OPEN SCIENCE

A PRACTICAL GUIDE FOR PHD STUDENTS

Index

Legend
Underlined text is explained in the glossary.
This triangle ▼ refers to tools which are given as examples.
Links can be found in 'Going Further'.

Introduction

This guide to Open Science is designed to accompany you at every step of your research while doing your PhD, and we hope beyond. From developing your academic approach to the dissemination of your research results, it provides a set of tools and best practices that can be directly implemented and is aimed at researchers from all disciplines.

Open Science & Scholarship focuses on making research accessible for all by removing as many technical or financial barriers which may hinder access to research as possible. It involves opening up the academic process itself, not just the outputs. From increasing transparency and sharing the data and methods used for publications to bringing the general public into the research process by means of Citizen Science, embracing Open Science is a great way to enhance your potential impact.

These changes to basic research processes help enhance research credibility, reproducibility and worldwide reach. This reach adds to the ultimate goal of Open Science & Scholarship: to create equity of access to research content worldwide, irrespective of language, location, ability, affiliation or economic status.

Open Science & Scholarship are regarded as internationally important and in the past 12 months UNESCO has adopted a series of recommendations on Open Science which it recommends that all 193 member states apply.

In the UK, most research funders including UKRI, Wellcome Trust and NIHR have their own Open Access or Open Science policies which require different aspects of Open practice. Similar policies can also be found across most major European and American funders. This is why this guide recommends embedding these principles into your work from day one. This has the potential to make meeting these requirements a much simpler part of the transition to your future career.

Ultimately, it is researchers like you whose commitments and practices embody and bring Open Science & Scholarship to life. As you progress through your PhD programme and begin to prepare your thesis we ask you to put these principles into practice and drive these goals forward. The principles will inform the research environment in the future so will provide the framework for your research career whether it is within or beyond academia. We all wish to make our research count and to encourage wider society to become more involved and to enjoy the fruits of the discoveries that we make.

We hope this guide will motivate you and provide the means for you to realise the ambitions of Open Science & Scholarship by understanding all of the ways you can share your research with as many people as possible.

UCL Office for Open Science & Scholarship team

August 2022

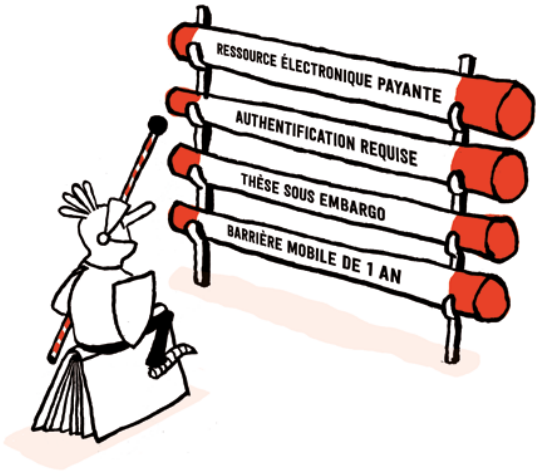
1

Embracing open science principles

Using freely accessible resources

You are beginning work on your thesis.
You need access to the publications, articles and data already produced in your field. This is the beginning of your obstacle race...

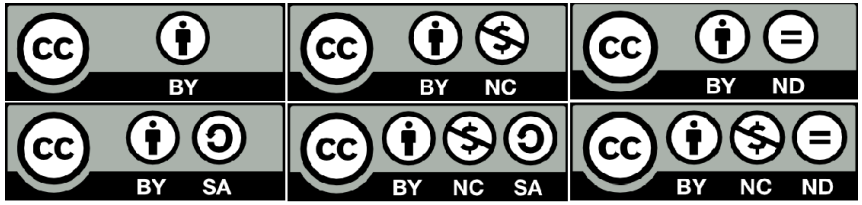
Access and reuse



The open science movement's aims are to **facilitate access to academic content and encourage its reuse**.
Often we speak of:

- Open-access resources: thanks to their author and/or publisher, these resources can be freely accessed without researchers or their institutions having to pay.
- Free resources: as well as being open access these are reusable depending on the distribution license involved.

For example, Creative Commons licenses allow reuse as long as conditions set by the author are respected.
For more information, visit ▼**creativecommons.org**.



REMEMBER

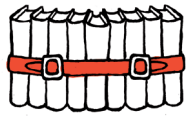
The fact that a resource is open access is not a guarantee of quality in itself. It needs to be critically evaluated before being used like all document resources.

Where should you look for resources?

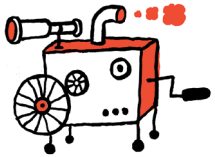
Open access platforms: Open access journals can have diverse business and editorial models. The ▼**Directory of Open Access Journals (DOAJ)** contains over 17,000 verified OA journals for you to search, and ▼**Directory of Open Access Books (DOAB)** collects Open Access books.

Open repositories: These may be institutional or subject based and outputs are deposited in the archives by researchers themselves for free. Explore ▼**Core (core.ac.uk)** - a search engine specialising in open access academic publications.

Platforms for preprints or working papers:
The availability of these unpublished articles means researchers can quickly discover the latest research approaches. Investigate ▼**OSF** and ▼**ArXiv** - two of many platforms for Preprints.



Databases for theses or academic works:
These databases collect the digital thesis collections of universities and research centres. Visit ▼**Dart-Europe** or the ▼**E-Thesis online service (EThOS)**.



Data and code repositories:
These may be multidisciplinary or specialised. Searches can be run for different types of data and they enable the deposit, conservation and sharing of research data. ▼**FAIRsharing.org** is a good resource for standards, policies and databases. ▼**GitHub** is the home of Open Source code and software.

Specialist search tools
These aggregate open access content to make them easier to discover. ▼**Unpaywall** and ▼**OpenAccess Button** can both be used to seek out freely available versions of resources that are behind paywalls.

Planning data management

WHAT ARE RESEARCH DATA?
Research data are the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form (e.g. print, digital, or physical). It isn't limited just to scientific projects.

Data may include, for example, statistics, collections of digital images, sound recordings, transcripts of interviews, survey data and fieldwork observations with appropriate annotations, an interpretation, an artwork, archives, found objects, published texts or a manuscript. (source: UKRI)

Why manage research data?

From the very start of your research, you will collect, produce and use data in some form. Research Data Management (RDM) is part of the research process. It covers all activities involved in **collecting, describing, storing, processing, analysing, archiving and accessing data**.

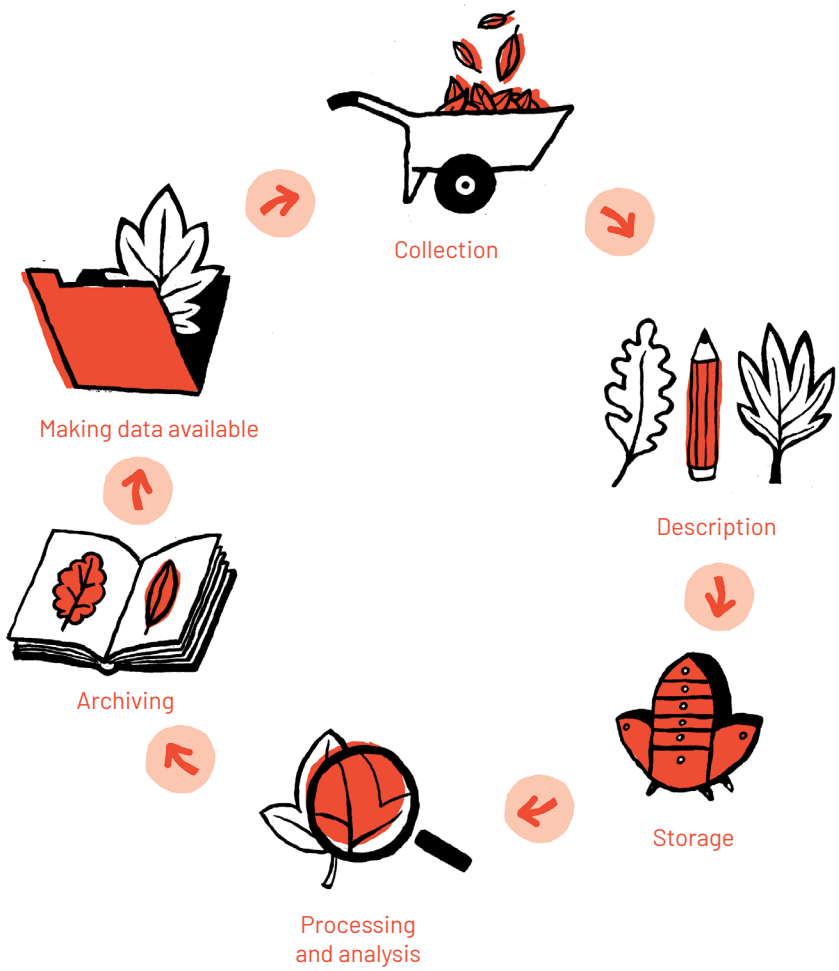
Planning Research Data Management in advance helps you to ensure **data quality**, **minimise risks**, **save time** and **comply** with legal, ethical, institutional and funders' requirements.

How to manage research data?

Data management needs to be considered at the very beginning of a project by creating a Data Management Plan (DMP).

This document helps you think about how to organise your data, files and other supporting documents during and after the project. Planning up front will allow you to think through and plan to deal with as many problems you may come across as possible. Many research funding agencies including those that form part of UK Research & Innovation (UKRI) require you to provide a DMP so it is a good habit to get into.

A DMP is an ongoing document which needs to be updated throughout your research project.



Good data management is useful for you and for others. As well as making your project more organised and easier to manage as you find results and write up, in the longer term, it makes it easy to find your data and make them accessible and reusable by others. At the end of the project, it facilitates the archiving and dissemination of data.

What should a Data Management Plan contain?

Institutions and funders often provide templates to guide the production of your DMP but in general it should include the following:

Collection and documentation: describe the type, format and volume of the data you will be collecting. The format of the data produced is usually related to the software you use and has consequences for the possibilities of sharing and long-term archiving. This initial description enables the creation of documentation (metadata) that is useful for understanding your data and that you will continue enriching during the production phase.

Security and backups: how and where will the data be stored and saved during your research? Where will it be backed up and how often? Who will be responsible for data recovery if there is an incident?

Legal and security issues: what are the protection rules which apply to your data? What methods will you be using to ensure the protection of personal data or other sensitive data? In the UK you should particularly find out about the General Data Protection Regulation (GDPR) rules and the Data Protection Act 2018, but this may vary where you are.

Sharing data and long-term preservation:
This relates to what happens to your data at the end of your project. Think about:

- which people might wish to use your data,
- the criteria for choosing the data to be shared,
- the duration of the data's preservation,
- the data repository you could deposit the data in,
- the way to identify your data (persistent identifier/DOI).

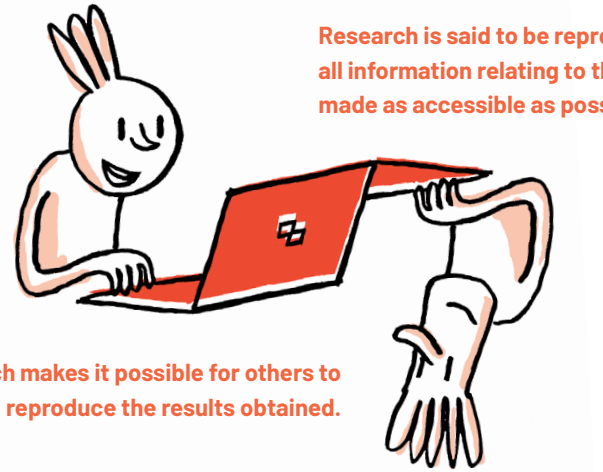
Responsibilities and resources: specify the roles and responsibilities of the people working on the project especially in the case of collaborative projects involving many researchers, institutions and groups with different working methods.

▼**DMP Online** and the ▼**Digital Curation Centre** provide help with creating data and software management plans.

Also see "Making research data open", page 23

Working in a reproducible way: for yourself, for others

What are we talking about?



Research is said to be reproducible if all information relating to the project is made as accessible as possible...

... which makes it possible for others to reproduce the results obtained.

Practices related to Reproducibility vary according to discipline and the methods used. Embracing reproducibility as much as you can in your research improves the accuracy of the methods used and documents all stages of the research process to ensure its transparency and traceability.

For example, in sciences, you need sufficient documentation to enable experiments to be repeated to achieve the same conclusion, or in social sciences to reproduce the statistical processing, or in humanities to reconstruct the stages of analysis of a corpus of images or texts.

WORTH KNOWING

Reproducibility of research is neither an end in itself nor a guarantee of quality as research may be 100% reproducible while of low inherent value. Methods are a core issue in the debate about reproducibility. It is important that researchers come to an agreement on common data analysis methods which go further than explanations provided in the methodological sections of articles.

The advantages of a reproducible approach

Errors are easier to identify and correct. You trace and record how your project documents, data or code etc evolve from the very start of the project and with each modification. It is much harder and less safe if you have to reconstruct these developments at the end of a project during writing.

The results you obtain can be more easily explained and justified to peers. When submitting an article for publication, it will be easier for you to respond to any requests from your reviewers.

Future work is made less uncertain. You give yourself the possibility of reusing data, code, documents, etc. in the future.

How to put this approach into practice

Manage your bibliographical references by using a management tool like ▼**Zotero**. Working according to a reliable bibliographic standard is a common requirement in all disciplines.

Produce a Data management plan and document your processes thoroughly**,** including:

- **Organising data, files and folders:** apply file naming conventions, construct trees with a consistent, scalable structure, separate raw data from analysed data, etc.
- **Version control** even if your actual research does not require coding skills. Being able to go back to a particular version of a document written over a period of several years can be highly valuable.
- **Document your code and data:** what is clear when working may be less clear months later even when you are the author.

Automate certain recurring tasks where possible. You will be able to increase the reliability of your results and make writing articles easier because you can vary parameters more easily.

Think about using collective approaches! Train yourself in collaborative working methods; take part in a project with another team; use public data if it exists.

Opt for open-source solutions for greater transparency and better access.

For further information and guidance, refer to the ▼**Manifesto for Reproducible Science** (Munafo et al), https://doi.org/10.1038/s41562-016-0021

The ▼**UK Reproducibility Network (UKRN)** provide a wide range of support and resources to support your research.

Citizen Science

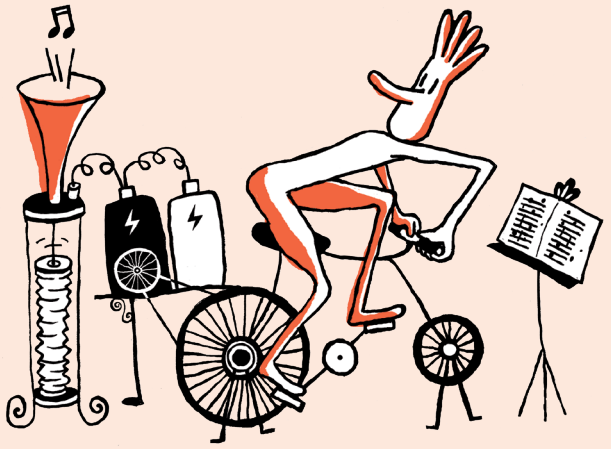
Citizen science is a common name for a wide range of activities and practices. The principles of Citizen Science are intrinsically linked to Open Science as it encourages collaboration and widening participation in research from across society, beyond simply sharing outputs.

Citizen Science as a group of methodologies is found in different academic disciplines – from the natural sciences to the social sciences and the humanities. Within each discipline, the interpretation of citizen science can be slightly different to match the disciplinary practices but they all have similar characteristics, such as including participants in the research process in a meaningful way.

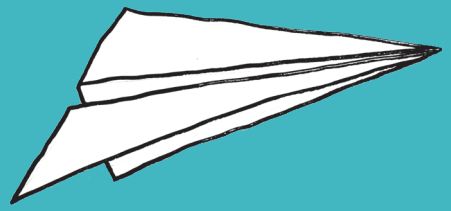
Despite these differences, citizen science is a rapidly growing area of research and practice, with evolving standards on which different stakeholders are developing methodologies, theories and techniques

Harnessing the advantages of the internet, openly available software packages and combining with local knowledge, embracing citizen science can bring about a change in the way research is conducted – one no longer limited to academic researchers, it encourages active collaboration from groups across society, making members of the public fellow researchers.

For further information and guidance, refer to the ▼ **Nature Reviews Primer (Fraisl et al)**, https://doi.org/10.1038/s43586-022-00144-4



2



Disseminating research

Disseminating your publications open access

Open access dissemination involves the **immediate, free and permanent availability of academic outputs on the Internet**.

There are several options to disseminate your work open access - you can publish open access directly, in a journal or a book, or deposit your work straight into an open access repository.

These are not mutually exclusive practices and you can combine them to ensure maximum dissemination of your work while respecting intellectual property. regulations.

Publishing open access

In the traditional academic journal business model, access to articles is reserved for individuals or institutions who have paid subscriptions, in much the same way that books or e-books need to be purchased individually. Publishing open access gives everyone free and immediate access to your article, book, or other work.

If you choose to publish open directly, different funding models exist to cover publication costs, and you should investigate these locally at your institution. There are two main routes which are constantly evolving:

- **Cost to you :** publication costs, often called Article Processing Charges (APCs) or Book Processing Charges (BPCs), are charged. These costs may be paid by your department or home organisation, so its worth asking.
- **No cost to you:** there are no charges for the author to pay. The journal costs are financed in advance by the organisation that publishes or distributes the journal or on the basis of various different funding mechanisms (institutional financing, freemium, subscription, etc.).

HOW MUCH DOES IT COST?

The costs of an article varies from several hundred to several thousand pounds according to the journal. The **▼Directory of Open Access Journals (DOAJ)** is a worldwide directory which indicates whether a journal charges APCs and, if so, the cost.

HYBRID JOURNALS

Many publishers retain a traditional subscription access model while also offering the option of paying to publish articles as open access. This model is known as hybrid open access. If you are not able to pay for Open Access, you can always upload a version of your article to a Repository, but should get advice from your local repository team about which version to use.

Some publishers have adopted a new method of delivering open access in subscription journals, known as Transformative Agreements. These agreements fund open access for authors at an institution, and encourage publishers to transition journals to become fully open access. If your institution or home organisation has a Transformative Agreement with a publisher, the costs of open access may be covered for you.

////////////////// Be aware! //////////////////

PREDATORY PUBLISHERS: BEWARE OF APPEARANCES

The development of digital technology has led to the emergence of publishers with dubious practices. They contact you to promise that your work will be rapidly published, but they do not guarantee editorial quality and an effective peer review process. They often charge a fee for publication.

As well as the financial costs, your academic credibility could also be damaged. It is sometimes difficult to spot a predatory journal but certain tools can help you to do so. There are also predatory conferences organised in a similar way.

▼Think. Check. Submit.: This website gives access to a set of checklists to help you assess the reliability of the journal which you plan to publish your work in.

Depositing in a Repository

Depositing your outputs in a Repository allows you to share your academic work, whether or not it has been published. It is not a substitute for the process of publishing in a journal, but it supports permanent preservation and broad accessibility. This is not the case with academic social networks like ResearchGate and Academia.edu.

Repositories are not limited to articles – you can also deposit a thesis, book chapter, poster, data, report, lecture, conference paper, a dissertation or a report.

Open access repositories can be disciplinary, institutional or national. If you have not been given specific guidelines by a funder, ask your library team for advice on choosing the most appropriate repository.

▼**OpenDOAR** lists deposits on open archives worldwide.

▼**CORE** aggregates open access research papers from repositories

▼**Figshare** and ▼**Zenodo** are freely available repositories open to deposit from individuals without institutional affiliations

WORTH KNOWING

Open access repositories generally provide help for those depositing work in the form of online guides or video tutorials.

You can deposit different versions of an article:

- The preprint or author's version (as submitted for publication): the version sent to a journal by the authors prior to the peer-review process.
- The author accepted manuscript (AAM): the version including revisions resulting from the peer-reviewing process, but without the publisher's final layout.
- The version of record (final published version, publisher's PDF): the article with its final, copyedited layout as published in the journal. The publisher may have exclusive rights to the distribution of this version under the terms of the publishing contract you signed.

What are my rights?

When depositing in a repository, you must make sure that you respect intellectual property and copyright rules.

- As author, you hold the rights to the text of your output until you sign a publishing contract that assigns some of your rights to the publisher.
- Most journal publishing agreements allow you to share your preprint and author accepted manuscript in an open access repository. Publishers often set an embargo period during which the accepted manuscript cannot be openly available.
- An embargo is normally up to 12 months for articles in science, technology and medicine, and up to 24 months for articles in the humanities and social sciences.
- You can use ▼**Sherpa Romeo** to find out about journals' open access policies. For other outputs, you may need to ask the publisher about their policy.
- Instead of signing away your rights to the publisher, you can negotiate. Tell the journal that you want to retain the right to make your author accepted manuscript open access on publication, without embargo. To allow wide re-use, request permission to apply the Creative Commons Attribution licence (CC BY), or another Creative Commons licence when you deposit this version in a repository. Many institutions and research funders, especially those that support ▼**Plan S**, encourage this.

In the field

Ségolène V., a young doctor in archaeology
University of Paris 1

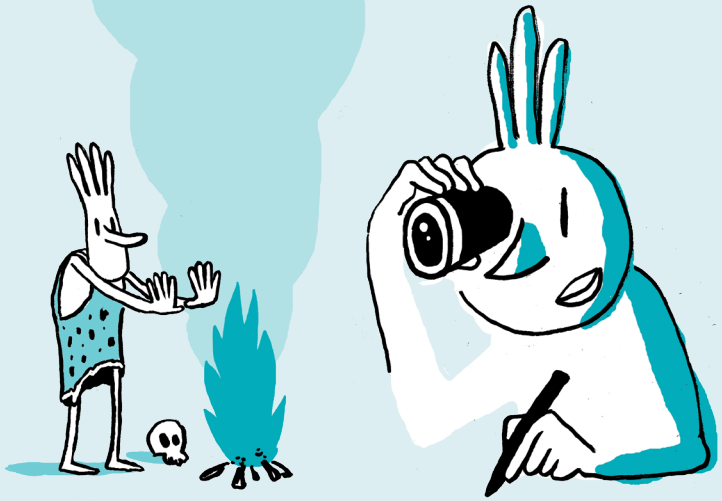
I specialise in geoarchaeology and study the human occupation of caves based on the presence of soot in limestone concretions.

While I was working on my thesis, I developed a pioneering method in my discipline called fuliginochronology which makes it possible to establish very precise records (with an annual resolution) of the occupation of cavities.

I can also archive my work on a permanent basis which also means I can prove the anteriority of my work in this new field if necessary.

Sharing my thesis openly means that I can share research that has not been published, and make this little-known method much more visible. I like being able to access other people's theses, so I make mine available too.

I'm currently working for my laboratory along with other researchers to develop a collaborative database on the ArchéoScopie platform. I classify my photos taken under a microscope. Sharing this data enhances the visibility of the researchers who produced them, enables everyone to reuse the images (rather than having to endlessly reinvent the wheel!), provides examples for teaching purposes and acts as a platform for academic exchanges and discussions.



Making your thesis freely accessible

Sharing your thesis

After your thesis has been submitted and processed, it will likely be made freely available online, usually on the institutional repository, but under certain circumstances it can be restricted.

You can defer the online publication of your thesis by putting it under an embargo during which the file will be restricted. Theses may also be restricted if the contents are deemed confidential.

Sharing of your thesis requires you to respect:

- the intellectual property rights: If your thesis contains works created by others, you must obtain authorisation from the rights holders to reuse and disseminate these, otherwise they may need to be redacted.
- the General Data Protection Regulation (GDPR): any personal data concerning third parties in your thesis must be anonymised or pseudonymised.

Why choose open access for your thesis?

Sharing of a thesis is considered good practice and is very much encouraged.

- Your work will be much more visible and downloaded and cited more by other researchers and students.
- A thesis that is more widely disseminated is much better protected against plagiarism than a work with more restricted dissemination.
- Download statistics can encourage publishers to propose publishing projects.
- Your thesis will have a permanent and validated online consultation address.

CAN I DISSEMINATE MY THESIS IF I HAVE AN OFFER OF PUBLICATION?

Sharing your thesis is not always an obstacle to publication but it is always good to get advice if you are planning to publish as you may wish to wait before opening access to your thesis.

Publishing a monograph or book based on your thesis is a long-term project which often requires significant rewriting and adaptation. The original version of a thesis is rarely published in its native form which means that disseminating may not be a problem. Ask your thesis supervisor or library team for advice.

The stages of depositing and disseminating

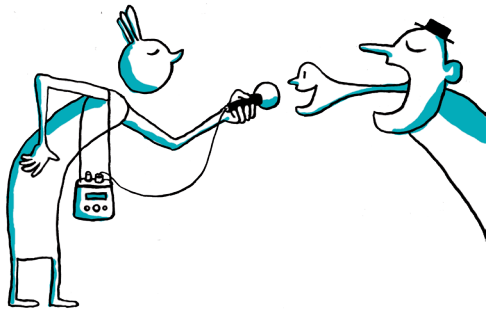
1. You defend your thesis.

2. You make all requested corrections and have your thesis approved.

3. You submit the final version of your thesis to your institution and make decisions about your desire to publish, embargoes and confidentiality.

4. The university library takes care of processing your thesis and depositing it in a repository.

5. You can then link to the online version of your thesis to promote it, share it with potential publishers, or if it is under an embargo, continue to publish papers based on it.



Making research data open

A key principle to keep in mind is that data should be "as open as possible and as closed as necessary". In other words, data should be widely shared but restrictions on access may be justified in certain situations.

Why share my research data?

FOR SHARED AND TRANSPARENT RESEARCH

Making data open allows all researchers to **re-use data produced by others**. For example, one way that openly available data can be exploited is for them to be combined and analysed en masse using data mining technologies (also called Text & Data Mining or TDM).

The cost of creating, collecting and processing data can be very high. Bad data practices are estimated to cost €10 billion per year in Europe (European Commission, 2018).

Research financed with public funding must be open to all where possible. Opening up data increases citizens' trust and enables them to get involved, especially in participatory science.

Putting your data online helps **increase the visibility of your work and enables you to be cited more often**. According to a study published in the journal PLOS ONE, the dissemination of data linked to a publication increases the citations of the article by 25% (Colavizza, Hrynaszkiewicz, Staden et al., 2020).

Making your data available provides **better potential for research transparency** and helps you guard against errors and fraud.

TO COMPLY WITH OBLIGATIONS

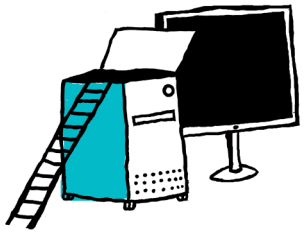
The funders of your research work may also stipulate that your data must be disseminated openly or a publisher may require your data to be deposited in a data repository so that it can be linked to your publication.

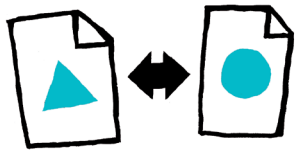
How to disseminate your data

RESPECT THE FAIR PRINCIPLES

The aim of the **Findable** principle is to facilitate the discovery of data by humans and computer systems and requires the description and indexing of data and metadata.

The **Accessible** principle encourages the long-term storage of data and metadata and facilitating their access and/or download by specifying the conditions of access (open or restricted) and use (license).





The **Interoperable** principle can be broken down as follows – data should be downloadable, usable, intelligible and combinable with other data by humans and machines.

The **Reusable** principle highlights the characteristics that make data reusable for future research or other purposes (education, innovation, reproduction/transparency of science).

Source: *Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18*

DATA CURATION

The term data curation covers the organization and integration of data collected from various sources throughout a project. It includes annotation, organisation, publication and presentation of the data such that the value of the data is maintained over time, and the data remains available for reuse when preserved.

TO CHOOSE THE RIGHT DATA REPOSITORY YOU SHOULD CHECK:

- the types and formats of data accepted,
- whether they provide a persistent identifier,
- whether it is possible to deposit several versions of data,
- the assurances offered in terms of archiving and long-term access,
- the conditions of use including licenses.

Depositing in a data repository can be free or may come at a cost depending on the specific business model. In some cases, institutions have their own data repositories or badged versions of commercial ones.

▼ **Zenodo**: generalist repository for European research
▼ **Figshare**: a general repository for research
▼ **Dryad**, datadryad.org for life sciences, agronomy, geosciences

YOU CAN ALSO OPT TO PUBLISH YOUR DATA IN A DATA PAPER

A data paper is an academic article devoted to the description of a set of raw data. These can be published in a data journal which only publishes such articles or in a conventional journal. In both cases, it is subject to peer review and links to the data stored in a repository.

The objective of publishing a data paper is to make the academic community aware of the existence of original data which can be reused by other researchers and in other academic contexts.

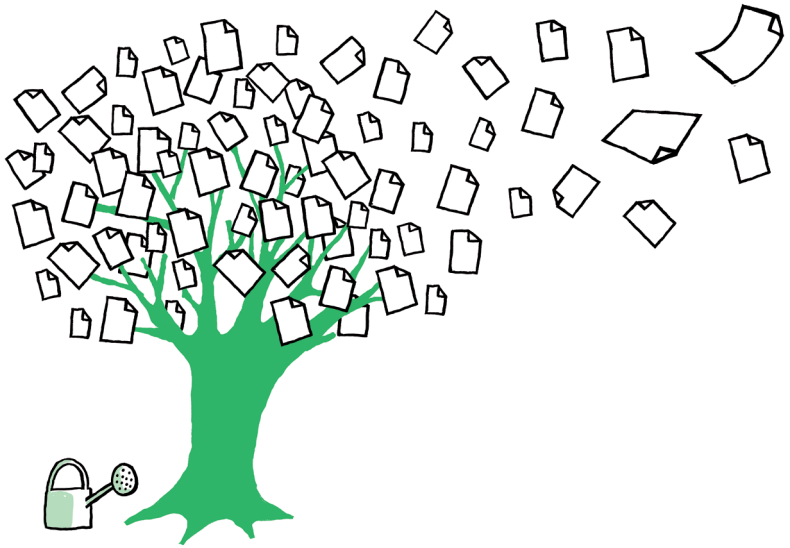
A data paper generally includes the following elements:

- access to the data itself in the form of attached files or a permanent link to a data repository;
- a detailed description (metadata) of the data (production, context, authors, rights attached, etc.).

Preparing for after your thesis

Understanding public policies

Open science officially began in the 2000s thanks to an initiative led by committed researchers and is now a firmly rooted component of public policies worldwide.



The UK open science journey!

2005, RCUK and Wellcome establish Open Access policies
These required funded research be published in an open access format, and provided funds to support this.

2012, UK Government funds Open Access
The UK Government announced funding of £10 million 'to kick start the process of developing policies and setting up funds to meet the costs of article processing charges (APCs)' in response to the Finch Report.

2016, Open Access Required for the REF
'Policy for open access in the post-2014 Research Excellence Framework', announced in 2014, came into force - requiring researchers to deposit publications into their institutional repository within three months of acceptance in order to be eligible for the REF.

And beyond?

▼Horizon Europe
This funding programme started on January 1st 2021 and consolidates the open science policy introduced under the Horizon 2020 programme. It includes the obligation to disseminate open access publications and strongly encourages the dissemination of data according to FAIR principles and an associated data management plan.

▼Plan S
'Plan S' is a product of cOAlition S which brings together research funding agencies committed to the development of open science. Its guiding principle is the free and immediate dissemination of publications funded by these agencies in journals, on open access platforms or open archives. It came into effect as of January 2021.

▼UNESCO Recommendation on Open Science
Adopted in November 2021, this international framework for open science policy and practice recognizes disciplinary and regional differences in open science perspectives, takes into account academic freedom, gender-transformative approaches and the specific challenges of scientists and other open science actors in different countries and in particular in developing countries, and contributes to reducing the digital, technological and knowledge divides existing between and within countries.

PERSISTENT IDENTIFIERS

A persistent identifier is a long-lasting reference to a digital entity: a person, place, or thing, like a journal article, book or dataset. Unlike URLs, which may break, a persistent identifier reliably points to a digital entity.

An **▼ORCID iD** is an example of a persistent identifier for a person. Having an ORCID iD is a great way to reliably and easily connect you with your research and affiliations. It is used by numerous journals and funders and ensures all research outputs and activities are correctly attributed to you and not someone else with the same name.

A **▼Digital Object Identifier** is a persistent identifier for digital entities such as journal articles, books, images and datasets. **▼Crossref** and **▼DataCite** are the main organizations assigning DOIs in scholarly communication.

Evaluating research

Open science represents a profound change in science and research which means some evaluation practices are changing.

Changing peer review

Peer review is often a prerequisite for publication, especially in a journal. This process is designed to assess the validity, quality and originality of research in a paper. The process is usually organised by the journal or the publisher and the manuscript is shared with other researchers in the same field, ideally who the author doesn't know. However, this system isn't perfect. For example, some reviewers may be competitors of the author or reviewers sometimes work in the same field but on research themes which are too far removed from the article needing review. Generally speaking, peer review is not an infallible solution – since the 2000s there has been an increase in the publication of fraudulent or questionable articles because of data manipulation or plagiarism (Grieneisen & Zhang, 2012; Fang, Steen & Casadevall, 2013).

Open science has brought about **the emergence of Open Peer Review practices** with two main modes. Firstly, making the names of reviewers public or the review is carried out on a platform that enables all users to comment on the article, subject matter expert or not. This practice has been made easier by the existence of pre-publication platforms like arXiv and bioRxiv which journals can use to collect comments.

Example
The ▼**Peer Community in** platform organises the peer review of pre-publications deposited in an open access repository which can lead to the issue of a certificate of validation. Journals can thus publish articles freely without having to ask for reviewers.

In certain disciplines such as biology and health, the pre-registration of hypotheses and protocols (registered reports) in registers or journals has led to practices changing. Peer review is carried out in two stages which reduces the effect of publication bias (a tendency to publish only positive results) and spotlights the research process, rather than just the outcome.

The evaluation of researchers with regard to open science

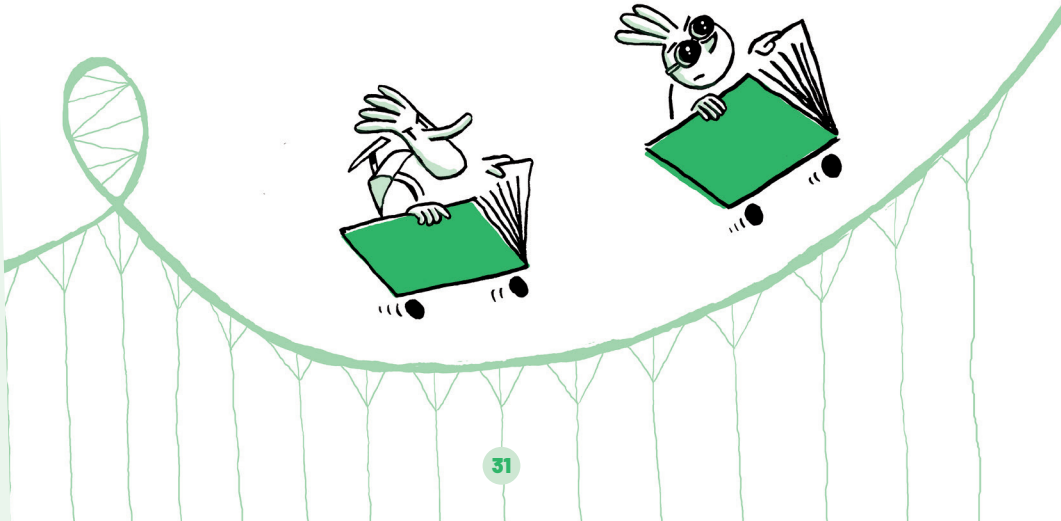
There is a growing movement to update and improve the way research is assessed. Traditionally this has been heavily focused on measures like journal impact factors, which do not reflect the quality of the research itself. The movement for responsible metrics tries to set out ways to assess research so that:

- research is assessed on its own merit
- and all forms of research are assessed fairly.

Example
The Wellcome Foundation requires all the organisations it funds to make a public commitment to using metrics responsibly - by assessing research outputs on their own merits, and discouraging the use of inappropriate metrics like the impact factor.

These new approaches can support open science by recognising the value of the following new practices:

- the open access and preprint dissemination of results - including negative results;
- the publication of work which reproduces earlier results;
- the publication of a research notebook describing the stages in the construction of its academic approach;
- making data available to support papers, or as freestanding datasets;
- submitting research protocols to peer review prior to experiments, etc.



The San Francisco Declaration on Research Assessment (DORA)

The San Francisco Declaration on Research Assessment (DORA), which particularly emphasised the problems with overuse of the impact factor, brought the issue of responsible metrics to wide attention in 2013. It was followed by the Leiden Manifesto for Research Metrics, published by a group of bibliometric researchers in Nature in 2015, which gave some broad recommendations for best practice. In particular, it highlighted issues with the h-index as well as with impact factors.

JOURNAL IMPACT FACTOR

The impact factor is one of the most widely-used – and most controversial – metrics. It estimates the impact of a journal on the basis of the average citations received by papers published in the last two years. While simple to calculate, it only shows a retrospective average for the whole journal, and there is a growing consensus that it should never be used to judge the quality or impact of a specific paper.

The same year, a review of the REF (Research Excellence Framework) in the UK published the Metric Tide report. This was the first to propose the term "responsible metrics", recommending the research community develop a more "sophisticated and nuanced" approach to metrics and their limitations. It argued that metrics could have a place in large-scale assessments like the REF, but that they should still be used with caution and could not fully replace peer review.

Since then, there has been a growing push for UK institutions and funders to adopt responsible metrics policies, and a large number of universities have now done so. Most prominently, the Wellcome Foundation has made it a requirement that institutions in receipt of funding have both a commitment to using metrics responsibly, and a plan to put that into place. An example set of principles is shown on the next page.

The principles overleaf were adopted by the ▼**UK Reproducibility Network** in 2020 as a model statement on Responsible Research Evaluation for use by a variety of institutions. They are available under a CC-BY license, and were based on those originally produced by UCL for their own statement on the responsible use of metrics

PRINCIPLES FOR RESPONSIBLE RESEARCH EVALUATION.

1. Quality, influence, and impact of research are typically abstract concepts that prohibit direct measurement. There is no simple way to measure research quality, and quantitative approaches can only be interpreted as indirect proxies for quality.

2. Different fields have different perspectives of what characterises research quality, and different approaches for determining what constitutes a significant research output (for example, the relative importance of book chapters vs journal articles). All research outputs must be considered on their own merits, in an appropriate context that reflects the needs and diversity of research fields and outcomes.

3. Both quantitative and qualitative forms of research assessment have their benefits and limitations. Depending on the context, the value of different approaches must be considered and balanced. This is particularly important when dealing with a range of disciplines with different publication practices and citation norms. In fields where quantitative metrics are not appropriate nor meaningful, we should not impose their use for assessment in that area.

4. When making qualitative assessments, we should avoid making judgements based on external factors such as the reputation of authors, or of the journal or publisher of the work; the work itself is more important and must be considered on its own merits.

5. Not all indicators are useful, informative, or will suit all needs; moreover, metrics that are meaningful in some contexts can be misleading or meaningless in others. For example, in some fields or subfields, citation counts may help estimate elements of usage, but in others they are not useful at all.

6. Avoid applying metrics to individual researchers, particularly those that do not account for individual variation or circumstances. For example, the h-index should not be used to directly compare individuals, because the number of papers and citations differs dramatically among fields and at different points in a career.

7. Ensure that metrics are applied at the correct scale of the subject of investigation, and do not apply aggregate level metrics to individual subjects, or vice versa (e.g., do not assess the quality of an individual paper based on the JIF of the journal in which it was published).

8. Quantitative indicators should be selected from those that are widely used and easily understood to ensure that the process is transparent and they are being applied appropriately. Likewise, any quantitative goals or benchmarks must be open to scrutiny.

9. If goals or benchmarks are expressed quantitatively, care should be taken to avoid the metric itself becoming the target of research activity at the expense of research quality itself.

10. New and alternative metrics are continuously being developed to inform the reception, usage, and value of all types of research output. Any new or non-standard metric or indicator must be used and interpreted in keeping with the other principles listed here for more traditional metrics. Additionally, consider the sources and methods behind such metrics and whether they are vulnerable to being gamed, manipulated, or fabricated.

11. Metrics (in particular bibliometrics) are available from a variety of services, with differing levels of coverage, quality and accuracy, and these aspects should be considered when selecting a source for data or metrics. Where necessary, such as in the evaluation of individual researchers, choose a source that allows records to be verified and curated to ensure records are comprehensive and accurate, or compare publication lists against data from internal systems.

Act now

When you can, submit your publications to open access journals.

Deposit your publications in an open access repository:

- Keep the latest version approved by peers but not yet formatted by the publisher (Author Accepted Manuscript).
- Ask your co-authors for approval.
- Deposit the latest version approved by the peer reviewers in a repository

Document and share research data and/or the source code you developed:

- Store data using a system or format in compliance with your team or institution's policy.
- Document the data with metadata so that they are reusable.
- Deposit the data and code associated with your publications in an online repository.

Follow the evolutions of open science and get involved!

ANY QUESTIONS ?

This guide to open science has provided you with an overview of all the relevant issues and avenues to be explored for your thesis work. If you wish to take this further, please do not hesitate to contact your university's departments who can answer your questions, provide you with individual support and offer you training in these subjects. Your contacts:

- your doctoral school and/or your home research group;
- your institution's university library (for training, depositing your thesis, management of the repository, access to documentation, etc.);

Further resources

GENERAL RESOURCES

UNESCO Recommendation on Open Science
Approved by the 41st session of the General Conference of the United Nations Educational, Scientific and Cultural Organization (UNESCO), from 9 to 24 November 2021.
https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en

UKRI Open Access policy
This document presents the development and context of the UK Research and Innovation Open Access policy.
https://www.ukri.org/manage-your-award/publishing-your-research-findings/

FOSTER Open Science
A portal providing online training on open science (in English) created by FOSTER, a consortium of academic and research institutions in 6 European countries, funded by the European Union. https://www.fosteropenscience.eu

Open Science Framework (OSF)
The Open Science Framework (OSF), developed by the Centre for Open Science in Zurich is a free, open platform for sharing research data and other outputs. It can also be used to discover projects, data, materials, and collaborators on OSF that might be helpful to your own research. https://osf.io/

MANTRA (Research data: digital learning)
MANTRA is a free online course for those who manage digital data as part of their research project. https://mantra.ed.ac.uk/

SPECIFIC RESOURCES

Creative Commons
Creative Commons is a nonprofit organization that helps overcome legal obstacles to the sharing of knowledge and creativity to address the world's pressing challenges.
https://creativecommons.org/

Reproducibility
The UK Reproducibility Network supports improving the reproducibility and reliability of research in the UK by providing a network of experts, training resources and best practice guidance. https://www.ukrn.org/

Open Researcher & Contributor ID (ORCID)
ORCID provides a persistent digital identifier (an ORCID iD) that you own and control, and that distinguishes you from every other researcher. https://orcid.org/

Finding Open Access resources
DOAJ and DOAB are discovery services that collate Open Access content that has met their standards. https://doaj.org/ ; https://www.doabooks.org/

Glossary

APC (article processing charges)
Charges for publishing open access which are billed to the author (or his or her institution).

Article: preprint or author's version
The version of an article sent to a journal by the authors prior to the peer review process.

Article: authour accepted manuscript
The version of an article including revisions resulting from the peer-reviewing process but without the publisher's final layout.

Article: version of record
The version of an article with the editor's final layout as published in the journal.

BPC (book processing charges)
Charges for publishing open access which are billed to the author (or his or her institution).

Data Management Plan (DMP)
An ongoing plan written at the start of a research project which sets out how the data will be managed covering collection, documentation, storage and sharing as well as managing sensitive data, etc.

Data/code repository
These may be multidisciplinary or specialised in one disciplinary field and data are deposited in them. A repository provides better archiving and wider access to data than a laboratory server or other local solutions.

Data Protection Act (2018)
Research that collects personal data in the UK is subject to terms of both GDPR and the Data Protection Act.

Distribution license
The license chosen by an author at the time of publication. These set conditions for distribution and reuse (example: Creative Commons).

Embargo
Period during which an academic output cannot be shared openly. Authors of theses may define an embargo period during which the thesis is only available within the academic community or completely closed for a short time.

FAIR (principles)
The aim of FAIR is to make data findable, accessible, interoperable and reusable.

General Data Protection Regulation (GDPR).
Sets out the framework for the management of personal data across Europe.

Intellectual Property (IP).
Ideas, information and knowledge as the results and outcomes of research. Intellectual property rights are the rights given to persons over these outcomes.

Metadata
Data which enables the description of other data. Good metadata is essential for the sharing and reusability of data, publications, and other research outputs that are shared.

Open access
Open Access content should be accessible to all with no barriers (such as authentication requirements, resources under an embargo, paid access, etc.). This may concern academic publications, data, code, etc.
There are several ways of disseminating open access publications. For example, self-archiving academic production in an open archive (sometimes called green open access) or opting to publish in an open access journal with or without APCs (sometimes called gold open access).

Open access platform
A platform that specialises in collating or publishing open access material. The Directory of Open Access Journals (DOAJ) and the Directory of Open Access Books (DOAB) are both examples of these platforms.

Open access publication
A journal or book that was been directly published in open access. These books and journals have varied business models.

Open access repository
Can be institutional or subject based and outputs are deposited in the repository by researchers for free. They also preserve and facilitate sharing of outputs.

Persistent identifier

A unique stable reference for a digital object such as a dataset or article, for example a digital object identifier (DOI) for an article or an ORCID for a researcher.

Personal data
Data concerning a living person who is identified or identifiable.

Predatory publishers or journals
Publishers or journals with dubious business or peer reviewing practices.

Repository
Researchers deposit their work directly in thematic or institutional repositories so that it can be consulted by all with no barriers.

Reproducibility
The capacity of another researcher to obtain the same results using the same methods or data. This highlights the importance of the methods used to produce such results.

Research data
All the information collected, observed, generated or created to validate research results.

Original Sources

ADDA, Gilles, ASKENAZI, Philippe, GANASCIA, Jean-Gabriel, *et al. Les publications à l'heure de la science ouverte*. https://www.ouvrirlascience.fr/wp-content/uploads/2020/02/COMETS_Les-publications-a-lheure-de-la-science-ouverte_Avis-2019-40-1.pdf

BROWN, Titus. *Software and Workflow Development Practices*. http://ivory.idyll.org/blog/2020-software-and-workflow-dev-practices.html

CHRISTENSEN, Garret, FREESE, Jeremy, MIGUEL, Edward. *Transparent and Reproducible Social Science Research: How to Do Open Science*. https://www.ucpress.edu/book/9780520296954/transparent-and-reproducible-social-science-research

CNRS. *Feuille de route du CNRS pour la science ouverte*. https://www.cnrs.fr/sites/default/files/press_info/2019-11/Plaquette_ScienceOuverte.pdf

COLAVIZZA, Giovanni, HRYNASZKIEWICZ, Iain, STADEN, Isla, *et al. The citation advantage of linking publications to research data*. https://doi.org/10.1371/journal.pone.0230416

Comité pour la science ouverte. *Types de documents, productions et activités valorisées par la science ouverte et éligibles à une évaluation*. https://www.ouvrirlascience.fr/types-de-documents-productions-et-activites-valorisees-par-la-science-ouverte-et-eligibles-a-une-evaluation

DESQUILBET, Loïc, GRANGER, Sabrina, HEJBLUM, Boris, *et al. Vers Une Recherche Reproductible*. https://hal.archives-ouvertes.fr/hal-02144142

DoRANum. *Données de la recherche : Apprentissage numérique*. https://doranum.fr

Euraxess Spain. *The European Charter & Code for Researchers*. https://www.euraxess.es/spain/services/charter-code-researchers#Charter

European Commission. 2019. *Horizon Europe - Investing to shape our future*. https://ec.europa.eu/info/files/horizon-europe-investing-shape-our-future_en

European Commission. 2018. *Cost of not having FAIR research data*. https://www.ouvrirlascience.fr/wp-content/uploads/2019/03/Cost-Benefit-analysis-for-FAIR-research-data_KI0219023ENN_en.pdf

European Science Foundation. *Addendum to the cOAlition S Guidance on the Implementation of Plan S : Principles and Implementation*. https://www.coalition-s.org/addendum-to-the-coalition-s-guidance-on-the-implementation-of-plan-s/principles-and-implementation/

FANG, Ferric. C., STEEN, R. Grant, CASADEVALL, Arturo. *Correction for Fang et al. Misconduct accounts for the majority of retracted scientific publications*. https://doi.org/10.1073/pnas.1220649110

GRIENEISEN, Michael. L., ZHANG, Minghua. *A Comprehensive Survey of Retracted Articles from the Scholarly Literature*. https://doi.org/10.1371/journal.pone.0044118.t002

INRAe. *Produire des données FAIR*. https://www6.inrae.fr/datapartage/Produire-des-donnees-FAIR/

Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique, article 30. https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000033202746&categorieLien=id#JORFARTI000033202841

LOWNDES, Julia S. Stewart, BEST, Benjamin D., SCARBOROUGH, Courtney, *et al. Our Path to Better Science in Less Time Using Open Data Science Tools*. https://doi.org/10.1038/s41559-017-0160

MARKOWETZ, Florian. *Five Selfish Reasons to Work Reproducibly*. https://doi.org/10.1186/s13059-015-0850-7

MAUREL, Lionel. *L'ouverture des données de recherche : un retour aux sources de l'Ethos de la Science*. https://scinfolex.com/2019/06/05/louverture-des-donnees-de-recherche-un-retour-aux-sources-de-lethos-de-la-science/

PINEAU, Joelle. *How the AI community can get serious about reproducibility*. https://ai.facebook.com/blog/how-the-ai-community-can-get-serious-about-reproducibility/

POLDRACK, Russell A. *The Costs of Reproducibility*. https://doi.org/10.1016/j.neuron.2018.11.030

WHITAKER, Kirstie, AINSWORTH, Rachael, ARNOLD, Beckie, *et al. The Turing Way*. https://the-turing-way.netlify.com//introduction/introduction.html

Additional Sources - UK

Fraisl, D., Hager, G., Bedessem, B. *et al.* Citizen science in environmental and ecological sciences. Nat Rev Methods Primers 2, 64 (2022). https://doi.org/10.1038/s43586-022-00144-4

Munafò, M., Nosek, B., Bishop, D. et al. A manifesto for reproducible science. Nat Hum Behav 1, 0021 (2017). https://doi.org/10.1038/s41562-016-0021

Department for Business, Innovation & Skills and The Rt Hon David Willetts. 2012. Government to open up publicly funded research. https://www.gov.uk/government/news/government-to-open-up-publicly-funded-research

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Credits - UK Version

Editor
Kirsty Wallis

Contributors
Dominic Allington-Smith, Paul Ayris, David Bogle, Andrew Gray, Muki Haklay, James Houghton, Christiana McMahon, Catherine Sharp

Graphic Design Support
Janine Clayton

Thanks

Feedback on the first UK draft
UCL Open Science Committee
UCL Office for Open Science & Scholarship Steering Group

The digital version of this guide is available on UCL Research Data Repository:
https://doi.org/10.5522/04/20585898

Original Credits

Direction of publication
Ministry of Higher Education, Research and Innovation

Editorial coordination
University of Lille

Scientific council
The Skills and Training College of The Committee for Open Science

Project leader
Jennifer Morival

Writers
Johann Berti, Marin Dacos, Gabriel Gallezot, Madeleine Géroudet, Sabrina Granger, Joanna Janik, Claire Josserand, Jean-François Lutz, Christine Okret-Manville, Sébastien Perrin, Noël Thiboud

The working group used Foster Open Science's taxonomy of open science to create this guide: https://www.fosteropenscience.eu/resources

Translator
Richard Dickinson, Inist-CNRS
Proofread by Katherine Kean, Inist-CNRS

Graphic design
Studio 4 minutes 34
Studio Lendroit.com

Printing
L'Artésienne, Liévin

Thanks

The PhD students from Lille's doctoral schools who took part in discussions on the first version of the guide
Kaouther Azouz, Paul Cambourian, Olivier Capra, Camille Cornut, Quentin Devignes, Clémentine Garric, Frédéric Géhanno, Marie Genge, Mathilde Greuet, Juliette Le Marquer, Margaux Lyonnet, Caroline Maupas, Meriam Meziani Mekki, Nadia Moujtahid, Jonathan Olivier, Alizée Petitmangin, Elena Tagliani, Hélène Villain, Marie Winter

The PhD students who shared their experiences of open science
Sacha Hodencq, Minh-Châu Nguyên, Ségolène Vandevelde

Experts consulted
Laurence Crohem, Romain Féret, Jacques Lafait, Yvette Lafosse, Maïté Roux.

The digital version of this guide is available on the www.ouvrirlascience.fr site

This guide to Open Science is designed to accompany you at every step of your research, from developing your academic approach to the dissemination of your results. It provides a set of tools and best practices that can be directly implemented and is aimed at researchers from all disciplines.

We hope this guide will motivate you and provide the means for you to realise the ambitions of open science by sharing your research results and data with as many people as possible.

UK version created by

UCL OPEN SCIENCE

Original supported by

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR, DE LA RECHERCHE ET DE L'INNOVATION
Liberté
Égalité
Fraternité

OUVRIR LA SCIENCE !

Université de Lille