Table of Contents

The Open Science Training Handbook

A group of fourteen authors came together in February 2018 at the TIB (German National Library of Science and Technology) in Hannover to create an open, living handbook on Open Science training. High-quality trainings are fundamental when aiming at a cultural change towards the implementation of Open Science principles. Teaching resources provide great support for Open Science instructors and trainers. The Open Science training handbook will be a key resource and a first step towards developing Open Access and Open Science curricula and andragogies. Supporting and connecting an emerging Open Science community that wishes to pass on their knowledge as multipliers, the handbook will enrich training activities and unlock the community's full potential.

Sharing their experience and skills of imparting Open Science principles, the authors (see below) produced an open knowledge and educational resource oriented to practical teaching. The focus of the new handbook is not spreading the ideas of Open Science, but showing **how** to spread these ideas most effectively. The form of a book sprint as a collaborative writing process maximized creativity and innovation, and ensured the production of a valuable resource in just a few days.

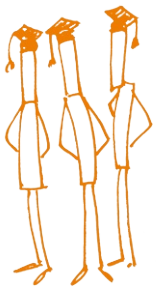
Bringing together methods, techniques, and practices, the handbook aims at supporting educators of Open Science. The result is intended as a helpful guide on how to forward knowledge on Open Science principles to our networks, institutions, colleagues, and students. It will instruct and inspire trainers how to create high quality and engaging trainings. Addressing challenges and giving solutions, it will strengthen the community of Open Science trainers who are educating, informing, and inspiring themselves.

Help us making the handbook better

We welcome comments and feedback from everyone, irrespective of their expertise or background. The easiest way to do this is to use hypothes.is. Also, you can create pull requests, either from within the Gitbook website or app, or with any tool you like. The handbook's content is maintained as [this GitHub repository] (https://github.com/Open-Science-Training-Handbook).

Let's run an Open Science training together

Are you interested in running or attending trainings or webinars that make use of the Open Science Training Handbook? Get in touch with us at elearning@fosteropenscience.eu - we'd love to hear from you.



How to refer to the handbook

Please consider citing the handbook when using the content. To cite the book, we recommend that you either refer to

- https://book.fosteropenscience.eu/, which is the most friendly way to read the book (also available as PDF and ePub), to comment and to suggest changes, *or*
- https://doi.org/10.5281/zenodo.1212496, which is a citable DOI refering to a (hardly comprehensible) archived dump of the book.
- If you are looking for other languages or formats you can go to the FOSTER portal, where we linked everything on this page.

The Authors and the Book Sprint facilitators

Learn more about the authors and the book sprint facilitators, their experiences and inspiration, as well as their affiliation, contact information, Twitter and ORCID profiles, in the Handbook's last chapter.

Thank you to

- Gwen Franck (EIFL, Belgium) for covering social media during the book sprint & keeping us motivated with energizers
- Patrick Hochstenbach (University of Gent, Belgium) for drawing the awesome cartoons and images
- Vasso Kalaitzi (LIBER, Netherlands) for recording the really nice videos
- Matteo Cancellieri (Open University, UK) for supporting us with all technical issues and creating the gitbook
- Simon Worthington (TIB, Hannover, Germany) for providing advice with maintaining and converting bibliographic metadata

Copyright statement

Funding

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 741839.

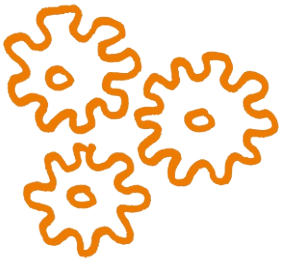
Purpose of the book

"When all researchers are aware of Open Science, and are trained, supported and guided at all career stages to practice Open Science, the potential is there to fundamentally change the way research is performed and disseminated, fostering a scientific ecosystem in which research gains increased visibility, is shared more efficiently, and is performed with enhanced research integrity." Open Science Skills Working Group Report (2017)

Open Science, the movement to make scientific products and processes accessible to and reusable by all, is about culture and knowledge as much as it is about technologies and services. Convincing researchers of the benefits of changing their practices, and equipping them with the skills and knowledge needed to do so, is hence an important task.

This book offers guidance and resources for Open Science instructors and trainers, as well as anyone interested in improving levels of transparency and participation in research practices. Supporting and connecting an emerging Open Science community that wishes to pass on its knowledge, the handbook suggests training activities that can be adapted to various settings and target audiences. The book equips trainers with methods, instructions, exemplary training outlines and inspiration for their own Open Science trainings. It provides Open Science advocates across the globe with practical know-how to deliver Open Science principles to researchers and support staff. What works, what doesn't? How can you make the most of limited resources? Here you will find a wealth of resources to help you build your own training events.

Building on the authors' cumulative experience and skills of imparting Open Science principles, this handbook is oriented towards practical teaching in an open knowledge and educational setting. In other words, the focus of this handbook does not lie on spreading the idea of Open Science, but on *how to* support Open Science practices most effectively.



Who is this book for?

This handbook is intended for anyone who wishes to host Open Science training events or introduce Open Science concepts to discipline-specific training events, in order to foster the uptake of open research practices. This includes researchers, librarians, infrastructure providers, research support officers, funders, policy makers and decision makers. This handbook is also meant for all those who have regular or occasional contact with researchers (and other stakeholders) and wish to share their Open Science knowledge,

either as part of their regular working duties or as an extra investment of time. Importantly, it will be of use to those who wish to host training events to foster reuse, participation, efficiency, equity, and sharing in research, regardless of whether they ascribe to (or even wish to use) the term Open Science.

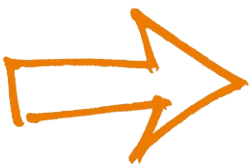
In this handbook, we define "trainer" as any person wishing to run an Open Science training event, regardless of their levels of experience. Importantly, this includes those who would feel uncomfortable or do not wish to use the Open Science label in their teaching. The book contains advice on teaching concrete skills and concepts to improve the work of researchers. And while most fall under the umbrella term "Open Science", they needn't be taught as such. Wariness of the label "Open Science" might mean that "Open Science" training only attracts a particular segment of researchers, whereas "How to publish your data" training attracts a more diverse group. Part of a trainer's job is to define their target audience and how best to reach them, and so such decisions are best made by you!

What is Open Science?

According to the FOSTER taxonomy, "Open science is the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society." It can be defined as a grouping of principles and practices:

- **Principles**: Open Science is about increased transparency, re-use, participation, cooperation, accountability and reproducibility for research. It aims to improve the quality and reliability of research through principles like inclusion, fairness, equity, and sharing. Open Science can be viewed as research simply done properly, and it extends across the Life and Physical Sciences, Engineering, Mathematics, Social Sciences, and Humanities (Open Science MOOC).
- **Practices**: Open Science includes changes to the way science is done - including opening access to research publications, data-sharing, open notebooks, transparency in research evaluation, ensuring the reproducibility of research (where possible), transparency in research methods, open source code, software and infrastructure, citizen science and open educational resources.

A note on language: As the English word "science" traditionally does not include the humanities and social sciences, more explicitly inclusive terms like "open scholarship" or "open research" are often used. As "Open Science" is the more common term, we shall use it here, but it should be read as referring to research from all scholarly disciplines.



How to use the book

This handbook is designed in a modular way. Feel free to choose chapters and skip others that might not be relevant to you or your training.

In **Chapter 2** "Open Science Basics" you will dive into the content of your training. All topics pertaining to Open Science are presented and explained in this part of the handbook. Already familiar with one or two topics? Great, then have a look at other aspects you might not have heard of yet. Even if you are not planning to run training events on those exact topics, you will likely find them of use - there is a lot of overlap between Open Science topics.

If you have no or little prior knowledge about training in general, please have a look into **Chapter 3** "On Learning and Training". It gives you an overview of training techniques as well as practical tips for designing your training. If you already have some experience you can also use it to learn about different teaching approaches and for refreshing your knowledge.

Bigger workshops and information events can require a lot of planning. Making your event a success will involve a lot of decisions, from the small to the large, which are time-sensitive. **Chapter 4** "Organizational Aspects" provides helpful information about organizational aspects. It also offers a useful checklist to aid in planning your training.

Lively and interactive training events need engaging activities. Our example exercises and additional resources will engage your audience, give practical insight about theoretical topics, or provide you with feedback from your participants. **Chapter 5** "Examples and Practical Guidance" offers you a range of tested and approved exercises and resources by Open Science training experts. Feel free to test, reuse, and adapt them!

Like any other emerging field, Open Science uses quite a lot of sometimes difficult terminology. Some of it you may not be familiar with. Don’t lose heart! The **"Glossary"** will explain most of the less familiar terms and concepts.

This handbook was created to be a living resource. This means it will regularly be updated due to new developments in Open Science, as well as in response to feedback and suggestions from other Open Science trainers and our general audience. Please feel free to add your best practices, examples, resources, opinions or experiences via GitHub.

We hope you will enjoy reading this handbook and wish you all the best for your future Open Science training!

Open License and Credits

We have tried to acknowledge all of our sources. If for some reason we have forgotten to provide you with proper credits it has not been done with malicious intent. Feel free to contact us at elearning@fosteropenscience.eu for any corrections.

Open Science Basics

This chapter aims to provide concrete context as well as the key points for the most relevant aspects of Open Science. Starting from the core concepts and principles of Open Science, the chapter continues to address components such as Open Research Data, Open Access, Open Peer Review and Open Science Policies, together with more practical aspects such as Reproducible Research, Open Source Software and Open Licensing and File Formats.

Each section is structured so that it includes a short description of the topic, an explanation of the relevance to Open Science, the key learning objectives that should be highlighted within the context of a training session, the major components (knowledge and skills) that should be involved, some frequent questions/obstacles/misconceptions that are encountered for that topic, and finally the expected outcomes of a training session and some further reading.

Chapters

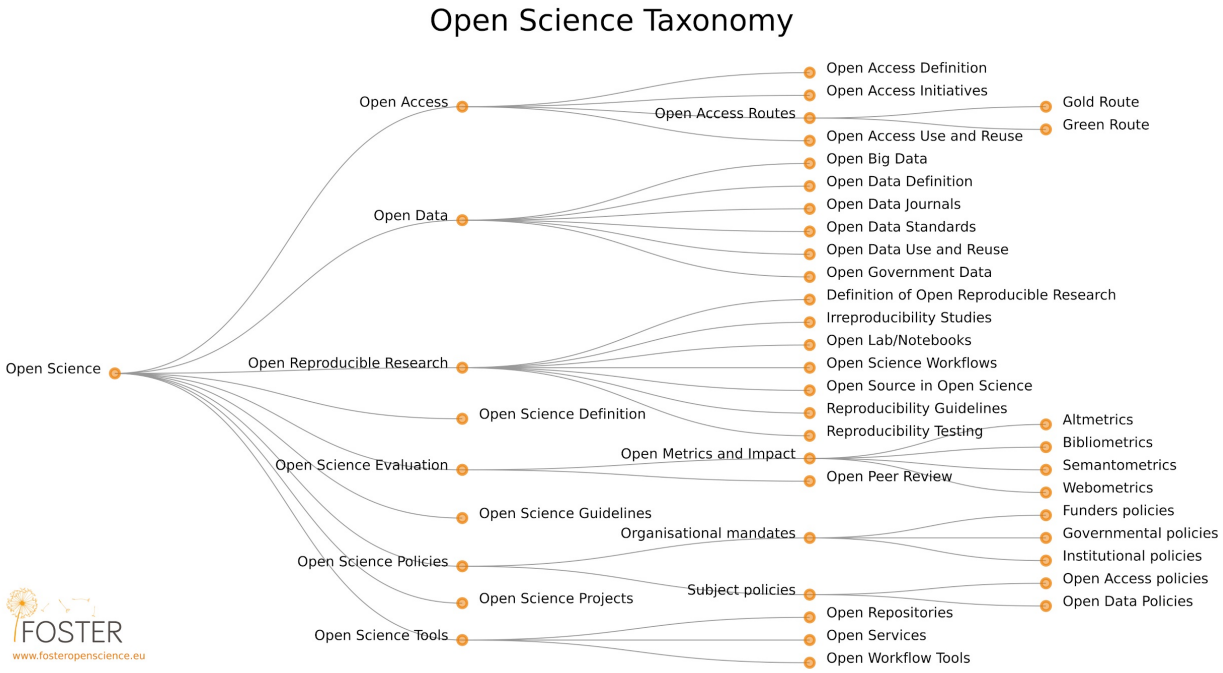
1. Open Concepts and Principles

What is it?

Open Science is the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods (FOSTER Open Science Definition). In a nutshell, Open Science is transparent and accessible knowledge that is shared and developed through collaborative networks (Vicente-Sáez & Martínez-Fuentes 2018).

Open Science is about increased rigour, accountability, and reproducibility for research. It is based on the principles of inclusion, fairness, equity, and sharing, and ultimately seeks to change the way research is done, who is involved and how it is valued. It aims to make research more open to participation, review/refutation, improvement and (re)use for the world to benefit.

There are several definitions of "openness" with regards to various aspects of science; the Open Definition defines it thus: “Open data and content can be freely used, modified, and shared by anyone for any purpose”. Open Science encompasses a variety of practices, usually including areas like open access to publications, open research data, open source software/tools, open workflows, citizen science, open educational resources, and alternative methods for research evaluation including open peer review (Pontika et al., 2015).



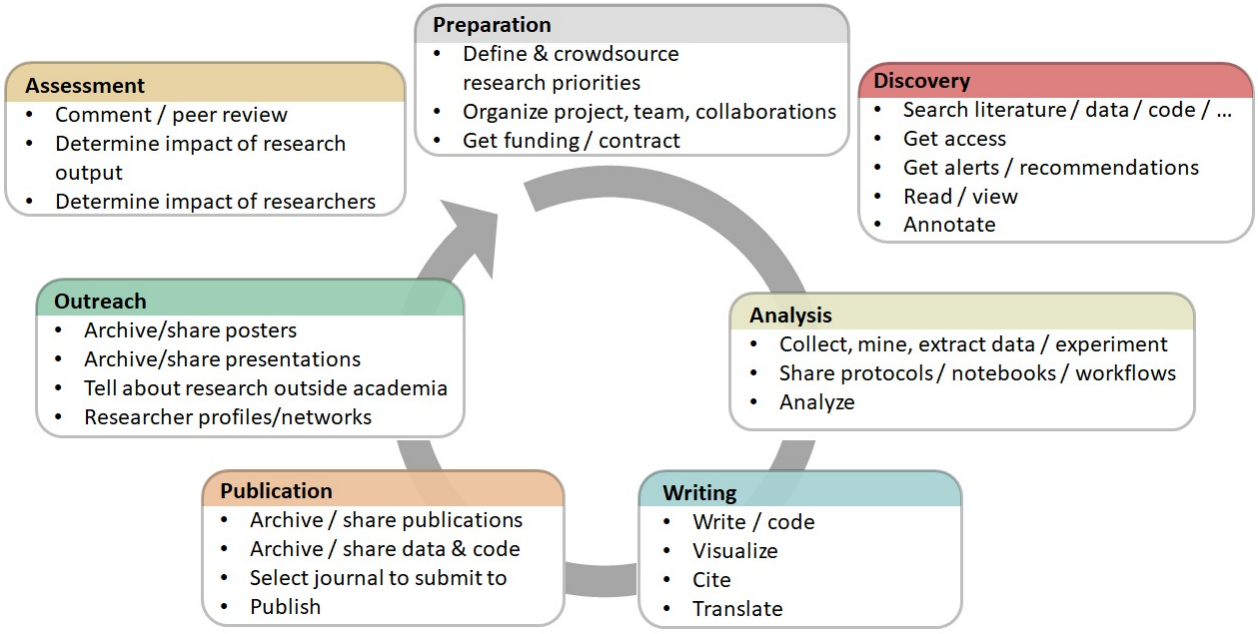
Pontika et al. (2015)

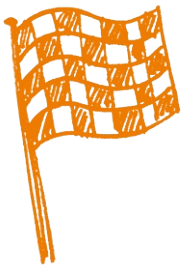
The aims and assumptions underlying the push to implement these various practices have been analysed by Fecher & Friesike (2013), whose analyses of the literature found five broad concerns, or "schools of thought". These are:

- **Democratic school:** Believing that there is an unequal distribution of access to knowledge, this area is concerned with making scholarly knowledge (including publications and data) available freely for all.
- **Pragmatic school:** Following the principle that the creation of knowledge is made more efficient through collaboration and strengthened through critique, this area seeks to harness network effects by connecting scholars and making scholarly methods transparent.
- **Infrastructure school:** This thread is motivated by the assumption that efficient research requires readily available platforms, tools and services for dissemination and collaboration.
- **Public school:** Based on the recognition that true societal impact requires societal engagement in research and readily understandable communication of scientific results, this area seeks to bring the public to collaborate in research through citizen science, and make scholarship more readily understandable through lay summaries, blogging and other less formal communicative methods.
- **Measurement school:** Motivated by the acknowledgement that traditional metrics for measuring scientific impact have proven problematic (by being too heavily focused on publications, often only at the journal-level, for instance), this strand seeks "alternative metrics" which can make use of the new possibilities of digitally networked tools to track and measure the impact of scholarship through formerly invisible activities.

Rationale

Open Science, as defined above, encompasses a huge number of potential structural changes to academic practice, whose culture can often be hierarchical and conservative. Moreover, even where researchers are sympathetic to the aims of Open Science, they might not yet see the worth in taking them up, as existing incentive mechanisms do not yet reflect this new culture of openness and collaboration. As a consequence, convincing researchers of the need to change their practices will require a good understanding not only of the ethical, social and academic benefits, but also of the ways in which taking up Open Science practices will actually help them succeed in their work. This section will describe some of the core concepts, principles, actors, and practices in Open Science, and how these fit within a broader research ecosystem.

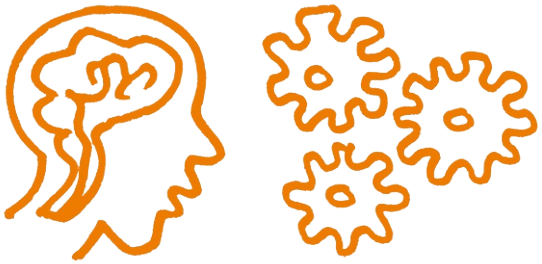


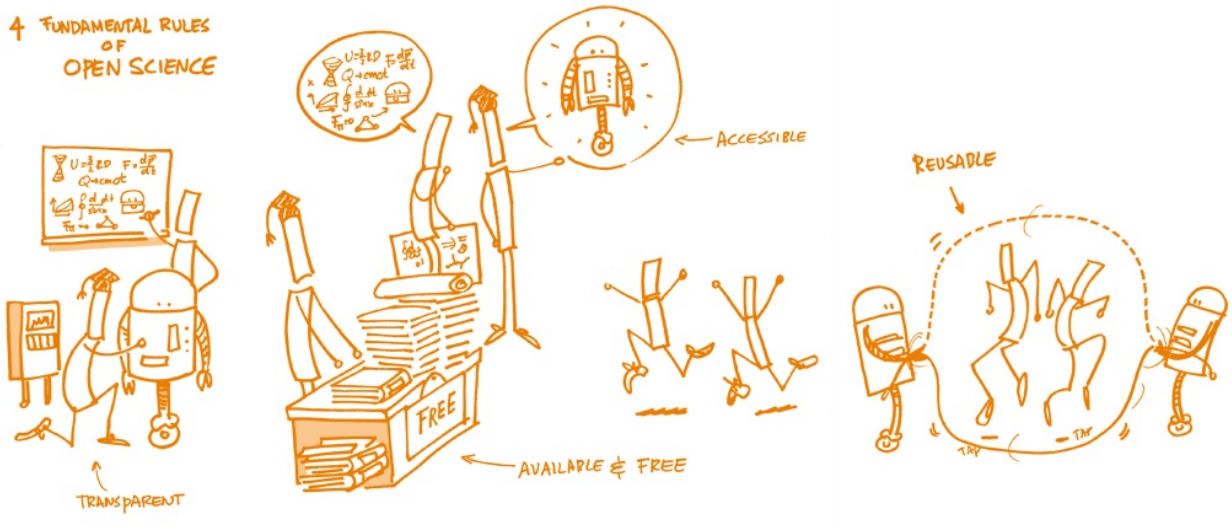


Learning objectives

1. Understand the social, economical, legal, and ethical principles and concepts underpinning Open Science.
2. Become familiar with the history of Open Science, and the disparity and diversity of views from different research communities, disciplines and cultures.
3. Gain insight into the developments around Open Science, and the personal impact these can have on researchers, research, and society more broadly.

Key components





Knowledge & Skills

- Open Science is the movement to help make the results of scholarly research more accessible, including code, data, and research papers.
    - It encompasses many different but often related aspects impacting the entire research lifecycle, including open publishing, open data, open source software, open notebook science, open peer review, open dissemination, and open materials (see glossary for definitions).

- History of Open Science, and the motivations behind the movement.
  - The origins of academic publishing began in the 17th century with the first academic journals.
  - Increasing motivation to share resources between research disciplines, as well as increased transparency for greater efficiency, rigour, accountability, sustainability for future generations, and reproducibility.
  - Ethical cases whereby increased transparency can reduce fraud, data manipulation, and selective reporting of results.
- Present state arose from pressure from research academies and governments for publicly-funded research to be shared more openly, often for the purpose of accelerated societal or economic growth and innovation.
  - Publicly funded research outputs should be publicly available.
  - Need to drive cultural change in research and amongst researchers.
  - Embracing of Web-based tools and technologies to facilitate scientific collaboration.
- Differences and commonalities within Open Science practices, principles and communities.
  - It is generally accepted that Open Science leads to increased impact associated with wider sharing and re-use (e.g., the so-called "open access citation advantage").
  - Open Science could increase trust in science and in the reliability of scientific results.
- Open Science and relations to licensing, copyright issues.
  - Typically, open research outputs are openly licensed in order to maximize re-use while allowing the creator to retain ownership and receive credit for their work.

Questions, obstacles, and common misconceptions

Q: "What is the difference between Open Science and ‘science’?"

A: Open Science refers to doing traditional science with more transparency involved at various stages, for example by openly sharing code and data. Many researchers do this already, but don’t call it Open Science.

Q: "Does ‘Open Science’ exclude the Humanities and Social Sciences?"

A: No, the term Open Science is inclusive. Indeed, the case is that sometimes Open Science is more broadly referred to as ‘Open Research’ or ‘Open Scholarship’ to be more inclusive of other disciplines, principles and practices. However, Open Science is a commonly used term at multiple levels and so it makes sense to adopt it for communication purposes, with the proviso that it includes all research disciplines.

Q: "Does Open Science lead to misuse or misunderstanding of research?"

A: No, the application of Open Science principles is in fact a safeguard against misuse or misunderstanding. Transparency breeds trust, confidence and allows others to verify and validate the research process.

Q: "Will Open Science lead to too much information overload?"

A: It is better to have too much information and deal with it, than to have too little and live with the risk of missing the important parts. And there are technologies such as RSS feeds, machine learning and artificial intelligence that are making content aggregation easier.

Learning outcomes

1. Be able to explain the core underlying academic, economic, and societal principles and concepts supporting Open Science, and why this matters to you in terms of broader impacts.
2. Develop an understanding of the numerous dimensions of Open Science, and some of the tools and practices involved in this.
3. Be familiar with the present state of Open Science, and the diversity of perspectives that this encompasses.

Further reading

- European Commission's Directorate-General for Research & Innovation (RTD) (2016). Open innovation, Open Science, open to the world - a vision for Europe. ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe
- Fecher and Friesike (2014). Open Science: One Term, Five Schools of Thought. doi.org/10.1007/978-3-319-00026-8_2
- High Level Group (2017). Europe's future. Open innovation, Open Science, open to the world: reflections of the Research, Innovation and Science Policy Experts (RISE). doi.org/10.2777/79895
- Masuzzo and Martens (2017). Do you speak Open Science? Resources and tips to learn the language. doi.org/10.7287/peerj.preprints.2689v1
- Watson (2015). When will ‘Open Science’ become simply ‘science’?. doi.org/10.1186/s13059-015-0669-2

2. Open Research Data and Materials

What is it?

Open research data is data that can be freely accessed, reused, remixed and redistributed, for academic research and teaching purposes and beyond. Ideally, open data have no restrictions on reuse or redistribution, and are appropriately licensed as such. In exceptional cases, e.g. to protect the identity of human subjects, special or limited restrictions of access are set. Openly sharing data exposes it to inspection, forming the basis for research verification and reproducibility, and opens up a pathway to wider collaboration. At most, open data may be subject to the requirement to attribute and sharealike (see the Open Data Handbook).

Rationale

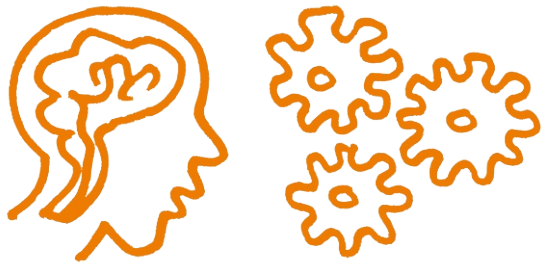
Research data are often the most valuable output of many research projects, they are used as primary sources that underpin scientific research and enable derivation of theoretical or applied findings. In order to make findings/studies replicable, or at least reproducible or reusable (see Reproducible Research And Data Analysis) in any other way, the best practice recommendation for research data is to be as open and FAIR as possible, while accounting for ethical, commercial and privacy constraints with sensitive data or proprietary data.

Learning objectives

1. Gain an understanding of the basic characteristics and principles of open and FAIR research data, including appropriate packaging and documentation, to enable others to understand, reproduce, and re-use in alternative ways.

2. Familiarity with the sorts of data that might be considered sensitive, and the restrictions or constraints on openly sharing them.
3. Be able to convert a ‘closed’ dataset into one which is ‘open’ by implementing the necessary measures in a data management plan, with appropriate data stewardship and metadata.
4. Be able to use research data management plan and to make your research results findable and accessible, even if it contains sensitive data.
5. Understand the pros and cons of openly sharing different types of data (e.g., privacy, sensitivity, de-identification, mediated access).
6. Understand the importance of appropriate metadata for sustainable archiving of research data.
7. Understand the basic workflows and tools for sharing research data.

Key components



Knowledge & Skills

FAIR principles

In 2014, a core set of principles were drafted in order to optimize the reusability of research data, named the FAIR Data Principles. They represent a community-developed set of guidelines and best practices to ensure that data or any digital object are **F**indable, **A**ccessible, **I**nteroperable and **R**e-usable:

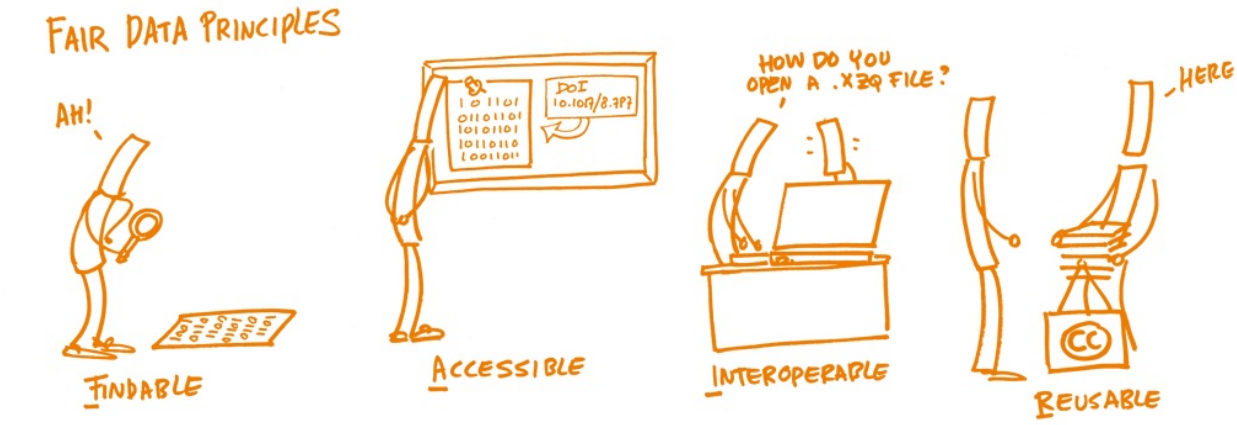
Findable: The first thing to be in place to make data reusable is the possibility to find them. It should be easy to find the data and the metadata for both humans and computers. Automatic and reliable discovery of datasets and services depends on machine-readable persistent identifiers (PIDs) and metadata.

Accessible: The (meta)data should be retrievable by their identifier using a standardized and open communications protocol, possibly including authentication and authorisation. Also, metadata should be available even when the data are no longer available.

Interoperable: The data should be able to be combined with and used with other data or tools. The format of the data should therefore be open and interpretable for various tools, including other data records. The concept of interoperability applies both at the data and metadata level. For instance, the (meta)data should use vocabularies that follow FAIR principles.

Re-usable: Ultimately, FAIR aims at optimizing the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings. Also, the reuse of the (meta)data should be stated with (a) clear and accessible license(s).

Distinct from peer initiatives that focus on the human scholar, the FAIR principles put a specific emphasis on enhancing the ability of machines to automatically find and use data or any digital object, in addition to supporting its reuse by individuals. The FAIR principles are guiding principles, not standards. FAIR describes qualities or behaviours that are required to make data maximally reusable (e.g., description, citation). Those qualities can be achieved by different standards.



Data publishing

Most researchers are more or less familiar with Open Access publishing of research articles and books (see chapter 5). More recently, and for the reasons mentioned above, data publishing has gained increasing attention. More and more funders expect the data produced in research projects they finance to be findable, accessible and as open as possible.

There are several distinct ways to make research data accessible, including (Wikipedia):

- Publishing data as supplemental material associated with a research article, typically with the data files hosted by the publisher of the article.
- Hosting data on a publicly-available website, with files available for download.
- Depositing data in a repository that has been developed to support data publication, e.g., Dataverse, Dryad), figshare, Zenodo.
- A large number of general and domain or subject specific data repositories exist which can provide additional support to researchers when depositing their data.
- Publishing a data paper about the dataset, which may be published as a preprint, in a journal, or in a data journal that is dedicated to supporting data papers. The data may be hosted by the journal or hosted separately in a data repository. Examples of data journals include Scientific Data (by SpringerNature) and the Data Science Journal (by CODATA). For a comprehensive review of data journals, see Candela et al.

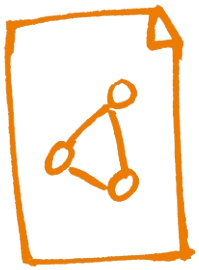
The CESSDA ERIC Expert tour guide on Data Management provides an overview of pros and cons of different data publication routes. Sometimes, your funder or another external party requires you to use a specific repository. If you are free to choose, you may consider the order of preference in the recommendations by OpenAIRE:

1. Use an external data archive or repository already established for your research domain to preserve the data according to recognised standards in your discipline.
2. If available, use an institutional research data repository, or your research group’s established data management facilities.
3. Use a cost-free data repository such as Dataverse, Dryad, figshare or Zenodo.
4. Search for other data repositories in re3data. There is no single filter option in re3data covering the FAIR principles, but considering the following filter options will help you to find FAIR-compatible repositories: access categories, data usage licenses, trustworthy data repositories (with a certificate or explicitly adhering to archival standards) and whether a repository gives the data a persistent identifier (PID). Another aspect to consider is whether the repository supports versioning.

You should consider where to deposit and publish your data already in your research data management plan. CESSDA offers some practical questions, which are recommended to be considered. For example: Which data and associated metadata, documentation and code will be deposited? How long does the data need to be retained? For how long should the data remain reusable? How will the data be made available? What access category will you choose? For more questions check Adapt your DMP: part 6. On the other hand don't forget to check if a chosen repository meets requirements of your research and of your funder. Some repositories have already gained certification, like CoreTrustSeal, which certifies them to be trustworthy and to be able to meet Core Trustworthy Data Repositories Requirements. It is worth mentioning that some domain specific repositories may accept only high-quality data with a potential for reuse and that can be publicly shared.

Since there are several routes to publish your data, you should note that for a dataset to "count" as a publication, it should follow a similar publication process as an article (Brase et al., 2009) and should be:

- Properly documented with metadata;
- Reviewed for quality, e.g. content of the study, methodology, relevance, legal consistency and documentation of materials;
- Searchable and discoverable in catalogues (or databases);
- Citable in articles.



Data citation

Data citation services help research communities discover, identify, and cite research data (and often other research objects) with confidence. This typically involves the creation and allocation of Digital Object Identifiers (DOIs) and accompanying metadata through services such as DataCite, and can be integrated with research workflows and standards. This is an emerging field, and involves aspects such as conveying to journal publishers the importance of appropriate data citation in articles, as well as enabling research articles themselves to be linked to any underlying data. Through this, citable data become legitimate contributions to the process of scholarly communication, and can help pave the way for new metrics and publication models that recognize and reward data sharing.

As an initial step towards good practice for data citation, the Data Citation Synthesis Group of FORCE11 has put forward the Joint Declaration of Data Citation Principles, targeted at both researchers and data service providers. Adhering to these principles, data repositories usually provide researchers with a reference they can use when referring to a given dataset.

Data packaging

Data packages are containers for describing and sharing accompanying data files, and typically comprise a metadata file describing the features and context of a dataset. This can include aspects such as creation information, provenance, size, format type, field definitions, as well as any relevant contextual files, such as data creation scripts or textual documentation. From the Data Packaging Guide:

- Data are forever: Datasets outlive their original purpose. Limitations of data may be obvious within their original context, such as a library catalog, but may not be evident once data is divorced from the application it was created for.
- Data cannot stand alone: Information about the context and provenance of the data--how and why it was created, what real-world objects and concepts it represents, the constraints on values--is necessary to helping consumers interpret it responsibly.
- Structuring metadata about datasets in a standard, machine-readable way encourages the promotion, shareability, and reuse of data.

Sharing sensitive and proprietary data

With appropriate data management planning much sensitive and proprietary data can be shared, reused, and FAIR. The metadata can almost always be shared. Guidance and best practices for sharing sensitive data are necessarily region-specific because of differing regulations (see for example UKDS'Companion material for Managing and Sharing Research Data handbook). International Association for Social Science Information Services and Technology keeps a list of international guidance in data management that is a good starting point. There are several approaches and initiatives to help researchers achieve this. DCC's DMPonline tool includes a number of templates for funders. The CESSDA Expert Tour Guide on Data Management provides information and practical examples on how to share personal data and on copyright and database issues across the European countries. The Tour Guide also gives an overview on the impact of the GDPR which will harmonize personal data legislation in Europe (May 2018), and provides an updated overview on EU diversity on data protection.

Data brokers

Data brokers are knowledgeable, independent parties who act as data stewards for sensitive data. Researchers can transfer their sensitive data and jurisdiction over access to that data to the broker. This is especially common with patient-level data from clinical studies. Brokers provide a level of independence in the evaluation of whose data requests are scientifically valid and will not violate the privacy of research participants. Examples of data brokers include The YODA Project, ClinicalStudyDataRequest.com, National Sleep Research Resource and Supporting Open Access for Researchers (SOAR).

Analysis portals

Analysis portals are platforms that allow approved analysis of data without allowing full access (viewing or downloading) or controlling where and who gets access. Some data brokers also use analysis portals. Analysis portals control what additional datasets can be pooled with the sensitive data as well as what analyses can be run to ensure that personal information is not revealed during reanalysis. Examples of virtual analysis portals include Project Data Sphere, Vivli, RAIRD, Corpuscle, and INESS.

Social science and other researchers with sensitive data use a single-site analysis portal that can be accessed only under controlled regime. Approved researchers can access the data on-site, in a safe room, for scientific purposes. However, the metadata describing the data should be openly available and adhering to the FAIR principles.

De-identified and synthetic data

Many datasets containing participant-level private information can be shared once the dataset has been de-identified (Safe Harbor method) or a expert has determined that the dataset is not individually identifiable (Expert Determination method). Consult with your Research Ethics Board / Institutional Review Board to learn how to do this with your data. We also recommend the CESSDA Expert Tour Guide on Data Management, which provides information and practical examples on how to share personal data. However, some datasets cannot be safely de-identified and shared. Researchers can still improve the openness of research on such data by creating and sharing synthetic data. Synthetic data is similar in structure, content, and distribution to the real data and aims to attain "analytic validity": statistical analysis will return the same results for the synthetic data as the real data. The United States Census Bureau, for example, uses synthetic data and analysis portals in combination to allow reuse of highly sensitive data.

DataTags

DataTags is a framework designed to enable computer-assisted assessments of the legal, contractual, and policy restrictions that govern data sharing decisions. The DataTags system asks a user a series of questions to elicit the key properties of a given dataset and applies inference rules to determine which laws, contracts, and best practices are applicable. The output is a set of recommended DataTags, or simple, iconic labels that represent a human-readable and machine-actionable data policy, and a license agreement that is tailored to the individual dataset. The DataTags system is being designed to integrate with data repository software, and it will also operate as a standalone tool. DataTags is being developed at Harvard University. In Europe, DANS is working on adjusting DataTags to European legislation / General Data Protection Regulation (GDPR) (cf. DANS GDPR DataTags).

As mentioned above, the ultimate goal of data sharing your research data is to make them maximally reusable. To that end, before sharing your data you should manage them according to best practice. This includes, i.a., documentation and the choice of open file formats and licenses. You can read more about these issues in Section 4: Reproducible Research and Data Analysis as well as Section 6: Open Licensing and File Formats.

Open Materials

In addition to data sharing, the openness of research relies on sharing of materials. What materials researchers use is discipline-specific and sometimes unique to a lab. Below are examples of materials you can share, although always confer with peers in your discipline to identify which repositories are used. When you have materials, data, and publications from the same research project shared in different repositories, cross-reference them with a link and a unique identifier so they can be easily located.

Reagents

A reagents is a substance, compound or mixture that can be added to a system in order to create a chemical or other reaction. Reagents can be deposited with repositories like Addgene, The Bloomington Drosophila Stock Center, and ATCC to make them easily accessible to other researchers. License your materials so they can be reused by other researchers.

Protocols

A protocol describes a formal or official record of scientific experimental observations in a structured format. Deposit virtual protocols for citation, adaptation, and reuse using Protocols.io.

Notebooks, containers, software, and hardware

Reproducible analysis is aided by the use of literate programming, container technology, and virtualization. In addition to sharing your code and data, also share your Jupyter notebooks, Docker images, or other analysis materials or software dependencies. Share notebooks with Open services such as mybinder that allow for public viewing and execution of the entire notebook on shared resources. Containers and notebooks can be shared with Rocker or Code Ocean. Software and hardware used in your research should be shared following best practices for documentation as outlined in Section 3. Read-only protocols should be deposited in your disciplines registry such as ClinicalTrials.gov and SocialScienceRegistry or a general registry like Open Science Framework. Many journals, such as Trials, JMIR Research Protocols, or Bio-Protocol, will publish your protocol. Best practices for publishing your protocol open access are the same as publishing your report open access (see Section 5).

Questions, obstacles, and common misconceptions

Q: "Is it sufficient to make my data openly available?"

A: "No—openness is a necessary but not sufficient condition for maximum reuse. Data have to be FAIR in addition to open."

Q: "What do the FAIR principles mean/imply for different stakeholders/audiences?"

A: "This is a great topic for discussion!"

Obstacle: Researchers may be reluctant to share their data because they are afraid that others will reuse them before they have extracted the maximum usage from them, or that others might not fully understand the data and therefore mis-use them.

(suggested) A: You may publish your data to make them findable with metadata, but set an embargo period on the data to make sure that you can publish your own article(s) first.

Q: "Is making my data FAIR a lot of extra work?"

A: "Not necessarily! Making data FAIR is not only the responsibility of the individual researchers but of the whole group. The best way to ensure that your data is FAIR is to create a Data Management Plan and plan everything beforehand. During the data collection and data processing follow the discipline standards and measures recommended by a repository.

Q: "I want to share my data. How should I license them?"

A: "That’s a good question. First of all think about who owns the data? A research funder or an institution that you work for. Then, think about authorship. Applying a suitable license to your data is crucial in order to make them reusable. For more information about licensing, please see 6. Open Licensing and File Formats.

Q: "I cannot make my data directly available—they are too large to share conveniently / have restrictions related to privacy issues. What should I do?"

A: "You should talk to experts in domain specific repositories on how to provide sufficient instructions to make your data findable and accessible."

Learning outcomes

1. Understand the characteristics of open data, and in particular the FAIR principles.
2. Be familiar with some of the arguments for and against open data.
3. Be able to differentiate and address sensitive data and opFAIR data; these two categories are not necessarily incompatible.
4. Be able to transform a dataset into one that is sufficient for open sharing (non-proprietary format), meets the standards of the FAIR principles, and is designed for maximized accessibility, transparency and re-use by providing sufficient metadata.
5. Know the difference between raw and processed (or cleaned) data, and the importance of version labels.
6. Know commonly used file formats and community standards for maximum re-usability.
7. Be able to write a data management plan.

Further reading

- Averkamp et al. (2018). Data packaging guide. github.com/saverkamp/beyond-open-data/blob/master/DataGuide.md.
- Barend et al. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. doi.org/10.3233/ISU-170824
- Brase et al. (2009). Approach for a joint global registration agency for research data. doi.org/10.3233/ISU-2009-0595
- Candela et al. (2015). Data journals: A survey. doi.org/10.1002/asi.23358
- CESSDA Training Working Group (2017-2018a). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC. cessda.eu/DMGuide

- CESSDA Training Working Group (2017-2018b). CESSDA Data Management Expert Guide: Citing your data. Bergen, Norway: CESSDA ERIC.cessda.eu/DMGuide/citingdata
- FAIRsharing.org (2016). FAIR. The FAIR Principles. doi.org/10.25504/FAIRsharing.WWI10U
- Force 11 (n.y.). Guiding principles for Findable, Accessible, Interoperable, and Re-usable data publishing Version B1.0. force11.org/fairprinciples
- Gorgolewski et al. (2013). Making data sharing count: a publication-based solution. doi.org/10.3389/fnins.2013.00009
- Kratz and Strasser (2015). Making Data Count. doi.org/10.1038/sdata.2015.39
- Piwowar and Vision (2013). Data reuse and the open data citation advantage. doi.org/10.7717/peerj.175
- Wilkinson et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. doi.org/10.1038/sdata.2016.18
- Wilkinson et al. (2918). A design framework and exemplar metrics for FAIRness. doi.org/10.1038/sdata.2018.118

Initiatives and projects

- DANS GDPR DataTags. zingtree.com
- FAIR Metrics. fairmetrics.org
- GO FAIR Initiative. go-fair.org
- The FAIR Data Principles explained. go-fair.org
- 5★ OPEN DATA. 5stardata.info

3. Open Research Software and Open Source

What is it?

Open research software, or open-source research software, refers to the use and development of software for analysis, simulation, visualization, etc. where the full source code is available. In addition, according to the Open Source Definition, open-source software must be distributed in source and/or compiled form (with the source code available in the latter case), and must be shared under a license that allows modification, derivation, and redistribution.

Rationale

Modern research relies on software, and building upon—or reproducing—that research requires access to the full source code behind that software (Barnes, 2010; Morin et al., 2012; Ince et al., 2012; Prins et al. 2015; Lowndes et al., 2018). As Buckheit and Donoho put it, paraphrasing Jon Claerbout, ‘‘An article about a computational result is advertising, not scholarship. The actual scholarship is the full software environment, code and data, that produced the result’’ (Buckheit & Donoho, 1995). Open access to the source code of research software also helps improve the impact of the research (Vandewalle, 2012).

Sharing software used for research (whether computational in nature, or that relies on any software-based analysis/interpretation) is a necessary, though not sufficient, condition for reproducibility. This is due to the unavoidable ambiguity that arises when trying to fully describe software using natural language, e.g., in a paper (Ince et al., 2012). Furthermore, many (if not most) software programs may contain some undetected errors (Soergel, 2015), so even a "perfect" written description of software would not be able to account for all results.

In addition to reproducibility, sharing software openly allows developers to receive career credit for their efforts, either through direct citation (Smith et al., 2016) or via software meta-articles published in, e.g., the Journal of Open Research Software or the Journal of Open Source Software (Smith et al., 2018). Neil Chue Hong maintains a list of many domain-specific journals that publish software articles.

Learning objectives

1. Learn the characteristics of open software; understand the ethical, legal, economic, and research-impact arguments for and against open software, and further understand the quality requirements of open code.
2. Learn how to use existing open software and appropriately attribute (cite) it.
3. Learn how to use common tools and services for sharing research codes openly.

4. Be able to choose the appropriate license for their software, and understand the difference between permissive and non-permissive licenses.

Key components

Knowledge

There are several different platforms that support open sharing and collaboration on software, research or otherwise. First of all, you can use this checklist to evaluate openness of existing research software:

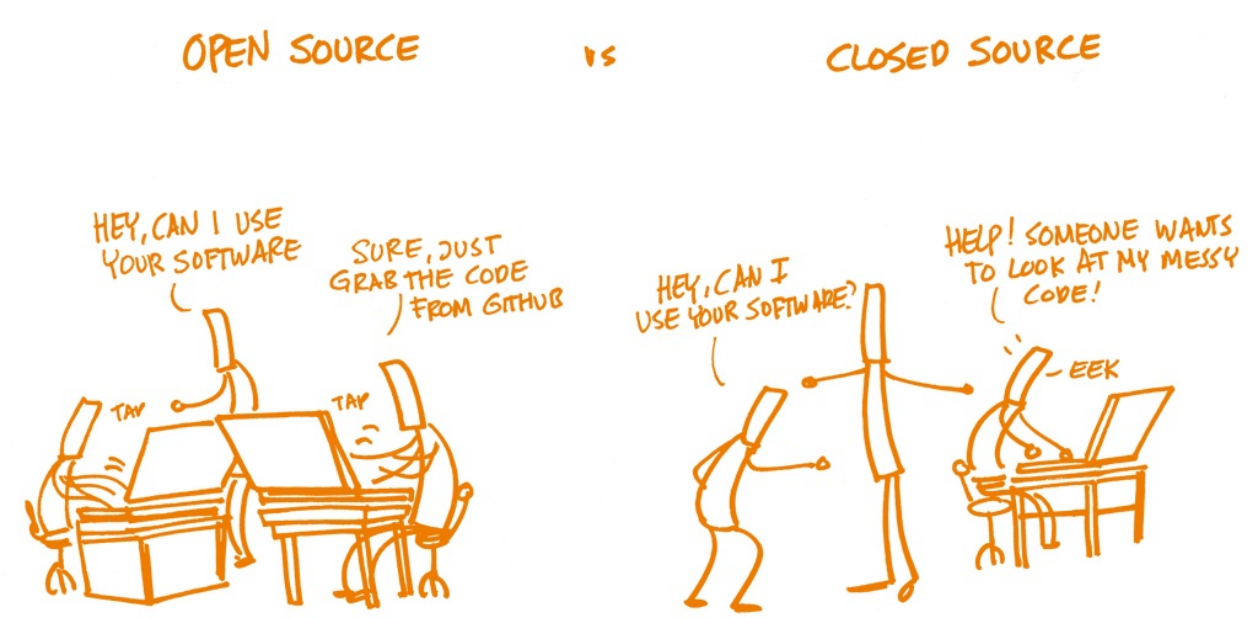
- Is the software available to download and install?
- Can the software easily be installed on different platforms?
- Does the software have conditions on the use?
- Is the source code available for inspection?
- Is the full history of the source code available for inspection through a publicly available version history?
- Are the dependencies of the software (hardware and software) described properly? Do these dependencies require only a reasonably minimal amount of effort to obtain and use?

These qualities relate to and build on the Open Source Definition.

GitHub is a popular tool that allows version control: management and overall tracking of changes in a particular piece of software. Services such as GitHub, GitLab, Bitbucket, and others provide an interface to the tool as well as remote storage services that can be used to maintain, share, and collaborate on research software. As a tool it is quite widespread and, although it has an initial learning curve, it has proven invaluable to establishing an open and reproducible research workflow.

Having the research software on GitHub is just the first part; it is equally important to have a published and persistent identifier associated with it, such as a DOI. There are several ways of associating a DOI with a GitHub repository; the easiest one is to employ Zenodo (a free, open catch-all repository created by OpenAIRE and CERN) to do the assignment, although other repositories for archiving software and obtaining a DOI do exist, such as Figshare. Zenodo integrates with GitHub to archive the software and provide a DOI when developers make a formal release on GitHub.

Publicly shared software is not actually open source unless accompanied by a suitable license, because by default software (along with any other creative work) falls under exclusive copyright to the creators, meaning no one else can use, copy, distribute, or modify your work (choosealicense.com). (If you truly want to share your code with no restrictions whatsoever, you can dedicate it to the public domain.) Instead, you should choose an appropriate license for your software, based on what you would prefer to let others do (or prevent them from doing) with your code; the choosealicense.org site is a helpful resource to differentiate between licenses, although it does not feature every available or popular open-source license. Once you select a license, put the text—edited to include the author name(s) and year—in the software repository as a plaintext LICENSE file.



Although sharing software in any form is better than not sharing it, your software will have more impact and be more easily used by others—and your future self!—if you include documentation. This can include helpful comments in the code that explain **why** you did something (rather than what you did, which should be evident), an informative README file that describes what your software does and gives some helpful information (e.g., how to install, how to cite, how to run, important dependencies), tutorials/examples, and/or API documentation (which may be automatically generated from properly formatted comments in the code).

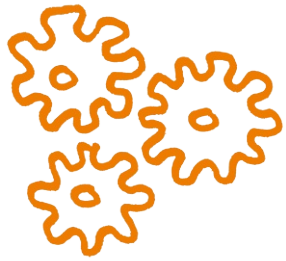
Missing or inaccessible dependencies or insufficient documentation of the computational environment are very common barriers to reuse and reproducibility. One approach to address these barriers is to share your code with your computational environment using container technology. Containers package the code with the dependencies and computational environment so others can more easily run your analysis. Examples of container implementation in research include Rocker, Binder, and Code Ocean.

When you use software — whether you wrote it, or someone else did and made it available — appropriate citation is important for reproducibility (discussed more in Section 4; briefly, the version used can change your results or interpretation) and giving credit to the developers of the software (Niemeyer 2016, Smith 2016). The decision of when to cite software is up to you as the researcher, but we recommend a citation whenever the software did some work integral to your results, interpretation, or conclusions. The best way to make *your* code easily citable is to use the GitHub–Zenodo integration described before and provide the resulting DOI in an obvious place like the software's README, perhaps along with a suggested citation format. When citing any software, you should include at minimum the author name(s), software title, version number, and unique identifier/locator (Smith 2016). If you use someone else's software and they provided a DOI, then you can easily use that to identify and point to the software; if they did not archive their software, then you should include a URL where the software can be found and the version number or (e.g.) commit hash.

Additional, more complicated concepts include automated testing and continuous integration of software, packaging of software in binary formats, and governance and management of multi-person open-source projects (i.e., codes of conduct, contributing guides). Some of these topics are described by Scopatz and Huff (2015).pdf). Wilson et al. (2017) also provide a practical guide to best practices for scientific computing that includes advice specifically on research software development.

Open Source Hardware

The open source principles above extend to hardware. Researchers often use proprietary instrumentation or hardware in their research that is not freely accessible, reusable, or adaptable. Scientific hardware includes everything from sequencing tools and microscopes to specialized testing equipment and particle colliders. Open Science Hardware (OScH) community, for example, is leading a push for the open source movement to include scientific tools, hardware, and research infrastructures through their Global Open Science Hardware Roadmap.



Skills

- Create a repository on GitHub, and enable the integration with Zenodo. Mint the first release of the software.
- Choose a software license using (e.g.) choosealicense or the Open Source Initiative.
- Create documentation for a software package, including README, comments, and examples.
- Appropriately cite software used for a paper.

Questions, obstacles, and common misconceptions

Q: "I can’t share my software—it’s too messy / it doesn’t have good documentation / I didn’t leave good comments!"

A: Developers of research software around the world empathize with this feeling—people rarely feel like their code is "ready" to publicly share or that it is “finished”. However, as Barnes (2010) put it, “if your code is good enough to do the job, then it is good enough to release—and releasing it will help your research and your field.” In other words, if you feel comfortable enough with your software to publish a study or report results, then the code is sufficiently developed to share with your colleagues. (In the other direction, if you don’t feel comfortable sharing the code, then perhaps it requires more development or testing before using in a publication). Plus, sharing your code allows others to improve and build upon it, leading to even greater impact and innovation (and citations for you!).

Q: "What if someone takes the code I have shared and uses it for nefarious purposes, or claims they wrote it?"

A: Selecting an appropriate license for your software will help protect you from any uses of your software by others; for example, the common MIT License includes both limitations of liability and states that no warranty is provided. If someone else tries to claim that they wrote the software you made available, then you can point to the timestamps on your repository or archived versions as proof of your prior work.

Q: "If I share my code in an online repository, I will be deluged with requests for user support."

A: Although potential users may ask you for help, either via email or (e.g.) issues filed on the online repository, you are under no obligation to provide support if you prefer not to or cannot do so. An appropriate license even provides you with legal protection for this (e.g., the no-warranty clause of the MIT License).

Common misconception: simply putting code online makes it open-source software. In fact, unless the software is accompanied by a license that grants permission for others to use, copy, modify, and/or distribute, then the developer(s) retain exclusive copyright. A open-source license needs to accompany the code to make it open-source software.

Learning outcomes

1. Be able to share software under the most appropriate license (i.e., both the tools and the licensing).
2. Be able to upload, version, and register a piece of code under a persistent identifier.
3. Be able to cite software used for a research article.

Further reading

- Balasegaram et al. (2017). An open source pharma roadmap. doi.org/10.1371/journal.pmed.1002276
- Dryden et al. (2017). Upon the Shoulders of Giants: Open-Source Hardware and Software in Analytical Chemistry. doi.org/10.1021/acs.analchem.7b00485
- Ince et al. (2012). The case for open computer programs.doi.org/10.1038/nature10836
- Iskoujina and Roberts (2015). Knowledge sharing in open source software communities: motivations and management. PDF
- Jiménez et al. (2017).Four simple recommendations to encourage best practices in research software. doi.org/10.12688/f1000research.11407.1
- Martinez-Torres and Diaz-Fernandez (2013).Current issues and research trends on open-source software communities PDF
- Morin et al. (2012). Shining Light into Black Boxes. PDF
- Oishi et al. (2018). Perspectives on Reproducibility and Sustainability of Open-Source Scientific Software from Seven Years of the Dedalus Project. arXiv:1801.08200v1 [astro-ph.IM]
- Scacchi (2010). The Future of Research in Free/Open Source Software Development. PDF
- Sandve et al. (2013). Ten simple rules for reproducible computational research doi.org/10.1371/journal.pcbi.1003285

- Shamir et al. (2013).Practices in source code sharing in astrophysics. arXiv:1304.6780v1 [astro-ph.IM]
- Steinmacher et al. (2014). A systematic literature review on the barriers faced by newcomers to open source software projects. PDF
- Stodden (2010). The Scientific Method in Practice: Reproducibility in the Computational Sciences.PDF
- Vandewalle (2012). Code Sharing Is Associated with Research Impact in Image Processing. PDF

4. Reproducible Research and Data Analysis

What is it?

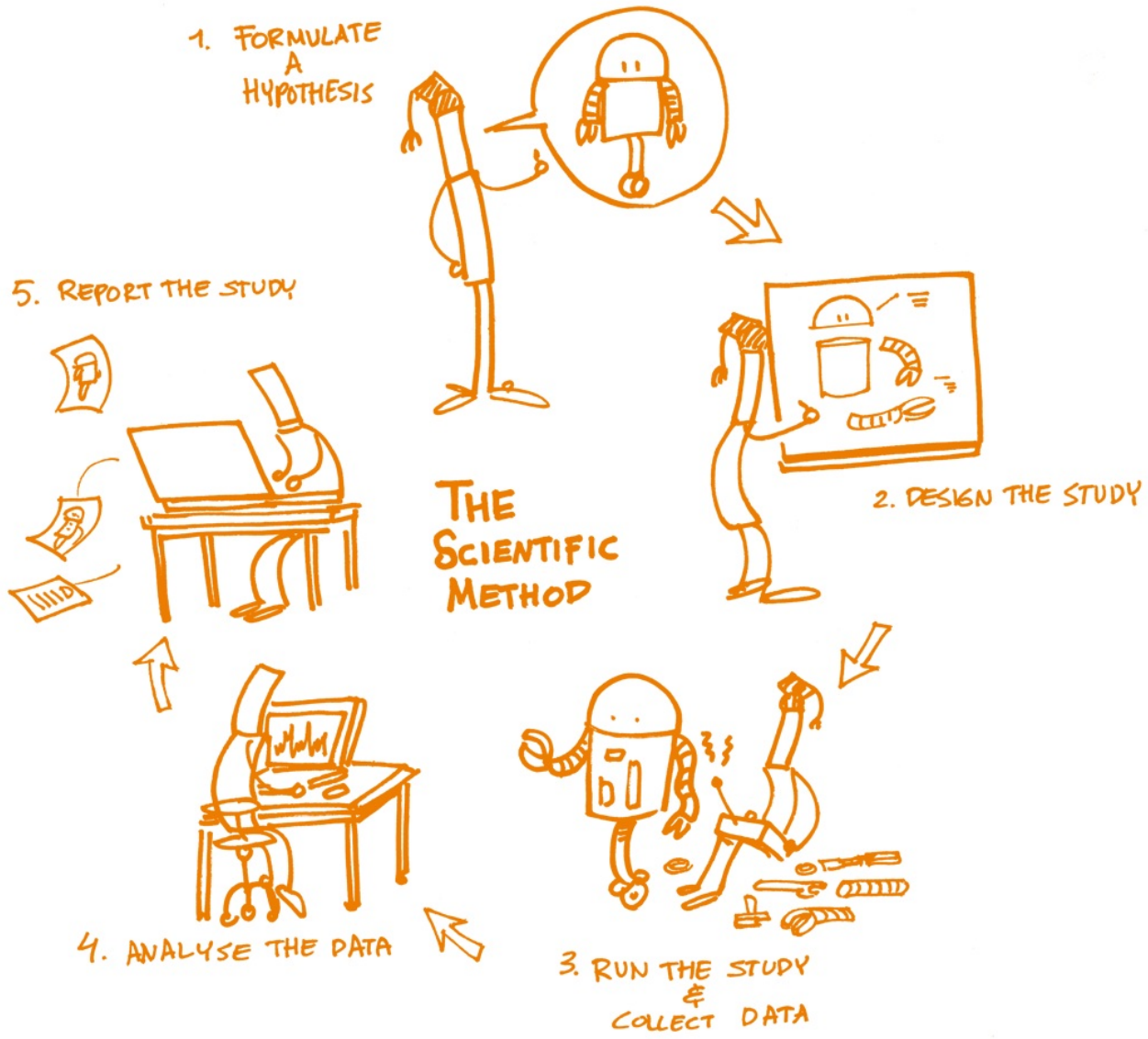
Reproducibility means that research data and code are made available so that others are able to reach the same results as are claimed in scientific outputs. Closely related is the concept of replicability, the act of repeating a scientific methodology to reach similar conclusions. These concepts are core elements of empirical research.

Improving reproducibility leads to increased rigour and quality of scientific outputs, and thus to greater trust in science. There has been a growing need and willingness to expose research workflows from initiation of a project and data collection right through to the interpretation and reporting of results. These developments have come with their own sets of challenges, including designing integrated research workflows that can be adopted by collaborators while maintaining high standards of integrity.

The concept of reproducibility is directly applied to the scientific method, the cornerstone of Science, and particularly to the following five steps:

1. Formulating a hypothesis
2. Designing the study
3. Running the study and collecting the data
4. Analyzing the data
5. Reporting the study

Each of these steps should be clearly reported by providing clear and open documentation, and thus making the study transparent and reproducible.



Rationale

Overarching factors can further contribute to the causes of non-reproducibility, but can also drive the implementation of specific measures to address these causes. The culture and environment in which research takes place is an important ‘top-down’ overarching factor. From a ‘bottom-up’ perspective, continuing education and training for researchers can raise awareness and disseminate good practice.

While understanding the full range of factors that contribute to reproducibility is important, it can also be hard to break down these factors into steps that can immediately be adopted into an existing research program and immediately improve its reproducibility. One of the first steps to take is to assess the current state of affairs, and to track improvement as steps are taken to increase reproducibility even more. Some of the common issues with research reproducibility are shown in the figure below.

Source: Symposium report, October 2015. Reproducibility and reliability of biomedical research: improving research practice PDF.

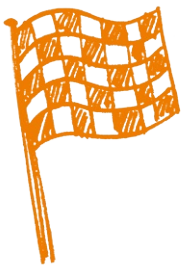
Goodman, Fanelli, & Ioannidis (2016) note that in epidemiology, computational biology, economics, and clinical trials, reproducibility is often defined as:

"the ability of a researcher to duplicate the results of a prior study using the same materials as were used by the original investigator. That is, a second researcher might use the same raw data to build the same analysis files and implement the same statistical analysis in an attempt to yield the same results."

This is distinct from replicability: *"which refers to the ability of a researcher to duplicate the results of a prior study if the same procedures are followed but new data are collected."* A simpler way of thinking about this might be that reproducibility is methods-oriented, whereas replicability is results-oriented.

Reproducibility can be assessed at several different levels: at the level of an individual project (e.g., a paper, an experiment, a method or a dataset), an individual researcher, a lab or research group, an institution, or even a research field. Slightly different kinds of criteria and points of assessment might apply to these different levels. For example, an institution upholds reproducibility practices if it

institutes policies that reward researchers who conduct reproducible research. On the other hand, a research field might be considered to have a higher level of reproducibility if it develops community-maintained resources that promote and enable reproducible research practices, such as data repositories, or common data-sharing standards.



Learning objectives

There are three major objectives that need to be addressed here:

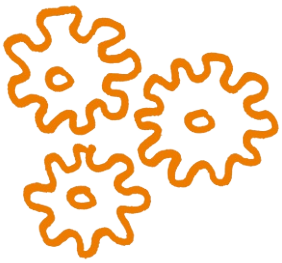
1. Understand the important impact of creating reproducible research.
2. Understand the overall setup of reproducible research (including workflow design, data management and dynamic reporting).
3. Be aware of the individual steps in the reproducibility process, as well as the corresponding resources that can be employed.

Key components

Knowledge

The following is an indicative list of take-away points on reproducibility:

- What is the ‘reproducibility crisis’, and meta-analyses of reproducibility.
- Principles of reproducibility, and integrity and ethics in research.
- What are the computing options and environments that allow collaborative and reproducible set up.
- Factors that affect reproducibility of research.
- Data analysis documentation and open research workflows.
- Reproducible analysis environments (virtualization).
- Addressing the "Researcher Degrees of Freedom" (Wicherts et al., 2016).



Skills

There are several practical tips for reproducibility that one should have in mind when setting out the particular skills necessary to ensure this. Best practices in reproducibility borrow from Open Science practices more generally but their integration offers benefits to the individual researcher themselves, whether they choose to share their research or not. The reason that integrating reproducibility best practices benefits the individual researcher is that they improve the planning, organization, and documentation of research. Below we outline one example of implementing reproducibility into a research workflow with references to these practices in the handbook.

1. Plan for reproducibility before you start

Create a study plan or protocol.

Begin documentation at study inception by writing a study plan or protocol that includes your proposed study design and methods. Use a reporting guideline from the Equator Network if applicable. Track changes to your study plan or protocol using version control (reference to Version Control). Calculate the power or sample size needed and report this calculation in your protocol as underpowered studies are prone to irreproducibility.

Choose reproducible tools and materials

Select antibodies that work using an antibody search engine like CiteAb. Avoid irreproducibility through misidentified cell lines by choosing ones that are authenticated by the International Cell Line Authentication Committee. Whenever possible, choose software and hardware tools where you retain ownership of your research and can migrate your research out of the platform for reuse (see Open Research Software and Open Source).

Set-up a reproducible project

Centralize and organize your project management using an online platform, a central repository, or folder for all research files. You could use GitHub as a place to store project files together or manage everything using a electronic lab notebook such as Benchling, Labguru,or SciNote. Within your centralized project, follow best practices by separating your data from your code into different folders. Make your raw data read-only and keep separate from processed data (reference to Data Management).

When saving and backing up your research files, choose formats and informative file names that allow for reuse. File names should be both machine and human readable (reference to Data Management). In your analysis and software code, use relative paths. Avoid proprietary file formats and use open file formats (see 6 Open Licensing and File Formats).

2. Keep track of things

Registration

Preregister important study design and analysis information to increase transparency and counter publication bias of negative results. Free tools to help you make your first registration include AsPredicted, Open Science Framework, and Registered Reports. Clinical trials should use Clinicaltrials.gov.

Version control

Track changes to your files, especially your analysis code, using version control (see Open Research Software and Open Source).

Documentation

Document everything done by hand in a README file. Create a data dictionary (also known as a codebook) to describe important information about your data. For an easy introduction, use: Karl Broman's Data Organization module and refer to Data Management.

Literate programming

Consider using Jupyter Notebooks, KnitR, Sweave, or other approaches to literate programming to integrate your code with your narrative and documentation.

3. Share and license your research

Data

Avoid supplementary files, decide on an acceptable permissive license, and share your data using a repository. Follow best practices as outlined in the Open Research Data and Materials chapter.

Materials

Share your materials so they can be reused. Deposit reagents with repositories like Addgene, The Bloomington Drosophila Stock Center, and ATCC to make them easily accessible to other researchers. For more information, see the Open Materials subsection of Open Research Data and Materials.

Software, notebooks, and containers

License your code to inform about how it may be (re)used. Share notebooks with services such as mybinder that allow for public viewing and execution of the entire notebook on shared resources. Share containers or notebooks with services such as Rocker or Code Ocean. Follow best practices outlined in Open Research Software and Open Source.

4. Report your research transparently

Report and publish your methods and interventions explicitly and transparently and fully to allow for replication. Guidelines from the Equator Network, tools like Protocols.io, or processes like Registered Reports can help you report reproducibly. Remember to post your results to your public registration platform (such as ClinicalTrials.gov or the SocialScienceRegistry) within a year of finishing your study no matter the nature or direction of your results.

Questions, obstacles, and common misconceptions

Q: "Everything is in the paper; anyone can reproduce this from there!"

A: This is one of the most common misconceptions. Even having an extremely detailed description of the methods and workflows employed to reach the final result will not be sufficient in most cases to reproduce it. This can be due to several aspects, including different computational environments, differences in the software versions, implicit biases that were not clearly stated, etc.

Q: "I don’t have the time to learn and establish a reproducible workflow."

A: In addition to a significant number of freely available online services that can be combined and facilitate the setting up of an entire workflow, spending the time and effort to put this together will increase both the scientific validity of the final results as well as minimize the time of re-running or extending it in further studies.

Q: "Terminologies describing reproducibility are challenging."

A: See Barba (2018) for a discussion on terminology describing reproducibility and replicability.

Learning outcomes

1. Understand the necessity of reproducible research and its reasoning.

2. Be able to establish a reproducible workflow within the context of an example task.
3. Know tools that can support reproducible research.

Further reading

- Button et al. (2013). Power failure: why small sample size undermines the reliability of neuroscience. doi.org/10.1038/nrn3475
- Karl Broman (n.y.). Data Organization. Choose good names for things. kbroman.org

5. Open Access to Published Research Results

What is it?

Open Access to publications means that research publications like articles and books can be accessed online, free of charge by any user, with no technical obstacles (such as mandatory registration or login to specific platforms). At the very least, such publications can be read online, downloaded and printed. Ideally, additional rights such as the right to copy, distribute, search, link, crawl and mine should also be provided. Open Access can be realised through two main non-exclusive routes:

- Green Open Access (self-archiving): The published work or the final peer-reviewed manuscript that has been accepted for publication is made freely and openly accessible by the author, or a representative, in an online repository. Some publishers request that Open Access be granted only after an embargo period has elapsed. This embargo period can last anywhere between several months and several years. For publications that have been deposited in a repository but are under embargo, usually at least the metadata are openly accessible.
- Gold Open Access (Open Access publishing): The published work is made available in Open Access mode by the publisher immediately upon publication. The most common business model is based on one-off payments by authors (commonly called APCs – article processing charges – or BPCs – book processing charges). Where Open Access content is combined with content that requires a subscription or purchase, in particular in the context of journals, conference proceedings and edited volumes, this is called hybrid Open Access.

Rationale

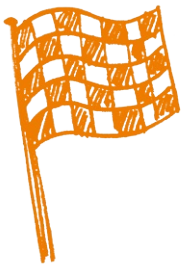
One of the most common ways to disseminate research results is by writing a manuscript and publishing it in a journal, conference proceedings or book. For many years those publications were available to the public under a payment by means of a subscription fee or individually. However, at the turn of the 21st century a new movement appeared with a clear objective: make all the research results available to the public without any restriction. This movement took the name of Open Access and established two initial strategies to achieve its final goal. The first strategy was to provide tools and assistance to scholars to deposit their refereed journal articles in open electronic repositories. The second one was to launch a new generation of journals using copyright and other tools to ensure permanent open access to all the articles they publish. As a result of the first strategy we see self-archiving practices: researchers depositing and disseminating papers in institutional or subject based repositories. And as a result of the second strategy we have seen the creation of the open access journals that provide free access to readers and allow reuse of their contents without almost any restriction.

Beyond those two strategies established in the Budapest Open Access Initiative in 2002, we have seen the growth of new methods of dissemination. Among them, we find the publication of preprints through institutional repositories and preprint servers. Preprints are widely used in physical sciences and now emerging in life sciences and other fields. Preprints are documents that have not been peer reviewed but are considered as a complete scientific publication in a first stage. Some of the preprints servers include open peer review services and the availability to post new versions of the initial paper once reviewed by peers. Following this trend of including open peer review processes in preprint servers we have seen the development of new publishing platforms supported by funders like the Wellcome Trust or the Bill and Melinda Gates Foundation . Even the European Commission is planning to to launch a publishing platform for the Horizon 2020 funded projects.

The choice of a journal or a publishing platform may affect the availability and accessibility of the research results. There are several options for researchers when deciding where, when, and how to publish their findings. It is fundamental to know all the implications to avoid future problems.

The rise of many business models around open access journals poses a lot of misunderstandings and uncertainties to the researchers when deciding where to publish. Moreover, paywalled journals offer individual open access models, the so-called hybrid model, that brings more complexity when deciding where and how to publish.

Regarding self archiving, researchers are confused by the different requirements established by the publishers in relation with version of a paper that they can deposit in a repository and when this version can be available to the public. This delay in allowing public access to the full text is often called embargo period and it is not uniform for all the journals. Institutions who provide a repository for its researchers should facilitate self archiving practices by digesting all those publisher requirements.



Learning objectives

1. Learn about the different options a researcher has when deciding where to publish a paper, including funder requirements.
2. Be able to decide if a paper can be published before peer review, for example in a preprint server. Trainees will learn how to determine which options they have according to their disciplines/journal policies, and if there would be consequences afterwards that might jeopardize final publication in a peer-reviewed journal.
3. Trainees will learn how to discover the differences between policies of peer-reviewed journals, particularly when submitting something available as a preprint. They will learn the differences among open-access journals, such as which require a fee for submission/publication and which licenses they use.
4. Trainees will learn about the implications of publishing in paywalled journals for future self-archiving in a repository, and the publisher requirements in terms of version and embargo. Trainees will also learn about hybrid open-access journals.
5. (optional depending on audience) Trainees will learn about open-access opportunities when publishing in books, since this is the main avenue of dissemination for some disciplines.
6. Trainees will learn about different business models used by open-access journals, and opportunities for obtaining funds to support publishing if needed.

Key components

Knowledge

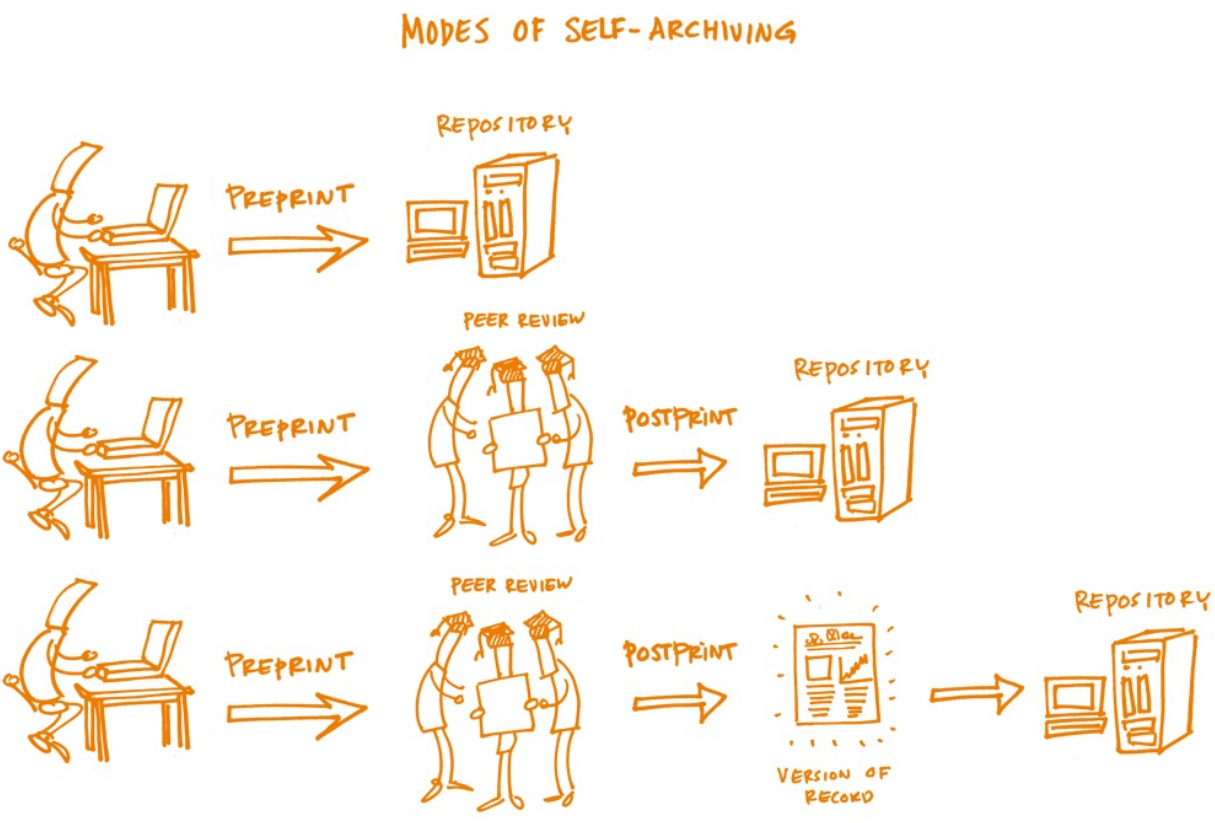
Repositories and self-archiving

At the beginning of 2018 more than 4600 repositories are available for researchers to self-archive their publications according to the Registry of Open Access Repositories. In this list we can find institutional repositories, subject based or thematic repositories and harvesters. The first ones are generally managed by research performing institutions to provide to their community a place to archive and share openly papers and other research outputs. Subject based repositories are usually managed by research communities and most of the contents are related to a certain discipline. Finally, harvesters aggregate content from different repositories becoming sites to perform general searches and build other value-added services. It is fundamental for a repository to be harvested to acquire more visibility. For that purpose, repository managers need to follow standard guidelines regarding the use of metadata and the values of these metadata. Moreover, institutional repositories can be linked with other information databases to increase discoverability, for example PubMed offers the possibility to link its registers by the linkout project. Repositories have always been seen as an alternative way to access to scientific publications when accessing to the original source is not affordable. Currently there are tools like the Unpaywall browser extension that facilitates this alternative.

When choosing a journal to publish research results, researchers should take a moment to read the journal policy regarding the transfer of copyright. Many journals still require for publication that authors transfer full copyright. This transfer of rights implies that authors must ask for permission to reuse their own work beyond what is allowed by the applicable law and unless there are some uses already granted. Among those granted uses we can find teaching purposes, sharing with colleagues, and especially how researchers can self-archive their papers in repositories. Sometimes there a common policy among all the journals published by the same publishers but in general journals have their own policy, especially when they are published on behalf of a scientific society. When looking at the self-archiving conditions we must identify two key issues: the version of the paper that can be deposited and when it can be publicly available.

Regarding the version, some journals allow the dissemination of the submitted version, also known as preprint, and they allow its replacement for a reviewed version once the final paper has been published. Due to the increase of policies requiring access to research results, most of the journals allow to deposit the accepted version of the paper, also known as the author manuscript or postprint. This version is the final text once the peer review process has ended but it has not the final layout of the publication. Finally some journals allow researchers to deposit the final published version, also known as the version of record.

In relation to the moment to make the paper publicly available, many journals establish a period of time from its original publication: the embargo period, which can range from zero to 60 months. Some journals include or exclude embargoes depending on the versions. For instance the accepted version could be made publicly available after publication but the published version must wait 12 months.



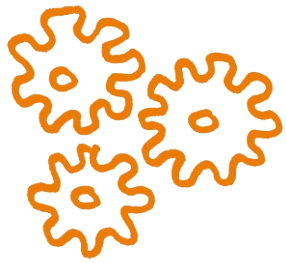
Open Access publishing

The number of Open Access Journals has increased during the last years becoming a real option for researchers when deciding where to publish their findings. According to the Directory of Open Access Journals (DOAJ), currently there are more than 11,000 journals. Nevertheless is important to remark that an open access journal must provide free access to its contents but it also must license them to allow reusability. No legal notice must be legally understood as "all rights reserved". Although the definition of an open access journal does not include any condition about the business model, there is a fact that those journals are commonly known as journal where you have to pay to publish. This misconception is due to the fact that the most successful journals and the ones that got the highest impact follow this model. Nevertheless, a recent study shows that the majority of journals registered in DOAJ do not charge any fee for publication (Data available here).

Currently many paywalled journals offer individual open access options to researchers once the paper is accepted after peer review. Those options include the publication under a free content license and free accessibility to anyone since its first publication. This model is commonly known as the hybrid model because in the same issue of a journal, readers can find open access and paywalled contributions. Usually publishers ask for a fee to open individual contributions. Recent studies show that the hybrid fees are higher than the average of the article processing charges in some pure open access journals (Jahn & Tullney 2016). One of the reasons researchers choose the hybrid model is to fulfil some of the requirements of funders policy, especially the ones requiring immediate public access to research results or short embargo periods.

Some funders, have decided to establish their own publishing platforms to provide their grantees with a place to release their findings. In general, to publish in those platforms costs around 1000 € and all the materials are released under a CC BY license. The publication is not limited to papers, researchers can include, for instance, data and software. There is no previous peer review process and therefore researchers publish documents that only pass through a limited editorial review to check the format but there is not an evaluation on the content. Peer review is done in a transparent way allowing anyone to see who wrote it and what the comments were. After the open peer review, authors can upload updated versions of their papers accordingly.

Some disciplines prefer to use other formats than journals to publish results, for instance books. Initially, publishers were very reluctant to allow researchers to self archive a full book or even a book chapter. However, some publishers have begun to adopt policies to facilitate it. On the other hand, some university presses have shifted their publication model to open access to increase the visibility of their contents, especially monographs. This change can be explained as an answer to the cuts in some of the expenditures in monographs due to the restrictions in library budgets. A common model for this open access university presses is to provide a free version in PDF and sell paper or epub versions (see for instance UCL). Moreover the creation of the Directory of Open Access Books have increased their discoverability. In a similar way than other journal initiatives, there have appeared some projects to join forces to establish a common fund to build open access monographs, for instance Knowledge Unlatched.



Skills

- Choose a suitable repository or server to post a preprint according to your discipline
- Self archive a publication in a suitable repository, institutional or subject-based, following the possible restrictions posed by the publisher, mainly related to the allowed version to be deposit and the embargo period
- Choose among the options of open-access journals and publishing platforms available
- Find available funds or discounts to publish in open-access journals if needed

Questions, obstacles, and common misconceptions

Q: "If I publish my work as a preprint, it won't be acknowledged - I will only receive credit for a peer-reviewed journal article."

A: Many funders are acknowledging the growing presence of preprint publishing in their policies: Howard Hughes Medical Institute (HHMI), Wellcome Trust, the Medical Research Council (UK) and the National Institutes of Health (NIH) announced policies allowing researchers to cite their own preprints in grant applications and reports (Luther 2017). In addition, preprints help establish priority of results and may increase the impact - and citation count - of a later peer-reviewed article (McKiernan 2016).

There are still some researchers reluctant to deposit other versions than the final published version. It is important to inform them about the copyright implications when they sign a transfer document.

Avoid the misconception of understanding an open-access journal as a journal where authors must pay to publish. The author-pay model is just one of the existing business models for an open-access journal. You might show data about the number of journals that do not ask for a publication fee (for example, as of 31 January 2018, DOAJ reports that 71% of the 11,001 open-access journals listed require no publishing charge). You may want to show other business models like the SCOAP3 Initiative, the LingOA project, or the Open Library of Humanities.

The use of publishing platforms has implications for research evaluation, the peer-review process, and the role of publishers. There are still many research assessments based on journal metrics and therefore this new way of publishing challenges those evaluations. Moreover the fact that peer review is completely transparent allows readers to identify reviewers and track the versioning of the paper. Finally, if those platforms become the common tool to publish results, publishers would need to redefine their role in the scholarly communication process.

The hybrid model is very controversial and it could raise a lot of questions about the costs, possible double-dipping, and the use (or lack) of licensing.

You may discuss the future of the scholarly communication by presenting some of the offsetting models or transition projects like OA2020 global alliance proposed by the Max Planck Society.

Learning outcomes

1. Trainees will be able to choose where to publish their research paper, describing the implications and consequences of this choice.
2. Trainees will be able to determine the self-archiving policy of a journal where they want to publish based on the information available at the corresponding website or at any of the portals that provide general information as Sherpa/Romeo, Dulcinea, and Heloïse.

3. Trainees who want to establish a new open-access journal will be able to describe their own self-archiving policy, license, and business model.
4. Trainees who manage repositories will be able to describe the tools and services that allow researchers to self-archive.

Further reading

- Björk (2017). Growth of hybrid open access, 2009–2016. *PeerJ* 5:e3878 doi.org/10.7717/peerj.3878
- Piwowar H, Priem J, Larivière V, Alperin JP, Matthias L, Norlander B, Farley A, West J, Haustein S. (2018) The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ* 6:e4375 https://doi.org/10.7717/peerj.4375
- The Open Access Directory. oad.simmons.edu/oadwiki

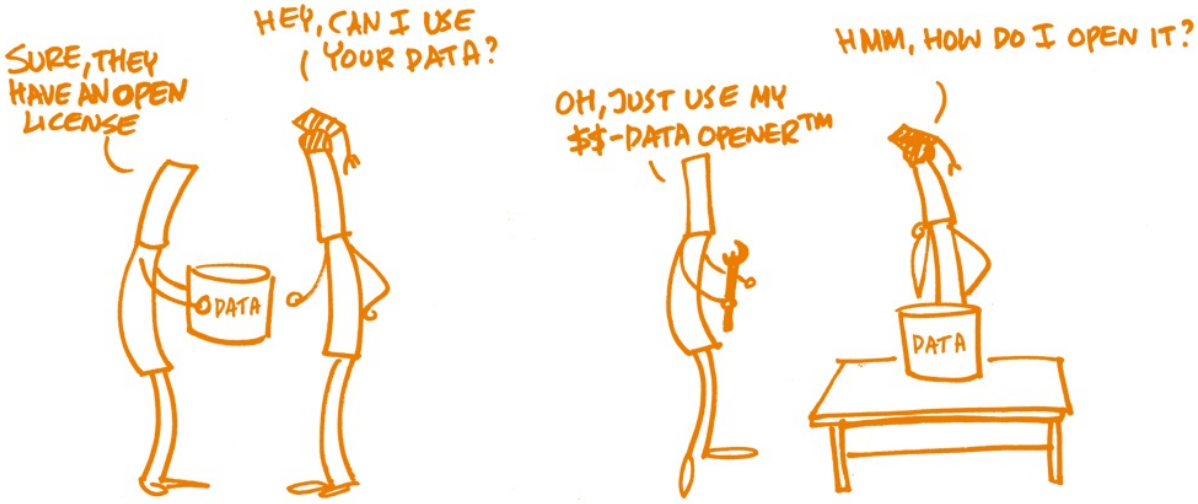
6. Open Licensing and File Formats

What is it?

A license is a legal document that grants specific rights to user to reuse and redistribute a material under some conditions. Any right that is not granted by default by the licensor through the license can be asked. Licenses can be applied to any material (e.g., sound, text, image, multimedia, software) where some exploitation or usage rights exist.

Free content licenses are licenses that grant permission to access, re-use, and redistribute material with few or no restrictions. Those licenses range from very open to very restrictive. The more restrictions, the more difficult it becomes to combine differently licenses content—thus potentially preventing interoperability.

A file format is a standard way that information is encoded for storage in a computer file; however, not all formats have freely available specification documents, partly because some developers view their specification documents as trade secrets.

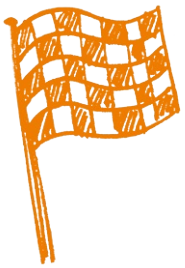


Rationale

Applying an open license to a scientific work (whether it is an article, dataset or other type of research output) is a way for the copyright holder to express the conditions under which the work can be accessed, re-used and modified.

It is important to know that a license builds on existing copyright regulations. In other words: you can only license content if you are the rights owner, and you cannot license any forms of reuse if they do not fall under existing copyright regulations.

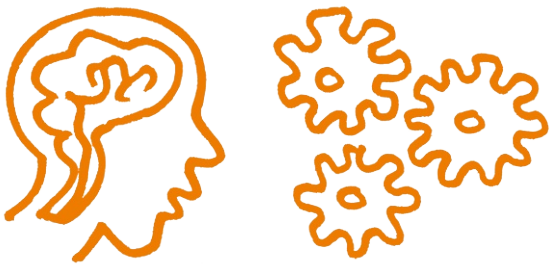
When sharing any open content it is not enough to attach a license you must take into account the format. A choice of a non-open file format may make impossible to reuse the content. For that reason is important to know the options available when deciding in which format you want to share your content.



Learning objectives

1. Participants should learn about differences among licenses and how they can suit some open-science definitions, open-science requirements, or how they fit into different research outcomes.
2. Learn about the different building blocks of licenses, such as attribution, (non-)commercial, derivatives, etc.
3. Learn the importance of defining who holds the copyright or related rights of research output.
4. Learn about the differences between proprietary and open file formats, and how these can prevent or facilitate reusability and interoperability.

Key components



Knowledge & Skills

Basic concepts of copyright are needed in order to understand how the licenses work. Since copyright laws are not internationally harmonized you must refer to the applicable laws in your context.

Among the range of free content licenses there are the copyleft licenses, originated in the free software community, that allow a broad reuse of materials under the condition that any new material build upon the existing one must be licensed under the same license. This fact has brought some interoperable problems that newer versions overcomed by stating that the derived materials should be licensed under the same terms of the original license.

The most used licenses for scientific content are Creative Commons licenses. In general, a CC BY license (requiring only attribution) is a good option for works such articles, books, working papers, and reports while a dedication to the public domain using CC Zero (CC0) is recommended for datasets and databases (NOTE: In the US and EU, individual facts cannot be copyrighted, although collections of facts that underwent some creative selection or organization may be copyrighted. Additionally, in the EU there is a sui generis right granted to the maker of a database for the investment made in its compilation, even when this does not involve any creativity.). Creative Commons licenses should not be used for licensing software because they were not designed for that purpose, as the organisation states. Instead, software developers should use appropriate licenses like those collected by the Open Source Initiative or Free Software Foundation. You can check your options at choosealicense.

CC0 was originally created as a legal tool to release scientific databases without any restriction, and especially to overcome the different treatments of legal protection when publishing a database. CC0 has been seen as a tool for dedicating works to the public domain but it is more than a simple waiver. CC0 is a three-step instrument built to allow its use in jurisdictions where a full public domain dedication is not possible (for instance in many continental Europe countries). First, by using CC0, the copyright holder waives any right to the maximum extent allowed by applicable law. Second, if there is any remaining unwaivable right, CC0 acts as a license to grant any of

those remaining rights without any restriction or obligation. And finally, the copyright holder asserts not to enforce any right that could not been possible to waive or grant by the applicable law. The idea behind CC0 is to convince researchers to follow community norms instead of using licenses in materials as a database where, in many cases, its contents are uncopyrightable.

As a trainer, you may show the differences among licenses and how they can suit some of the Open Science definitions, the Open Science requirements or how they fit into different research outcomes. Depending on the prior knowledge of your audience, you can give an overview of the different building blocks (attribution, (non)commercial, derivatives, etc.) of the licenses in general or provide a detailed analysis of each building block and its effects on re-use and interoperability. As copyright rules vary greatly per jurisdiction (common law vs. civil law countries, but also within the European Union), usability of licenses can vary greatly. This can be discussed in detail if the audience has previous knowledge about licensing, but if they are relatively new to the subject this should not be discussed in detail.

Core licensing items to consider (from the Data Packaging Guide):

- Choosing an open license.
- Stating the chosen license clearly and prominently, preferably in a machine readable format.
- Explain the liberations/limitations of the chosen license, and what barriers or restrictions may apply.
- Let users know where they can find more information about this license.
- Explain that the license applies to the data, and not the content that the data represents (an open license on the metadata is not the same as the content itself being open, out of copyright, or able to be used freely).
- Explain why this license was chosen.

Training should provide an overview of intellectual property policies in universities and public research institutions. It is important to stress the need to define who holds copyright or any other related rights of the research output. The copyright holder is the one who can decide to lift restrictions if they are not lifted by default through the licenses. Regarding research outputs, the copyright holder can be a researcher, a publisher, a scientific society, a research institution, a funder, etc.

Within the context of Open Science, and for optimal long-term archiving, files should not be compressed and should avoid proprietary or patent-encumbered formats and in favor of open formats based on documented standards. This ensures the access and re-usability of the content. Only unencrypted files should be published and archived. Examples for open file formats are:

- Text: TXT, ODT, PDF/A, XML
- Tabular data: CSV, TSV
- Image: TIFF, PNG, JPG 2000, SVG, WebP
- Audio: WAV, FLAC, OPUS
- Video: MPEG2, Theora, VP8, VP9, AV1, Motion JPG 2000 (MJ2),
- Binary hierarchical data: HDF5

Some file formats cannot be converted to open formats, but are nonetheless archived. They are often device-specific, but have a broad user community. Check if the repository where you want to deposit a publication has a list of preferred formats.

Questions, obstacles, and common misconceptions

Q: "Why should I use the CC-BY license for my written/creative content?"

A: The CC-BY license is the most permissive license that also retains some rights for the creators—the only requirement is that someone who uses, modifies, or distributes the content attributes the original creator. Other attributes of Creative Commons licenses include No Derivatives (ND), Non Commercial (NC), and Share Alike (SA), which add additional restrictions that may limit the potential use and impact of your work. Preventing derivatives with ND strongly limits the impact and use of your work, since no one else will be able to build on what you have done. Similarly, while many researchers may prefer the NC limitation to prevent companies from commercializing or making money on their work, strictly defining commercial use is challenging. Furthermore, the intent of much publicly funded research is to lead to economic development through (ventual) commercial use, which would be prevented by this license. Using an SA license allows reuse and distribution, but requires downstream works to apply the same license, limiting use and combination with other works.

A common fear when using CC0 is that the attribution requirement is dropped—however, proponents state that attribution is a key element in good scientific practice, regardless of copyright status of license conditions of the quoted work. Some repositories applying CC0 explicitly mention attribution, cf., e.g., this example from Dataverse: "Our Community Norms as well as good scientific practices expect that proper credit is given via citation. Please use the data citation above, generated by the Dataverse."

Obstacle: different countries have different copyright laws, which may limit the ability to choose any license or dedicate work to the public domain. For example, in Germany and other European countries it is not possible to fully waive copyright, and thus fully dedicating work to the public domain is not legally possible. Instead, the CC0 license can be used as an "effective" public domain license that allows unrestricted use.

Interoperability of licenses: be aware that sometimes when you mix content licensed differently it may be impossible to release the derivative work. For example, material distributed with an SA license can only be combined with other SA-licensed content.

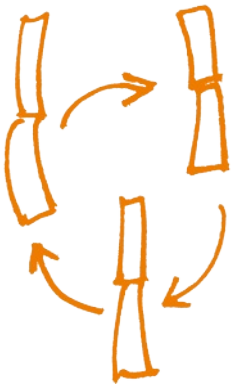
Suitability of licenses: for instance, CC licenses should not be used for software, there are specific licenses for databases (Open Data Commons), and CC licenses were not suitable for databases before version 4.0.

Learning outcomes

1. Will be able to use existing resources to choose an appropriate license for written research work, based on the desired freedom/limitation for others to use/reuse.
2. Will be able to use existing resources to choose an appropriate license for data, based on the desired freedom/limitation for others to use/reuse.

Further reading

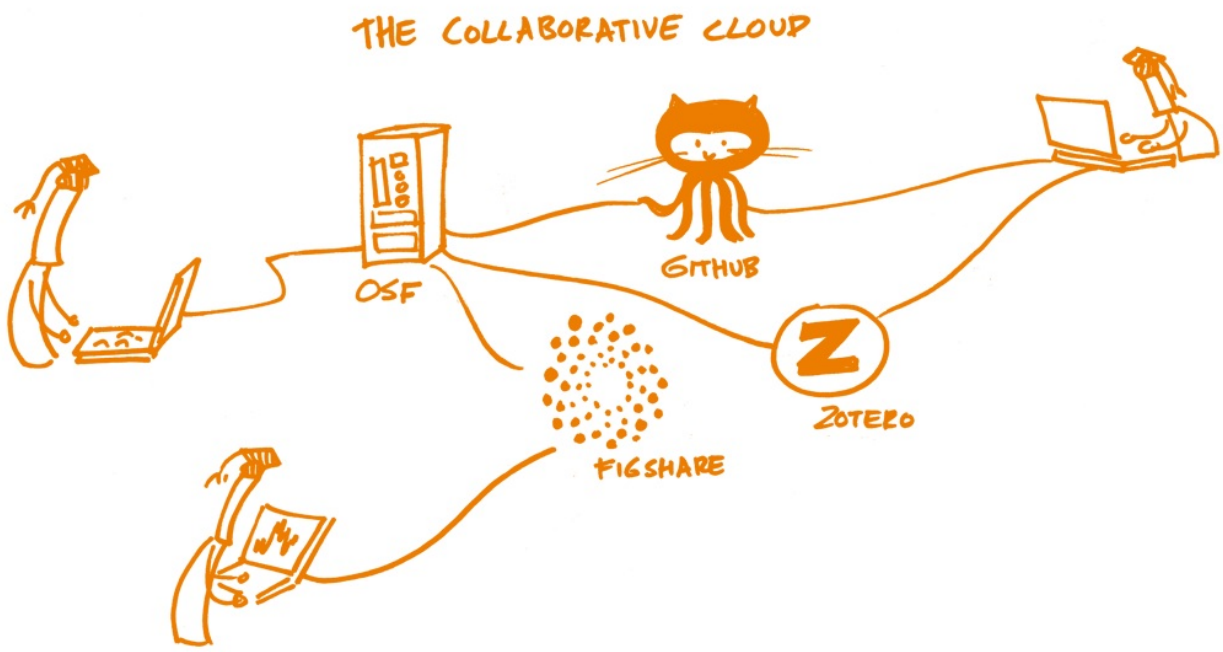
- Creative Commons License Picker. creativecommons.org
- How to License Research Data. dcc.ac.uk
- Klimpe (2012). Free knowledge thanks to Creative Commons Licenses - Why a non-commercial clause often won‘t serve your needs. Original PDF in German, English translation PDF
- Kreutzer (n.y.). Validity of the Creative Commons Zero 1.0 Universal Public Domain Dedication and its usability for bibliographic metadata from the perspective of German Copyright Law. PDF
- List of open formats. Wikipedia
- Open Content - A Practical Guide to Using Creative Commons Licences/The Creative Commons licencing scheme. meta.wikimedia.org
- Open Definition. Licenses. opendefinition.org
- Open Source Licensing. opensource.org/licenses
- Redhead (2012). Why CC-BY?. Open Access Scholarly Publishers Association. oaspa.org/why-cc-by
- World Intellectual Property Organization. Universitites and Intellectual Property. wipo.int



7. Collaborative Platforms

What is it?

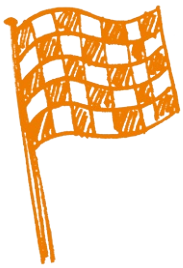
Online collaborative platforms connect geographically-dispersed researchers to enable them to cooperate seamlessly on their research, sharing research objects as well and ideas and experiences. Collaborative platforms are usually online services that provide a virtual environment to which multiple people can concurrently connect and work on the same task. These can range from extensive virtual research environments (VREs) which encompass a host of tools to facilitate sharing and collaboration, including web forums and wikis, collaborative document hosting, and discipline-specific tools such as data analysis or visualisation, right down to single specific tools which enable researchers to work together in real time on specific aspects of research (such as writing or analysis).



Rationale

Research collaboration is growing exponentially and teams are becoming ever more interdisciplinary as researchers increasingly work in international and cross-disciplinary consortia to enable a multitude of perspectives on specific research questions. Fostering national and international collaborative research is increasingly a funder priority. It lies, for example, at the heart of EC Research Commissioner Carlos Moedas' strategy, i.e., "Open Science, open innovation, open to the world".

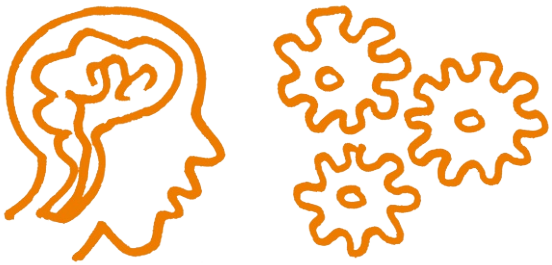
Virtual Research Environments (VRE) and collaborative platforms enable collaboration across continents, time zones and disciplines. In this module you will develop an understanding of collaborative platforms that work today, and how they can greatly enhance your research workflows.



Learning objectives

1. Learn what major types of collaborative platforms are available and what the use cases for each might be.
2. Learn the advantages of such systems.
3. Identify any possible shortcomings of collaborating via such platforms and how to overcome them.

Key components



Knowledge & Skills

Virtual research environments (VREs)

Virtual research environments have been defined as "innovative, dynamic, and ubiquitous research supporting environments where scattered scientists can seamlessly access data, software, and processing resources managed by diverse systems in separate administration domains through their browser" (Candela, Castelli and Pagano, 2013).

An important aspect here is the disciplinary-specific nature of many of these tools. The European Commission has funded a range of community-specific VREs under its eInfrastructure funding stream to enable researchers to collaboratively perform complex tasks such as integrating heterogeneous data from multiple sources, modelling, simulation, data exploration, mining and visualisation:

- VI-SEEM - VRE for regional Interdisciplinary communities in Southeast Europe and the Eastern Mediterranean
- MuG - Multi-Scale Complex Genomics
- OpenDreamKit - Open Digital Research Environment Toolkit for the Advancement of Mathematics
- BlueBRIDGE - Building Research environments for fostering Innovation, Decision making, Governance and Education to support Blue growth
- VRE4EIC - A Europe-wide Interoperable Virtual Research Environment to Empower Multidisciplinary Research Communities and Accelerate Innovation and Collaboration
- West-Life - World-wide E-infrastructure for structural biology

Some libraries already offer personalised VREs for specific projects. For example, Leiden University library offers VREs for all externally-funded projects of more than five persons.

An especially important collaborative platform in the context of Open Science is the Open Science Framework (OSF). Based on open source technologies and created by the not-for-profit Center for Open Science, the OSF brands itself as "a scholarly commons to connect the entire research cycle". The OSF enables researchers to work on projects privately with a limited number of collaborators and make any part or the whole of their project public. It connects directly with many other collaborative systems like dropbox, GitHub and Google Docs, and can be used to store and archive research data, protocols, and materials.

Collaborative writing platforms

Especially in the currently-predominant "publish or perish" culture of research, writing is a core task in the life of researchers. Several online tools and platforms now enable researchers to work together on documents in real-time, and so avoid the versioning-hell of emailing Word documents back and forth. Platforms include Overleaf, Authorea, Fidus Writer, ShareLaTeX and Google Docs. Note that many of these tools are based on proprietary technologies and some require payment for advanced features.

Reference management & discovery

There are plenty of tools which enable groups to store and manage references. Examples include Zotero, Citavi and CiteUlike. Mendeley incorporates a sharable reference manager, as well as a social network and article visualization tools. Relatedly, BibSonomy allows researchers to share bookmarks and lists of literature.

Annotation and review

The power of the Web enables new modes of post-publication collaborative review through services like PubPeer and Academic Karma, as well as annotation tools like Hypothes.is and PaperHive.

Academic social networks

Researchers have long made use of the Web for social networking - either via mainstream social networks like Twitter, Facebook and Linkedin or dedicated academic social networks like ResearchGate, Academia.edu and Loop.

Questions, obstacles, and common misconceptions

Q: "Why should I add another layer of complexity to my collaboration process? Sharing the doc file is sufficient!"

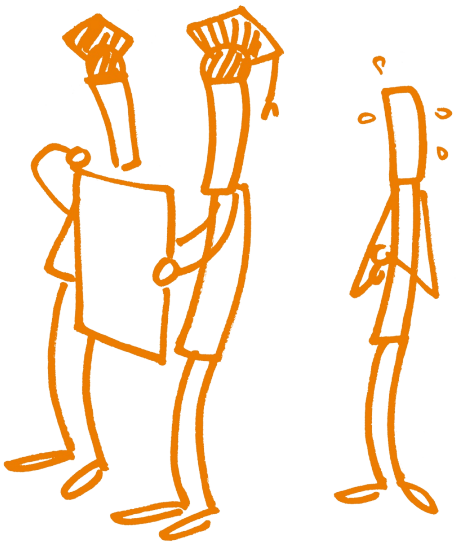
A: This is incorrect; although it may seem that you are introducing additional tools and platforms into your usual working approach, they are actually resolving communication issues that you were probably not aware of in the first place. For example, using just a doc file (with or without track changes), only shows the higher level of information and usually only at the tail of the entire scientific process. Working in the context of a collaborative environment, from design to reporting, establishes both clear communication and adequate provenance.

Learning outcomes

1. The researcher will become familiar with the range of options available to aid greater collaborative research.
2. After deciding what works optimally for their workflow, the researcher will be able to use collaborative tools such as GitHub and the Open Science Framework for increased collaboration for the research process, writing/authoring, and sharing your research outputs.
3. The researcher will be able to collaborate with colleagues to write documents collaboratively, annotate articles and share this discussion.

Further reading

- Candela et al. (2013). Virtual Research Environments: An Overview and a Research Agenda. Data Science Journal. 12, pp.GRDI75–GRDI81. doi.org/10.2481/dsj.GRDI-013
- Open Science Framework. The promise of Open Science collaboration. osf.io
- Voss and Procter (2009). Virtual research environments in scholarly work and communications, Library Hi Tech, Vol. 27 Issue: 2, pp.174-190. doi.org/10.1108/07378830910968146



8. Open Peer Review, Metrics, and Evaluation

What is it?

To be a researcher is to find oneself under constant evaluation. Academia is a "prestige economy", where an academic's worth is based on evaluations of the levels of esteem within which they and their contributions are held by their peers, decision-makers and others (Blackmore and Kandiko, 2011). In this section it will therefore be worthwhile distinguishing between evaluation of a piece of work and evaluation of the researcher themselves. Both research and researcher find themselves evaluated through two primary methods: peer review and metrics, the first qualitative and the latter quantitative.

Peer review is used primarily to judge pieces of research. It is the formal quality assurance mechanism whereby scholarly manuscripts (e.g., journal articles, books, grant applications and conference papers) are made subject to the scrutiny of others, whose feedback and judgements are then used to improve works and make final decisions regarding selection (for publication, grant allocation or speaking time). Open Peer Review means different things to different people and communities and has been defined as "an umbrella term for a number of overlapping ways that peer review models can be adapted in line with the aims of Open Science" (Ross-Hellauer, 2017). Its two main traits are “open identities”, where both authors and reviewers are aware of each other’s identities (i.e., non-blinded), and “open reports”, where review reports are published alongside the relevant article. These traits can be combined, but need not be, and may be complemented by other innovations, such as “open participation”, where members of the wider community are able to contribute to the review process, “open interaction”, where direct reciprocal discussion between author(s) and reviewers, and/or between reviewers, is allowed and encouraged, and “open pre-review manuscripts”, where manuscripts are made immediately available in advance of any formal peer review procedures (either internally as part of journal workflows or externally via preprint servers).

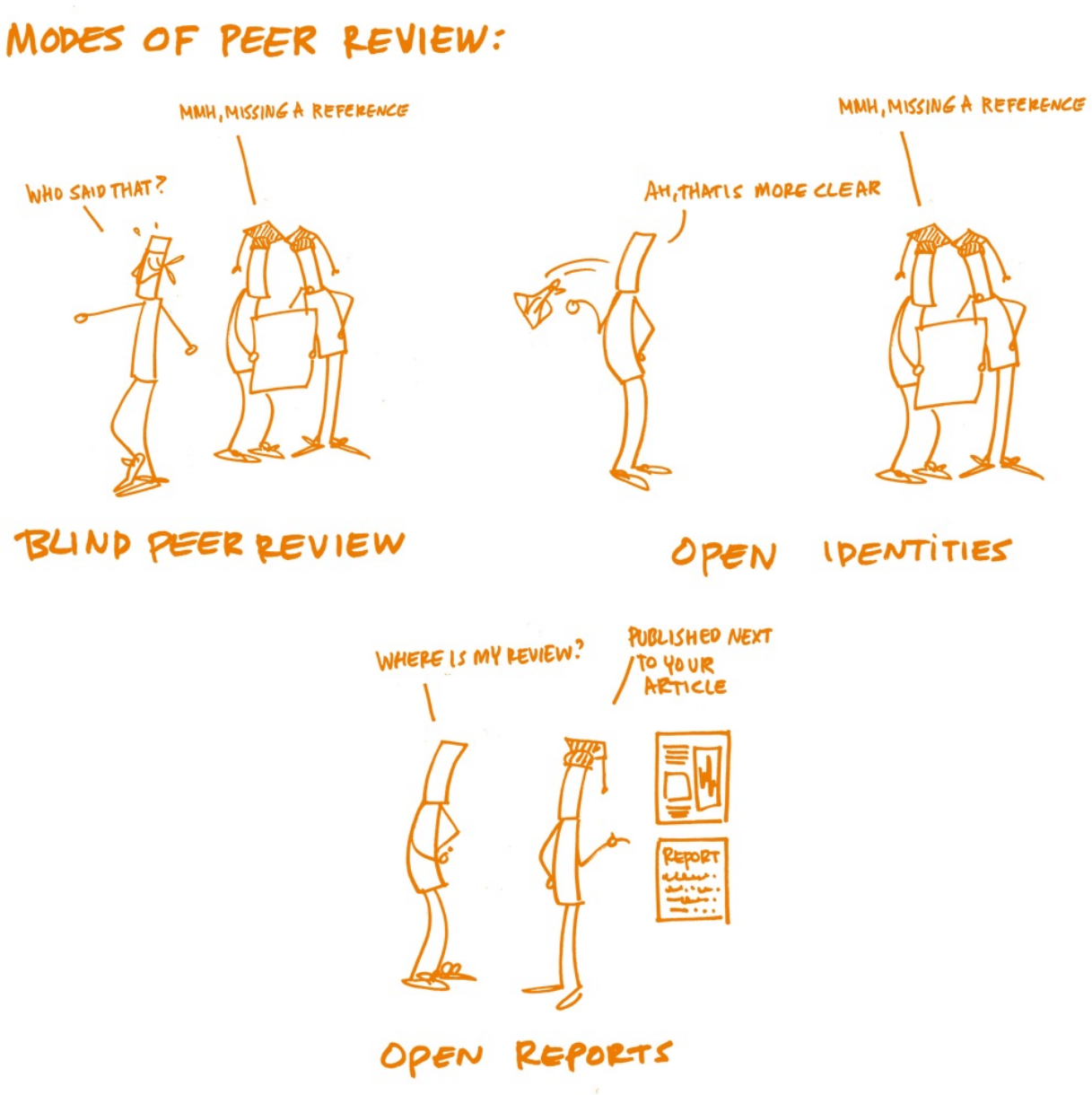
Once they have passed peer review, research publications are then often the primary measure of a researcher's work (hence the phrase "publish or perish"). However, assessing the quality of publications is difficult and subjective. Although some general assessment exercises like the UK's Research Excellence Framework use peer review, general assessment is often based on **metrics** such as the number of citations publications garner (h-index), or even the perceived level of prestige of the journal it was published in (quantified by the Journal Impact Factor). The predominance of such metrics and the way they might distort incentives has been emphasised in recent years through statements like the Leiden Manifesto and the San Francisco Declaration on Research Assessment (DORA).

In recent years “Alternative Metrics” or altmetrics have become a topic in the debate about a balanced assessment of research efforts that complement citation counting by gauging other online measures of research impact, including bookmarks, links, blog posts, tweets, likes, shares, press coverage and the like. Underlying all of these issues with metrics is that they are very produced by commercial entities (e.g., Clarivate Analytics and Elsevier) based on proprietary systems, which can lead to some issues with transparency.

Rationale

Open peer review

Beginning in the 17th century with the Royal Society of London (1662) and the Académie Royale des Sciences de Paris (1699) as the privilege of science to censor itself rather than through the church, it took many years for peer review to be properly established in science. Peer review, as a formal mechanism, is much younger than many assume. For example, the journal Nature only introduced it in 1967. Although surveys show that researchers value peer review they also think it could work better. There are often complaints that peer review takes too long, that it is inconsistent and often fails to detect errors, and that anonymity shields biases. Open peer review (OPR) hence aims to bring greater transparency and participation to formal and informal peer review processes. Being a peer reviewer presents researchers with opportunities for engaging with novel research, building academic networks and expertise, and refining their own writing skills. It is a crucial element of quality control for academic work. Yet, in general, researchers do not often receive formal training in how to do peer review. Even where researchers believe themselves confident with traditional peer review, however, the many forms of open peer review present new challenges and opportunities. As OPR covers such a diverse range of practices, there are many considerations for reviewers and authors to take into account.

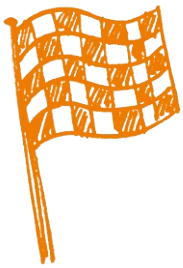


Regarding evaluation, current rewards and metrics in science and scholarship are not (yet) in line with Open Science. The metrics used to evaluate research (e.g. Journal Impact Factor, h-index) do not measure - and therefore do not reward - open research practices. Open peer review activity is not necessarily recognized as "scholarship" in professional advancement scenarios (e.g. in many cases, grant reviewers don’t consider even the most brilliant open peer reviews to be scholarly objects unto themselves). Furthermore, many evaluation metrics - especially certain types of bibliometrics - are not as open and transparent as the community would like.

Under those circumstances, at best Open Science practices are seen as an additional burden without rewards. At worst, they are seen as actively damaging chances of future funding and promotion as well as tenure. A recent report from the European Commission (2017) recognizes that there are basically two approaches to Open Science implementation and the way rewards and evaluation can support that:

1. Simply support the status quo by encouraging more openness, building related metrics and quantifying outputs;
2. Experiment with alternative research practices and assessment, open data, citizen science and open education.

More and more funders and institutions are taking steps in these directions, for example by moving away from simple counts, and including narratives and indications of societal impact in their assessment exercises. Other steps funders are taking are allowing more types of research output (such as preprints) in applications and funding different types of research (such as replication studies).



Learning objectives

1. Recognise the key elements of open peer review and their potential advantages and disadvantages
2. Understand the differences between types of metrics used to assess research and researchers
3. Engage with the debate over the way in which evaluation schema affect the ways in which scholarship is performed

Key components

Knowledge

Open peer review

Popular venues for OPR include journals from publishers like Copernicus, Frontiers, BioMed Central, eLife and F1000research.

Open peer review, in its different forms, has many potential advantages for reviewers and authors:

- Open identities (non-blinded) review fosters greater accountability amongst reviewers and reduces the opportunities for bias or undisclosed conflicts of interest.
- Open review reports add another layer of quality assurance, allowing the wider community to scrutinize reviews to examine decision-making processes.
- In combination, open identities and open reports are theorized to lead to better reviews, as the thought of having their name publicly connected to a work or seeing their review published encourages reviewers to be more thorough.
- Open identities and open reports enable reviewers to gain public credit for their review work, thus incentivising this vital activity and allowing review work to be cited in other publications and in career development activities linked to promotion and tenure.
- Open participation could overcome problems associated with editorial selection of reviewers (e.g., biases, closed-networks, elitism). Especially for early career researchers who do not yet receive invitations to review, such open processes may also present a chance to build their research reputation and practice their review skills.

There are some potential pitfalls to watch out for, including:

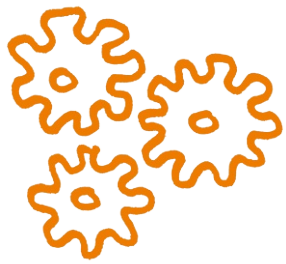
- Open identities removes anonymity conditions for reviewers (single-blind) or authors and reviewers (double-blind) which are traditionally in place to counteract social biases (although there is not strong-evidence that such anonymity has been effective). It's therefore important for reviewers to constantly question their assumptions to ensure their judgements reflect only the quality of the manuscript, and not the status, history, or affiliations of the author(s). Authors should do the same in receiving peer review comments.

- Giving and receiving criticism is often a process fraught with unavoidably emotional reactions - authors and reviewers may subjectively agree or disagree on how to present the results and/or what needs improvement, amendment or correction. In open identities and/or open reports, the transparency could exacerbate such difficulties. It is therefore essential that reviewers ensure that they communicate their points in a clear and civil way, in order to maximise the chances that it will be received as valuable feedback by the author(s).
- Lack of anonymity for reviewers in open identities review might subvert the process by discouraging reviewers from making strong criticisms, especially against higher-status colleagues.
- Finally, given these issues, potential reviewers may be more likely to decline to review.

Open metrics

The San Francisco Declaration on Research Assessment (DORA) recommends moving away from journal based evaluations, consider all types of output and use various forms of metrics and narrative assessment in parallel. DORA has been signed by thousands of researchers, institutions, publishers and funders, who have now committed themselves to putting this in practice. The Leiden Manifesto provides guidance on how to use metrics responsibly.

Regarding Altmetrics, Priem et al. (2010) advise that altmetrics have the following benefits: they accumulate quicker than citations; they can gauge the impact of research outputs other than journal publications (e.g. datasets, code, protocols, blog posts, tweets, etc.); and they can provide diverse measures of impact for individual objects. The timeliness of altmetrics presents a particular advantage to early-career researchers, whose research-impact may not yet be reflected in significant numbers of citations, yet whose career-progression depends upon positive evaluations. In addition, altmetrics can help with early identification of influential research and potential connections between researchers. A recent report by the EC's Expert Group on Altmetrics (Wilsdon et al. (European Commission), 2017) identified challenges of altmetrics, including lack of robustness and susceptibility to 'gaming'; that any measure ceases to be a good measure once it becomes a target ('Goodhart's Law'); relative lack of social media uptake in some disciplines and geographical regions; and a reliance on commercial entities for the underlying data.



Skills

Example exercises

- Trainees work in groups of three. Each individually writes a review of a short academic text
- Review a paper on a pre-print server
- Use a free bibliometrics or altmetrics service (e.g. Impactstory, Paperbuzz, Altmetric bookmarklet, Dimensions.ai) to look up metrics for a paper, then write a short explanation of how exactly various metrics reported by each service are calculated (it's harder than you'd assume; this would get at the challenges of finding proper metrics documentation for even the seemingly most transparent services)

Questions, obstacles, and common misconceptions

Q: Is research evaluation fair?

A: Research evaluation is as fair as its methods and evaluation techniques. Metrics and altmetrics try to measure research quality with research output quantity, which can be accurate, but does not have to be.

Learning outcomes

1. Trainees will be able to identify open peer review journals
2. Trainees will be aware of a range of metrics, their advantages and disadvantages

Further reading

- Directorate-General for Research and Innovation (European Commission) (2017). Evaluation of Research Careers Fully Acknowledging Open Science Practices: Rewards, Incentives and/or Recognition for Researchers Practicing Open Science. doi.org/10.2777/75255
- Hicks et al. (2015) Bibliometrics: The Leiden Manifesto for research metrics. doi.org/10.1038/520429a, leidenmanifesto.org
- Peer Review the Nuts and Bolts (2012). A Guide for Early Career Researchers. PDF

Projects and initiatives

- Make Data Count. makedatacount.org
- NISO Alternative Assessment Metrics (Altmetrics) Initiative. niso.org
- Open Rev. openrev.org

- OpenUP Hub. openuphub.eu
- Peer Reviewers' Openness Initiative. opennessinitiative.org
- Peerage of Science. A free service for scientific peer review and publishing. peerageofscience.org
- Responsible Metrics. responsiblemetrics.org
- Snowball Metrics. Standardized research metrics - by the sector for the sector. snowballmetrics.com

9. Open Science Policies

What is it?

We could define Open Science policies as those strategies and actions aimed at promoting Open Science principles and at acknowledging Open Science practices. Those policies are usually established by research performing institutions, research funders, governments or publishers. The initial policies were aimed at requiring an open dissemination of the research results based on the idea that results achieved from publicly funded research should be available to the public without any restriction. However, now the scope of the policies has grown and we may find national policies fostering Open Science practices at any point of the research level. Moreover, we might find specific provisions in new and existing laws, regulations or directives.

Rationale

Since one of the main drivers to Open Science are the current policies established by institutions, funders, governments and publishers, it is important to know how they affect any researcher. If you are planning to design a policy aimed at the adoption and acknowledgement of Open Science practices is important to know the existing policies in order to avoid any overlapping or contradiction. Therefore researchers and policy makers should have a knowledge of the current policies and should be able to understand how they affect them.

Learning objectives

1. Depending on your audience the objectives of the training session would be different. We can make a broad division between researchers (in a broad sense) and policy makers (within an institution or funders - in a broad sense).
2. If your training program is addressed mainly to researchers, including all "levels", then the main objective is to review how Open Science policies affect them.
3. If your training program is addressed to policy makers, you might focus in designing and implement a policy to foster Open Science.
4. If we want to train funders or policy makers within an institution then it should be important to show how to design, develop, implement and monitor a policy

Key components

Knowledge

You must review all the policies that affect your training audience. First of all check all the institutional policies at institutional level, for instance; copyright, intellectual property, open access, research data.

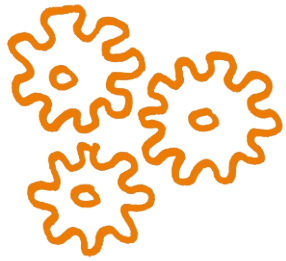
Secondly you may review any national policy or law that can affect researchers when performing Open Science, for instance laws with open access provisions or decrees affecting PhD dissertations , Call for projects.

At the national level there could be some laws or decrees that directly or indirectly could influence a policy or pose some requirements. For instance you could review the national open access policies in Europe available at OpenAIRE.

Since science is international, then we should review any international policy that could affect your audience, mainly coming from international funders. At the European level we have the policies coming from the H2020 research Framework regarding the dissemination of research outputs but we could have other policies affecting other parts of the research cycle.

Also at the international level, some publishers have introduced new policies, especially regarding the publication of research data when submitting a paper.

If your training audience is willing to develop a roadmap or agenda to implement a national Open Science policy it could be advisable to benchmark what has been done in other parts. As a starting point, the 2016 Amsterdam Call for Action could show some of the issues that must be taken into account and to whom are addressed. Examples of the Netherlands, Portugal, or Finland can help to plan national policies, outline some actions and find how to measure their implementation.



Skills

Trainees would need to identify the main features of each policy mainly: to whom is addressed, what are the requirements, how they overlap with each other.

You might show how researchers can fulfill with the different policies: where are the services, the tools that the institution can provide but also where they can find alternatives. For instance, an institution might not provide an infrastructure for depositing and publishing research data; but it can point out external solutions that fulfill policy requirements. It is also useful to compare those solutions with other external options with not desired features.

When designing an Open Science policy, trainees would need to be able to define the main purposes of having such a policy and to establish the goals or changes they are pursuing. Once defined, they must be able to find key performance indicators to measure if the policy have achieved its goals and they must be able to review and update the policy if the goals are not achieved.

Questions, obstacles, and common misconceptions

The main question coming from researchers in training sessions on policies is how they can fulfill the requirements without losing any freedom on deciding where to publish, for instance. You as a trainer, may describe all the available options researchers have because in general, Open Science policies provide a range of options.

Another question often raised is what happens if researchers don’t fulfill the requirements. In this case you may give examples of projects monitored by funders or warnings received by researchers.

A common misconception regarding research data policy is that researchers should share all data openly. To overcome it, you must highlight the different excerpts in the text of a policy where there are explanations about which is the data affected by the policy and when it must be shared. We might also remark all the opt-out choices that policies include. A good resource to clarify those issues can be an infographic like the one available from Horizon 2020.

When planning a policy is important to know what do you intend to achieve or solve. Sometimes policies are created following other initiatives without thinking if there is a need for another one and if your new policy will overlap other existing ones. The main challenge when creating a policy is to align it with other initiatives and to avoid contradictions with laws or regulations.

Learning outcomes

1. Trainees would be able to identify the requirements of any policy that could affect them when performing Open Science.
2. They would be able to distinguish among general policies like copyright or data protection and specific policies related to Open Science, for instance regarding how to disseminate research outputs.
3. They would be able to outline the steps to fulfill a certain policy.
4. Trainees attending a session aimed at policy making would be able to plan an Open Science policy, establishing objectives and indicators to measure its implementation.

Further Reading

- EC Working Group on Education and Skills under Open Science (2017). Providing researchers with the skills and competencies they need to practise Open Science. ec.europa.eu
- Open Research Funders Group & SPARC. Open Policies 101. PDF from orfg.org
- Model Policy for Research Data Management (RDM) at Research Institutions/Institutes. In: Leaders Activating Research Networks (LEARN) (ed.) LEARN Toolkit of Best Practice for Research Data Management. (pp. 133-136). learn-rdm.eu
- Guidance for Developing a Research Data Management (RDM) Policy. In: Leaders Activating Research Networks, LEARN Project (ed.) LEARN Toolkit of Best Practice for Research Data Management. (pp. 137-140). learn-rdm.eu

Projects and initiatives

- FOSTER. Designing Successful Open Access and Open Data Policies: Introductory. fosteropenscience.eu
- FOSTER. Designing Successful Open Access and Open Data Policies: Intermediate. fosteropenscience.eu
- LEARN Project 2015-2017. Toolkit of Best Practice for Research Data Management learn-rdm.eu
- Pasteur4OA. pasteur4oa.eu

10. Citizen Science

What is it?

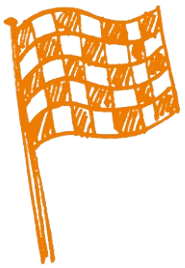
Citizen Science is the involvement of the non-academic public in the process of scientific research – whether community-driven research or global investigations (citizenscience.org). Citizens do scientific work—often working together with experts or scientific institutions. They support the collection, analysis or description of research data and make a valuable contribution to science. The first documented Citizen Science project took place at Christmas in 1900 in the USA, when the National Audubon Society carried out a Christmas Bird Count. "Galaxy Zoo" with over 150,000 participants who classified galaxies in one year is probably the so far most successful Citizen Science project.

Citizen science is essentially a direct product of successful science communication or public engagement. In the age of digital networked technologies, researchers have a wealth of channels through which to disseminate their work to wider non-academic audiences. Whereas research has been traditionally disseminated narrowly via conference papers, research articles and book publications, researchers now can use blogs, social media, video-hosting sites, and a wide range of social digital networks to target and broaden their dissemination activities.

Rationale

Citizen science is both an aim and enabler of Open Science. It can refer to citizens actively and openly participating in the research process itself, often through crowdsourcing activities.This includes aspects such as data collection, data analysis, volunteer monitoring, and distributed computing. Alternatively, it can also mean greater public understanding of science facilitated through greater access to information about the research process, including the ability to use open research data and to access to journal articles openly available. The latter (aka Do-It-Yourself Science) involves examples such as patient innovation, patient activism/advocacy, NGOs and Civil Rights Groups. This leads to a clearer classification by distinguishing scientist and non-scientist led activities (see Outside the Academy – DIY Science Communities). The public can also be engaged in policy making through, for example, agenda-setting for research systems' (see the European Commission's Open Science Monitor).

"*Citizen Science and Open Science together can address grand challenges, respond to diminishing societal trust in science, contribute to the creation of common goods and shared resources, and facilitate knowledge transfer between science and society to stimulate innovation. The issues of openness, inclusion and empowerment, education and training, funding, infrastructures and reward systems are discussed regarding critical challenges for both approaches. You might consider Citizen Science and Open Science jointly, to strengthen synergies by building on existing initiatives, launching targeted actions regarding education and training, and infrastructures*". Extracted from the Policy Brief on Citizen Science and Open Science by the European Citizen Science Association (ECSA)



Learning objectives

1. Understand the different aspects of citizen science (collaborative versus DIY).
2. Understand the basic concepts and viewpoints of a variety of stakeholders in science communication.
3. Management of intellectual property in citizen science projects. A guide for this is available here.
4. Management of citizen science data.
5. Identify the best strategies in establishing clear and concise communication of scientific principles.
6. What are the best ways to communicate your research/story, with whom, and using what tools.

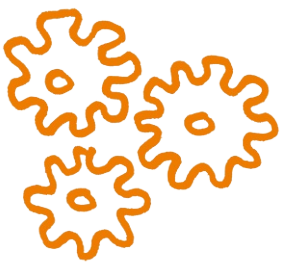
Key components

Knowledge

The European Citizen Science Association (ECSA) created a best practice guideline on what constitutes good citizen science and wrote the 10 principles of Citizen Science. This statement has been translated into many languages. Those 10 principles offers a guidance of best practices for any project based on Citizen Science.

When starting a citizen science project there are a few key elements that must be take into account: how are you going to engage citizens? how are you going to ensure data quality? how are you going to deal with ethics and legal issues?

Although there is still an open debate on how to assess some citizen science activities there are already some examples that can be included as societal impact in evaluation reports like the cases studies extracted from the UK Research Excellence Framework.



Skills

- Be able to differentiate in different citizen science project approaches: projects where citizens just provide data versus projects where the citizen engagement is along the research project.
- Be able to provide advice on legal and ethical aspects regarding the collection of data, including personal data from citizens.
- Be able to provide different solutions on sharing research outputs.

Questions, obstacles, and common misconceptions

- One of the controversies that usually arise in citizen projects is how researchers make data gathered by citizens publicly available. Researchers should be aware on how this data can be shared taking into account legal and ethical aspects.
- The lack of rewards for citizen science practices if they do not end in a "traditional" research output: paper, proceeding etc. is a common issue when training on citizen science. Probably a good way to overcome this issue is to start a conversation on how participants would like to get rewarded and which methods they propose.

Learning outcomes

1. Trainees will be able to know the different approaches of citizen science projects and how to deal with legal and ethical aspects, especially in relation with data management.
2. Participants in the training sessions would learn how to engage citizens in their research at any point of their research activities.

Further reading

- Bonn et al. (2016): Green Paper Citizen Science Strategy 2020 for Germany. Bürger Schaffen Wissen (GEWISS) publication. PDF from buergerschaffenwissen.de
- Citizen Science Cost Action. Training Schools. cs-eu.net

- Community Places (2014). Community Planning Toolkit - Community Engagement PDF from communityplanningtoolkit.org
- Grey et al. (2016). Citizen science at universities. Trends, guidelines and recommendations. leru.org
- Socientize consortium (2014). White Paper on Citizen Science for Europe. socientize.eu
- Pettibone et al. (2016). Citizen science for all – a guide for citizen science practitioners. Bürger Schaffen Wissen (GEWISS) publication. PDF from buergerschaffenwissen.de
- Quality Criteria for Citizen Science Projects on 'Österreich forscht'. fosterscience.eu
- Overview of citizen science projects:
    - Socientize Project. socientize.eu
    - ZOONIVERSE - People-powered research. zooniverse.org
    - Crowdcrafting scifabric. crowdcrafting.org
    - German Citizen Science Projects. citizen-science.at

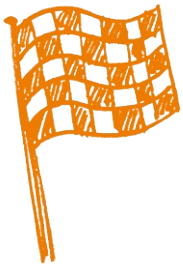
11. Open Educational Resources

What is it?

Open Educational Resources (OER) are defined as "teaching, learning and research materials in any medium – digital or otherwise – that reside in the public domain or have been released under an open license that permits no-cost access, use, adaptation and redistribution by others with no or limited restrictions" (William and Flora Hewlett Foundation definition). Open educational resources include full courses, course materials, modules, textbooks, streaming videos, tests, images, software, and any other tools, materials, or techniques used to support access to knowledge.

Rationale

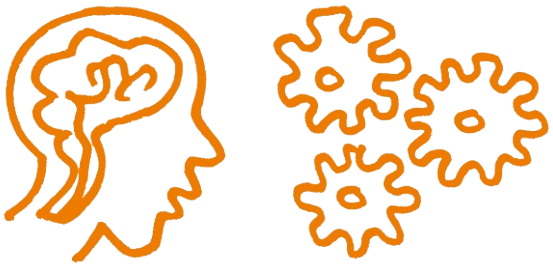
In many cases open educational resources are build upon research findings. If you are an Open Science practitioner it makes sense that your educational resources maintain the level of openness of your research. Moreover other instructors could use your material to elaborate new resources or adapt existing ones. In fact the creation of educational resources can be seen as a cycle similar to the research cycle: find, compose, adapt, use, and share (wikieducator.org/OER_Handbook/educator/OER_Lifecycle).



Learning objectives

1. Participants should learn the difference between open and non-open educational resources.
2. Licensing is an essential part and indicates how to easily use and combine OER.
3. Participants should know where to find and place created OER resources.

Key components



Knowledge and Skills

Open Educational Resources are only OER, if they have an open license. However, there is no clear guideline for the choice of license for your resource. So what kind of license is appropriate? In practice, Creative Commons (CC) licenses are most often used for OER. Open Creative Commons licenses are CC0 (Public Domain Dedication), CC BY (Attribution) and CC BY-SA (Attribution-ShareAlike), which can be used for most educational resources. For the distribution of databases under a free license, Creative Commons is not ideal. Rather, choose a specially suitable open license such as ODbl, ODC-BY or PDDL to be legally compliant.

It is important to stress the need to define who holds copyright or any other related rights of the research output. The copyright holder is the one who can decide to lift restrictions if they are not lifted by default through the licenses. Licenses should therefore be explained in detail to properly attribute authors and to create true OER. This also includes the combination of different license types and its consequences.

Training should provide an overview of OER platforms and their intended use. OpenCourseWare (OCW) is one of the first open educational resource platforms and one of the key initiators of the open educational resources movement. Initiated at the Massachusetts Institute of Technology (MIT) in 2002, the Open Education Consortium now provides materials from all over the world in form of courses under free licenses. Other pioneers were UNESCO and the William and Flora Hewlett Foundation which are still committed to open educational resources.

Examples for OER platforms are:

- Creative Commons Search for image, audio, and video files
- Open Education Consortium for open course material
- OERCommons for educational resources

Questions, obstacles, and common misconceptions

Q: How can you ensure quality of the materials?

A: This is not always a given. So far there is no quality seal for OER materials. Open user comments, peer review, and the publication of materials on platforms of established institutions like e.g. universities can provide a first indication of quality. Just as with printed text materials, quality can though not be guaranteed. This unsettles many users. The actuality and adaptability of the materials nevertheless speaks for the use of OER. At the end of the day, you only know yourself whether the selected material is suitable for the intended purpose and whether its content is correct.

Learning outcomes

1. Trainees will be able to distinguish between copyrighted and free materials.
2. The combination of different license types and their effects will be known.
3. They will be able to find, use and create Open Educational Resources.

Further reading

- Butcher (2015). A Basic Guide to Open Educational Resources (OER). hdl.handle.net
- Miao et al. (2016). Open Educational Resources: Policy, Costs and Transformation. hdl.handle.net
- OECD (2007). Giving Knowledge for Free: The Emergence of Open Educational Resources. OECD Publishing, Paris. doi.org/10.1787/9789264032125-en
- Open Knowledge Foundation (2014). Open Education Handbook 2014. education.okfn.org

12. Open Advocacy

What is it?

Advocacy in all its forms seeks to ensure that people, particularly those who are most vulnerable in society, are able to:

- Have their voice heard on issues that are important to them. Advocacy means giving voice to a group.
- Defend and safeguard their rights.
- Have their views and wishes genuinely considered when decisions are being made about their lives.

Advocacy include actions of defending, influencing, changing, decision-making, persuading, lobbying, attracting attention.

Open Advocacy focuses on the movement to promote Open Science at various levels of stakeholders, highlighting and stressing the societal, professional and personal advantages that it entails.

Rationale

Trainings (workshops, seminars, presentations) can be used as advocacy tools. The structured approach to advocacy practices helps to address the main issues the trainer has to keep in mind if the training is connected to an Open Science advocacy program. how to use advocacy strategies as tools for effecting specific changes, and on building the basic skills necessary for employing advocacy tools (e.g., ad campaigns, meetings with policymakers). Training here is considered as a tool for effecting specific changes, and for building an Open Science advocate community.

Learning objectives

1. Understand the context and goals of the advocacy program
2. Be able to communicate effectively with audiences and draw community’s attention to an important issue and directing decision makers toward a resolution.

Key components

Knowledge

Objectives to achieve

SMART is a way of reminding you that your objectives should be:

Specific — by this we mean that you need to set a specific objective for your programmes.

Measurable — your objective should be measurable.

Achievable — the objective should be attainable or practicable.

Realistic — which also means credible.

Time-bound — and should be accomplished and achieved within a certain amount of time.

Objectives can be long term or short term. Long-term objectives usually focus on changing the policy or practice of institutions, whereas shorter-term objectives can focus on attitude changes, raising awareness, getting an issue on the agenda, building a constituency of support or movement for change. It may be necessary to achieve some of the short-term objectives before you can achieve the longer-term ones.

Main goals of advocacy program:

- To increase awareness among influential groups and the public
- To reduce stigma and fear
- To engage and mobilize key stakeholders within the community who will champion the development
- To expand advocacy groups, including community volunteers
- To mobilize resources to support the implementation of key priority (core) interventions
- To maintain the involvement of decision-makers and the public l by disseminating information on achievements to date and future challenges.

Steps to good advocacy

1. Define your goals
    i. What needs changing?
    ii. What do we want to ask for? Changing legislation, policy, regulation, programs, funding
2. Understand your audience: different strategies for each target
3. Build a profile of open access stakeholders and their attitudes
4. Craft your message: create compelling messages that appeal to stakeholders' interests
    i. Be clear on what we are asking for
    ii. Keep it simple and focussed
    iii. Use positive language
    iv. Use evidence - facts carry more weight than anecdotal evidence
    v. Economic arguments are important
5. Plan and develop your communication and advocacy campaign
6. Identify delivery methods:
    i. Advocacy is relationship building,

   ii. Tactics change by target audience
7. Identify Resources and Gaps:
   i. Do a SWOT (strengths, weaknesses, opportunities and threats) analysis
   ii. Build on existing resources and opportunities
8. Plan next steps, identify achievable goals that set stage for larger work: advocacy strategy/plan
9. Evaluate effectiveness regularly

Aspects of advocacy

- Advocating for your own rights as an author
- The basic steps for achieving local culture change (Kotter n.y.)
- Advocating to your peers: Writing letters and articles advocating for Open
- Talking to journal editors - having the OA conversation with your field
- Talking to policymakers

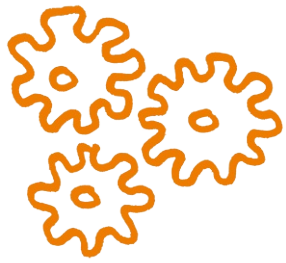
Tools and methods

Indirect: stimulate participants to take action on their own behalf

Direct: lobbying before decision makers by representatives on behalf of others

Campaigning: generating a response from the wider public and using a variety of techniques such as:

- Chain e-mail or letter
- Opinion pieces and letters to the editor in newspapers
- Newsletters
- Celebrity endorsements
- Media partnerships with newspapers, journalists and film-makers
- Web-based bulletins and online discussions
- Public events
- Large-scale advertising campaigns
- Use of social media (Twitter, Facebook)



Skills

- Write a letter for a newsletter or forum for your scholarly society about Open Access.
- Make your own email template reply about only reviewing for OA journals, etc. Reuse/base it on ones out there already.
- Outline concrete solutions and benefits Open Science can deliver for current headaches university administrators may struggle with.

- Find your local advocacy group and volunteer for them!

Questions, obstacles, and common misconceptions

Lack of interest from audiences. Lack of understanding the value.

The institution and/or senior management is concerned about the impact of the advocacy efforts.

Learning outcomes

The trainer will be able to consider the training event in the context of a program.

Further reading

- A Crowdsourced Resource by OpenCon attendees. Starting Open Projects From Scratch. CC Zero Google Doc
- Bolick et al. (2017). How open access is crucial to the future of science. doi.org/10.1002/jwmg.21216 (*comment by authors: rebuttal article written in the Journal of Wildlife Management after a misleading / fear mongering article about OA*)
- Clyburne-Sherin (FSCI2017). Advocating for transparency policies - a toolkit for researchers, staff, and librarians. github.com
- JISC Pathfinder project Pathways to Open Access (n.y.). Advocating Open Access - a toolkit for librarians and research support staff. PDF
- Jones (2015). Open science and its advocacy. fosteropenscience.eu
- Kotter (n.y.). Kotter's 8-Step Change Model of Managementt. study.com
- Lingua / Glossa articles on their move away from Elsevier - their advocacy as editors with a publishing organization Wikipedia)

- Mozilla Science Lab (2015). Open Science Leadership Workshop. Working OpenProject Guide. github.com
- Smith (2014). The Open Access Movement and Activism for the “Knowledge Commons”. asanet.org (*comment by authors: example of a letter to a scholarly society advocating for Open Access*)
- Smith (2015). Defending the global knowledge commons. opendemocracy.net
- SPARC*. Author Rights & the SPARC Author Addendum. Your work, your rights. sparcopen.org
- Webinar Report: Organising and advocating (2018). How can early-career researchers make their voices heard? eLife ECRwednesday webinar. elifesciences.org
- 8 Steps to Good Advocacy. PDF

Initiatives and projects

- FOSTER Plus Project (2017-2019). Fostering the practical implementation of Open Science in Horizon 2020 and beyond. fosteropenscience.eu
- PATH. Strengthen Advocacy. sites.path.org
- PASTEUR4OA. Advocacy Resources. pasteur4oa.eu
- Retraction Watch. retractionwatch.com

On Learning and Training

This chapter is providing context on training strategies, practical guidance in designing a course as well as an overview of pedagogical theories. It will focus on three key concepts in teaching and training:

1. Preparation
2. Execution
3. Reflection

Teaching and training is firstly about preparation before delivering a course. Preparation includes the choice of content, deciding on appropriate teaching methods and putting them into a sequence to maximise the effectiveness and impact of your training. Secondly, teaching is about delivering a course (i.e., how you act and interact with the participants). Even if you are feeling very confident on a particular topic, it is very advisable to avoid starting the delivery before having finished the preparation. Moreover, you may need to test your content, especially the practical exercises. Then, during the course delivery, you need a good portion of flexibility, because things rarely happen completely as you expect. Finally, teaching is also about evaluation and self-assessment once you have delivered a course. It is more than likely that you have to engage yourself in the same or a similar course several times, in particular if the evaluation shows that it was good.

To better prepare yourself for future events, you should reflect on what worked well and what did not work so well, and use this to iteratively define your preparations and delivery. Briefly said, there is a “before”, a “during” and an “after” class, i.e. activities in a cycle, similar to science. This chapter provides a practical guidance for the trainers on how to prepare and deliver a course to various audiences: what are the main obstacles one has to overcome and what are the main issues one needs to keep in mind when putting together a training.

Some reflections before you start

In the following part, we will focus mainly on the first aspect (preparation) and then give you guidance on how to plan and manage your course. To start with, we will speak about some theoretical issues which will provide you with an idea of what teaching and learning means and how teaching adults differs from teaching teenagers or children.

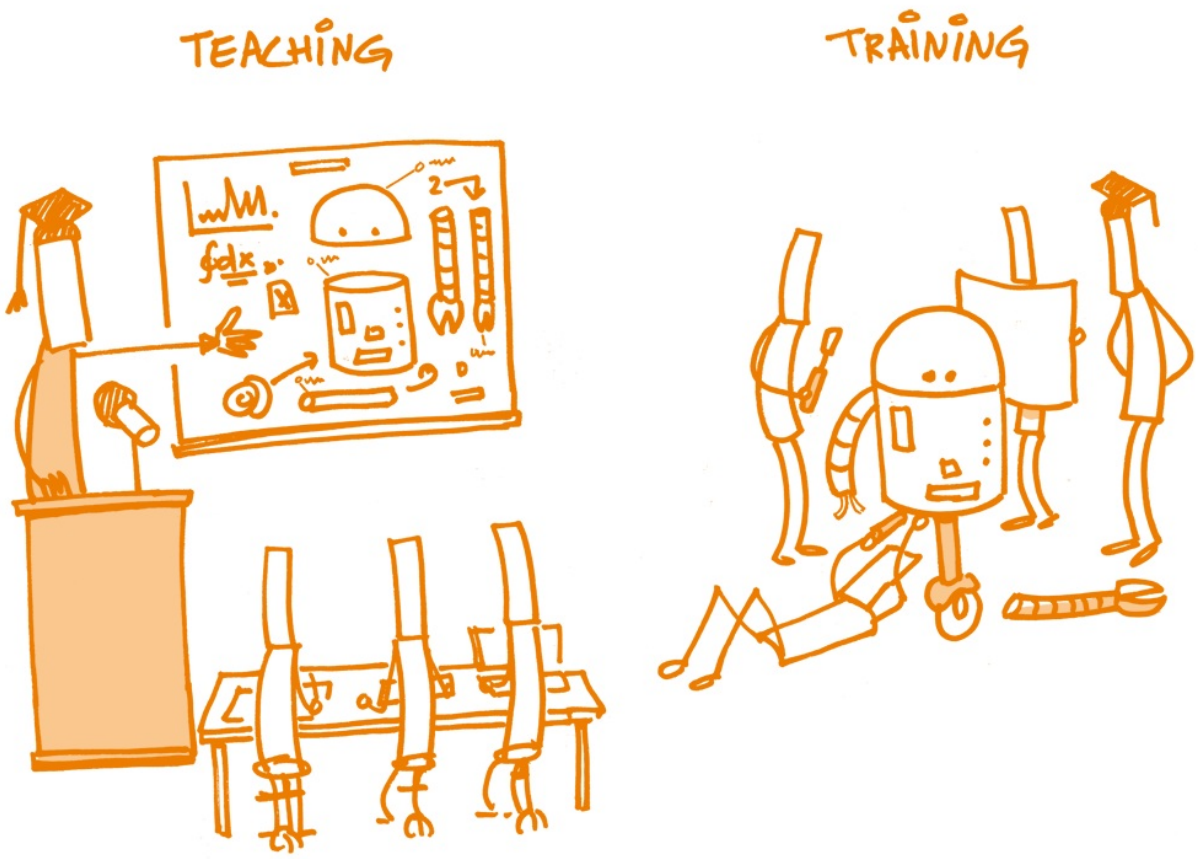
Training vs. Teaching

- Teaching is more related to theoretical concepts than training, which is related to the practical application of knowledge (i.e., development of skills).
- Teaching seeks to impart new knowledge while training equips the already knowledgeable with tools and techniques to develop a specific skill set.
- Teaching is, usually, done within the context of education and academic environments, while training is associated with post-high school and/or postgraduate short and intensive courses.

- Usually, teachers give feedback to their students, while trainers receive feedback from the learners.

However...

- Training is the process of teaching or learning a skill or job, and trainers do actually teach something. Therefore, training can be considered as a broader activity that may encompass teaching.
- Teaching may also include typical training activities and goals, such as practical sessions and demonstrations.
- Despite the fact that teaching and training techniques may sometimes vary, the difference between training and teaching is not related to the process itself but to the focus, with training generally having a more specific focus than teaching.
- In order to develop competencies as a professional, a person needs to attempt to understand the theoretical concepts as well as to have practical exposure. Therefore, teaching and training are equally important and complementary educational concepts.

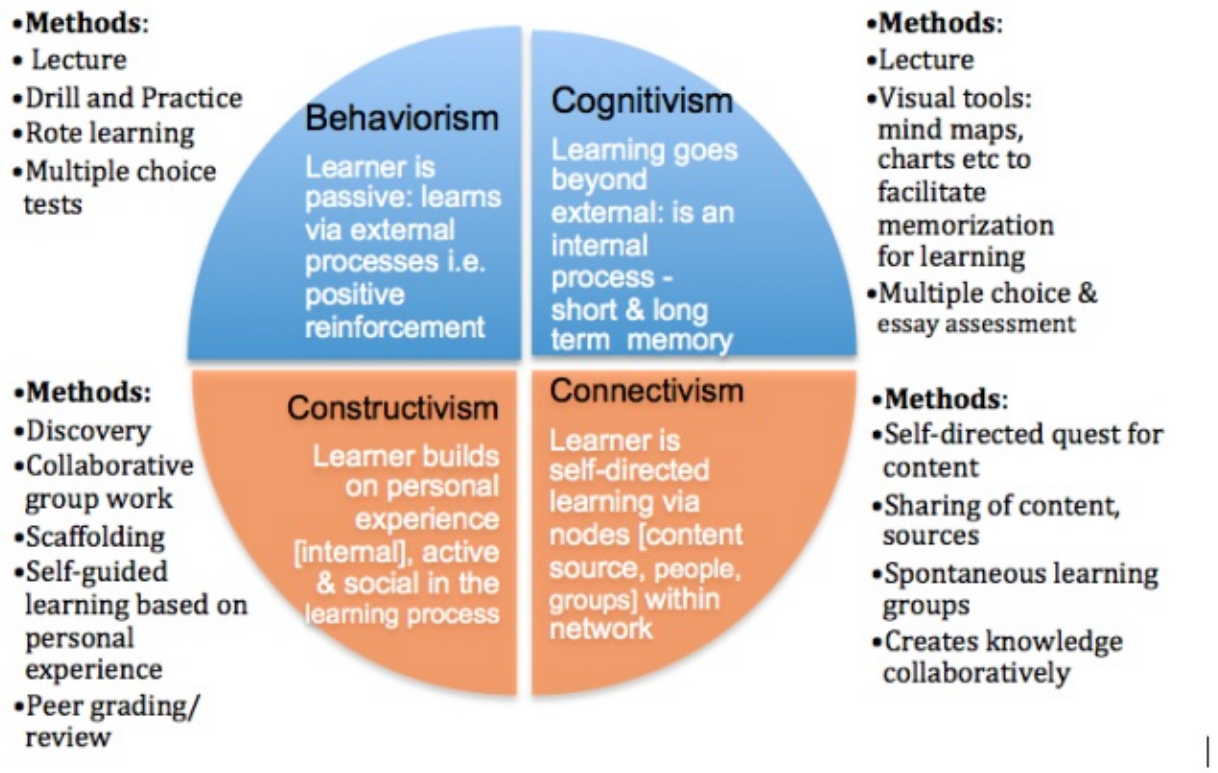


Strategies

There are different theoretical approaches to learning and training, which are sometimes also influenced by the culture you live in. Some people like to talk and give lectures. Others like to listen, others don’t. Some exercises are simple and look for clear answers. Other exercises are centered around problems and focus on giving the participants time and space to reflect on them and find solutions. Finally, some trainings are designed to give the participants maximum freedom and let them be as creative as possible. Success in trainings like these is more difficult to evaluate.

Four well known learning theories are behaviorism, cognitivism, connectivism and constructivism. They describe different perspectives on how people learn.

This simplified diagram summarises their main characteristics in very practical terms:



Transcribed from:

https://onlinelearninginsights.wordpress.com/2013/05/15/how-couse-design-puts-the-focus-on-learning-not-teaching/

The work done by the Software Carpentry also helps to understand learning processes: https://carpentries.github.io/instructor-training/

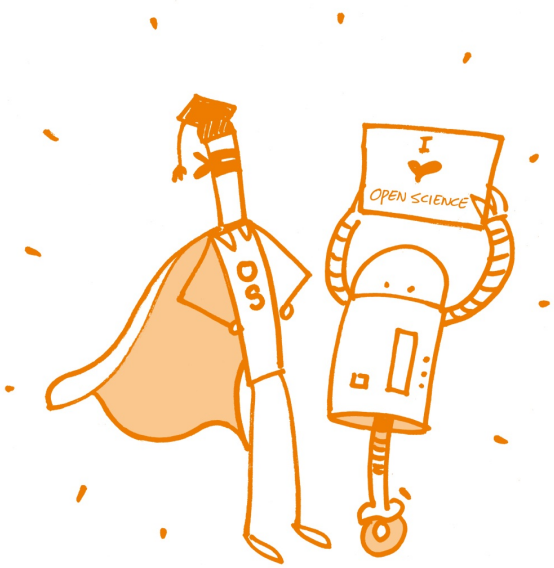
The Connected Curriculum Framework

The recent movement ‘Connected Curriculum Framework’ aims at modernizing learning approaches and adapting them to the 21st century learner. The general objective of the framework is to improve the relationships between student education and research practices by breaking down unnecessary divisions. The framework values rich dialogue, active inquiry, collaboration, and interactions between students and researchers as well as universities and wider communities. This carries interesting promises in the area of Open Science and Citizen Science, Crowdsourcing, etc. You can read the Connected Curriculum here: http://www.ucl.ac.uk/ucl-press/browse-books/a-connected-curriculum-for-higher-education

How is this relevant to you?

What is important to know, is that there are different approaches and you should not feel obliged to follow only one strategy, but rather decide at which point of your training you should apply which strategy to teach and evaluate.

In the end it is practice that matters and it may be helpful to check your content and practical exercises against one of the theoretical approaches in order to find out if they are appropriate at the given moment and for the target audience.



Expectations about a trainer

Everyone that comes to your training will come with expectations, conscious and unconscious ones. Among others (such as teaching methodology, content and prior knowledge) they will have specific expectations about the trainer.

Most learners will expect you to:

- Be enthusiastic about the topics that they are teaching.
- Have a general understanding of core scientific (or humanist) values, and recognise the role of ‘openness’ as an intrinsic, core element of this.
- Understand the importance of factors such as research transparency and reproducibility, and the broader societal implications of these.
- Show familiarity with the research process, including planning research, conducting research, producing research results, and communicating and publishing those results.
- Have knowledge about the different types of research processes and outputs that can be shared, including data, code and software, papers, communication, workflows, grant applications, and data management plans.
- Be aware of the policies, regulations and laws that could affect researchers when performing Open Science
- Understand the pressures that result from institutional policies, or lack of them, that shape the way in which researchers handle data and results, from the acquisition stage to the sharing and dissemination stages.
- Understand the expectations that are raised in the social fabric about the use of the resources and outcomes of scientific activities, such as its impacts in citizen science, the public understanding of science, the influence in the education providers, etc.
- Be able to teach and have a profound knowledge in Open Science. (In fact, this is what this book is about.)
- Provide links to online documents and resources that support newcomers.

Target audiences

A good way to get started with your Open Science training is to address audiences that have some idea and/or are interested in the topic. Generally, these people may be more open to the idea of Open Science. Starting your training with a motivated audience has several advantages:

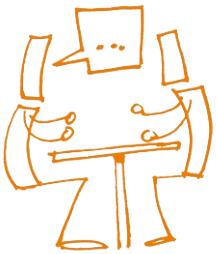
- Knowing that your audience really is interested in the topic may make you more comfortable diving into a new training area/topic. You may contemplate running a survey to assess this in advance.
- A motivated audience probably will contribute to discussion and provide you with useful input on how to further develop your training curriculum.
- Motivated audiences can become ambassadors of your training

Information you need to gather about your audience:

1. Maintaining an inclusive environment, and taking into consideration the diverse backgrounds of your potential attendees, is important for any successful training event. To learn how to make your workshop inclusive, see the Conference Planning Checklist by SPARC.
2. Whether the audience members know one another or not in advance will impact the group dynamic and the sorts of activities you might want to conduct.
3. Whether the participation is voluntary or not will influence their motivation.
4. The knowledge level of the audience regarding the planned discussion topics will affect the content and style of presentations.
5. Whether the audience is accustomed to a specific learning method might affect how the participants react to very different training format.
6. Audience size:
    i. set a target audience size, based on the available space/capacity and available time for practical work..
    ii. the size of the audience will impact on how well they engage together and interact with the process.
    iii. if you want a larger audience, consider break-out groups, and the logistical requirements that might come with that.
7. Consider whether your event will be open to the public or limited to those affiliated with the host institution. A public event may help increase and diversity attendance, while limiting it can help you focus on particular topics. In addition, attendees from the same institution are more likely to already know each other.
8. Consider using video-lectures, as you might reach a broader audience. Though with a small group of people attending an on-site event it is often easier to maintain their attention, and to create and use the feeling of an authentic connection.
9. Consider what the best way is to approach different target audiences (meeting, face to face workshops, webinar, newsletter, social media, etc.)
10. With a heterogeneous audience, keep in mind the different stakeholders involved in order to address their different needs, knowledge and/or responsibilities:
    i. funder, institution/employer, researcher (student, PhD student, researcher, project lead),
    ii. support (research office, library, IT)
    iii. commercial partners in a project

The outcome of the training should be that the trainees:

1. have a better practical understanding of the key concepts and corresponding applications for Open Science.
2. confidently use what was learned during the training, thus increasing their impact in their professional environment.
3. become able to network with advocates from multiple disciplines,and act in a global Open Science initiative.



Teaching adults

Scholarly research is practised by adults, as such, the participants of any training in Open Science will most likely be adults, often with a first or second degree in higher education. It is therefore interesting to see, how far teaching children or teenagers (pedagogy) differs from teaching adults (andragogy). The Canadian Literacy and Learning Network did some interesting work on this difference and recapitulated it in seven principles:

1. **Adults must want to learn.** This means that the inner motivation and added values are decisive and it might be worth to know them before starting the course.
2. **Adults will learn only what they feel they need to learn.** Adults are practical in their approach to learning; they want to know, "How is this going to help me right now?" You should therefore be practical and direct.
3. **Adults learn by doing.** This is true for children too, but active and immediate participation matters more for adults.
4. **Adult learning focuses on problems and the problems must be realistic.** The participants will often come with a problem and it will be your task to discover gaps and try to close them.
5. **Experience affects adult learning.** Adults have more experience than children, either negative or positive. You can make use of this experience by avoiding negative associations.
6. **Adults learn best in an informal situation.** School-age youngsters usually have to follow a curriculum. Often, adults learn only what they feel they need to know. You should therefore try to involve your audience in the learning process. This may happen by making the environment relaxed, informal and inviting.
7. **Adults want guidance.** Adults want information that will help them improve their situation or solve problems, but they do not want to be told what to do, but rather choose options based on their individual needs.

Therefore, you will need to

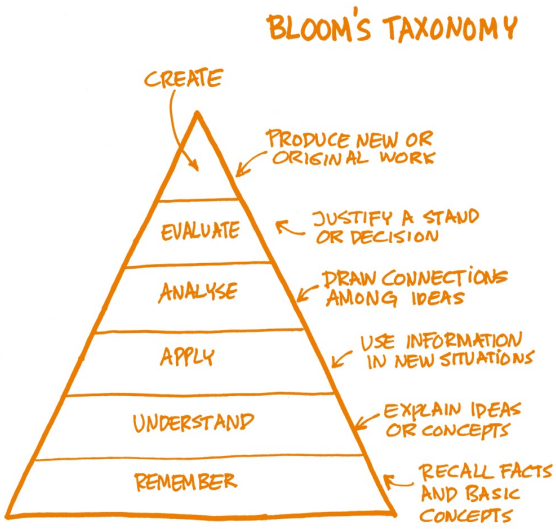
- provide the discovery points, tools and support where researchers will find them
- prepare online documentation with clear, understandable, and up to date guidance
- put together good usable (and discoverable) tools or templates to generate it.

In summary, adults have their interests focused on their own improvement and see training as a self-centered, capacity-building exercise. Adults like to be respected as such, and that their expectations are individually met, in an exhaustive way whenever possible.

Bloom’s Taxonomy

Learning outcomes are often the most specific way of establishing how a training instance is delivered, by tailoring whatever is needed so that the best part of the expected outcomes are met by the best part of the audience. Learners meet outcomes in a variety of ways, often amenable to a quantitative evaluation.

Specifying outcomes is part of handling training as a cognitive process. In 1956 Benjamin Bloom created a taxonomy of cognitive levels that has been modified through time. This is a very useful tool to build consistent and reusable learning outcomes in any subject matter. Transitions between non-contiguous levels of cognition is generally not acceptable. The taxonomy helps to detect potentially difficult situations where assessment can fail because the cognition level of the learning delivery is not the same as the cognition level of the assessment that is being used.



A present day version (since 2001) can also be found here] (https://thesecondprinciple.com/teaching-essentials/beyond-bloom-cognitive-taxonomy-revised/

Bloom's Taxonomy is a classification method with six levels. Using Bloom's Taxonomy is worth the effort because it represents a significant step towards a desire to build robust training and teaching. Together with Bloom's Taxonomy you can find several types of design aids such as annotated terminologies, verbs to use or to avoid in course planning and building assessment questions, etc.

Learning objectives & learning outcomes

These two terms are often used interchangeably by the training community. Objectives, comprising aims or goals, and Outcomes, comprising tangible results, may overlap, but are not genuinely the same.

When designing training, you should think primarily of objectives, then list what outcomes you want your audience to reach for. Do not worry if they seem to overlap here and there, or if, as in most cases, an objective encloses one or more outcomes. Design all your practical exercises around specific outcomes.

Note: you should avoid using the abbreviation LO as it becomes ambiguous.

Here is an attempt to clarify this situation and remove ambiguities:

Learning objectives

- Describe the goals and intentions of the instructor.
- State the purpose and goals of the course.
- Focus on content and skills important within the classroom or programme.
- May describe what the instructors will do.
- Should be specific and detailed.

Learning outcomes

- Student Learning Outcomes catalog the overarching "products" of the course and are the evidence that the goals or objectives were achieved.
- Learning Outcomes are statements that describe or list measurable and essential mastered content-knowledge—reflecting skills, competencies, and knowledge that students have achieved and can demonstrate upon successfully completing a course.
- Outcomes express higher-level thinking skills that integrate course content and activities and can be observed as a behavior, skill, or discrete usable knowledge upon completing the course.
- Outcomes are exactly what assessments are intended to show – specifically what the student will be able to do upon completing the course.
- An assessable outcome can be displayed or observed and evaluated against criteria.
- Outcomes are clear and measurable criteria for guiding the teaching, learning, and assessment process in the course.

(Adapted from http://provost.rpi.edu/learning-assessment/learning-outcomes/objectives-vs-outcomes)

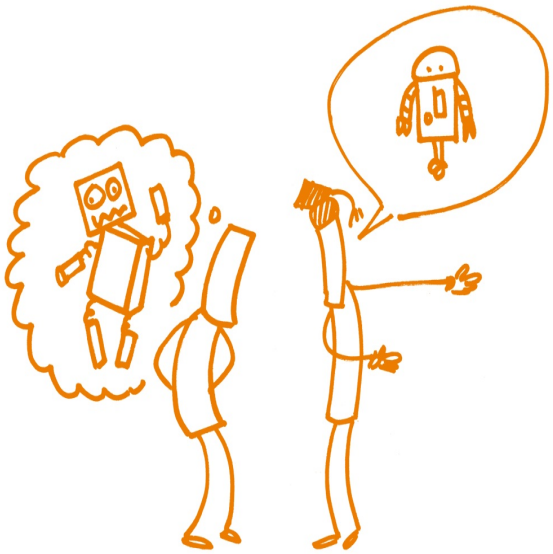
For Open Science Learning Objectives, see this FOSTER document: https://doi.org/10.5281/zenodo.15603 (see page 13 & 14)

Example of a training objective:

- "To learn how to use assessment and feedback in training with maximised effectiveness"

Example of a training outcome:

- "Upon completing the course, the learner will be able to design a training exercise and a strategy to evaluate its effectiveness"



Motivation & demotivation

One of the key components in a training event is to make sure that the lack of confidence that the participants might have when being introduced to a new field (Open Science, in this instance) does not discourage them from pushing onwards. Even if some participants are generally familiar with the concepts presented in the training event, it is important to acknowledge when people are becoming confused. Acknowledging that their misunderstandings are valid is key to encouraging a growth mindset and motivating them to accept and endorse the Open Science practices.

There are several strategies that can be employed throughout the training event that can motivate participants. (Taken from the Carpentry Instructor Training, https://carpentries.github.io/instructor-training/08-motivation/)

- **Strategies to establish value**

- Connect the material to the participants' interests or values.
- Provide authentic, real-world tasks and case studies, ideally matched to the participants background and immediate interests.
- Show relevance to the participants' current academic lives.
- Convey your own passion and enthusiasm for Open Science.

- **Strategies to build positive expectations**
  - Ensure alignment of objectives, assessments, and instructional strategies.
  - Provide early success opportunities by applying the concepts in hands-on exercises and tutorials.
- **Strategies for self-efficacy**
  - Provide participants with options and the ability to make choices.
  - Give participants an opportunity to reflect and make their own connections between Open Science and their particular work.

Practical guidance

You will find more information concerning the concrete planning and execution of a training on Open Science in the chapters on Organizational Aspects and the Examples and Practical Guidance.

Designing a course

The creation of your course will either be driven by planning on the course's objectives or on its outcomes.

Planning based on objectives, rather than outcomes

SMART is an interesting technique for specifying goals / objectives that is also used in project management. SMART is an acronym that stands for five criteria: Simple – Measurable – Ambitious – Realistic – Timed.

- Your goal is **simple** if it can be understood by a person not familiar with the topic. That is, you can explain to your students beforehand what they are going to learn. It is usually a good idea to present your goal at the beginning of a lesson. Simple means that the goal can be put into no more than one concise sentence.

- Your goal is **measurable** if you can determine objectively whether the goal has been reached. Measurability prevents imprecise goals like "students understand Open Science", which is too broad and difficult to measure as there are many different components. Instead, use verbs that are actionable: identify, draw, name, explain, calculate etc. Verbs for good teaching goals have been categorized by the Bloom's taxonomy of cognitive domains (clinton.edu/curriculumcommittee/listofmeasurableverbs.cxml). Measuring helps you and your students to assess or self-assess progress.
- Your goal is **ambitious** if you challenge your students. Is there a clear benefit for them? Do you want the lesson to broaden their horizon? In which way does it give them an edge? Being ambitious means having an answer to the question: What will students learn that they could not by other means? If you feel a desire to make a stand and defend your viewpoint, it probably is ambitious.
- Your goal is **realistic** if you sincerely believe your learning goal can be reached in the given timeframe. Being realistic involves homework: Do your students have the necessary background knowledge? What practical abilities do they need? What technical prerequisites are there? Are you prepared for unexpected questions? For instance, understanding all Creative Commons licenses in one hour may be realistic for one group, but out of reach for another.
- Your goal is **timed** if there is a concrete timeframe which the goal is to be reached. First-time teachers often overextend their time budget. Setting time limits for your learning goals helps you to structure your lesson, recognize and react to unexpected delays. A good form of planning time is having a detailed schedule or lesson plan.

Adapted from SMART Goals, How to create objective, measurable project goals by Kristian Rother.

Planning based on outcomes, rather than objectives

Use reverse instructional design, known as Backward design, a technique for planning lessons that emphasizes outcomes:

1. Start from your learning objectives.
2. Decide what constitutes evidence that these objectives have been met (summative assessment, see Post-training Evaluation below).
3. Choose the best format and design content to prepare the audience for what they will have to do during the summative assessment.
4. Sort the content in order of increasing complexity and then provide the content and motivation they need to close the gap between what they know and what they need to know to complete the summative assessment. (Software Carpentry Instructor Training)

Backward design challenges "traditional" methods of curriculum planning. In traditional curriculum planning, a list of content that will be taught is created and/or selected.[4] In backward design, the educator starts with goals, creates or plans out assessments and finally makes lesson plans. Supporters of backward design liken the process to using a "road map".[5] In this case, the destination is chosen first and then the road map is used to plan the trip to the desired destination. In contrast, in traditional curriculum planning there is no formal destination identified before the journey begins.

The idea in backward design is to teach toward the "end point" or learning goals, which typically ensures that content taught remains focused and organized. This, in turn, aims at promoting better understanding of the content or processes to be taught to students. The trainer is able to focus on addressing what the students need to learn, what data can be collected to show that the students have learned the desired outcomes (or learning standards) and how to ensure the students will learn.

Content

Content collection

Before starting to teach you will have to collect and prepare content. Content is nowadays available en masse, and the question is less about finding or creating content than rather about finding appropriate content or making the discovered content appropriate to the needs and capabilities of your target audience.

Please check the chapter on Examples and Practical Guidance which will contain helpful information on how to adopt, adapt and develop content.

Content reduction

One of the biggest challenges in designing training courses is the reduction of content to the training format. If you have only two hours, you need to provide the most important information on a topic during this time. As a trainer, however, you usually have much more knowledge that you would like to pass on. Try to reduce the content to the most important key points. What is really necessary to know and what are only details or marginal topics? Set thematic priorities, be transparent about omissions and inform your participants about these.

And try to keep enough time for open questions, discussions, sharing experience among participants. It will help you to get the "right" questions. Usually much more basic, than you expected or more detailed and specific than you planned.

Starting the training

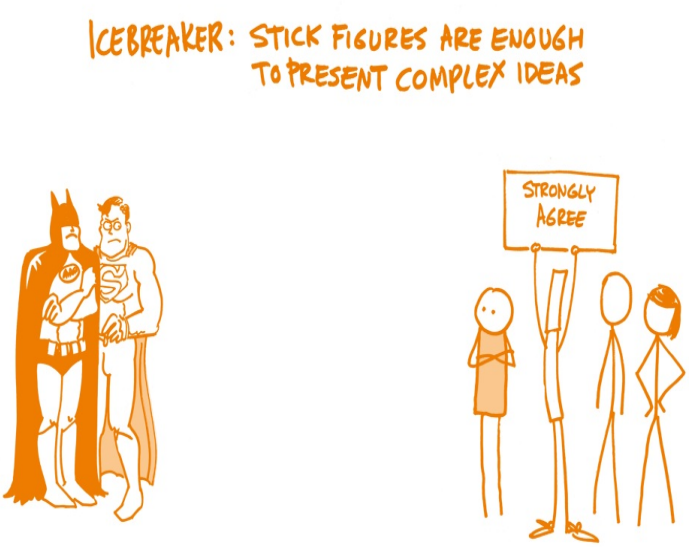
Introductions

At the beginning of the event, speakers should clearly and succinctly introduce themselves and their areas of expertise. Why should the attendees listen to you? What experience and skills do you have that are relevant to them? You should then give a general presentation of objectives, content, and outcomes for the training event - what participants will learn, and why. Projecting confidence as a figure is key here in order to establish trust.

Depending on the size of your audience, the amount of time available, and the degree to which audience interaction will be key to successful training outcomes, you may wish to begin by having participants introduce themselves briefly (although this is probably not recommended if the group is larger than 15-20 participants). This might be a good time to collect thoughts from participants on their own expectations and levels of experience (if not done before, e.g. with an online-questionnaire), and to gauge to what extent these match the intended outcomes and your overview of the intended or target audience for the training. If there is a large mismatch, now would be the time to consider ways to spontaneously adapt the programme. For example, if participants are more knowledgeable or experienced than anticipated, you may wish to move more quickly over the basics of particular areas of Open Science in order to spend more time on interactive discussion in which the participants' own questions and experiences are brought to the forefront.

Know that there is no absolute need to immediately adapt the content, just be clear by letting all participants know what will be covered or not.

Once more, the information delivered by Software Carpentry might be helpful to create the right ambiance.



Ice-breaker

In order to energize audience members and help them get to know the trainers and each other, many training sessions begin with an ice-breaker exercise. Creating a warm, welcoming, friendly and positive learning environment should enable attendees to better participate and learn, and help them to feel more comfortable.

While icebreaker games can help create a positive atmosphere, a poorly chosen icebreaker can do the opposite, making people feel nervous or uncomfortable. You should carefully consider your attendees and the potential group dynamics when choosing an icebreaker. People should not be made to feel embarrassed, or forced to reveal personal information they do not wish to share. Groups will differ in important ways - whether attendees are of different ages or statuses within an organization, from different cultural levels, or of differing levels of educational attainments, will all affect the amount of common ground that might already exist between them. Try to keep such exercises related to the intended learning outcomes. Please refer to the Further Reading section for examples.

During the training

Define the intended outcomes of the training and always give orientation to your trainees:

- Where are we?
- Where do we want to go?
- What will we cover?

Establish a balanced change of pure talks about the content to deliver (max. 20 minutes) and activity sessions to work with the content (Klaus Döring, 2008).

Always make the learners' voices sound as soon as possible or, in other words, go for active learning!

Active Learning

Active Learning is a process whereby learners are actively engaged in the learning process, rather than "passively" absorbing lectures. Active learning involves reading, writing, discussion, and engagement in solving problems, analysis, synthesis, and evaluation. Active learning often involves cooperative learning with other attendees.

Using active learning principles and implementation in training is, in general, a good idea. You are the second best judge for the benefits. Do remember that the first judge is the participant.

Active learning helps to bypass diversity in learning styles and other difficulties with audiences. While more efficient in reaching outcomes of higher levels, active learning also addresses cognition issues related to the nature of the content and the way to present it, as shown in the following diagram, commonly found in several textbooks and online resources, and known as the Cone of Learning. Active learning is best utilised at the top levels of Bloom’s Taxonomy (Analize, Define, Create, Evaluate), and that also corresponds to

the best strata of memorization: what you say, write or do - the bottom half of the Cone of Learning. Cognition issues arise with more ease when content spans several of these levels at a time and fails to address the intermediate levels as well. Checking your content against the Cone of Learning is an easy way of detecting these potential miss-outs while you deliver training. Likewise it allows you to decide to use more visual aids where you expect that the need for memorization is higher. So, when your audience gets behind you may use this technique to diagnose, try to locate the causes and pick the most effective remediation.

Gamification

The foundations of the methodology in Active Learning lie in modern learning theories (partly in Constructivism and some Connectivism) and add learning engagement techniques to break barriers and flatten as many obstacles as possible. For example, gamifying a learning instance can move learners away from passive acquisition of content to full engagement, leading to the repositioning of the learner as someone who steps back and observes the learning process and how it works. An example of gamification in training is given here: Key Terms, a learning game for conceptual consolidation.. An additional example can be found in CURATE: The Digital Curator Game.

Inclusive engagement

How to engage quiet participants? A good starting point might be to ask a question and wait at least 30 seconds for answers (Mary Budd Rowe, 1986). The result will be that more people engage in the discussion, the answers are of better quality and slow learners get a chance to answer.

Another method of achieving inclusive engagement is *progressive stacking*. A moderator chooses who speaks next from those participants who wish to speak and have not yet spoken, as usual. In addition, underrepresented voices, including underrepresented gender and racial identities, are chosen to speak first.

During discussions (in larger groups), you should avoid standing microphones with first-come-first-speak engagement, as it discourages inclusive engagement and encourages monologuing. Use a wireless microphone instead or raised hands to ensure that who speaks next can be selected by the moderator. The larger the group, the bigger the need for a moderator who monitors who is speaking and who is not. It will also be the moderator’s task to choose who speaks next from those participants who wish to speak, but have not yet spoken to avoid the workshop engagement to be dominated by just a few participants.

General recommendations

- Stay connected! Always try to keep the contact with the group, check your pace and those of the others.
- Be careful not to overload the participants with too much and/or too difficult content.
- Be open for feedback at any time but avoid or actively break-up never-ending discussions.

- Breaks: Always give enough space for breaks. The longer your course, the longer and more often your breaks.
- Prepare short, middle and long versions of your exercises to become flexible if the discussions are more or less intensive.
- Be prepared for difficult students and consult some troubleshooting guidance before the course.

(You may find some ideas in the MozFest2017 Facilitator Guide). You should in any case have an idea of what you do when a parallel conversation emerges or what to do when somebody is constantly rude or inattentive etc. Know that there are verbal and non-verbal ways to tackle this.

- Wrap-Up / Meta View: At the end of the training it might be worth to tell your participants what you did and why you did it. This will also make the evaluation easier.
- Enjoy the session yourself.

Instant feedback

At the end of each module, request feedback from participants in the form of a one-up/one-down (i.e. state one thing that was useful/good in the module and one thing that was unclear/could be improved). It can also be more graded / scaled. Here is an example feedback with 6 levels.

Another way for getting instant feedback, especially at predefined points, is through continuous polls. As an example, Slack can be employed to provide anonymous feedback on the pace, by giving the option for members of a channel to change their choice on a poll at any given time. Feedback counts should be shown to the participants. Showing totals or graphs can act as an incentive. Online, cloud based tools generate more engagement, especially because the dependence on devices such as clickers is disappearing. Learners can use internet connected mobile devices and feel empowered. Examples of this are abundant. You should test the methods before you use them with a real audience, and start with the systems that have smoother familiarisation steps, such as Socrative and Learning Catalytics, Polleverywhere, Directpoll.

Some more instant feedback strategies can be found under teachthought.com

Training evaluation

Successful Open Science training also needs evaluation phases. Especially when starting a course, it is useful to look at trainees feedback. An evaluation can provide you with valuable insights on your methods and content. Continuous evaluation and consideration of the feedback improves the quality of the training and the trainer's performance.

Types of feedback

There are different ways to get feedback from your participants:

Classic forms of evaluation

- Use an evaluation form in which you ask the participants for feedback on you as a teacher.
- Get interim statements during the course to check, if the course meets expectations. This gives you the opportunity to make adjustments before going on.

Verbal feedback

- Ask the trainees for a short summary of their course experience.

Self-Evaluation

- Make your own evaluation, what went well, what went wrong?

Long term feedback

- 6 months later, questions about plastic changes in behaviour, more generally about modifications in the attitude and its potential effects.

Peer to peer feedback

- Colleagues will help you with their experience to prepare your course, eventually attend themselves the course and exchange with you afterwards and will give you their feedback.

Metrics for training efficiency

In order to evaluate a course you should need to establish, first, what you want your learners to be familiar with, know, analyse critically or be able to explain . Why are you doing the course? Which goals do you want to achieve? And once the course has finished you should check if you reached those goals. There are different criteria on how to measure the success and efficiency of your course (Kirkpatrick & Kirkpatrick, 1994):

- Reaction (meeting expectations): Are the trainees satisfied with the course? Have the participants reached their learning goals? Were the expectations realistic? How did they react to the course? Was there a clear structure or a common thread?
- Learning: Did the attendees learn something new? Is it helpful in their current situation? Did they understand everything? Can they assign suggested tools/platforms to the respective Open Science practices? Do they meet the pre-specified learning objectives?
- Behaviour: Will they change their way of conducting research? What will they do with their acquired knowledge? Will they recommend the training/content to others?
- Results: which outcomes, when met, have a more positive impact towards the objectives? Which were the ones that brought more benefits?

Kirkpatrick’s Training Evaluation Technique

Kirkpatrick's Four-Level Training Evaluation Model is an standardised way to analyse the effectiveness and impact of your training.

Exercises

- Check the learning outcome with gap texts and quizzes.
- Run a simple exercise at the start and same exercise at the end. Then see if opinions have changed.
- Keywords: Prepare paper slips with different key aspects of Open Science. Divide the trainees into groups (at least 3 people) and let each of them explain 2-3 keywords to each other.
- Give the participants a printout of the general structure for the scientific method, and ask them to assign Open Science tools and methods that can be applied to each of them.
- Depending on time, you can also ask them to create an imaginary/simple research scenario and go ahead in establishing the Open Science protocols for it.

Rework your course

You will have had your own expectations before teaching the course and the experience of having done so will show you that things do not always work the way you planned. You should not be too disappointed, because a first time yield for all outcomes is almost impossible, but rather take the end of the course as a starting point to rework your material and rethink some of your methods and practical exercises.

Be aware that it might even take you three attempts until you will have the feeling that your course has the format it needs and will satisfy both the attendees as well as you as the instructor.

Learning outcomes of this chapter

After going through this chapter you should be able to respond to requests to plan and deliver training in Open Science to specific audiences.

Exercise

Consider the following hypothetical situation: You have been invited to train principal investigators at an engineering school. The training will be about the management of datasets that are shared between research groups in the school and their colleagues in Canada and New Zealand, in an Open Science context.

- In one paragraph describe the design strategy for your training session, in major steps, for example what would you plan to do before, during and after your training session
- List three questions that you are allowed to ask to characterise your audience
- List three learning objectives
- List three expected learning outcomes
- List three actions that you can use to break the ice and get your audience engaged
- List three questions that you would ask to check what the participants have learned
- List three questions that you would ask to check if the participants enjoyed the session.
- Be ready to react to genuine and spontaneously created wordcloud (sli.do or some other tool) : don’t be afraid to co-work with your audience, learn to play with what you know (and assume you also have to right for some perplexity)

Similar exercises can be applied to training different audiences, for which you may consider the same way of testing your knowledge.

Further reading

About Bloom's Taxonomy:

- Davis (2014). Using Bloom’s Taxonomy to Write Learning Outcomes. pearsoened.com
- Clinton Community College (1966-2017). List of Measurable Verbs Used to Assess Learning Outcomes. clinton.edu

Resources/Exercises for ice-breakers

- Mindtools. Ice Breakers. Easing Group Contribution. mindtools.com
- Students as Partners, Teaching, Learning and Support Office. Peer Support Icebreakers. documents.manchester.ac.uk
- The balance careers. The 10 Best Icebreaker Activities for Any Work Event. Activities for Meetings, Training, and Team Building Sessions. thebalance.com

References

- Ambrose, Bridges, DiPietro, Lovett, Norman and Mayer. How learning works. Highlights summarized by Brent and Felder. Seven research-based principles for smart teaching"; ISBN: 978-0-470-48410-4.
- Dale (1969). Audio-Visual Methods in Teaching, 3rd ed., Holt, Rinehart & Winston, New York, p.10.
- Döring (2008). Handbuch Lehren und Trainieren in der Weiterbildung. Beltz Verlag (Weinheim, Basel).
- Fung (2017). A Connected Curriculum for Higher Education. UCL Press. ucl.ac.uk
- Felder and Brent (n.y.). Active learning. An introduction. PDF
- Kirkpatrick and Kirkpatrick (1994). Evaluating Training Programs, Berrett-Koehler Publishers.
- Mazur (2014). Peer Instruction for Active Learning. Serious science. video
- Owen Wilson (2018). The Flipped Classroom. thesecondprinciple.com
- Prince (2004). Does Active Learning Work? A Review of the Research. PDF
- Rowe (1986). Wait Time: Slowing Down May Be A Way of Speeding Up! Journal of Teacher Education, 37(1), 43–50. doi.org/10.1177/002248718603700110
- Siemens (2006). Knowing Knowledge. PDF

For a deeper understanding of the matter:

- Knowles, Holton, and Swanson (2011). The Adult Learner: The Definitive Classic in Adult Education and Human Resource Development. Oxford: Butterworth-Heinemann.

Organizational aspects

This chapter will guide you through the main practical aspects of organizing a training event. Of course, what you need and will use will depend on the type of event you'll organize! The checklist should be adjusted accordingly. You will get information on preparation steps and necessary organizational tasks. This will provide you not only with valuable knowledge about event organization, but will reassure you while preparing your training. Note that most of the material in this chapter, and the whole handbook, is focused on training regarding practical workshops. Running a different type of event may require different decisions than the recommendations that follow.

Training event basics

Format

Deciding what type of event you want to coordinate is the first critical step in training exercises. Here are some points to consider:

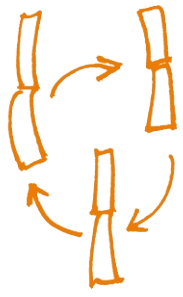
- Format of the training: live workshop, seminar, lecture, online training or mixture of online and in-person?
- Will it be participatory, formal, self-contained?
- Can the event be integrated into existing curricula?
- Do you need to invite any other external experts? What are the requirements for that (e.g., funding)?
- Is the training a requirement, or something participants are choosing to attend?
- Will attendees receive any form of accreditation for the training?
- What sort of venue type do you need for this format?

To provide you with initial guidance over possible types of training and their characteristics, see the table below for recommendations.

	TYPE OF TRAINING			
	Live workshop	Course/ class	Lecture	Online Training
Audience Size				
less than 20	x	x	x	x
less than 40		x	x	x
more than 40			x	x
Funds				
none			x	x
little	x	x	x	x
loaded	x	x		
Time				
less than ½ day	x	x	x	x
½ - 1 day	x			
1- 4 days	x	x		
more than 4 days			x (series)	x (series)
Training level				
Introductory			x	x
Aware of	x	x		x
Intermediate	x	x		x
Advanced	x	x	x	x

Audience, guest speakers, and partners

Before committing to the event be sure you defined your target audience and that you are aware of their needs. Consider your audience, its size and the number or area of competence of (guest) trainers.



Cooperating with others

Some forms of training require more than one instructor. Try to get support from colleagues or service units in your institution. Identify institutional support (e.g., funding, room(s), work time) and reach out to decision makers to ask for these things - for example, you could ask for help with registration, or contact the printing service or communication department regarding advertising. Make sure any volunteers are sufficiently briefed on all activities, and know what the aims and practicalities of the event are. Make them aware of the importance of encouraging participation from the attendees. You can also outsource some tasks, if the budget allows for this.

Consider partnering with other departments at your institution or other local institutions to pool resources and increase impact/collaborating with other projects or programs. These are the key points to work out prior to committing to, or announcing any event. Resolving these will help the training run smoothly for yourself and your participants. Also, consider integrating the training into a recognized conference or local/international event.

Identify other trainers or experts/guest speakers that could help with the event. Ideally, these will be other Open Science advocates at the institution or otherwise local to the event, but you may need to find suitable non-local trainers (who may need financial support for travel). Work to have diverse representation (see Representation below). According to The Carpentries, a workshop with 40 people needs at least two trainers (and possibly a third) who alternate between talking and supporting learners, including also one helper per 5 participants that will continuously monitor for any issues.

Representation

Maintaining an inclusive environment is important for any successful training event. Ensure that each component of your program includes a range of backgrounds. Your organizing team, speakers, and trainers should include representation across gender identities, different disciplines, underrepresented groups, diverse racial backgrounds, and geographic regions (if you intend to open your event to non-local participants).

Actively invite trainers and speakers from underrepresented groups. Make sure to discuss with them their specific goals and needs, and include these in the planning of the event. To learn more about trainers see On learning and training chapter, Expectations about a trainer subchapter. Ensure that a proportion of participant spots are reserved for attendees across ethnic backgrounds, gender identities, disciplines and geographic regions (see Inclusive engagement). To learn more about how to make your workshop inclusive and welcoming, see the Conference Planning Checklist by SPARC.

Venue

Before organizing a face-to-face training event consider few things related to a venue. It might help you to reduce several obstacles:

The venue should be easily accessible for the participants. The venue should have elevator access, accessible entrances and ramps as well as clear legible signs. Check if the venue is easily accessible by public transport or car (parking spaces) and that it’s not too far away from rail stations or the airport. For a checklist of what makes a workshop accessible, see the Accessible Meetings Toolkit from the American Bar Association and the Conference Planning Checklist by SPARC. Locate a place to greet your attendees and a place for them to circulate and socialize. A separate area for catering should be available. Also, check if the venue offers a maternity room, a prayer room and a gender-neutral washroom.

The training room should be sufficiently equipped (see equipment and media). The room should have sufficient WiFi and power access for every participant (possibly via power strips/extension cords). Check to see if furniture can be rearranged in order to suit your requirements. The presenter will need a high (or raisable) table for standing and a microphone (for recording and/or accessibility). An additional microphone for participant questions aids accessibility.

Timing

The length of the event depends on the content and depth of the training you intend to provide. You should have an estimate how much time each component will take. Make sure to define an agenda or time schedule, including any icebreakers and introductions. Allow enough time for lunch and coffee breaks. Be reasonable with your start and end times (see chapter Starting the training).

Before scheduling your event think about obstacles that might prevent or induce people to join and try to pick a suitable time and date of the event. Make sure to avoid conflict with any public holidays, religious holidays, or similar events. If your event is hosted at a university, keep class schedules in mind. Consider to place your training session along with a larger conference or meeting in order to bring more attention, increase the attendance and get the chance to bring any speaker attending the other event. A family friendly workshop should avoid evenings and weekends, provide childcare or childcare sponsorships, and ensure areas for breastfeeding mothers.

Budget

You may need financial support to help run your event, to pay for things like the venue (if the host institution cannot or will not provide this for free), travel support for non-local trainers/experts, refreshments, materials (e.g., name badges, USB drives) and swag. Most types of training will need at least a little money for material and equipment. Also, keep in mind that the costs associated with human resource are often the largest costs associated with running an event. It is good to identify time needed for staff to prepare materials and content which is often not budgeted for. These costs may be covered through as a core aspect of the job, but if not it might be wise to ensure funding to cover this aspect is sourced.

Consider different ways of creating a budget for your training. If possible, request funds from your institution. Otherwise, you might have to charge a fee from participants or look for scholarships and other ways of funding.

Fee

Collecting and managing funds or fees can be tricky. If you plan to do this, you should consider using an existing online registration service (e.g., Eventbrite, Event Smart) or your institution’s conference/event services. Although any cost impacts accessibility of the event, charging a nominal registration fee (e.g., $20–40 or €15–30) encourages those who register to actually attend—Software Carpentry found this reduced no-shows from almost a third to about 5% (Wilson 2016).

If planning to charge a fee of any sort, it is good to clarify with your institution's finance team how best to handle this. In some cases, the amount of time/effort required to set such things up can outweigh the value of charging - particularly if it is only a nominal fee being charged. Your institution will likely have specific financial processes and budget codes that need to be used, so speak to them early on to see what the best approach is. This is true even when using external services such as Eventbrite (you'll need an institutional budget center to allow the income to come into your institution).

If you do charge a fee, consider making a waiver available upon request for those unable to pay or creating scholarships. Scholarship allocation should be prioritized for groups that face the most barriers for self-funding.

Funding

You can get funding from a few difference sources: the host institution, external sponsors like companies, budgeted funds on faculty/principal investigator's grants, or through registration fees. Check if there are any internal sources for funding, or relevant local organizations who can sponsor your event. If you have found a potential partner, check the funding conditions. This could include advertising on your event website or at the event itself.

Consider different levels of sponsorship (bronze, silver, gold) in case of bigger events. You might also want to look at other projects or programs to co-organize and share costs with.

Organizational tasks

Equipment & Media

Long-term preparation

Here are some things to consider:

Will participants need access to WI-FI? Make sure that any requirements for access are dealt with ahead of time (e.g., by providing guest account details). Check if the venue has enough power outlets. Make sure to check with the venue owner in advance for availability of technical support. If you are planning on recording the event make sure you have the correct equipment, and that attendees are aware (and have consented) to being recorded. Think about how you are going to license any outcomes: will you apply a CC license to pictures, videos, and training materials? Are the authors ok with that?

Shortly before event

Making sure that all of your equipment, media, and materials are in fully functioning order can help to avoid any embarrassing hiccups during your training. Make sure that your laptop, or the device which is hosting your material, is compatible with the media technology in the venue. Ask guest lectures for their presentations in advance and store them all on the same laptop. This will make it easier to switch from one speaker to the other. Make sure to bring any relevant adaptors or extensions. Check WiFi strength and power outlets, as well as, if the speaker and projectors work in advance, and that your file formats are supported. Make sure there is an emergency contact for technical issues.

Make sure to print out any paper handouts in advance, and to have enough of them to go around. If you plan to hand out a lot of material, consider providing folders or binders to help with organization. Or, consider just making all your material available digitally via your event website.

Preparing a variety of media can help engage an audience with diverse learning styles. You should prepare any teaching aids in advance (e.g, flipcharts, practical exercises, games). Bring notepads, post-it notes, pens, thumbtacks. If participants need any other computer-based materials make sure these are well-organized and available in advance.

During the event

If your equipment fails, do not panic. Call the IT support and explain the problem to the attendees. Most people understand that. What might feel like hours to you are just a few minutes of lost time. If the equipment still not works try to work offline with flip charts for example. If you are relying heavily on media equipment and it is just a small group of participants suggest to reschedule the training.

Marketing & Advertising Strategy

Long before the event

Developing a strong marketing and communication strategy is fundamental to driving participation, as well as teaching you how to develop and refine your messaging.

Consider which kind of name your training will have. Think about your framing and messaging. What are the common values that you can appeal to? For example, will you run an "Open Access workshop", or a workshop on “How to get published”? How are you going to get people in the room? Remember, training is not unidirectional, and can be incentivized by framing it as a networking opportunity. For example, find some partners in Graduate Schools, Master Schools, Support Staff trainings, Valorisation Center etc.

Consider both digital and non-digital media. Use institutional mailing lists and social media (e.g., Twitter, Facebook, blog). Will you have dedicated social media profiles? What sort of content will you share on them? Think about relevant images and logos. This is more important if you want to run more than one event. If the event is being run with the sponsorship of, or in coordination with, an institutional organization (e.g., the library, a particular college/department), then you may want or need to use the profiles of the organization. This might require someone else to post the material, so keep that in mind. Several of these recommendations might require organizational sign off or additional budget support - start investigating these options as soon as possible.

Find out if you can post flyers or posters at your institution. Are you going to design a poster? What sort of logos, images, text, and information do you need to include? Make sure to clearly communicate the pre-defined objectives (skills and knowledge). Ask relevant organizations to help with advertising. Connect with relevant media, create a press release. Use existing communication channels, e.g. at the university library you might ask subject librarians to promote the event to their academic communities.

Shortly before the event

Send a reminder on social media and mailing lists. Put up signs so your attendees find the room.

During the event

Publish pictures and short videos from the event on the website and social media. Tell participants the hashtag for the training and ask them to send at least one tweet/message during the event. Collect reasons for attendance for advertising of future events.

Registration

Long before the event

Set up an event registration using a service like Eventbrite or Event Smart (which are free for free events, but may include fees if your event has a registration cost), or something like Google Forms to capture basic information. For smaller events you can also use registration via email. But don't forget to send them a confirmation, when they register and before the event to send a reminder.

Think about the fee you want/need to charge (see budget). Think about the credits students can get. Is a certificate needed (see certification of attendance)?

Be sensible and transparent about the information you collect. If you need to ask information like gender, age or nationality, keep into account that this is not always as straightforward as you might think - always offer the option of a blank field. Please do not use the distinction between Mrs. and Ms.

You can make a short poll to measure what do participants already know about the topic (their pre-knowledge). It can help you to prepare training material. Make clear what data is going to be shared/retained and why. Always offer people the option of opting out, and keep any information you do archive safely stored. Consider creating a list of interested participants for a newsletter or for keeping in touch, but be aware of data protection (like the EU General Data Protection Regulation (Regulation (EU) 2016/679)).

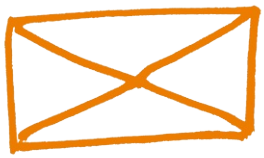
Shortly before the event

Depending on the size of the audience, provide a separate staffed registration desk. Make sure staff has all information including a participants list, and let them take care of badges and attendance sheets/certificates.

If there is no separate registration desk, prepare a cheat sheet with information to keep at hand (think: public transportation, emergency numbers, requests for certificates, safety during the event etc.).

During the event

Do you have consent from participants to re-use or share their contact information or to take pictures and publish them? Did all participants sign the participants list?



Communication

Long before the event

Prepare and send formal invitations to participants, guest and keynote speakers.

Create a website for the training event, such as using GitHub Pages or on an institutional website. [link to examples/template]

Make sure any key resources are visible and accessible if needed. If you want the participants to come with research outputs (e.g., papers, code, data) for exercises, let them know with plenty of time to prepare (and consider making this optional).

Shortly before the event

Communicate requirements to your audience in advance.

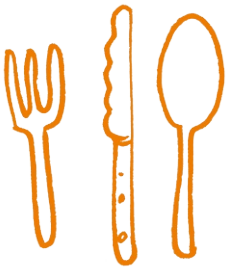
- Let them know if they need to bring laptops or other work materials.
- Make sure any prerequisites for software or programming abilities are communicated in advance.
- Provide basic contextual reading materials, so you don’t have to start at the beginning point.

Send a reminder email to your attendees a day or two in advance of the event, if possible (this may not be necessary if you are relying on a registration service).

Remind people about reachability and accessibility of the venue. Send detailed instructions for parking and public transport options.

During the event

Dedicate some time to housekeeping at the start of your event. Write down hashtags and WiFi passwords.



Catering

Long before the event

What refreshments will you either need to provide, or will people need to bring their own? If you provide refreshments, you may need to obtain funding or charge for registration.

If relevant, you can ask during registration in advance for dietary requirements - but keep in mind this might make it very complicated for you. Sometimes it’s better to ask the caterer to provide sufficient varieties (vegetarian, vegan, gluten free, etc.) and add one free field on your submission forms so that participants can fill in specific requests if necessary (e.g. intolerances and allergies).

Shortly before the event

Check the venue and inform the caterer where and when to deliver the refreshments.

During the event

Be sure you have the contact information of the caterer if the catering is not showing up, delivering the wrong lunch or forgot something.

Code of Conduct

Long before the event

To help ensure your workshop is a friendly, inclusive, and respectful environment for trainers and participants, identify or create a robust Code of Conduct (CoC) for your event. Make sure the Code of Conduct is communicated in advance—we recommend prominent placement on your event website (see task 2) and onsite. Participants should be asked to review and acknowledge the Code of Conduct while registering for the workshop. Included in your Code of Conduct should be clear consequences of violation (for example, removal from the workshop). Ensure the reporting process for violations is communicated clearly before and during the event and that at least one designated organizer is identified as the point of contact, who is easily accessible to receive reports of code of conduct violations. Examples you can borrow or adapt from include:

- The Mozilla Science Lab Code of Conduct
- Contributor Covenant Code of Conduct
- FORCE2017 Conference Code of Conduct
- The Carpentries Code of Conduct
- Mozilla Science Lab: Getting Started with Codes of Conduct

Shortly before the event

Make sure the Code of Conduct is clearly visible/accessible from the event website (if one exists); if your event does not have or need a website, print out the CoC and give it to participants.

During the event

Make sure there is a safe space for participants to report any breaches of the Code of Conduct. Communicate sanctions, and follow through if any breaches occur.

Certification of attendance

Long before the event

Prepare a template and assign who will keep records or monitor registration process.

Shortly before the event

Prepare a generic certificate of attendance with event or organiser’s logos and event information that can be distributed digitally when requested.

During the event

Ask participants if a certificate of attendance is needed.

If a signature sheet is required, make sure you do a check during the day or ask to complete it at registration.

Signs

Long before the event

Check the venue and define spots to be marked by signs to help participants to easily find a room.

Immediately before the event

Design, print and place the signs and leave useful information at the reception desk.

During the event

Remove the signs after the event.

Social media and notes

Long before the event

Plan your social media activities, ask colleagues from other departments and/or partner organization to help you in sharing information.

Immediately before the event

Prepare note documents (e.g. public Google Docs or etherpads). Make announcements on social media.

During the event

Ask your audience whether they are ok with being filmed, photographed and featured on social media. If it's a big audience, you might consider handing out stickers to those who do not want to be featured.

Assign note takers and people responsible for social media. Ideally, rotate heavily to avoiding slacking and loss of attention.

Event closure

Venue

Make sure you leave the venue neat and clean, unless your agreement for using it doesn't require this.

Debrief

Debrief with the other trainers/speakers to self-assess how the event went.

Evaluation

Send post-training assessment survey to participants (see Training evaluation) or distribute an evaluation form during the event and make sure people hand it in at the end.

Read and count the questions in the evaluation form. Make your self-evaluation.

Dissemination

Upload all the material used during the event (presentations, documents) if they were not available beforehand. Make sure to provide open licenses if possible, and make sure participants are not identifiable (e.g., within a notes document).

Prepare a report for your funder or institution and if needed make it public (e.g. blog, twitter, website).

Checklist

What	When and who?	Done?
Equipment/media		
Determine what technical equipment is needed		
Check if enough power outlets are available		
Order WiFi for participants		
Organize video recording and taking pictures		

Test equipment a few days before the training		
Print out handouts, feedback forms and material for exercises or publish them online		
Prepare flip charts and pinboards		
Venue		
Check elevator access, accessible entrances, ramps		
Check public transport and parking availability		
Locate maternity room, prayer room and gender neutral washrooms		
Clear, legible signs		
Brief your helpers before the event		
Marketing/advertising		
Identify communication channels		
Set up online presence		
Send event information to mailing lists		
Inform about your event in social media		
Registration		
Set up registration module		
Collect information on dietary needs and allergies		
Ask for childcare needs		
Provide hotel information for events over several days		
Send confirmations/invitations to attendees and provide clear text and image instructions to the venue		
Send a reminder 1 or 2 days before the event		
Prepare name tags and print participants list		
Prepare a registration desk		
Organize a wardrobe checkroom for larger events		
Catering		
Identify catering options and needs		
Order catering		
Check if meals are clearly labeled (especially regarding dietary needs and allergies)		
Communication during event		
Inform the participants where to find emergency exits, food/beverages and restrooms etc.		
Hand out consent forms for video recordings, live streaming and/or photos		
Post event dissemination		
Make photos of flip charts and other non-digital material or results		
Hand out or send certificates of attendance		
Provide or send training material (slides, notes, video recordings) to the attendees		
Provide a report for your funder or institution		
Evaluation		

Hand out or provide an online or printed form for feedback		

References

- Christodolou et al. (2014). How to conduct a successful workshop: The trainee's perspective (Arab Journal of Urology), 12(1), 12-14. doi.org/10.1016/j.aju.2013.08.004
- Commission on Disability Rights of American Bar Association (2016). Planning Accessible Meetings and Events. A toolkit. PDF
- Pavelin et al. (2014). Ten simple rules for running interactive workshops, *PLOS Biology*. doi.org/10.1371/journal.pcbi.1003485
- SPARC (n.y.). Diversity, Equity, and Inclusion. Conference Planning Checklist. sparcopen.github.io
- Wilson G. Software Carpentry: lessons learned [version 2; referees: 3 approved]. F1000Research. 2016;3:62. doi.org/10.12688/f1000research.3-6e 2.v2

Inspirations

- International Council on Archives (2010). Organising training workshops and seminars: Guidelines for professional association. PDF
- Software Carpentry. Workshop Operations. software-carpentry.org
- Software Carpentry. Teaching and Hosting. Admin Checklist. software-carpentry.org
- Wikihow. Conduct a Workshop. wikihow.com

Examples & Practical Guidance: adopt, adapt, develop

In this chapter, you will find a wealth of materials to help you actively engage your trainees in critically examining Open Science issues.

We recommend you approach all of these materials with the motto "Adopt, adapt, develop" in mind—meaning that its best to re-use what exists where possible. Hence, before you start developing training resources from scratch you should find out whether there are existing resources you may use. We give some example resources here, with tips for how they could be adapted for your purposes. We also provide links and strategies to help you find further material. In some cases, existing resources may be used as they are, so you may simply **adopt** them. An example at stake may be an openly available video tutorial about open file formats which you may point your audience to. In other cases, you may have to **adapt** existing resources somewhat in order to make them fit your purposes. For example, you may need to add/replace some institution- or country-specific references to an existing overview of Open Access requirements issued by research funders. Only as a last resort you should **develop** your own training resources from scratch. If you want to develop your own training materials, be sure to develop Open Educational Resources so that other trainers can reuse and adapt your materials.

Example training structures

Open Science Göttingen Meet-ups at the University Library at Uni Göttingen **(3 hours)**

The Open Science Network Göttingen, a group of researchers and librarians who support open science practices and knowledge exchange regularly organize these meet-up events where various open science related topics are discussed. The network unites people interested in Open Science topics at the Göttingen Campus and is open to everyone. They have become quite popular attracting scholars from different disciplines who are eager to discuss their experiences with open scholarship and to learn about new methods, tools, and practices. Invited speakers usually introduce the topics which is followed by small group discussions with a more in-depth view on related issues.

- More information: State and University Library Göttingen - Open Science

Mozilla Study groups (a series of 2–3 hour meetings)

Study groups are communities of peers (e.g., from the same institution) committed to learning and teaching each other. They’re fun, informal meetups allowing participants to share skills, experiences, and ideas around open science, open source, code, and community in research. The goal of the Mozilla Study Group Project is to support this kind of peer-to-peer study by providing a simple set of tools, template lesson plans, and access to an international community of like-minded researchers and avid learners in code (text adapted from science.mozilla.org/programs/studygroups)

Reproducible analysis and Research Transparency (a single full-day workshop)

Transparency, open sharing, and reproducibility are core values of science, but not always part of daily practice. A first iteration of this workshop took place within the context of the Open Science Tools, Data & Technologies for Efficient Ecological & Evolutionary Research event, organized by NIOO-KNAW and DANS-KNAW. It provides an overview of current status in reproducible analysis in order to provide transparency in research. The workshop covers methodological topics (such as the use of the Open Science Framework

and reporting guidelines) as well as software tools (such as Git, Docker, RMarkdown / knitr, and Jupyter). Going beyond simple listing and presentations, the second half of the workshop focuses on hands-on skill building, with exercises and tutorials covering most of the software aspects. Material and content is available here: reproducible-analysis-workshop.readthedocs.io

Open Science: what's in it for me? (1-2 days)

The aim of the workshop is to provide researchers and administrators with hands-on examples of Open Science tools and workflow examples across various disciplines, and to start applying and discussing these. For this, we present an overview of Open Science practices and tools that are used throughout the scientific workflow, with practical examples, audience polling and interactive discussions. The second day is oriented at application and sharing. In various rounds participants explore and where possible try out or apply tools and practices. They do this in small groups and individually and also in a lively marketplace. In a final session we have a discussion on obstacles and incentives for switching to open science in your own research.

- Open Science - what's in it for me (Vienna, 2017, workshop report)
- Open Science - what's in it for me (Torino, 2018, workshop program)

Carpentry workshops (2 days)

A Carpentry workshop is a hands-on two-day event that covers the core skills needed to be productive in a small research team. Short tutorials alternate with practical exercises, and all instruction is done via live coding. Software Carpentry was founded in 1998 and Data Carpentry was founded in 2013. Both focus on computational skills, run two-day workshops taught by volunteer instructors, and strive to fill gaps in current training for researchers. However, they differ in their content and intended audience. Data Carpentry workshops focus on best practices surrounding data. Its learners are not people who want to learn about coding, but rather those who have a lot of data and don't know what to do with it. Data Carpentry workshops are aimed at pure novices, are domain-specific, and present a full curriculum centered around a single data set. Software Carpentry workshops are intended for people who need to program more effectively to solve their computational challenges, are not domain-specific, and are modular—each Software Carpentry lesson is standalone.

- Software Carpentry
- Data Carpentry

EIFL Train-the-Trainer program (4 days)

EIFL organized a train-the-trainers program for five universities in EIFL partner countries (Ethiopia, Ghana, Zimbabwe, Tanzania, and Nepal) that have committed to integrating open access, open science and open research data into courses for PhD students. Day 1 covered open access and open data. Day 2 and 3 were dedicated to open science across the research workflow, including current practices at participant's universities. On Day 4, participants designed and prepared their own training program.

- EIFL Train-the-trainer program (Addis Ababa, 2017, program and materials)

Open Science summer schools (5 days)

Various universities across Europe organize weeklong summer schools on open science, primarily aimed at early career researchers. These events cover a variety of topics in five days, usually with many hands-on activities to apply open science into daily practice.

- EPFL Summer school Open Science in Practice (2017, program overview)
- Utrecht University Summer school Open Science and Scholarship (2017, program and materials)
- Essex Summer school in Social Science and Data Analysis - Introduction in Open Science (2017, program overview)
- LERU Doctoral Summer school on Data Stewardship (2016, description, learning objectives)

Monday	Tuesday	Wednesday	Thursday	Friday
Introduction	**Discovery**	**Writing**	**Outreach**	**Assessment**
• Introduction to OS & workflow model • Motivations & barriers • Cultural & disciplinary differences	• Searching for (OA) literature & data • Getting access • Open annotation/curation	• Collaborative writing • Plain language explanations/writing • Contributorship roles	• Defining your audience • Academic social networks • OA/OS advocacy	• (Alt)metrics for research objects • Narrative assessment • Open commenting • Open peer review, PPPR
Preparation	**Analysis**	**Publication**	**Outreach (cont'd)**	**Wrap-up**
• Funding • Publication strategy • Citizen science	• Documenting steps • Collaborating on code and documents • Replication and reproducibility	• CC-licenses and copyright • Preprints • Journal types and journal selection	• IRL outreach actions	• Motivations and barriers • Concrete actions, next steps • What did you miss & how to cover that
https://tinyurl.com/summerschoolopenscience				

Program schedule Summer School Open Science and Scholarship, Utrecht University 2017

Example Exercises

Master Template

- Format, time needed
- Topic (see Open Science Basics)
- Learning objectives
- Exercise description
- Materials and tools needed
- Level of prior knowledge needed
- Things to bear in mind
- How to adapt for other purposes

Use this Google form to suggest additional exercises!

Types of exercises

```
* quick warm-up / short break exercises
* small group exercises
    * role-play
    * discuss OS topics/statements
    * marketplace: exchange experiences/expertise
    * meeting with researchers / policy makers
    * ...
* plenary exercises
    * collaborative mapping
    * simulation game
    * inventorizing
    * card games
* presentations
    * role-play
    * present real-life cases/examples (also by participants)
    * one-minute presentations of a concept (by participants)
    * guest lecturers
    * ...
* hands-on exercises (individual or in pairs)
    * visualizing
    * explore / try out tools & platforms
    * implement an open science practice in your own research
    * check reproducibility of a research paper
    * …
```

Example exercises (including materials)

	Title	Topic	Type	Duration
1	Line up!	general	whole group	5-10 min
2	Prioritization of training needs	Open Concepts and Principles	whole group	10 min
3	Selection of Open Science practices	Open Concepts and Principles	whole group	1-1.5 hour
4	Open Science discussion topics	Open Concepts and Principles	small groups	20-30 min
5	LIBER Open Science café	Open Concepts and Principles	small groups	1.5 hour
6	What is research data for me?	Open Research Data and Materials	individual / pairs	15 min
7	Why not share data?	Open Research Data and Materials	small groups	20 min

8	"Open Data Excuse" Bingo	Open Research Data and Materials	whole group	20-30 min
9	Me and my data - Datagramms	Open Research Data and Materials	whole group	1-4 hours
10	Find your data publisher	Open Research Data and Materials	individual / pairs	10-15 min
11	What do you need for a data publication?	Open Research Data and Materials	whole group	10 min
12	Creating metadata	Open Research Data and Materials	individual / pairs	5 min
13	Get started with sharing software openly	Open Research Software / Open Source	individual / pairs	20-30 min
14	Establishing a Reproducible Data Analysis Workflow	Reproducible Research and Data Analysis	individual / pairs	4-8 hours
15	Choose the right version for the repository	Open Access to Published Research Results	individual / pairs	15-20 min
16	Open file formats	Open Licensing and File Formats	whole group	10-15 min
17	Creative Commons License matching	Open Licensing and File Formats	whole group	5-10 min
18	OER Remix	Open Licensing and File Formats Open Educational Resources	whole group	10-15 min
19	Open peer review - participants openly review each others' texts	Open Peer Review, Metrics, and Evaluation	small groups	90 min
20	Open peer review - your 2 cents	Open Peer Review, Metrics, and Evaluation	whole group	1.5 hour
21	Taking a stance	Open Science Policies	whole group	10 min
22	Plain language explanations (in progress)	Citizen Scientists and Science Communication Collaborative Platforms	small groups	2-3 hours
23	Devil's advocate - convincing the skeptics	Open Advocacy	small groups	30 min
24	Set up OSF project & link to other platforms (in progress)	Open Research Data and Materials	individually or in pairs	
25	The publishing trap (in progress)	Open Access to Published Research Results	small group exercise	2 hours
26	(in progress)	Open Research Data and Materials	small group exercise	4 days (5 hrs/day)
27	Train-the-trainer card game for Open Science training	Open Advocacy	small group exercise	2 hours

Example 1: Line up!

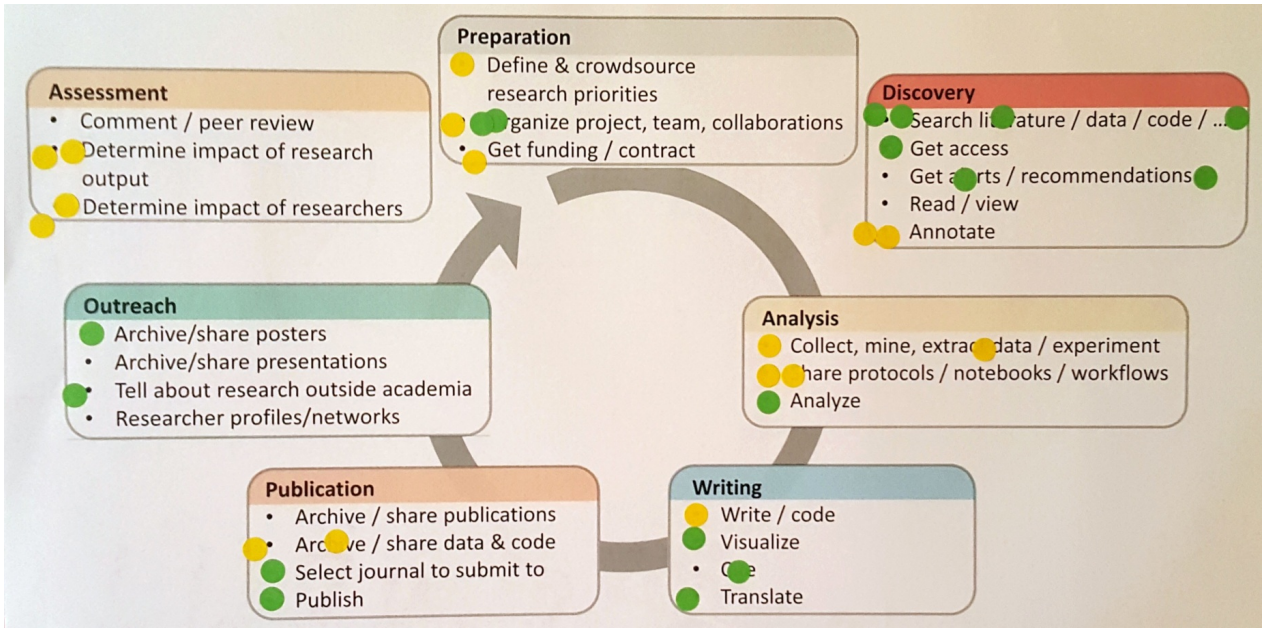
- Format, time needed
  - Group exercise, 5–10 minutes
- Topic
  - Icebreaker, can be on topic or unrelated
- Learning objectives
  - Get participants to loosen up
- Exercise description
  - Imaginary line in the room forms a spectrum between 'strongly agree' and 'strongly disagree'. One participant, or the

moderator, makes a statement (can be on topic ‘closed data should not be cited’ or off-topic ‘leggings are not trousers’. All participants have to position themselves along the imaginary line. The moderator asks some participants to explain their (literal) standpoint.

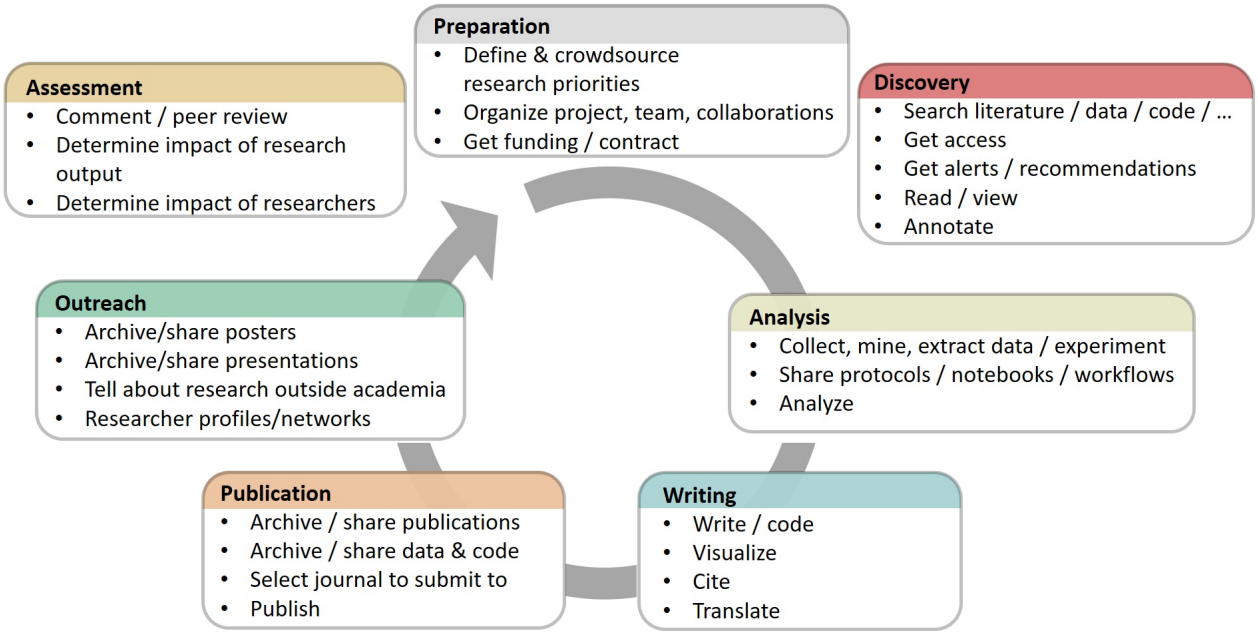
- Materials and tools needed
  - None
- Level of prior knowledge needed
  - None
- Things to bear in mind
  - Make sure not only the opinionated people are talking. Ask people who linger in the middle to explain their point of view.
- How to adapt for other purposes
  - Adapt the type of question to the situation. For a new group, allow people to make an off-topic or trivial statement, but the technique can also be used to test the waters on certain controversial subjects related to the topic of the workshop, especially with people who have been working together for a while already (e.g., on a second day of a workshop)

Example 2: Prioritization of training needs

- Format, time needed
  - Plenary, ~10 minutes
- Topic
  - Open Concepts and Principles
- Learning objectives
  - Identify knowledge gaps / areas participants feel they would most benefit from training in.
  - (optional) Identify areas participants feel knowledgeable about (and can thus share their own knowledge).
- Exercise description
  - Briefly introducing the research cycle and activities therein.
  - Ask participants to individually identify two to three activities they would most benefit getting training in (in relation to open science).
  - Optionally, also ask participants which two to three areas they already feel knowledgeable about (again, in relation to open science).
  - On individual printouts, participants add sticky dots for each question.
  - Participants then add similar sticky dots to the communal printout.
  - Discuss the results with the full group. Make sure people when seeing the dots also realize there may be a big opportunity for learning from other participants.



- Materials and tools needed
  - Printout of research cycle with activities: one for each participant and a communal one
  - Sticky dots in two colors

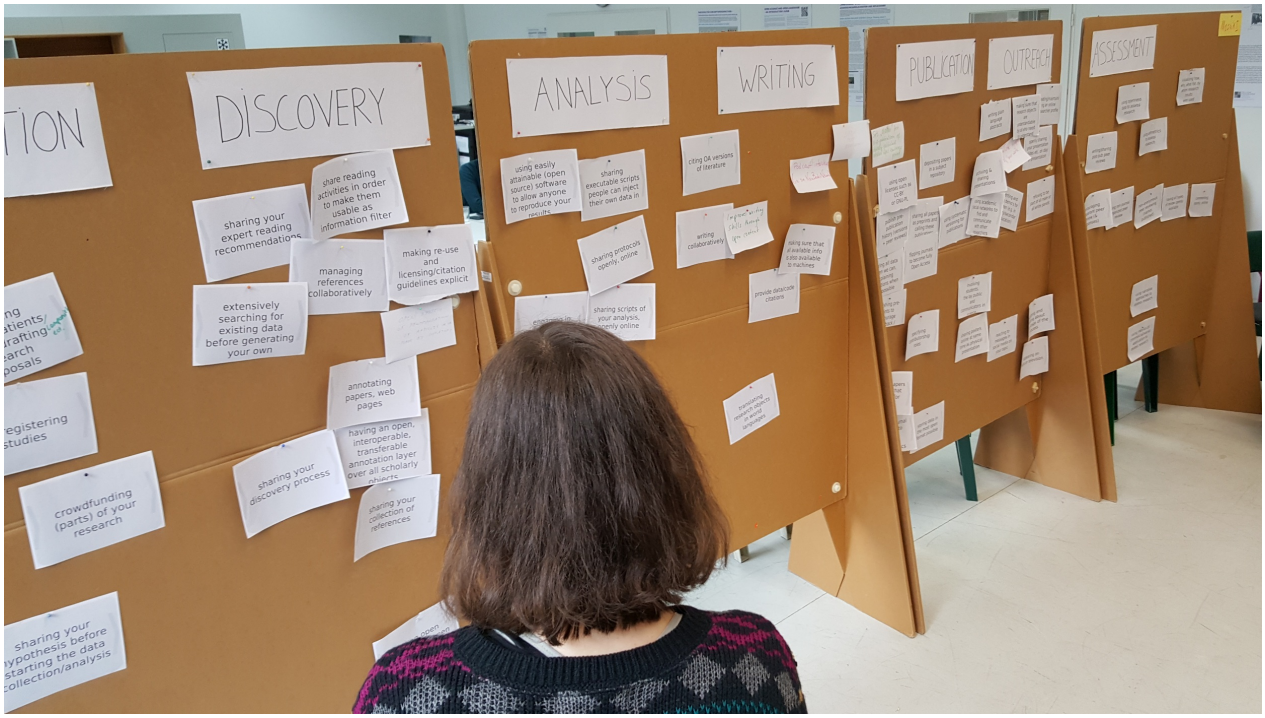


- Level of prior knowledge needed
  - None; some familiarity with the research cycle is helpful.
- Things to bear in mind
  - Best at the beginning of a longer training program where multiple topics will be covered.
  - For the sticky dots, choose a combination that is colour-blind friendly.
  - The number of activities to choose depends on the number of participants (e.g., three for smaller groups, two for larger groups).
  - Individual printouts are used to prevent peer pressure / bias.
  - Individual printouts can be kept for reference during the remaining of the training.

- How to adapt for other purposes
  - This exercise can easily be adapted to prioritize other topics.

Example 3: Selection of open science practices

- Format, time needed
  - Plenary, 1–1.5 hours
- Topic
  - Open Concepts and Principles
- Learning objectives
  - See the spectrum of open science practices across the full research workflow.
  - Assess which practices would the most feasible and effective to focus on.
- Exercise description
  - Prior to the exercise, sort the cards according to research phase/activity and spread them across the room (e.g., on tables, or on a large section of the floor).
  - Mark a large section of a wall (windows or pinboards can also be used) with the different phases of the research cycle (e.g., preparation, discovery, analysis, writing, publication, outreach, assessment).
  - Ask participants to select practices they feel are really important for open science, and hang them on the wall, grouped by research phase.
  - Encourage people to add research practices that are not included in the cards.
  - Divide participants in seven groups.
  - Each group looks at the selected practices for one research phase, and chooses the two practices that they feel are *most feasible to implement* and *most effective to make research more open*. Either move these cards higher up on the wall, or remove the other cards.
  - The small groups explain their choice to all participants.
  - Together, the selected research practices can form a blueprint of an open science workflow.
  - As a follow-up exercise, participants can discuss possible steps to implement these practices:
    1. what tools/platforms can be used
    2. what potential incentives and barriers would be
    3. what support would be needed
    4. what policy changes would be needed



involving public/patients etc. in drafting research proposals

extensively searching for existing data before generating your own

translating research objects in world languages

acknowledging contributor roles in a publication

refusing to be part of all male or all white panels

using metrics of commercial/social application to assess research

- Materials and tools needed
  - Large wall, windows, or multiple pinboards to hang materials on
  - Enough room to move around
  - Printed cards with open science practices (also available as editable powerpoint slides or in a Google spreadsheet)
  - Empty cards, pens / markers
  - Pins or tape
- Level of prior knowledge needed
  - None, some familiarity with the research process is helpful
- Things to bear in mind
  - Depending on the number of participants, small groups can prioritize practices for more than one research phase.
  - Test tape on windows / walls first, some types are really hard to remove :-)
  - The whole group may not agree with the small group's selection of practices for a given research phase. Decide beforehand whether to stick with the choices made, or whether there is room for discussion and consensus-based swapping of practices.
- How to adapt for other purposes
  - The exercise could be modified to focus on specific activities / a specific phase of the research cycle (e.g., publication or assessment).
  - Other selection criteria could be used, e.g. practices participants use themselves, or practices that would be most ideal (independent of feasibility/efforts needed).

Example 4: Open Science discussion topics

- Format, time needed
  - Small groups, 20–30 minutes
- Topic
  - Open Concepts and Principles
- Learning objectives
  - Confront own experiences and opinions on open science with perspectives from others.

- Exercise description
  - Divide participants in groups of four or five and distribute discussion topics (e.g., printed out on paper).
  - Have groups discuss the topics from participants' own perspectives.
  - (optional) Have each group summarize most important points that came up for the whole group .
  - Suggestions for discussion topics:
    1. "Working in an Open Science manner makes research more fun"
    2. "Scooping is a real and existing problem that makes Open Science a hard choice"
    3. "APCs (article processing charges) are the main obstacle to publishing more in Open Access"
    4. "We need more explicit support for Open Science from funders and the government"
    5. "Engaging in open peer review is problematic for young researchers that want to make a career"
    6. "We should take citizen scientists more seriously, and also not just see them as data suppliers"
    7. "Impact factors are a symptom and not the cause of the publishing rat-race"
    8. "There is absolutely no reason we should not publish a paper as a preprint as soon as it is ready"
    9. "Just sharing our data is fine, but to speed up science we need to also work on interoperability and reusability of those data"
    10. "Sharing ideas and projects through ResearchGate is a good way of doing outreach for our research"
    11. "Demands of our PIs are probably the main reason why young researchers do not engage more in Open Science"
    12. "We should strive to create a kind of 'commons' where we share all our research outcomes/objects to foster collaboration and reuse"
- Materials and tools needed
  - Printouts of discussion topics
- Level of prior knowledge needed
  - Some familiarity with the research system.

- Things to bear in mind
  - This exercise is best suited to researchers (rather than support people), because they can directly relate to their own situation and speak from their own experience .
- How to adapt for other purposes
  - By changing the discussion statements, this exercise can be adapted to other topics.

Example 5: LIBER Open Science café

- Format, time needed
  - small groups, 1.5 hour
- Topic
  - Open Concepts and Principles
- Learning objectives
  - Have knowledge of different aspects of open science.
  - Connect different stakeholders to discuss statements and topics.
- Materials and tools needed
  - The LIBER Science Café card deck, or a prepared stack of written statements based on World Café
  - one table per 6-8 persons
- Exercise description
  - The set-up: 6-8 people gather around a table with 1 moderator and 1 note taker. To initiate conversations, they are provided with a deck of cards with statements and questions related to open science and the involved projects. These statements serve as conversation starters. Someone can pick a card, the group talks about it for some time, and then they can move on to the next card. In this way, people learn from each other and start to think about the bigger picture. Meanwhile, you can collect valuable input from different stakeholders.
  - The note taker: collects interesting points of the conversation in two different ways:
    1. The mindmap cards: You can use these cards for topics that get a lot of attention in the conversation. If things go too fast, don’t be afraid to stop the conversation and ask people to provide input for this mindmap. Write down the main topic in the centre, and work from there. Is it hard to find connections? You can also collect random thoughts and statements here.
    2. Brilliant quotes and ideas: Sometimes someone says something that’s just WOW, just spot on or somehow very useful. For this you have the ‘brilliant quote and ideas’ card. You only have one, so here you have to be very selective. Make a point of it if you think something is so good that it deserves to go on this card.
  - After 20-30 minutes, have the group change tables. Moderators and note takers remain seated.
  - At the end, each moderator reports on what has been said by the different groups at their table.

Example 6: What is research data for me?

- Format, time needed
  - Individual/pairs, 15 minutes
- Topic
  - Open Research Data and Materials
- Learning objectives
  - Know their own research data and data in their field of research
- Exercise description

    - Let the participants think about the last articles they wrote/read. Was there supplementary material (e.g., tables, images)? Let them write down examples and types of research data in their field of work. What information or data would they need in order to reanalyze the study? What would be needed for their own dissertation/article to be understood properly? Let them present their results either in pairs/groups and then in the plenary
- Materials and tools needed
    - A piece of paper and a pen
- Level of prior knowledge needed
    - No prior knowledge needed
- Things to bear in mind
    - Give the participants enough time to brainstorm
- How to adapt for other purposes:

You can shorten the activity by skipping the pair/group work and just discuss in the plenary

Example 7: Why not share data?

- Format, time needed
    - Small groups, ~20 minutes
- Topic
    - Open Research Data and Materials
- Learning objectives:
    - Get participants thinking about the ethical and practical barriers to data sharing, and to critically examine their beliefs in this area.
- Exercise description
    - In pairs or small groups, participants have five minutes to make a list as long as possible of all the reasons why researchers might not wish to share their data. Participants then report back on their reasons, discussing whether these are valid reasons or not, and strategies for how to overcome legitimate concerns. The team with the most reasons listed wins (prize optional).
- Materials and tools needed
    - Note taking equipment (pen, paper, or online document); optional: prize.
- Level of prior knowledge needed
    - Working knowledge of working with data
- Things to bear in mind
    - The exercise should be fun, and participants should be encouraged to come up with fun as well as serious examples.
- How to adapt for other purposes
    - The same format could easily be adapted for many other elements of Open Science, e.g., Open Access (why not publish OA, etc.)

Example 8: "Open Data Excuse" Bingo

- Format, time needed
    - Group exercise, 20–30 minutes
- Topic:
    - Open Research Data and Materials
- Learning objectives:
    - Being able to recognize stereotypes that prevent sharing research data and understand the advantages of opening research data.
- Exercise description:

    - This exercise should be used at the beginning of the training event. Participants split at least in two groups or more (depends on the group size). A trainer takes care that one group will develop pro and the other contra arguments. In small groups participants discuss excuses already defined at the "Open Data Excuse" Bingo, these are common arguments used by researchers when explaining why they can't share their data. For the last 10 minutes the groups should confront their arguments. A trainer helps participants to develop arguments for open their data and to better understand the idea of sharing their data.
- Materials and tools needed:
    - Printed sheets of "Open Data Excuse" Bingo
- Level of prior knowledge needed:
    - The participants should have experience with creating/collecting research data.
- Things to bear in mind:
    - Go around and try to help with arguments if needed, especially in the group, which supposed to develop strong arguments for sharing data. Extra help might be needed for these participants to be stronger later in the confrontation with participants from the other group.
- How to adapt for other purposes:
    - This exercise can be adapted to other topics (material would need to be adapted also)

Example 9: Me and my data - Datagramms

- Format, time needed
    - Group exercise, 1–4 hours (if done as part of a workshop)
- Topic
    - Open Research Data
- Learning objectives
    - Understanding what data is and what type of repository of archive is needed to store them properly
- Exercise description
    - Participants are asked to think about the last scientific work done in relation with a thesis (Bachelor, Master, or Ph.D.) and to reflect about the kind of data they produced.
    - They will then create a datagramm, i.e., write down on a card
        - the subject discipline
        - the title of the thesis
        - a bunch of letters, indicating
            - the format (like pdf, doc, csv, or similar)
            - the size (kb, mb, gb, tb, etc.)
            - the medium (like a for analogue, d for digital, i.e., digitized and b for born digital, or combinations of the three)
            - and finally the type of data, differentiating roughly between O for observations, E for experiments, S for simulations, D for derivations, R for references and D for digitized data, or combinations of them.
    - In several steps, all cards are finally clustered on a wall according to the letters (format, size, medium, and type)
    - The group discusses the different clusters and reflects about the requirements for an open data repository or archive.
- Materials and tools needed
    - Cards and flipcharts, or better a wall and material to fix the cards on the wall
- Level of prior knowledge needed
    - None as long as the exercise is started with some explanations on how to describe and differentiate data. Basic knowledge of

research data, repositories, and archives may be helpful.

- Things to bear in mind:
  - Make it a step by step approach
- How to adapt for other purposes
  - not yet applied

Example 10: Find your data publisher

- Format, time needed:
  - Individual / pairs, 10–15 minutes
- Topic:
  - Open Research Data
- Learning objectives:
  - Becoming aware of appropriate subject-specific data repositories and their characteristics and standards
- Exercise description:
  - The participants have to find a data repository for their research data. They go to re3data.org and search/browse by subject and/or content type. Let them limit their search to data repositories with DOI assignment. Give them time to have a look at the repository description and let them write down relevant repositories. Afterwards their success and experiences are discussed.
- Materials and tools needed:
  - Computer with internet access for every participant (can also be in pairs if necessary)
- Level of prior knowledge needed:
  - The participants should know which kind of research data they produce
  - Not applicable for bachelor students
- Things to bear in mind:
  - Some people might not find an appropriate repository, so prepare a list of generic and institutional repositories that can be used and show/hand it out afterwards
- How to adapt for other purposes:
  - You can adapt this exercise for Open Access by using the Directory of Open Access Journals [DOAJhttps://doaj.org) website

Example 11: What do you need for a data publication?

- Format, time needed:
  - Group exercise, 5–10 minutes (depending on group size)
- Topic:
  - Open Research Data
- Learning objectives:
  - Remembering the necessary steps for data publication
- Exercise description:
  - This exercise should be used at the end of the training. Let the participants play "I'm packing my suitcase" where they have to name necessary elements for a data publication (e.g., Research data (files), metadata, keywords, documentation, license, ORCID, repository, good title, references/sources, data citation, time, and courage!)
- Materials and tools needed:
  - No material needed
- Level of prior knowledge needed:
  - The participants know basic elements of data publishing through the course

- Things to bear in mind:
  - If participants forget an element, try to help or give pointers
  - Name as last element "courage"
- How to adapt for other purposes:
  - Can also be adapted for open access publishing process

Example 12: Creating metadata

- Format, time needed:
  - Individual / pairs, 5 minutes
- Topic:
  - Open Research Data
- Learning objectives:
  - Being able to create metadata for research data
- Exercise description:
  - Let the participants select a file they are currently working on. Let them answer the following questions on a piece of paper: Who created the content? What is the content? When was the content created? How was the content created? Why was the content created? Then discuss with them their results. Was it easy or difficult? Can they repeat this task for all the files in their research process?
- Materials and tools needed:
  - A piece of paper (or prepared form) and a pen
- Level of prior knowledge needed:
  - No prior knowledge needed
- Things to bear in mind:
  - To make the exercise faster prepare a form and print it out or make it available online.
  - For bigger projects with a lot of files offer a data dictionary template
- How to adapt for other purposes:
  - Can also be adapted as a documentation exercise

Example 13: Get started with sharing software openly

- Format, time needed
  - Individual / pairs, 20–30 minutes
- Topic
  - Open Research Software and Open Source
- Learning objectives
  - Learn how to use common tools and services for sharing research codes openly.
  - Be able to choose the appropriate license for their software, and understand the difference between permissive and non-permissive licenses
- Exercise description
  - This exercise is meant for any researchers that will use software/code for their research, whether they perform purely computational or experimental work (the latter use software for analysis, etc.).

    - First, have everyone sign up for a free GitHub account if they do not already have one. This free account will be sufficient for working with exclusively open/public code, although you may let them know that students, educators, and researchers can request a waiver for a free professional account.
    - In addition, have participants register for a Zenodo account, and link this to their GitHub account.
    - Next, have everyone create a new public repository, choosing an appropriate license based on the desired permissions (choosealicense.org can be helpful here). On Zenodo, enable the GitHub–Zenodo integration for this repository.
    - Have participants add their source file(s) to the repository, and add some description of the program/script to the README file. Once these files are added, choose a version number and create a release of the software.
    - Head to Zenodo, and obtain the DOI that has been generated for your software.
    - Congratulations, your software is now citeable! You can add a section to the README file with the DOI and suggested citation, or even add the DOI badge that Zenodo provides.
- Materials and tools needed
    - Individuals need to have a computer with internet connection.
    - Participants should have some code, script, or program ready—even if it is "messy"—that they will publicly share.
- Level of prior knowledge needed
    - None
- Things to bear in mind
    - None
- How to adapt for other purposes:
    - Not applicable

Example 14: Establishing a Reproducible Data Analysis Workflow

- Format, time needed
    - Individually and as a group, 4–8 hours (example here)
- Topic
    - Reproducible Research and Data Analysis
- Learning objectives
    - Use a (small) computational task relevant to your discipline/background, and establish it as an open and reproducible workflow.
    - Understand the key concepts, tools and services that are useful in the context of reproducibility.
- Exercise description
    - Each participant selects a dataset and corresponding data analysis process that is relevant to their field. Both dataset and the analysis process should be short enough that it concludes within a few minutes. Moreover, for the purposes of this exercise, the programming language should be Python or R, but other languages can be accommodated with slight changes in the underlying tools.
    - The participant initially runs the process in the traditional form, and then asks one of the other participants to re-run it with no external help. Identify both the time required for another person to run this, as well as the obstacles encountered.
    - Apply the same process using the Jupyter / Git / MyBinder approach; write the process as a Jupyter notebook, upload dataset and notebook to a repository on GitHub, and then connect the repository to mybinder. After than, ask again the same person to re-run this. Identify the change in time and accessibility.
- Materials and tools needed
    - Jupyter and Git are necessary (including an account on GitHub). Depending on the language, additional Jupyter kernels might

need to be installed. Finally, the trainer can decide on whether to provide a common example for all participants to use, or ask the participants to bring their own. The difference lies to the amount of time required for preparation, as well as on the uniformity of the participants' background.

- Level of prior knowledge needed
  - The workshop can be performed to different levels of expected prior knowledge, adapting for time. For example, a short basic introduction to Git can be included, but in all cases, the participants should be aware of the computational requirements of their own analysis.
- Things to bear in mind
  - The overall concept is straightforward, but has an initial learning curve of the individual components. Therefore you may consider spending some extra time in the beginning discussing each tool, before connecting them all together.
  - You should consider giving the participants a detailed explanation of the installation process (e.g., for Jupyter and Git), before the workshop, in order to minimize potential technical issues.
- How to adapt for other purposes
  - The workshop can be extended to introduce additional concepts of Open Science, such as Persistent Identifiers for software (such as assigning a DOI from Zenodo to the Git repo), as well as integrating all of the aspects under a common platform (such as the OSF).

Example 15: Choose the right version for the repository

- Format, time needed
  - Individual / pairs, 15–20 minutes
- Topic
  - Open Access to Published Research Publications
- Learning objectives
  - Being able to decide which is the version allowed to be deposit in a repository and state its copyright regime
- Exercise description
  - This exercise could be addressed to repository managers. Choose five different publications and ask participants to select which is the version that could be allowed in a repository and which would be the copyright notice they would include: who is the copyright holder and which copyright regime would hold: all rights reserved, a license, public domain. Discuss with them their results and show them the key elements that define the solutions.
- Materials and tools needed
  - The exercise can be performed with a piece of paper (or prepared form) and a pen
  - Individuals/pairs need to have an internet connection to access the papers and check policies. You may provide physical copies of the articles, too.
- Level of prior knowledge needed
  - Basic copyright notions
  - Knowledge on the different versions of a research paper
- Things to bear in mind
  - The exercise can be translated to an online version if you prepare a set of polls.
  - Use a range of publications including for instance papers published under hybrid models in order to show participants that is not enough to look up at sites with default self archiving policies.
  - The number of cases will determine the time of the exercise.
- How to adapt for other purposes:
  - Can be adapted to training sessions with researchers using their own papers.

Example 16: Open file formats

- Format, time needed:
  - Group exercise, 10–15 minutes
- Topic:
  - Open Licensing and File Formats
- Learning objectives:
  - Becoming aware of file formats used daily and their openness
- Exercise description:
  - Let the participants write down on post-its all the file formats they use in their daily work. Then get the post-its and stick them to the whiteboard or flipchart. Try to cluster them as best as you can into categories or groups (text, tabular, statistical, video, image, etc.). Then discuss the results with the audience. Talk about the openness of these file formats and possible alternatives.
- Materials and tools needed:
  - A few stacks of post-its, pen and a whiteboard or flipchart
- Level of prior knowledge needed:
  - No prior knowledge needed
- Things to bear in mind:
  - Prepare for "exotic" file formats that are subject-specific or machine-dependent or let the participants describe them.
- How to adapt for other purposes:
  - You can also use web tools like PINGO for the collection of file formats or let them write down their file formats on a piece of paper and collect those, if you don’t want to use post-its

Example 17: Creative Commons License matching

- Format, time needed:
  - Group exercise, 5–10 minutes
- Topic:
  - Open Licensing and File Formats
- Learning objectives:
  - Being able to differentiate between different Creative Commons licenses and to be able to combine them for works.
- Exercise description:
  - The participants have to combine two licenses. Let the group guess which Creative Commons license is created by the combination. Repeat the exercise with other combinations. Integrate a combination that is not possible (for example, CC BY-SA and CC BY-NC) and point out pitfalls. Discuss the results with the participants.
- Materials and tools needed:
  - Computer with projector, whiteboard, flipchart, or piece of paper for all attendees
- Level of prior knowledge needed:
  - The participants should know all Creative Commons licenses and/or have a paper to look at
- Things to bear in mind:
  - Wait more than three seconds before taking the answer. This enables participants to think it through and you are able to integrate even weak participants.
- How to adapt for other purposes:
  - First create pairs and let them solve the combinations, then discuss the solutions in the group
  - Use other licenses

Example 18: OER Remix

- Format, time needed:
  - Group exercise, 10–15 minutes
- Topic:
  - Open Licensing and File Formats
  - Open Educational Resources
- Learning objectives:
  - Being able to distinguish the different elements of the Creative Commons licenses
  - Being able to build content remixing previous works with multiple licenses including public domain and all rights reserved works and determine which will be the resulting license
- Exercise description:
  - There is an online version and a printed version
  - There is a set of cards marked with a type of content: text, image, music, and video, and each card carries a copyright sign that ranges from all rights reserved to public domain including the set of Creative Commons licenses and the GNU Free Documentation License.
  - One person of the group takes 12 cards and the rest of the group has to combine them building a material with the four type of content: text, image, music and video. Once they choose a right combination they have to decide which is a possible license for this new work.
- Materials and tools needed:
  - For the online game: computer with beamer
  - For the printed game: the set of cards is available at opencontent.org or you can create a set of cards yourself
- Level of prior knowledge needed:
  - The participants should know the elements of all Creative Commons licenses and have a basic notion of copyright issues including the notion of copyleft
- Things to bear in mind:
  - If you use the online version you might do the exercise with all your audience allowing multiple possible answers.
- How to adapt for other purposes:
  - You can adapt it to research elements, for instance to software licensing
  - You can use other licenses, include new kind of contents or define which contents should have the final work

Example 19: Open peer review - participants openly review each others' texts

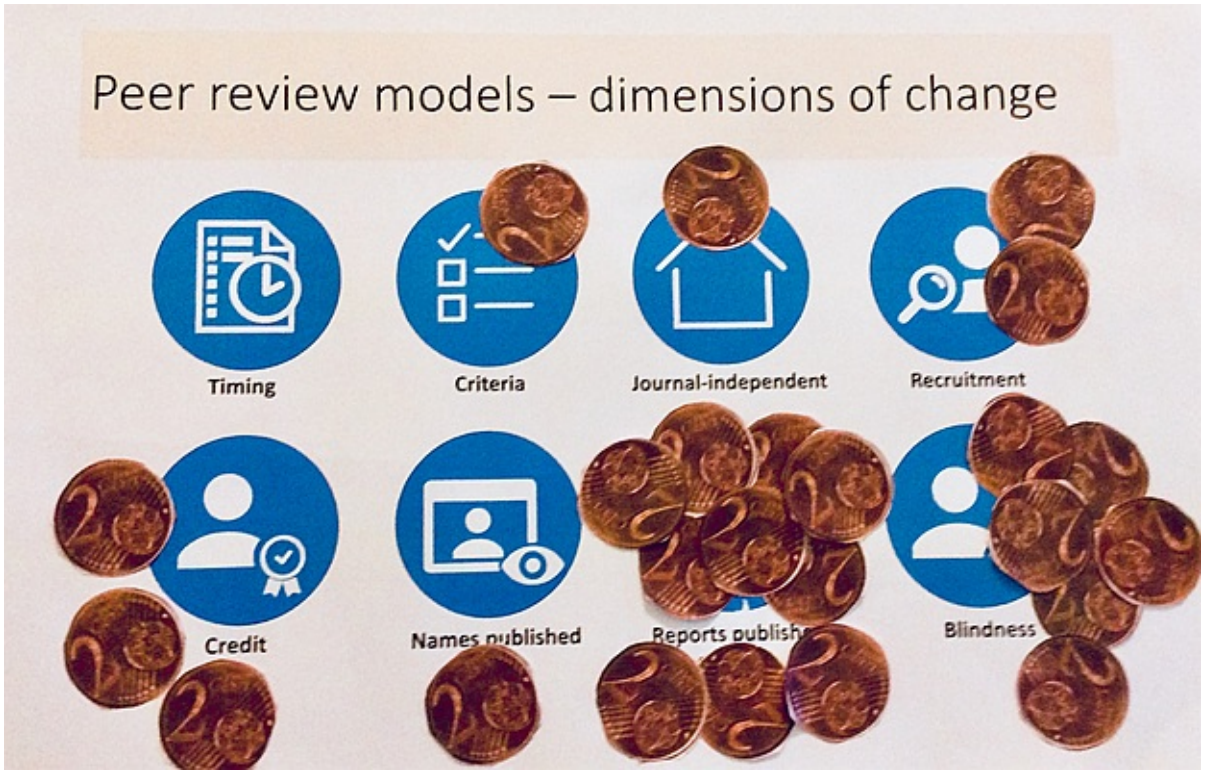
- Format, time needed
  - Small groups, 90 mins
- Topic
  - Open Peer Review, Metrics and Evaluation
- Learning objectives
  - Practise in writing constructive peer reviews
  - Critical reflection on the advantages and disadvantages of open peer review
- Exercise description
  - Participants work in groups of three. Each participants writes a short text (~300 words) giving their thoughts on open peer

review as discussed in the foregoing workshop. They then pass the text to the person on their left, who writes a brief peer review of the work. The text and the review are then passed to the next person on the left, so each now has a text and a review which they did not write. This person then gives feedback on the review—was it constructive, critical, what could have been better, etc. The group then reads all the texts and reflects on how open identities, open reports, etc. affected how they wrote their reviews, and reflects on the critical feedback from the others.

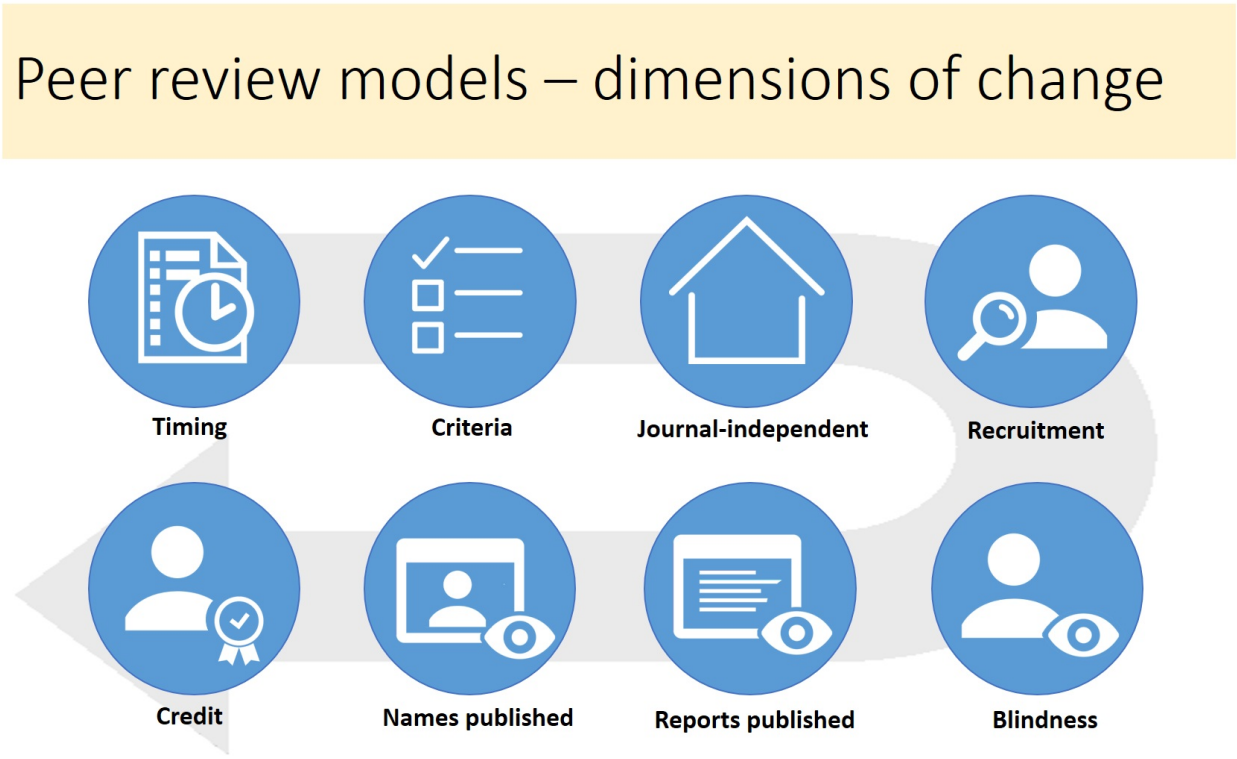
- Materials and tools needed
  - Pen and paper
- Level of prior knowledge needed
  - None, although the texts will require the knowledge gained in the foregoing workshop.
- Things to bear in mind
  - This exercise requires participants to make criticisms of each other's work—bear in mind that some people might be uncomfortable doing so, or that some may have difficulty accepting such critique. Where these issues occur, encourage participants to discuss them in the final discussion round.
- How to adapt for other purposes
  - Where this example is being used in a training workshop with a wider focus than just open peer review, it could be used to consolidate learning about other Open Science themes by asking participants to first write a text about those themes instead.
  - Instead of pen and paper, this exercise could also be done using a collaborative writing tool, such as Google Docs, Authorea, or Overleaf/ShareLaTeX.

Example 20: Open peer review - your 2 cents

- Format, time needed
  - Plenary, ~1.5 hour with discussion
- Topic
  - Open Peer Review, Metrics and Evaluation
- Learning objectives
  - Realize there are many aspects to open peer review and have knowledge of those different aspects of open peer review
  - Form an opinion on which aspects of open peer review would most benefit science
  - Have insights in the benefits and possible drawbacks of different aspects of open peer review, from the perspective of the reader, author and reviewer
- Exercise description
  - Introducing different aspects of peer review, including some examples of journals/platforms where they are put in practice
  - Ask participants to individually identify two to three aspects of open peer review they feel would contribute most to open science.
  - On a large printout, participants place a two-cent coin on each of the aspects they selected in the previous step
  - The results are viewed together and the most often chosen aspects identified
  - In small groups, participants then take the role of reader, author, or reviewer (all should be present in each group). They then discuss one of the aspects of open peer review from the perspective of their taken roles. What are the benefits and potential drawbacks?
  - Small groups then report back to the whole group, and additional perspectives/viewpoints can be discussed.



- Materials and tools needed
  - Large printout of dimensions of peer review: one for each participant and a communal one (presentation with animated slides also available)
  - Two-cent coins (if available in your monetary system, otherwise any low-denomination coins will do)

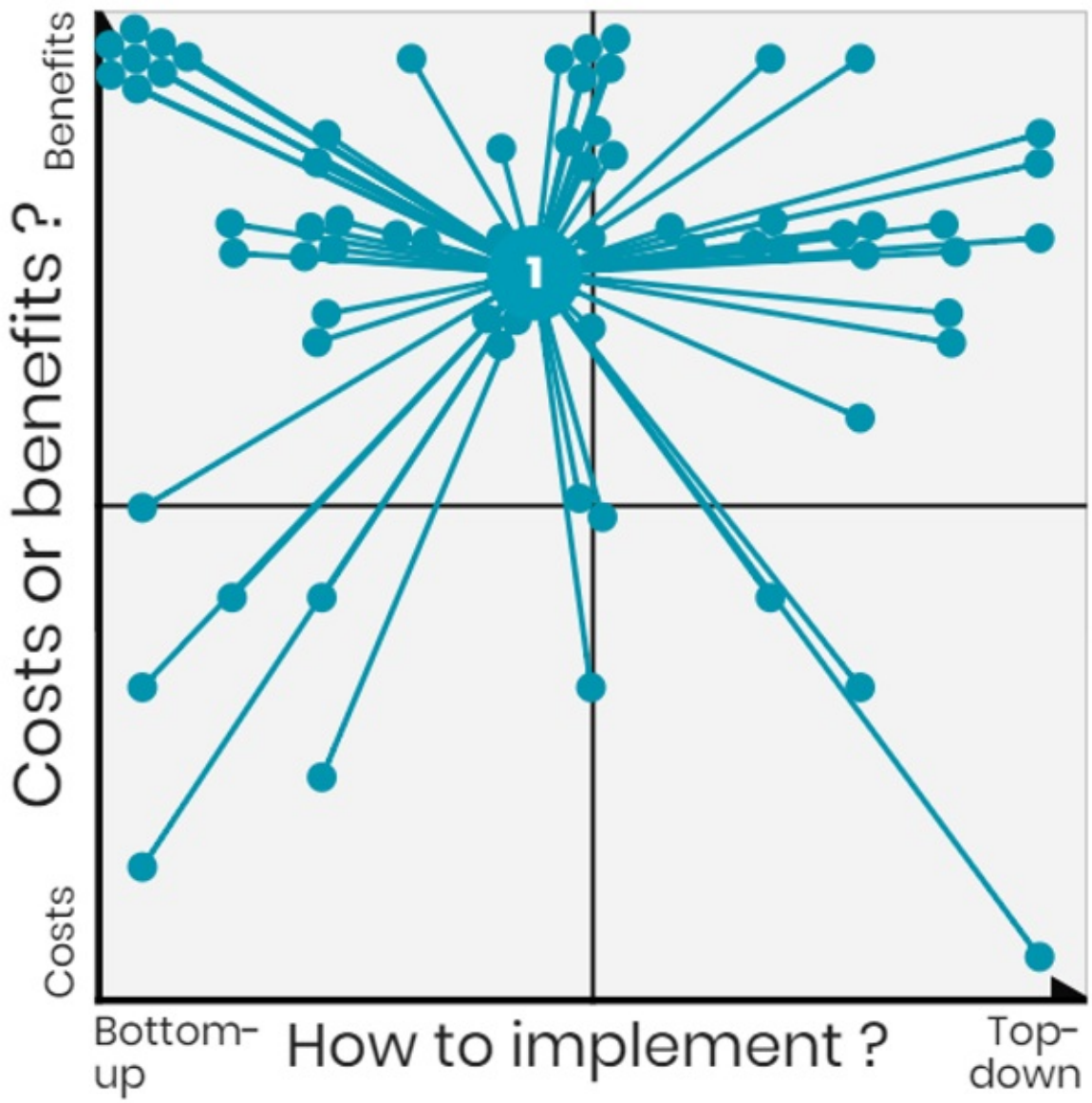


- Level of prior knowledge needed
  - None, some familiarity with the traditional process of peer review is helpful
- Things to bear in mind

    - For people not familiar with developments in open peer review, some aspects may require more explanation—plan enough time for that
    - In discussions, it can be hard for people to separate their personal opinion from their assigned role. Encourage and remind people to stick to their role where necessary.
    - The number of coins per person depends on the number of participants (e.g., three for smaller groups, two for larger groups)
- How to adapt for other purposes
    - The concept of voting with coins ("your two cents") can be applied to other topics, as can the assignment of roles in small group discussions

Example 21: Taking a stance

- Format, time needed
    - Plenary, 15 minutes
- Topic
    - Open Science Policies
- Learning objectives
    - Get participants to take a stance on Open Science policies or principles
    - Show similarity or diversity of opinions across participants
- Exercise description
    - Ask participants to express their opinion on two questions about Open Science policies or principles.
    - Responses should lie on a linear scale between two extremes (e.g., strongly disagree–strongly agree)
    - Participant vote using an online tool, or by placing sticky dots on a sheet of paper with axes representing the two answer ranges
    - Results are shown to the group, and the similarity or diversity of responses is discussed, e.g., by asking one respondent from each quadrant to explain their opinion.
    - Example question and results:
        1. For individual researchers, does Open Science have more costs or benefits?
        2. Should Open Science be organized bottom-up or top-down?



- Materials and tools needed
  - Access to an online tool like Mentimeter; a paid account allows export of the results but is not required for this exercise
  - For each participant, access to smartphone, tablet, or computer with internet access
  - Offline alternative: large paper with axes printed or drawn, sticky dots
- Level of prior knowledge needed
  - None; some background knowledge on the topic is useful to get informed opinions rather than gut feelings (although the latter may be useful to collect too)
- Things to bear in mind
  - If done on paper, it might make sense to have people mark down their answer individually first, before placing their dot on the map. This prevents peer pressure / bias.
- How to adapt for other purposes
  - This exercise can be adapted to many different questions and topics
  - An alternative online tool (that is also open source) for these kind of exercises is SimpleVote (https://simplevote.ml)

    - If the audience is heterogeneous (i.e., researchers, research support people, policy makers) it is informative to distinguish between the different groups, e.g., by creating a separate question for each (in Mentimeter), or using different color sticky dots (on paper)
    - For sticky dots, choose a combination that is colour-blind friendly

Example 22: Plain language explanations - in progress

- Format, time needed
    - Small groups, 2–3 hours
- Topic
    - Citizen Science
    - Collaborative Platforms
- Learning objectives
- Exercise description
- Materials and tools needed
- Level of prior knowledge needed
- Things to bear in mind
- How to adapt for other purposes

Example 23: Devil’s advocate - convincing the skeptics

- Format, time needed
    - Small groups, 30 minutes
- Topic
    - Open Advocacy
- Learning objectives
    - Formulate arguments against common objections to open science practices
    - Practice discussion with people questioning the value of open science
- Exercise description
    - In small groups of three or four, have one or two person(s) assume the role of open science skeptic and the others the role of open science advocate.
    - Have the "open science advocates" try to convince the “open science skeptics”
    - After 10 minutes, have participants switch roles and have another discussion (not repeating the same arguments)
    - After two rounds, gather as a group as share experiences. Which arguments were the hardest to refute? Which arguments worked best to convince the skeptics? Do participants feel these be arguments would be useful in real-life situations as well?
- Materials and tools needed
    - none; flexible room setup is useful to allow groups to spread across the room
- Level of prior knowledge needed
    - Familiarity with open science concepts
- Things to bear in mind
    - Encourage the open science skeptics to get into their role as much as possible. Often, people really enjoy taking on this role!
    - Be sure to switch roles to give everyone the chance to experience this exercise from both perspectives.

- How to adapt for other purposes
    - This exercise could be focused on specific aspects of open science

Example 24: Set up OSF project & link to other platforms - in progress

- Format, time needed
    - Individually or in pairs
- Topic
    - Open Research Data and Materials
- Learning objectives
- Exercise description
    - Create an OSF collaborative environment from data to publication.
        - Connect your OSF project to GitHub.
        - Upload any raw code, images, data, tables to project.
        - Obtain a DOI and ARK identifier for your project.
- Materials and tools needed
- Level of prior knowledge needed
- Things to bear in mind
- How to adapt for other purposes

Example 25: The publishing trap - in progress

- Format, time needed
    - Small group exercise, 2 h
- Topic
    - Open Access to Published Research Results
- Learning objectives
    - "The game lets you explore the impact of scholarly communications choices and discuss the role of open access in research by following the lives of four researchers, from doctoral research to their academic legacies." blogs.kent.ac.uk
- Exercise description
    - "It is played by four teams of up to four people – sat around a game board and using a playbook to guide the decisions the teams must make. The workshop leader acts as a host and presents the scenarios to the teams during each round. Each round involves making three decisions about publishing choices. After hearing the scenario, each team chooses from the pre-determined options. At the end of each round, the teams discuss the decisions they have reached and are asked to justify their choices." copyrightliteracy.org
- Materials and tools needed
    - The board, cards, booklets, points and other object has to be downloaded, printed and cut out. They plan to also have a professionally produced game available to purchase. Materials are available here: copyrightliteracy.org
- Level of prior knowledge needed
    - "The Publishing Trap is aimed at early career researchers and academics, as well as anyone who has a vested interested in understanding how access to information works and how the whole scholarly communication system in higher education operates." copyrightliteracy.org
- Things to bear in mind
    - Maybe stimulate discussions during the game play
- How to adapt for other purposes

- Licensing conditions
  - The beta version of the game is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 Licence.

Example 26: - in progress

- Format, time needed
  - Small groups, 4 days (5 hours/day)
- Topic
  - Open Research Data and Materials
- Learning objectives
  - The participants understand the basics of open data and reproducible research, understand the stages to setup a research data management plan, and can build their own data repository.
- Exercise description
- the knowledge about repository and licensing
- data assessment: types, sum, sensitivity
- setting up a research data management plan using DMPtool
- setting up an OSF repository
- using Git for version control
- integrating GitHub, Google Drive, and other services to OSF project
- using R, R Studio, and R Markdown to create a reproducible research
- exercise in creating a citizen science project
- Materials and tools needed
  - Registration of: ORCID, OSF, GitHub, and DMPTool
  - Downloading and installing: Git, R, and R Studio
- Level of prior knowledge needed
  - A basic knowledge in R, R Studio, and Git would be a plus.
- Things to bear in mind
  - He/she may have to put more time in explaining the concept of open data and why people should do it. Most debates occur in this preliminary stage.
- How to adapt for other purposes
  - The exercise is made especially for geo/spatial sciences, but most part of it can used for any science, including citizen science project.
  - Licensing conditions
    - CC BY - Dasapta Erwin Irawan, INArxiv, Institut Teknologi Bandung; Willem Vervoort, The University of Sydney; Gene Melzack, The University of Sydney

Example 27: Train-the-trainer card game for Open Science training

- Format, time needed
  - Small groups, 2 hours
- Topic
  - Open Advocacy
- Learning objectives
  - Trainers can use this game to facilitate ‘train-the-trainer’ workshops. Participants design a usable framework for a training – which will they deliver themselves at a later stage - on (a) topic(s) of their choice. The card game offers the participants the option to preselect audience type, audience size, training type and audience knowledge level. In addition, two ‘unforeseen’ circumstances can be added: audience mood, and ‘trouble’ (uh-oh!). Apart from going home with a usable design for a

training, the audience of this workshop will also benefit from the input and experience of the other participants.

- Exercise description
  - INSTRUCTIONS:
    - Have each group pick a card (blind) determining: audience type, audience size, audience knowledge level and training type. It is possible that the different cards turn out a training situation that is impossible or that is not in line with to the interests of the group. It can be useful to allow some flexibility and allow people to change cards or switch cards with another group. In most cases, there is an empty card or an 'other' card available as well, allowing people to modify the exercise according to their own needs.
    - Hand out the persona pages: every member of the group should create one persona according to the conditions laid out on the cards (no longer than 15 mins).
    - The group has 1,5 hour to prepare the training according to the conditions laid out by the cards, keeping in mind their target audience(s), with help of the persona pages created.
    - Have each group present their training (take note: they don't have to give the actual training, they only have to describe what they will do!) . Ask the other groups to give feedback afterwards: Is the proposed training suitable for the conditions laid out by the cards? What would they do differently? Do they have any experiences that they can share?
    - Extra challenge: How would each group deal with unexpected/unpredictable circumstances during the training? Right before their presentation, each group picks (blind) an 'audience mood' card and a 'trouble' card and gives them to the moderator, who will either during the presentation or afterwards discuss these cards with the entire group – allowing the audience to learn from the experience of their colleagues.
- Materials and tools needed
  - https://www.fosteropenscience.eu/node/2570
  - You can download the files in pdf and png format via this public dropbox link: https://www.dropbox.com/sh/k314ebvqpb6mqq8/AAABEcJqYF_2PYJxqmYf3mmna?dl=0&fbclid=IwAR0DBmnArU8raKlaoJa7RKPEGRNEv2y74PQRR2Ft_y4Oy7DLfdawF_n5LbQ
- Level of prior knowledge needed
  - Participants are expected to be knowledgeable about the topic(s) they will create the training about
- Things to bear in mind
  - Timekeeping is essential; limit the time people will work on persona's and training design. Clarify that the presentation should be a description of all the elements of the training they have designed, not actually giving the training. When evaluating with the group, make sure everybody gives input.
- How to adapt for other purposes
  - In principle, all parameters can be adapted and changed to suit a specific training, by creating new cards, new categories, or by removing existing ones.
- Licensing conditions
  - CC BY-SA 4.0. Creator: Gwen Franck

Resources

What tools & platforms to use / recommend?

There are many tools and platforms that support Open Science practices (see figure below for a selection). Which tools and platforms to use (or advise) depends on many factors, for example: whether the tool is available (either free of at low cost or licensed to your institution), whether it works in your browser or for your operating system, whether it is available in your language, and whether it meets your security and privacy requirements. In addition to these more technical criteria, consider whether a tool fits with the way you work. Does it work well with other tools and platforms that you use? Do the people you collaborate with use the same tool for the same practice, or at least one that is compatible with the one you use? Also consider the learning curve: do you need to invest a lot of time into learning the new tool, and if so, is that worth it for you? Do you have support (either in real life or online) that can help you learn to use the tool?

Perhaps the best advice is to first consider what it is you would like to do: what is the open science practice you'd like to implement? Then explore which tools/platforms are available, which ones the people in your community use, and why (ask around!). Then make your own decision. Don't be afraid to experiment and try out something new!

A final remark: many tools and platforms support open science practices without themselves being fully open. For example, many commonly used tools are not open source, even though they provide access to content (publications, data) that are open. You will have to follow your own judgement as to whether you will consider such tools and platforms or not. Another consideration is whether you can export all your data when you'd want to switch to another tool, or whether they are locked in? And do you know what will happen to your data when the platform closes down or is sold to a(nother) company?

Some resources listing research tools and platforms:

- Connected Researchers (all disciplines)
- DIRT Directory (Humanities)
- ResearchStash (Science, Technology and Medicine)
- 400+ Tools and innovations in scholarly communication (all disciplines)
- Tool combinations (which tools are commonly used together) [colour-blind safe]

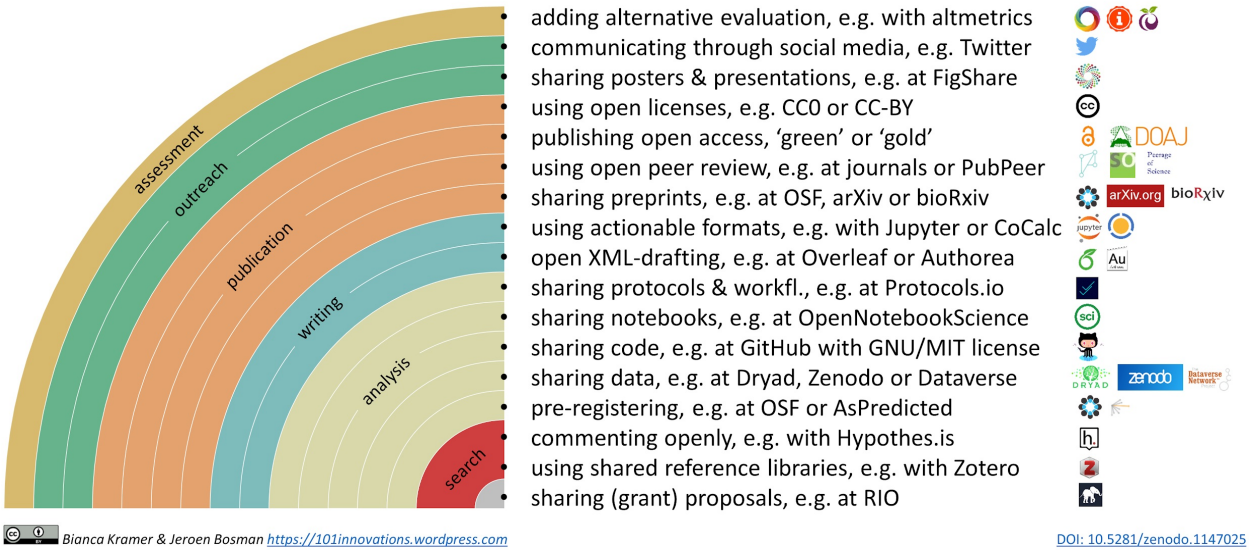


Figure x - Rainbow of open science practices (available on Zenodo in different formats, including as editable slide:10.5281/zenodo.1147025)

Other resources

- Ask Open Science. ask-open-science.org
- Digital Curation Centre. Because good research needs good data. dcc.ac.uk
- Fernandes and Rutger (2017). Open Science, Open Data, Open Source. 21st century skills for the life sciences. osodos.org
- Forschung und Daten managen (German information website about research data management).forschungsdaten.info
- MANTRA - Research Management Training. mantra.edina.ac.uk
- Materials for ELIXIR-EXCELERATE Train The Trainer workshops and courses. github.com/TrainTheTrainer/EXCELERATE-TtT (*comment by authors: A complete repository of materials and methods, selected for training instructors, only a small part is specific to Bioinformatics*)
- Open Science MOOC. opensciencemooc.eu
- Open Science Training Initiative. Graduate Training in Open Science. opensciencetraining.com
- Research Data E-Learning Platform. (German and French) researchdatamanagement.ch

- Research Data Management Educational Efforts. docs.google.com
- Research data management (RDM) open training materials. Zenodo Community
- Sewell (2017). Research Data Management: Activity Cards. doi.org/10.17863/CAM.10074.
- Tips on how to build and publish a versionised e-book are given in the github repository github.com/Pfern/OSODOS - Also available in GitHub Pages as a website pfern.github.io/OSODOS/SUMMARY. PDF, e-Pub and Mobi versions were made available by Unglue.it

Longlist of exercises - selection to be put in template format

Awaiting some formatting to comply with the template

PF - 1 Mind and Concept Maps

The conceptualisation of higher complexity subject matter can benefit a lot from visualizing recently acquired knowledge or skills. A great deal of enthusiasm can be raised when simple open source tools are used, individual and collectively. The general name for this set of techniques is idea and concept mapping. A relatively simple software like X-Mind is a good basis to start with.

Figure X An example of an idea map to represent content in a training course

Note: we might replace this by one made for Open Science or a related subject

Learner engagement raises sharply as learners understand the power of visualising ideas, connecting them in diagrams, comparing diagrams between learners in the same group, comparing different groups, comparing learners with instructor maps, etc.

Glossary

Altmetrics

- Altmetrics are alternative ways of recording and measuring the use and impact of scholarship. Rather than solely counting the number of times a work is cited in scholarly literature, alternative metrics also measure and analyze social media (e.g., Facebook, Twitter, blogs, wikis, etc.), document downloads, links to publishing and unpublished research, and other uses of research literature, in order to provide a more comprehensive measurement of reach and impact.

Audience

- The group addressed by a communication (e.g., those in attendance of an Open Science training). The target audience is a group of individuals that will be addressed or affected by the training.

Behaviorism (Learning Theory)

- Behaviorism means that learning is governed by drill-and-practice and is best done with the use of stimuli to which the learners respond. This generally means that you ask the learner to do an exercise for which there is a clear answer or a clear path to follow. Evaluation is clear and can easily be done with the help of simple metrics.

Cognitivism

- Cognitivism is based on the interaction between the outer world and what the reflecting brain makes out of the information perceived in combination with the knowledge that it has already stored. Cognitivism concentrates therefore on problem solving.

Connectivism

- Connectivism is the integration of principles explored by chaos, network, complexity and self-organization theories. Connectivism is driven by the understanding that decisions are based on rapidly altering foundations, as new information is continually being acquired.

Constructivism

- Constructivism in the strict sense means the world is not as it is. Instead the world is primarily the product of our individual experiences and minds. In the context of teaching and learning this means that learners themselves create the path of learning. The focus is hence on the learner’s creativity and evaluation of progress is not based on the differentiation between right or wrong.

Copyright

- The aspect of Intellectual property that grants creators the right to permit (or not permit) the reproduction of their creations. It is distinct from trademark rights or moral rights.

Creative Commons

- A suite of standardized licences that allow copyright holders to grant some rights to users by default. CC licences are widely used, simple to use, machine readable, and have been created by legal experts. There are a variety of CC licences, each of which use one or more clauses. Some licences are compatible with Open Access in the Budapest sense (CC0 or those carrying the BY, SA, and

ND clauses), and some are not (carrying the NC clause).

Curriculum

- Curriculum refers to the lessons and other training content taught in a school or in a specific course or program within a defined structure.

Data

- Data in the sense used here are all digitally available objects (simple or complex) that emerge or are the result of the research process.

Data Mining

- An analytic process designed to explore data in search of consistent patterns or systematic relationships between variables, transforming data into information for future use.

Digital Object Identifier (DOI)

- A unique text string that is used to identify digital objects such as journal articles, data sets or open source software releases. A DOI is one type of Persistent Identifier (PID).

Documentation

- A documentation is detailed information as well as background and methodological approach about the data or code (e.g., description of the project, variables, and measuring instruments).

FAIR Data

- FAIR Data (according to FORCE11 principles and published in Nature Scientific Data) are Findable, Accessible, Interoperable, and Re-usable, in order to facilitate knowledge discovery by assisting humans and machines in their discovery of, access to, integration and analysis of, task-appropriate scientific data and their associated algorithms and workflows.

Gamification

- The use of game design elements and game mechanics in non-game contexts, such as education where it can be used to bring extra engagement.

GDPR

- (General Data Protection Regulation) seeks to create a harmonised data protection law framework across the EU. It aims to restitute the control of personal data to citizens, whilst imposing strict rules on those hosting and 'processing' these data, anywhere in the world. The Regulation also introduces rules relating to the free movement of personal data within and outside the EU.

Impact Factor

- A numerical measure that indicates the average number of citations to articles published over the previous two years in a journal. It is frequently used as a proxy for a journal's relative importance. Its transfer to the impact of individual articles published in a journal is considered to be problematic.

Intellectual Property

- A legal term that refers to creations of the mind. Examples of intellectual property include music, literature, paintings, sculpturing, video and other artistic works; discoveries and inventions; and phrases, symbols, and designs.

Journal

- A series of published research articles. Historically divided into volumes and issues.

License

- A license allows a third party to perform certain actions with a work or data. The license informs about the usage rights of a resource (e.g. text, data, source code).

Metadata

- Metadata provide a basic description of the data, often including authorship, dates, title, abstract, keywords, and license information. They serve first and foremost the findability of data (e.g. creator, time period, geographic location).

Open Access

- Open Access refers to online, free of cost access to peer reviewed scientific content with free reusability regarding copyright restrictions.

Open Data

- Open Data are online, free of cost, accessible data that can be used, reused and distributed provided that the data source is attributed.

Open Evaluation

- The development of a fair evaluation system or protocol for research proposals, based on transparency of the process and those involved.

Open Lab Notebooks

- A concept of writing about research on a regular basis, such that research notes and data are accumulated and published online as soon as they are obtained.

Open Materials

- Sharing of research materials, for example, biological and geological samples, is another Open Science practice.

Open Peer Review

- An umbrella term for a number of overlapping ways that peer review models can be adapted in line with the aims of Open Science, including making reviewer and author identities open, publishing review reports and enabling greater participation in the peer review process.

Open Science

- Open science is the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society.

Open Source

- Availability of source code for a piece of software, along with an open source license permitting reuse, adaptation, and further distribution.

Peer Review

- A process by which a research article is vetted by experts from the community before publication.

Persistent Identifier (PID)

- A persistent identifier (also PID) is a unique and stable denomination (reference) of a digital resource (e.g. research data) through allocation of a code that can be persistently and explicitly referenced on the internet.

Persistent/Preferred File Formats

- Non-proprietary formats that follow documented international standards, are commonly used by the research community, use standard character encoding (e.g. ASCII, UTF-8), and were compression, if used at all, is lossless.

Preprint

- A manuscript draft that has not yet been subject to formal peer review, distributed to receive early feedback on research from peers.

Preregistration

- Researchers have the option or are required to submit important information about their study (for example: research rationale, hypotheses, design and analytic strategy) to a public registry before beginning the study. Preregistration can help counter reporting bias.

README file

- File where you document your research data. The documentation should be sufficient to enable other researchers to understand, replicate or reproduce the data or reuse them in any other way.

Reporting Bias

- Reporting bias occurs when certain aspects of a study are systematically not reported transparently, creating wastage and redundancy through selective reporting or non-publishing.

Repository

- Repository is defined as the infrastructure and corresponding service that allows for the persistent, efficient and sustainable storage of digital objects (such as documents, data and code).

Reproducible Research

- Reproducibility is a spectrum and instructors should choose the definition most used by their audience. Generally speaking, reproducible research makes it possible to obtain similar results of a study or experiment and independent results obtained with the same methods but under different conditions (i.e., pertains to results). Some break the definition into levels of reproducibility, including computationally reproducible (also called "reproducible"): where code and data can be analyzed in a similar manner as in the original research to achieve the same results, and empirically reproducible (also called “replicable”): where an independent researcher can repeat a study using the same methods but creating new data.

Research Impact

- Involve academic, economic and societal aspects, or some combination of all three. Impact is the demonstrable contribution that research makes in shifting understanding and advancing scientific, method, theory and application across and within disciplines, and the broader role that this plays outside of the research system.

Research Funder

- An institute, corporation or government body that provides financial assistance for research.

Scholarly Communication

- The creation, transformation, dissemination, and preservation of knowledge related to teaching, research, and scholarly endeavors; the process of academics, scholars and researchers sharing and publishing their research findings so that they are available to the wider academic community. The creation, transformation, dissemination, and preservation of knowledge related to teaching, research, and scholarly endeavors; the process of academics, scholars and researchers sharing and publishing their research findings so that they are available to the wider academic community.

Sharing

- The joint use of a resource or space. A fundamental aspect of collaborative research. As most research is digitally-authored & digitally-published, the resulting digital content is non-rivalrous and can be shared without any loss to the original creator.

Subscription

- A form of business model whereby a fee is paid in order to gain access to a product or service - in this case, the outputs of scholarly research.

Trainer

- The moderator and instructor of a training, whose role is to ensure the training objectives are met, run the practice, and ensure no one is left out.

Training

- Training is any organised activity that teaches, informs, or transfers skills or knowledge on specific useful competencies through active, engaged learning.

Training Format

- A conventionally named, standardised delivery method that is applied by a trainer and includes any number of the pedagogical tools necessary (i.e. motivation/demotivation, hands-on approaches, etc).

Version Control

- Version control is the management of changes to documents, computer programs, large web sites, and other collections of information in a logical and persistent manner, allowing for both track changes and the ability to revert a piece of information to a previous revision.

Additional Resources

- Open Research Glossary, hosted by the R2RC.
- FOSTER Taxonomy
- Open Definition
- Lexicon-of-Learning (ASCD)

References

> *Note:* This bibliography is collaboratively maintained by the *Open Science Training Handbook* community at https://www.zotero.org/groups/2114699/open_science_book/items/collectionKey/XIUNJBME.

Balasegaram, Manica, Peter Kolb, John McKew, Jaykumar Menon, Piero Olliaro, Tomasz Sablinski, Zakir Thomas, Matthew H. Todd, Els Torreele, and John Wilbanks. ‘An Open Source Pharma Roadmap’. *PLOS Medicine* 14, no. 4 (18 April 2017): e1002276. doi.org/10/gbrb4b

Barba, Lorena A. ‘Terminologies for Reproducible Research’. *ArXiv:1802.03311 [Cs]*, 9 February 2018. arxiv.org/abs/1802.03311

Barnes, Nick. ‘Publish Your Computer Code: It Is Good Enough’. *Nature* 467, no. 7317 (14 October 2010): 753–753. doi.org/10/cj8t6n

Björk, Bo-Christer, Patrik Welling, Mikael Laakso, Peter Majlender, Turid Hedlund, and Guðni Guðnason. ‘Open Access to the Scientific Journal Literature: Situation 2009’. *PLOS ONE* 5, no. 6 (23 June 2010): e11273. doi.org/10/csjg36

Blackmore, Paul, and Camille B. Kandiko. ‘Motivation in Academic Life: A Prestige Economy’. *Research in Post-Compulsory Education* 16, no. 4 (1 December 2011): 399–411. doi.org/10/fqrkft

Buckheit, Jonathan B., and David L. Donoho. ‘WaveLab and Reproducible Research’. In *Wavelets and Statistics*, edited by Anestis Antoniadis and Georges Oppenheim, 103:55–81. New York, NY: Springer New York, 1995. doi.org/10.1007/978-1-4612-2544-7_5

Candela, Leonardo, Donatella Castelli, and Pasquale Pagano (2013). Virtual Research Environments: An Overview and a Research Agenda. Data Science Journal. 12, pp.GRDI75–GRDI81. doi.org/10.2481/dsj.GRDI-013

Christodoulou, Michail, Stefanos Kachrilas, Ahmed Dina, Andreas Bourdoumis, Junaid Masood, Noor Buchholz, and Athanasios Papatsoris. ‘How to Conduct a Successful Workshop: The Trainees’ Perspective’. *Arab Journal of Urology*, Teaching and Training in Urology, 12, no. 1 (1 March 2014): 12–14. doi.org/10/gcbmkm

Cobb, Matthew. ‘The Prehistory of Biology Preprints: A Forgotten Experiment from the 1960s’. PeerJ Inc., 22 August 2017. doi.org/10.7287/peerj.preprints.3174v1

Crosas, Mercè. ‘Joint Declaration of Data Citation Principles - FINAL’. FORCE11, 30 October 2013. force11.org/datacitationprinciples

Dryden, Michael D. M., Ryan Fobel, Christian Fobel, and Aaron R. Wheeler. ‘Upon the Shoulders of Giants: Open-Source Hardware and Software in Analytical Chemistry’. *Analytical Chemistry* 89, no. 8 (18 April 2017): 4330–38. doi.org/10/gc5sjm

Goodman, Steven N., Daniele Fanelli, and John P. A. Ioannidis. ‘What Does Research Reproducibility Mean?’ *Science Translational Medicine* 8, no. 341 (1 June 2016): 341ps12-341ps12. doi.org/10/gc5sjs

Haklay, Muki. ‘Citizen Science and Policy: A European Perspective’. *Washington, DC*, February 2015, 76.

Ince, Darrel C., Leslie Hatton, and John Graham-Cumming. ‘The Case for Open Computer Programs’. *Nature* 482, no. 7386 (22 February 2012): 485–88. doi.org/10/hqg

Iskoujina, Zilia, and Joanne Roberts. ‘Knowledge Sharing in Open Source Software Communities: Motivations and Management’. *Journal of Knowledge Management* 19, no. 4 (13 July 2015): 791–813. doi.org/10/f7htj8

Jahn, Najko, and Marco Tullney. ‘A Study of Institutional Spending on Open Access Publication Fees in Germany’. *PeerJ* 4 (9 August 2016): e2323. doi.org/10/bnqm

Jiménez, Rafael C., Mateusz Kuzak, Monther Alhamdoosh, Michelle Barker, Bérénice Batut, Mikael Borg, Salvador Capella-Gutierrez, et al. ‘Four Simple Recommendations to Encourage Best Practices in Research Software [Version 1; Referees: 3 Approved]’. *F1000Research* 6 (13 June 2017): 876. doi.org/10/gbp2wh

Knowles, Malcolm S, Elwood F Holton, and Richard A Swanson. *The Adult Learner: The Definitive Classic in Adult Education and Human Resource Development.* Oxford: Butterworth-Heinemann, 2011.

Kreutzer, Till. ‘Validity of the Creative Commons Zero 1.0 Universal Public Do-Main Dedication and Its Usability for Bibliographic Metadata from the Perspective of German Copyright Law’, 2011. PDF

LEARN. *LEARN Toolkit of Best Practice for Research Data Management*. Edited by LEARN. Leaders Activating Research Networks (LEARN), 2017. dx.doi.org/10.14324/000.learn.00

Lionelli, Sabina. ‘Implementing Open Science: Strategies, Experiences and Models’. Thematic Report. Mutual Learning Exercise: Open Science – Altmetrics and Rewards. European Commission, n.d. rio.jrc.ec.europa.eu

Lowndes, Julia S. Stewart, Benjamin D. Best, Courtney Scarborough, Jamie C. Afflerbach, Melanie R. Frazier, Casey C. O’Hara, Ning Jiang, and Benjamin S. Halpern. ‘Our Path to Better Science in Less Time Using Open Data Science Tools’. *Nature Ecology & Evolution* 1, no. 6 (23 May 2017): 0160. doi.org/10/gc4jb3

Luther J. The Stars Are Aligning for Preprints. The Scholarly Kitchen. 2017. [December 2018]. scholarlykitchen.sspnet.org

Martinez-Torres, M.R., and M.C. Diaz-Fernandez. ‘Current Issues and Research Trends on Open-Source Software Communities’. *Technology Analysis & Strategic Management* 26, no. 1 (2 January 2014): 55–68. doi.org/10/gc5sjj

McKiernan, Erin C, Philip E Bourne, C Titus Brown, Stuart Buck, Amye Kenall, Jennifer Lin, Damon McDougall, et al. ‘How Open Science Helps Researchers Succeed’. *ELife* 5 (7 July 2016). doi.org/10/gbqsng

McQuilton, Peter, Alejandra Gonzalez-Beltran, Philippe Rocca-Serra, Milo Thurston, Allyson Lister, Eamonn Maguire, and Susanna-Assunta Sansone. ‘BioSharing: Curated and Crowd-Sourced Metadata Standards, Databases and Data Policies in the Life Sciences’. *Database* 2016 (1 January 2016). doi.org/10/f8wzmc

Morin, A., J. Urban, P. D. Adams, I. Foster, A. Sali, D. Baker, and P. Sliz. ‘Shining Light into Black Boxes’. *Science* 336, no. 6078 (13 April 2012): 159–60. doi.org/10/m5t

Munafò, Marcus R., Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P. A. Ioannidis. ‘A Manifesto for Reproducible Science’. *Nature Human Behaviour* 1, no. 1 (January 2017): 0021. doi.org/10/bw28

Niemeyer, Kyle E., Arfon M. Smith, and Daniel S. Katz. ‘The Challenge and Promise of Software Citation for Credit, Identification, Discovery, and Reuse’. *Journal of Data and Information Quality* 7, no. 4 (6 October 2016): 1–5. doi.org/10/gc5sjd

Oishi, Jeffrey S., Benjamin P. Brown, Keaton J. Burns, Daniel Lecoanet, and Geoffrey M. Vasil. ‘Perspectives on Reproducibility and Sustainability of Open-Source Scientific Software from Seven Years of the Dedalus Project’. arXiv:1801.08200 [astro-ph.IM], 24 January 2018. arxiv.org/abs/1801.08200

Pavelin, Katrina, Sangya Pundir, and Jennifer A. Cham. ‘Ten Simple Rules for Running Interactive Workshops’. *PLOS Computational Biology* 10, no. 2 (27 February 2014): e1003485. doi.org/10/gc5sjq

Picarra, Mafalda, and Alma Swan. ‘Monitoring Compliance with Open Access Policies’, December 2015. pasteur4oa.eu

Piwowar, Heather, Jason Priem, Vincent Larivière, Juan Pablo Alperin, Lisa Matthias, Bree Norlander, Ashley Farley, Jevin West, and Stefanie Haustein. ‘The State of OA: A Large-Scale Analysis of the Prevalence and Impact of Open Access Articles’. *PeerJ* 6 (13 February 2018): e4375. doi.org/10/ckh5

‘Point of View: How Open Science Helps Researchers Succeed’, n.d. doi.org/10/gc5sjc

Pontika, Nancy, Petr Knoth, Matteo Cancellieri, and Samuel Pearce. ‘Fostering Open Science to Research Using a Taxonomy and an ELearning Portal’. In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-Driven Business*, 11:1–11:8. I-KNOW ’15. New York, NY, USA: ACM, 2015. doi.org/10.1145/2809563.2809571

Priem, Jason, D. Taraborelli, P. Groth, C. Neylon (2010). Altmetrics: A manifesto, 26 October 2010. altmetrics.org/manifesto

Prins, Pjotr, Joep de Ligt, Artem Tarasov, Ritsert C Jansen, Edwin Cuppen, and Philip E Bourne. ‘Toward Effective Software Solutions for Big Biology’. *Nature Biotechnology* 33, no. 7 (July 2015): 686–87. doi.org/10/f3mn4p

Ross-Hellauer, Tony. ‘What Is Open Peer Review? A Systematic Review [Version 2; Referees: 4 Approved]’. *F1000Research* 6 (31 August 2017): 588. doi.org/10/gc5sjh

Sandve, Geir Kjetil, Anton Nekrutenko, James Taylor, and Eivind Hovig. ‘Ten Simple Rules for Reproducible Computational Research’. Edited by Philip E. Bourne. *PLoS Computational Biology* 9, no. 10 (24 October 2013): e1003285. doi.org/10/pjb

Scacchi, Walt. ‘The Future of Research in Free/Open Source Software Development’. In *Proceedings of the FSE/SDP Workshop on Future of Software Engineering Research - FoSER’10*, 315. Santa Fe, New Mexico, USA: ACM Press, 2010. doi.org/10.1145/1882362.1882427

Scopatz, Anthony, and Kathryn D. Huff. *Effective Computation in Physics: Field Guide to Research in Python*. Sebastopol, CA: O’Reilly Media, 2015. PDF.pdf)

Sewell, Claire. ‘Research Data Management: Activity Cards’, 23 November 2017. doi.org/10.17863/CAM.10074

Shamir, Lior, John F. Wallin, Alice Allen, Bruce Berriman, Peter Teuben, Robert J. Nemiroff, Jessica Mink, Robert J. Hanisch, and Kimberly DuPrie. ‘Practices in Source Code Sharing in Astrophysics’. *Astronomy and Computing* 1 (February 2013): 54–58. doi.org/10/gc5sjk

Smith, Arfon M., Daniel S. Katz, Kyle E. Niemeyer, and FORCE11 Software Citation Working Group. ‘Software Citation Principles’. *PeerJ Computer Science* 2 (19 September 2016): e86. doi.org/10/bw3g

Smith, Arfon M., Kyle E. Niemeyer, Daniel S. Katz, Lorena A. Barba, George Githinji, Melissa Gymrek, Kathryn D. Huff, et al. ‘Journal of Open Source Software (JOSS): Design and First-Year Review’. *PeerJ Computer Science* 4 (12 February 2018): e147. doi.org/10/gc5sjf

Soergel, David A. W. ‘Rampant Software Errors May Undermine Scientific Results [Version 2; Referees: 2 Approved]’. *F1000Research* 3 (2015): 303. doi.org/10/gc5sjg

Steinmacher, Igor, Marco Aurelio Graciotto Silva, Marco Aurelio Gerosa, and David F. Redmiles. ‘A Systematic Literature Review on the Barriers Faced by Newcomers to Open Source Software Projects’. *Information and Software Technology* 59 (March 2015): 67–85. doi.org/10/f6z643

Stodden, Victoria. ‘The Scientific Method in Practice: Reproducibility in the Computational Sciences’. *SSRN Electronic Journal*, 2010. doi.org/10/fzmph2

Union, Publications Office of the European. ‘Evaluation of Research Careers Fully Acknowledging Open Science Practices : Rewards, Incentives and/or Recognition for Researchers Practicing Open Science.’ Website, 14 November 2017. publications.europa.eu

Vandewalle, Patrick. ‘Code Sharing Is Associated with Research Impact in Image Processing’. *Computing in Science & Engineering* 14, no. 4 (July 2012): 42–47. doi.org/10/gc5sjp

Vicente-Saez, Ruben, and Clara Martinez-Fuentes. ‘Open Science Now: A Systematic Literature Review for an Integrated Definition’. *Journal of Business Research*, January 2018. doi.org/10/gc5sjb

Wicherts, Jelte M., Coosje L. S. Veldkamp, Hilde E. M. Augusteijn, Marjan Bakker, Robbie C. M. van Aert, and Marcel A. L. M. van Assen. ‘Degrees of Freedom in Planning, Running, Analyzing, and Reporting Psychological Studies: A Checklist to Avoid p-Hacking’. *Frontiers in Psychology* 7 (2016): 1832. doi.org/10/gc5sjn)

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. ‘The FAIR Guiding Principles for Scientific Data Management and Stewardship’. *Scientific Data*, 15 March 2016. doi.org/10/bdd4

Wilsdon, James, Judit Bar-Ilan, Robert Frodeman, Elisabeth Lex, Isabella Peters, Paul Wouters (2017). Next-generation metrics: Responsible metrics and evaluation for open science. Directorate-General for Research and Innovation.European Commission Expert Group on Altmetrics. PDF

Wilson, Greg. ‘Software Carpentry: Lessons Learned [Version 2; Referees: 3 Approved]’. *F1000Research* 3 (2016): 62. doi.org/10/gc5sjr

Wilson, Greg, Jennifer Bryan, Karen Cranston, Justin Kitzes, Lex Nederbragt, and Tracy K. Teal. ‘Good Enough Practices in Scientific Computing’. Edited by Francis Ouellette. *PLOS Computational Biology* 13, no. 6 (22 June 2017): e1005510. doi.org/10/gbkbwp

Wilson, Paul M., Mark Petticrew, Mike W. Calnan, and Irwin Nazareth. ‘Disseminating Research Findings: What Should Researchers Do? A Systematic Scoping Review of Conceptual Frameworks’. *Implementation Science* 5 (22 November 2010): 91. https://doi.org/10/cprfmr.

Wilson, Paul M, Mark Petticrew, Mike W Calnan, and Irwin Nazareth. ‘Disseminating Research Findings: What Should Researchers Do? A Systematic Scoping Review of Conceptual Frameworks’. *Implementation Science* 5, no. 1 (December 2010). https://doi.org/10/cprfmr.

About the authors & facilitators

Authors at the sprint event

Sonja Bezjak

- University of Ljubljana, Slovenia
- sonja.bezjak@fdv.uni-lj.si
- @sonja_adp

In the Social Science Data Archives I am primarily engaged with issues related to open access to research data. One of my roles is to train different stakeholders on research data policy, research data management planning, data citation, data publications etc. As a member of CESSDA ERIC training group I try to share my knowledge and experience internationally.

I was taught about scientific values, including transparency and reproducibility while studying sociology. But only later from my friends, a physicist and an engineer, I learnt about the Open Source movement. I immediately understood the importance of spreading the idea of not hiding your findings and sharing your knowledge with others as soon as possible. Only when I started to work at the Social Science Data Archives (University of Ljubljana, Slovenia) and became heavily involved in the Open data project I realized how much effort was needed to change the culture and to be able to get over the barriers of not openly sharing research outputs. I hope this handbook will help in making science as open and understandable as possible.

Philipp Conzett

- UiT The Arctic University of Norway, Norway
- philipp.conzett@uit.no
- @philippconzett
- 0000-0002-6754-7911

Trained as a linguist, I only had a vague understanding of Open Science when I started to work as a research librarian at UiT The Arctic University of Norway back in 2014. Luckily, I soon was involved in developing and running research support services, including repositories for open research data, starting with a discipline-specific one (TROLLing), then an institutional one (UiT Open Research Data), and finally a nationwide one (DataverseNO). Participating in the Open Science book sprint has been a fruitful contribution to my training competence.

There are two major pitfalls for Open Science trainers, as I see it. One, novice trainers may feel so overwhelmed by the topics to cover, and the available resources that they don't get started. Two, experienced trainers promoting Open Science may turn their efforts too much into a movement only accessible for the initiated. I hope this book can help to overcome both obstacles.

Pedro L. Fernandes

- Instituto Gulbenkian de Ciência, Portugal
- pfern@igc.gulbenkian.pt
- @pfern
- 0000-0003-2124-0241

I run a training program in Bioinformatics at the Instituto Gulbenkian de Ciência, in Oeiras, PT since 1999. More than 5000 course participants in 19 years. Extending this activity with distance and e-learning, to better reach for 21st century learners. I am an advocate of Open Access, Open Data, Open Source and Open Science that takes any possible chance to put these causes through via training. I am conscious that this movement needs to scale-up and reach for non-scientists as well, so I am very interested in its amplification and diffusion.

Open Science is an attitude that requires a large but feasible education step. Advocates like me need to join forces and make it happen every day. Training in Open Science is needed at a wide range of levels. To address the entry level, together with Rutger A.Vos, we prepared the free e-book "Open Science, Open Data, Open Source" in 2017 (http://osodos.org). More advocacy and training to come.

Edit Görögh

- University of Göttingen, Germany
- goeroegh@sub.uni-goettingen.de
- @gorogh_edit
- 0000-0002-0766-418X

I am currently working at the University of Göttingen as a project officer for OpenUP, an EU funded project which aims at developing a cohesive framework for new methods, indicators and tools for peer review, dissemination of research results, and impact measurement. I have been in involved in knowledge management and open science/access related programs for more than 10 years.

Working for Open Science projects, I had the chance to get acquainted with both the diverse community of Open Science advocates and the reluctant, skeptical groups of researchers and decision makers, which both urged me to get more immersed in the Open Science discourse and follow developments and learn about the tools and methods to speak effectively about the benefits and challenges we face in the changing world of research communications.

Kerstin Helbig

- Humboldt-Universität zu Berlin, Germany
- kerstin.helbig@cms.hu-berlin.de
- @FrauHelbig
- 0000-0002-2775-6751

I am research data management coordinator at Humboldt-Universität zu Berlin, Germany. In my consultative capacity, I assist researchers in the management of their research data and organize training as well as information sessions.

For me the biggest challenge with open science training is to show researchers that open science is more than a political aim or a moral responsibility. It is essential to show that there are levels of open science. One can start with a little step without having to open up completely from one day to another. In my trainings, I especially like the mix of backgrounds, disciplines and prior knowledge. They make the training all the more interesting. I remember one training course in particular: one participant (a professor) registered an ORCID on the spot while I was talking about the advantages of persistent identifiers.

Bianca Kramer

- Utrecht University, Netherlands
- b.m.r.kramer@uu.nl
- @MsPhelps
- 0000-0002-5965-6560

By day, I am a librarian for life sciences and medicine. My after hours project is 101 Innovations in Scholarly Communication together with Jeroen Bosman. We do research, training and advocacy on open science, to make research more relevant, robust and equitable.

Training in open science is rewarding because it is not just about teaching people new skills, it's about discussing fundamental concepts and exchanging different viewpoints and opinions. As a participant in one of our courses said: 'I came to learn practical things to apply in my research, but I discovered I am now part of a movement'. To me, a successful training should be interactive and hands-on, to encourage people to explore and challenge their perceptions. That includes my own role as a trainer: always be open to try new things and learn from the people participating in your training.

Ignasi Labastida

- Universitat de Barcelona, Catalonia
- ilabastida@ub.edu
- @ignasi
- 0000-0001-7030-7030

PhD in Physics, Universitat de Barcelona (UB), 2000. Now, devoted to openness: Head of the Office for the Dissemination of Knowledge at the CRAI of the UB and public leader of Creative Commons in Spain since its beginning in 2003.

I hope in the near future there will be no need to train about open science because those practices, now described here, will be the default ones. There will be no need to attach the open tag anymore, and researchers would need to justify why they close some of their results or activities. I think this book may help to achieve this situation by showing a lot of robust examples and viable cases to perform research openly.

Kyle Niemeyer

- Oregon State University, USA
- kyle.niemeyer@oregonstate.edu
- @kyleniemeyer
- 0000-0003-4425-7097

I am an Assistant Professor of Mechanical Engineering at Oregon State University in Corvallis, Oregon, USA. My research group studies combustion and fluid flows using computer simulations, and develop numerical methods and parallel computing strategies. Open science advocate!

As a graduate student, I frequently faced roadblocks in my research due to software not being shared openly; now, as the leader of a research group, my students and I face data availability and formatting challenges when working with results in the literature. However, simply showing others how easy it can be to share research products openly can be enough to catalyze change, as can leading by example.

Fotis Psomopoulos

- Center for Research and Technology Hellas, Greece
- fpsom@issel.ee.auth.gr
- @fopsom
- 0000-0002-0222-4273

Fotis is a Bioinformatician at the Institute of Applied Biosciences (INAB|CERTH) in Thessaloniki, Greece. He was awarded his PhD in Electrical and Computer Engineering in 2010 with a focus on Bioinformatics and e-infrastructures, and a particular appreciation to open and reproducible methods. He spends significant time in training activities, both within formal academic structures as well as through the Carpentries as a certified Instructor and Trainer. He rambles about bits and pieces on his website.

Convincing people that spending the extra time to put together a Jupyter notebook with all the text, notes, scripts and data currently stored in various "dusty" and forgotten folders on their computer, will actually help them become a bit more organized. #smallvictories #reproducibility

Tony Ross-Hellauer

- Know-Center GmbH, Austria
- tross@know-center.at
- @tonyR_H
- 0000-0003-4470-7027

Tony Ross-Hellauer is Senior Researcher (Open Science) at Know-Center, Graz, Austria. He has a PhD in Information Studies (University of Glasgow, 2012) and is an enthusiastic advocate of Open Access and Open Science whose research interests include peer review, metadata, and the philosophy/history of technology.

Although creating and delivering training events is very daunting, training others not only to do Open Science, but also to see the value of it for their everyday research, is one of the most rewarding aspects of working in this area. As a trainer, when learners are engaged to share their own experiences and you can feel how they are able to relate their new knowledge to these experiences, it is very exciting.

René Schneider

- HES//SO - Geneva School of Business Administration, Switzerland
- rene.schneider@hesge.ch
- @datosestupendos
- 0000-0003-4897-8561

René Schneider is a professor in Information Science at Geneva School of Business Administration (being part of the University of Applied Sciences and Arts Western Switzerland). Originally trained as a computational linguist, he is mainly interested in data and all of its aspects.

I discovered the field of research data management quite lately and mainly got engaged because of the complexity and high potential of open science. After having managed a project on how to train librarians to become instructors for research data management (www.researchdatamanagement.ch), I experienced myself that open science open doors, leads to a better understanding and reuse of scientific outcomes and finally links the academic ivory tower to the world outside.

Jon Tennant

- Open Science MOOC, Germany
- jon.tennant.2@gmail.com
- @protohedgehog
- 0000-0001-7794-0218

Jon finished his award-winning PhD in Palaeontology at Imperial College London in 2017, and became the Communications Director of ScienceOpen for two years in 2015. Now, he is independently continuing his research into dinosaur evolution, while working on building an Open Science MOOC to help train the next generation of researchers in open practices. He has published papers on Open Access and Peer Review, is currently leading the development of the Foundations for Open Science Strategy document, and is the founder of the digital publishing platform paleorXiv. Jon is also an ambassador for ASAPbio and the Center for Open Science, a Mozilla Open Leadership mentor, and the co-runner of the Berlin Open Science meetup. He is also a freelance science communicator and consultant, and has written a kids book called Excavate Dinosaurs.

I think the most challenging aspect of Open Science is education. It is an enormously complex paradigm, with its own lexicon, practices, principles, and represents a quite high learning barrier in many cases. However, watching others develop their knowledge and skills is incredibly rewarding, and I find myself learning more with every new experience too. Ultimately, we all have the same thing in mind - a fairer, more equitable, transparent and rigorous system of scientific research, and watching the huge steps the global research community, and especially younger generations, are taking towards this is very inspiring.

Ellen Verbakel

- 4TU.Centre for Research Data, Netherlands
- p.m.verbakel@tudelft.nl
- @Ellen4TUData
- 0000-0002-8194-6724

Ellen is a librarian by education. She has a long experience in faculty librarianship at the TU Delft. After that she worked at the Delft University Press and organised peer review process for three journals. She also designed the open access or the journals, back in 2000! As from 2005 she developed the publication repository from TU Delft and she moved to 4TU.Centre for Research Data (at that time 3TU.Datacentrum) in 2009. In 2013 she co-designed the training Essentials 4 Data Support, she is since then an enthusiastic trainer.

Where would we be without training? We need to be aware of all aspects of Open Science and be able to enthusiasm many others! This Handbook helps educators to make their training more effective in order to make Open Science the standard.

Authors at the sprint event remotely

April Clyburne-Sherin

- Code Ocean, USA
- april.clyburne.sherin@gmail.com

- @april_cs & @methodpodcast
- 0000-0002-5401-7751

April is an epidemiologist, methodologist, and expert in open science tools, methods, training, and community stewardship. She holds an MS in Population Medicine (Epidemiology). Since 2014, she has focussed on training scientists in open and reproducible research methods (Center for Open Science, Sense About Science, SPARC). In her current role of Outreach Scientist, she trains scientists in computational reproducibility best practices using Code Ocean.

I have been lucky enough to make a living out of training other scientists how to science better. My community of support grows with each workshop and I hope this handbook might help grow the open research training community. Conversations about open research often occur in echo-chambers of well-meaning researchers (like myself) and librarians with similar worldviews. Training in open research can be similarly siloed with Western or Northern perspectives being taught as though universal. Adding context and new perspectives to open research conversations is the only way to make knowledge work for everyone. The content we captured during this sprint is limited by our own experiences, but as other authors add and edit based on their own experiences, we can aim for a handbook that can improve how we talk and train others in open research.

Facilitators on site

Helene Brinken

- University of Göttingen, State and University Library, Germany
- brinken@sub.uni-goettingen.de
- @helenebrinken
- 0000-0002-3278-0422

Responsible for Outreach and Advocacy in the FOSTER project at Göttingen University since May 2017. Background in Information Science with focus on e-learning and usability & user experience, now developing learning materials and facilitating workshops.

Before working for FOSTER I worked with young activists engaging for worldwide education and against social injustice. I learned how important group dynamics are and what can be achieved when combining forces. Culture change starts at the level of individuals. Bringing together the researchers interested in Open Science can be a great step forward to foster OS at your institution. If they get support, meet other enthusiasts and learn they can soon be multiplicators themselves.

Lambert Heller

- TIB - German National Library of Science and Technology, Hannover, Germany
- lambert.heller@tib.eu
- @Lambo
- 0000-0003-0232-7085

With a background in social sciences, I’m a librarian by training, working as a subject specialist at a university library for several years, and kicked of the Open Science Lab at TIB (German National Library for Science and Technology) in 2013, now running a number of grant projects. Facilitating and advising book sprints since 2014. Helped to make VIVO, a free current research information system (CRIS) based entirely on Linked Open Data, popular in Germany. Kicked of a few discussions in libraryland and elsewhere, e.g. on Blockchain for Science.

When giving workshops (e.g., a half day workshop for PhD students and PostDocs from Leibniz Research Association in Germany in 2017, on the matter of scholarly profile and collaborative writing services) it’s always a pleasure to tap into the curiosity of learners. Even the busiest student has experiences, questions and imagines how things could work best for them. I love to make use of this positive energy! And it makes it much easier for a trainer to run a training session.

The handbook has been translated into Spanish and Portuguese:

- Portuguese translation
- Spanish translation

Currently, we are translating the book into more languages such as Greek, French and Italian. We will link them here as soon as they are availabe.