https://open-science-training-handbook.github.io/Open-Science-Training-Handbook_EN/

1. Introduction

"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 accessbile to and reusable by all, is about culture and knowledge as much as it is about technologies and services.
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. The handbook suggests training activities that can be adapted to various settings and target audiences

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. Aims to improve the quality and reliabilty of research through principles like inclusion, fairness, equity, and sharing.
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.

2. 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.

2.1 Open Concepts and Principles
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 Taxonomy
Figure Pontika et al., 2015

Openen 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.

2.2 Open Research Data and Materials
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.

FAIR principles
https://force11.org/info/the-fair-data-principles/
Findable: To make data reusable, the first thing should be that it is possible to find. It should be easy to find the data and the metadata for both humans and computers.
Accesible: 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).

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.

2.3 Open Research Software and Open Source

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.

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.

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?

2.4 Reproducible Research and Data Analysis

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.

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."

How to Reproducibility:

Keep track of things

Registration

Preregister important study design and analysis information to increase transparency and counter publication bias of negative results.

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.

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.

Report your research transparently

Report and publish your methods and interventions explicitly and transparently and fully to allow for replication.

2.5 Open Access to Published Research Results

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.

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.

2.8 Open Peer Review, Metrics, and Evaluation

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).

2.9 Open Science Policies

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.

2.11 Open Educational Resources

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.