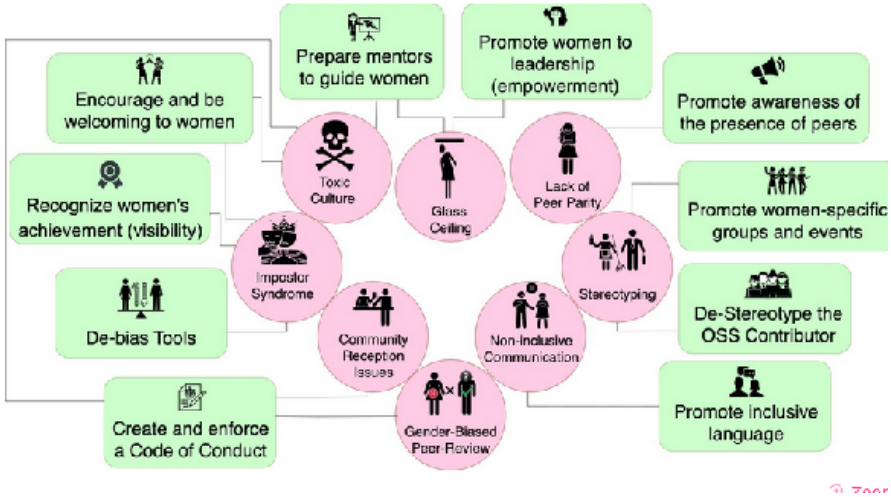
- Cite your software in publications, and encourage others to do the same.

- Indeed, you may wish to publish a software paper — the *Journal of Open Research Software* may be especially suitable, but there are other options as well.

### Building contributor communities

Perhaps surprisingly, one demographic that disproportionately does not participate in open source software development is "women"

— even though women comprise 15–22% of software engineers, the percentage of female contributors to open source software is between 7.5% and 10.4%. If you're building a software community, it's recommended to adopt strategies that encourage women's participation, as per this infographic:




Open software communities (including research software communities) are often quite friendly to newcomers, and there are many Open Educational Resources that help people get started with software, including a collection of curated resources on OER commons (https://oercommons.org/curated-collections/508). Free training that is specifically devoted to developing research software is available from the Software Sustainability Institute. Some developer communities have made their own small grants available to support innovation. Larger open software specific grant funding may also be available. Moreover, software is covered by the UKRI Open Access Policy, where it is classified as being part of the underlying research materials that form the "research data" that should be made openly available unless there's a good reason not to do so. Specific guidance is available from the FAIR for Research Software (FAIR4RS) Working Group concerning scientific data management and stewardship of research software.

⚠ **Practice Example:** *Here at Brookes, good settings to discuss open software include AIDAN and the Local Open and Reproducible Research Network.*

S.J. Hettrick et al, UK Research Software Survey 2014", DOI:10.5281/zenodo.1183562

Nangia, Udit; Katz, Daniel S. (2017): Understanding Software in Research: Initial Results from Examining Nature and a Call for Collaboration. doi:10.1109/eScience.2017.78

**And notes to harvest for more links**

Gun, Arkaprabha, and Tushar Garg. “Prerequisite for reproducible science: a call to embrace code sharing.” The Lancet Regional Health-Southeast Asia 29 (2024).

“Women’s Participation in Open Source Software: A Survey of the Literature” — https://arxiv.org/abs/2105.08777

https://opensciencemooc.eu/modules/open-research-software-and-open-source/

https://open-science-training-handbook.github.io/Open-Science-Training-Handbook_EN/02OpenScienceBasics/03OpenResearchSoftwareAndOpenSource.html

There’s even a journal for “Open Research Software” https://openresearchsoftware.metajnl.com/

https://www.geeksforgeeks.org/difference-between-free-software-and-open-source-software/
https://en.wikipedia.org/wiki/Free_and_open-source_software

How does Open Software align with the FAIR principle? (Findable, Accessible, Interoperable and Reusable - https://zenodo.org/records/6623556#.YqCJTJNBwlw; https://www.nature.com/articles/sdata201618

http://oss-watch.ac.uk/

OpenChain compliance advice from open source licence analysis through to DevOps process design and implementation.
https://orcro.co.uk/

- What is open software in the context of open research?
- What are the benefits and disadvantages/risks associated with using open software in open research?
- Is open software related to Open Source Software (OSS)?
- How do we distinguish between Open and Closed Source Software (OSS vs CSS)?
- What are the FAIR principles of Open Research Software/ How does Open Software align with the FAIR principle? (Findable, Accessible, Interoperable and Reusable - https://zenodo.org/records/6623556#.YqCJTJNBwlw; https://www.nature.com/articles/sdata201618)

# Open Code: Notebooks &c.

Welcome to this short talk on “open code”. This talk looks at research in which computational thinking is particularly relevant, and considers how ‘openness’ can make this kind of work more effective. We’ll look at some more practical points related to sharing open code via mechanisms like “open notebooks”. The talk concludes with some further reflections on research replicability, and how fields that use code can learn from fields that don’t, and vice versa.

Many researchers who would not describe themselves as software developers use “code” in their work. Often this takes the form of data analysis scripts written in general purpose languages, like R and Python. Research communities are increasingly pushing for such code to be shared openly. For example,

“Even before the introduction of the code policy in 2021, PLOS Computational Biology authors voluntarily shared code at a high rate (61% in 2020).”
https://theplosblog.plos.org/2022/05/uphold-the-code/

This short talk will briefly describe the relationship between open research and computational data analysis. PLoS defines “open code” as follows: “Open Code refers to custom, author-generated code used in a scientific research study—often during data collection, interpretation or analysis—and subsequently made publicly available under a [suitable] **license** via a linked repository, or as Supporting Information.”

Assuming you publish your article under open access terms, the associated open code may or may not be published under the same license. An open source software licence — rather than a Creative Commons licence — is typically recommended. Some journals like Elemental Microscopy are now publishing peer reviewed “computational articles”, which use a notebook interface to include ‘open code’ in the body of the article. (Naturally, you can also publish ‘open code’ without publishing a journal article.)

### Computational epistemologies

Before diving into the practicalities of sharing code as an open research practice, it is worth saying a few words about how current computational research tends to work. The basic narrative is that data analysis is carried out to convert raw data into “information” on the way to becoming “knowledge”. It’s worth remembering, though, that the reason that some particular data were collected or selected (i.e., and not others) almost certainly reflects certain biases. These

background considerations may be much harder to inspect and reason about than the data analysis code itself! In short, being computational in nature doesn't make a research project "better" than a project that isn't computation — but shared code can, because of its computational nature, contribute to certain aspects of replicability.

### Broader openness makes open code even better

From an open research perspective, open code ideally comes with the raw data, and includes any data cleaning and processing code, in addition to the analysis routines and the cleaned-up data. Documenting choices at this level supports a healthy and robust discourse about the research finding. Further contextualisation concerning data selection, and the interpretation of findings, as well as guidance for people who would like to build on the work can improve robustness, replicability, and impact.

### Notebook science as an example

Notebook environments are a form of so-called "literate programming" that can make it easy for people to interact with open code. Notebooks generally contain a mixture of code, examples, results, and interpretations which make them great for learning. "Jupyter Notebooks" are one particularly popular format. There are other notebook environments out there that work with a further variety of languages and computational backends. (Some platforms, like Nextjournal allow you to mix languages within one notebook.) Online hosting is available from various providers, which means that you don't have to install any software locally to develop your own notebook or to run someone else's, which can be an easy way to get started. Github, for example, can be used to share and collaborate on notebooks.

### New directions in open code

When used well, computational notebooks can help support reproducibility. However, they don't have FAIR principles built in; accordingly, attention has been given to aligning notebooks with the FAIR principles ("Can workflows in notebooks be made FAIR?"). "Provbook" and "Persist" are two mechanisms for tracking interactions with notebooks in computational terms, and which help document the transformation of raw data to controlled data. The Jupyter Kernel Gateway can be used to support strictly computational interaction with a notebook, allowing it to be used as a web service. It's also worth mentioning that Jupyter has a collaborative editing mode that you can use to

co-develop code in real time as part of a team. (All of these features may require some setup, but they show the direction that the tools are heading in.)

### Reproducible code, analogues, and extensions

If you want to build open code that others can use, it may be worth reviewing relevant guidance on writing **reproducible code**, available, for instance, from the British Ecology Society. In some fields of research, reproducibility isn't necessarily the most relevant criterion (even when software is being used). Matthew Hanchard has recommended that qualitative researchers adopt a practice centred on **re-renderability**, which refers to a way of working that makes it transparent to readers how and why the authors have rendered their interpretations, and that allows new researchers to 're-render' the data to make sense of it in their own way.

If working with code isn't your cup of tea, you may still gain some inspiration from reflecting on the role that code plays, and using this to inspire your own open research practice..

"Some qualitative papers include full coding structures and detailed quotes (e.g. as supplemental files), allowing the reader a similar level of insight to the researchers' workings as the sharing of quantitative statistical analysis code." — 10.1177/13591053241237620

### Putting it all together

Best "open code" practice will combine aspects of both code sharing and insight into the qualitative aspects of the code and a lot more. In the short paper:

"Open Code is not enough: Towards a replicable future for geographic data science"

... Levi Wolf and co-authors outline several further points that can contribute to the replicability of coding projects. Here are some take-home points from the article that, if widely taken up, would help build a robust open code ecosystem.

1. Use a "developer contract" to provide consistency across aligned (but separate) projects.
2. Institute active, robust scientific code review for papers, programs, and pedagogical
material.
3. Create easy, effective, and open technical spaces for contributors to write and maintain
documentation for fellow developers.

4. Build space, tools, and practices to help junior developers get invested and involved.
5. Develop mentoring so interested users can eventually become developers themselves.
6. Ensure each problem is solved exactly one time.
7. Provide simple, effective, and usable documentation for external scientific users.

— https://osf.io/preprints/socarxiv/3hbnt

**Closing note**

⚠ **Practice Example:** *At Brookes, if you want to try these tools, in addition to using hosted software (like Google Colab), you can use Jupyter notebooks on your computer after installing Anaconda with AppsAnyware. Recordings of tutorials developed in the 2024 Data Challenge hosted by the AIDA network are also available.*

Lamb D, Russell A, Morant N, Stevenson F. The challenges of open data sharing for qualitative researchers. *Journal of Health Psychology*. 2024;29(7):659-664. doi:10.1177/13591053241237620

# Open Data

Welcome to this talk on open data. The talk covers the benefits of open data, the circumstances under which open data isn't appropriate, and the relationship between open data and "web scraping". We will touch on both the history of open data, and current work that is helping to build the future of open data. The talk concludes with some practical pointers about how to get your hands on relevant open data for use in secondary analysis, how to prepare new data sets for deposit, and some reflections about what kinds of re-use different data formats support.

'Open data' describes data sets which are freely available to anyone to use and republish as they wish, without restrictions. Typically this implies that the data that is available over a computer connection — though you could potentially find legally-unencumbered data in out-of-copyright books and other historical records. Open Data is one of the key tenets of Open Research, though, as of now, it has not had uniform uptake: depending on the discipline, the rate at which data is available to support publications can range from 9% to 76%. Furthermore, in comparison with the development of open licences for publications, the development of legal mechanisms for sharing data has been slower, partly because the rights that are attached to research data are typically not a simple 'copyright'. Funders also have differing

requirements. Agencies that disperse **public** funds are more likely to require research data produced as part of a funded project to be made publicly available by default, with exceptions for valid privacy, confidentiality, or security reasons. Additionally, funders provide different levels of support for open data practices (Oxford University have published a useful guide).

### Why open data?

The Royal Society outlines these rationale for archiving data under terms that allow reuse.

- Firstly, it preserves your research contributions (after all: data that isn't archived in a sensible way can quickly be lost; duplication of the archive is a further strategy to avoid data loss).
- It allows others to build on your work
- It allows interested readers to replicate and/or verify the results of a study
- It can increase citation levels (this point is borne out in a 2020 article published in PLOS ONE)

### When is open data not a great idea?

Without being exhaustive, valid reasons for *not* making data open can include any or all of the following: the data may include personal information; informed consent for data sharing hasn't been secured; and possible uses of data may not benefit the people the data describes. To underscore this: just because you have gathered a data set does not necessarily mean you have the right to share it: the rights may be distributed across and among various institutions, data producers, and private citizens.

### Why not just 'scrape' data that's publicly available on the web?

In the United Kingdom, ***web scraping (i.e., programmatically gathering data from public web pages) is legal*** as long as it is done in a way that does not violate any other specific laws. In particular, this means that you'll need to comply with any Terms of Service that you've agreed to, obtain consent from individuals whose information you're gathering, and respect intellectual property rights. It is worth emphasising that under the terms of the GDPR, publicly available data is still personal data and cannot be repurposed without

the knowledge of the person to whom that data belongs. If you can scrape data in a way that complies with all of these requirements, that *doesn't* necessarily mean that you can (re-)publish the collected data set as open data. Your rights to publish material as 'open data' typically depend on licence information that has been published on the source site, or on explicit permission from the rightsholder in some other form. By contrast, a publicly available open data set should have cleared all of these hurdles already, but it behoves you to do at least some due diligence to be perfectly clear on permissions in any case.

### History of Open Data

An early conception of "open data" was referred to in the 1970s, with reference to international partners working on satellite operations, who would share information with NASA. In 1995, a National Academy of Sciences report "On the Full and Open Exchange of Scientific Data" expanded the purview of the term: alluding to the need for sharing data about various global phenomena that cross national boundaries, including astronomical data, marine data, aviation data, climatological data, satellite data, and various models that work with such data. Open data may be needed alongside open source software to make computational research sharable and scalable.

In the mid-'00s, **open government data** increasingly took off, through policy developments such as the "Memorandum on Transparency and Open Government" — signed by President Obama on his first day in office — followed soon thereafter by "Putting the Frontline First: Smarter Government", in the UK. Nevertheless, looking internationally, analysis on opendatabarometer.org shows that only 7% of key government data sets — such as land registries or government budgets — are actually "open".

### The Future of Open Data

In their book on the Future of Open Data, Pamela Robinson and Lisa Ward Mather highlight that **Publishing data often fails to achieve meaningful "awareness" or insight because making sense of data is not easy.** Universities may be especially well placed to address this gap, thanks in part to the diversity of perspectives that go into university research (see the video on "Research Diversity"). Writing on Govtech.com, Colin Worth argues that

"Governments aren't gaining the benefits of open data today because there's not been a rigorous effort to integrate the concept of openness into public-sector work."

Varied uptake of data sharing across academic disciplines suggests that there may be similar potential for value-gain in the research sector as well.

**Rather than producing open data, start by using some**

The simplest suggestion for people new to open data would be to look for existing open data sources that can support your research work.

**Guidance on how to prepare material for deposit**

When aiming to be "as open as possible" researchers may need to give more attention to questions like "How do we ensure anonymisation is effective?" before uploading data publicly. If this applies to you, you may want to check out recent Guidance from the ICO (linked below).

**What kind of open data? Depending on the expected use, more 'stars' might be better.**

A "five star" rating system for open data was proposed by Tim Berners-Lee, founder of the World Wide Web. The ratings are cumulative, with one star awarded for data that is available on the Web under an open licence (or more permissive terms, such as CC0). A second star is awarded for sharing structured data, and a third for data in a non-proprietary file format. An openly licensed file in the widely familiar CSV format would match this description. The fourth star is available only when meeting a somewhat more technical requirement: the data should use Uniform Resource Identifiers (or URIs) so that it is unambiguous. We mentioned several different kinds of URIs in the video on "Repositories and Persistent Research Identifiers" — such as "DOIs" and "ORCiDs", which are unique identifiers for documents and researchers, respectively. So, for example, if you were creating an open data set of researchers, you would be able to distinguish between the various researchers who have the same name by using their ORCiDs. This brings us to the fifth star: "Linked Open Data": for example, the records hosted on ORCiD.org link researchers both to the institutions where they work or have worked in the past, and to their publications — and all of this information is available in an unambiguous machine-readable format, both as a bulk download, and per user with a simple web command: ("curl -H"Accept:application/ld+json" -L https://orcid.org/0000-0003-1330-4698").

While "five star" open data may sound like it's always better, it's not always available (tabular data is much more common). You'll want to weigh the trade-offs involved when preparing a public data set.

You may find it helpful a useful walkthrough on the rationale for using Linked Open Data, published by the Open Data Informationsstelle Berlin (ODIS), linked below.

**Next steps**

⚠ **Practice Example:** *Here at Brookes, Nina Lewin, our Research Data Manager, can help with many aspects of research data, ranging from finding useful secondary data sources to creating data management plans.*

https://ico.org.uk/about-the-ico/ico-and-stakeholder-consultations/ico-call-for-views-anonymisation-pseudonymisation-and-privacy-enhancing-technologies-guidance/

# Open Publications and Transformative Publishing Agreements

Open Access is an important part of the Open Research agenda. Currently, approximately one third of articles published appear immediately and fully as Open Access. This talk will cover the different kinds of OA, and how each of the different models works in practice. It reviews the Transformative Publishing Agreements that UK universities have in place with publishers, which allow articles to be published without a per-article fee. While incorporating a representative selection of relevant critical perspectives (of which there are many), the primary purpose of the talk is to provide practical advice for academic authors.

If you're not familiar with Open Access, it may be worth having a look at the talk on "Open Licences" before diving into this one. That talk introduces the legal infrastructure through which Open Access is supported. The key take-away is that academic publications are increasingly available under permissive legal terms, which include being free access to read.

This stands in rather stark contrast with the traditional model of academic publishing, which is funded through fees paid by people who purchase the publications. Apart from textbooks and other publications that individuals buy for their personal libraries, typically, publishers' fees are paid through purchases and subscriptions taken up by University libraries. Indeed, getting access to a wide range of publications is one of the perks of working at a university. However, with the rise of Open Access, anyone with an internet connection can already access a wide range of books and articles for free. At least on

the face of it, if published material is available for free, this significantly reduces the incentive to purchase it for a fee. Not only does this suggest the need for different business models for publishers, it also hints at the evolving role of libraries and even universities. This talk will look at some of the alternative paths forward.

### A brief reminder about preprints.

Before looking at the different types of Open Access publications — which are associated with different publishing business models — it's worth briefly refreshing ourselves about preprints (which don't have a business model attached). A "preprint" is an article that has been uploaded to a suitable repository, perhaps at the same time the article is submitted to a journal, or, indeed, whenever the authors think that the material is ready to share with the world — even in the drafting stage if the work is happening in the open. By default, preprints are not reviewed, but they do make academic work available for public scrutiny. Most journals are happy to accept articles that have been posted as a preprint, though some do impose restrictions, so it's worth checking the relevant policies.

### Different types of Open Access publications

The primary types of OA publications are known as Green, Gold, and Diamond. While the names might sound a bit arcane, their differences between them are easy to understand.

#### Green

With the preprint model mentioned earlier in mind as a point of reference, Green Open Access is pretty straightforward. This model refers to sharing what's called a "postprint" — that is, a copy of a paper which has been sent to a journal, peer reviewed, modified as required, and accepted for publication. Publishers differ as to whether copyright must be transferred to them at this stage. Accordingly, what the author may do with the "Authors' Accepted Manuscript" depends on what rights they have. In some cases, it may be possible to make the AAM publicly available under a fully permissive open licence, but some publishers don't allow that. For example, Elsevier requires authors publishing their accepted manuscript to attach a **non-commercial Creative Commons** user licence (CC-BY-NC-ND). Some journals may additionally impose an **embargo period** (for example, 12 months) that temporarily delays the distribution of such 'postprints'. These restrictions allow the publisher to gather revenues associated with publishing the article using a subscription model. Once a postprint article is available, there's no cost for the

reader to access it. It's worth noting that some journals will waive the embargo period in exchange for what's referred to as an article development charge (ADC); and that some journals simply do not enforce an embargo period whatsoever. In principle, an ADC allows a journal to recoup costs associated with finding and organising peer reviewers and other relevant editorial work: it's worth remembering that the peer reviewers themselves typically work for free.

**Gold**

Gold Open Access describes a model that's similar to the ADC just mentioned, except that in this case, the article is published as Open Access directly by the journal itself. The author can of course archive it in a repository as well, but in this model that's a side issue. Gold Open Access means that there's no embargo period, and no cost for readers to view the article right away. However, there is typically a fee for authors — which typically will be covered by their institution or funder, rather than paid out of pocket. Furthermore, some publishers waive or reduce the fees for authors who cannot pay the standard charge: in the first instance this is decided by their country of residence.

**Diamond**

Diamond Open Access does away with fees for both authors and readers. Naturally, there are still costs involved in producing Diamond OA publications, however. These costs are typically underwritten by an academic society or charity, and may be supplemented by "suggested donations" for people submitting books or articles.

### What is the status of these different types of OA?

Each of these models has its proponents. Diamond OA might sound like the best option — after all, who doesn't like knowledge that's free for everyone, and that is community supported? However, it may also sound like the least plausible in practice. In fact, "as a far as the number of individual journals is concerned, diamond open access is the main form of open access publishing:" That said, Diamond OA only accounts for 8–9% of all scholarly articles — despite the number of Diamond journals, they tend to have lower throughput than bigger journals.

As for the other models, it's been argued that "Publishers convinced funders that the only realistic way to transition to open access was to embrace pay-to-publish OA, demoting green OA to an [']also ran[']." That said, while there has been a significant groundswell of interest in Gold OA, Green OA is having a resurgence as well, both through the uptake of Rights Retention policies, and also because some funders

(such as the Gates Foundation) have refused to fund Gold OA. More on all of these points, below.

### Transformative publishing agreements

As mentioned, in the "**Gold OA**", there is a fee to publish articles. Transformative agreements are arrangements between publishers and institutions — or funders — who pay for a certain number of OA publications upfront. (Some TAs are offered in an "uncapped" format.)

**Plan S** is (or was) a strategy whereby funder requirements on OA publications are to be implemented, in connection with TAs. https://www.coalition-s.org/plan_s_principles/ For enthusiasts of Open Access, this may sound like a great idea, but some are more cautious. In any case, 2024 was supposed to be the deadline by which Plan S would be delivered.

"From [...] 31 December 2024, cOAlition S will no longer financially support transformative agreements or transformative journals. Instead, funders will direct their efforts to innovative and community-led Open Access publishing initiatives such as the Diamond model of OA."

Some publishers, such as Cambridge University Press, have indicated that when authors don't have a Transitional Agreement in place, and cannot afford a APC, they will route the papers to Gold Open Access without charge.

### What about Green OA and "Rights Retention"?

Rights Retention is an agreement between the academic and their institution which relies on a contractual agreement to make academic works available under a CC-By licence. In this scenario, authors still own the copyright to their work as the default. The licensing arrangement is set up in such a way that it takes place prior to any subsequent copyright transfer.

Practically speaking, RRS is implemented by including a statement such as the following within the acknowledgements at the time of submission.

"This research was funded in whole or in part by [Funder] [Grant number]. For the purpose of Open Access, the author has applied a **CC BY** public copyright licence to any Author Accepted Manuscript (AAM) version arising from this submission."

Some publishers are known to question these policies, and there is ongoing dialogue around their use. Some publishers may route articles known to be under RRS to their for-fee journals.

⚠ **Practice Example:** *(At the time of writing, Oxford Brookes has not implemented a Rights Retention clause, though it is under discussion.)*

### If Diamond OA is so widespread, why isn't it the norm on a per-article basis?

For a long time, it's been relatively simple to publish material 'openly' on your own web page. Peer review and the prestige of publishing in competitive journals adds considerable value for academics — robust peer review is a crucial differentiator between academic publications and "vanity" or "self-published" texts. As indicated in the talk on "Open Peer Review", there is an increasing variety of peer review models out there. However, publishers — and this is especially true in the case of long form books — often provide other services as well, and these, typically, are not free. Martin Paul Eve, who runs the Open Library of Humanities, tabulated the fixed and per-article costs of running a small publishing business. These included fixed costs of £182K/annum, and nominal per-article costs of around £100. In short, there are costs involved in the publishing business that can't easily be defrayed, even with efficiency-boosting tools like open source software. Of course, these costs can be underwritten through noncommercial models, and this is what happens in Diamond OA. Nevertheless, as indicated earlier, most diamond open access journals are managed by academic institutions, communities or platforms, and don't have the same efficiencies that larger publishing houses have.

### Problems for research diversity

Marcel Homba has argued that because not all authors can afford Article Processing Charges, a cultural preference for Gold OA can distort the kind of work that is carried out — "amplify[ing] certain publication biases that favour topics and viewpoints that are backed by rich organisations and industries". This has implications for Research Diversity (see the video on that topic for relevant background on why this matters).

### Other problems

Green OA is not without criticism either. The basic argument is that Green OA relies on continuation of library subscription fees, which creates a free-rider problem: why pay a fee if you can just get the

articles for free. A counter-argument is that authors have the right to do what they like with their work, and publishers shouldn't try to secure a monopoly on OA publication. Whichever side you take on the matter, this discussion highlights that the "business aspects" of research (and open research in particular) can matter more than we might naively think they do.

### So what should I do as a researcher?

⚠ **Practice Example:**

### Are there any other practical things to do?

You could follow the lead of Deakin University and others who use the open source Open Journal Systems (OJS) software to set up and host a new Diamond journal — or find other ways to get involved with other innovations in OA, such as overlay journals.