

Problems of Ethics, Open Science and Open Innovation

Jakub Rydzewski

NCU Institute of Physics

Updated: September 9, 2022

Course Information

- ▶ Basic information: **USOS**
- ▶ GitHub repository: **<https://github.com/jakryd/7404-EONOI>** where the lecture and exercise are available.
- ▶ Lecture: 30 min + 60 min for the exercise.
- ▶ Send questions and remarks to **jr@fizyka.umk.pl**.
- ▶ (Don't worry about passing).

Outline

Open Science

- ▶ What is Open Science?
- ▶ Tools, workflows, methodology
- ▶ Licensing
- ▶ Example: (Narodowe Centrum Nauki) NCN application
- ▶ Exercise

Definition

“Open science is the movement to make scientific research (including publications, data, physical samples, and software) and its dissemination accessible to all levels of society, amateur or professional.”

Definition

Open scientific research mostly in STEM
(Science, technology, engineering, and mathematics)¹

- ▶ Methodology (workflow, protocol, methods)
- ▶ Source (software, codes, implementations)
- ▶ Data (results)
- ▶ Access (for anyone)
- ▶ Peer review (verification)

¹However, even literature writers use such protocols sometimes.

Definition

Advantages:

- ▶ Free and open access to research reports and data makes peer-review rigorous
- ▶ Publicly funded science should be publicly available
- ▶ Science is reproducible and transparent
- ▶ Lower chance of human error
- ▶ High-quality science
- ▶ Plagiarism detectable
- ▶ More impact

Tools

Available software and data servers:

- ▶ **Publications:** Overleaf, Google Docs, GitHub, GitLab
- ▶ **Data:** Zenodo, GitHub, GitLab
- ▶ **Software:** GitHub, GitLab
- ▶ **Preprint servers:** arXiv and its derivatives

Each allows for:

- ▶ Tracking changes
- ▶ Versioning
- ▶ Collaboration
- ▶ Public free access
- ▶ Easy licensing
- ▶ Obtaining DOI

Licensing

Open access requires licensing:

- ▶ Allows for retaining copyright (e.g., CC license)
- ▶ Access to source and important criteria or usage
- ▶ Collaboration
- ▶ Free redistribution
- ▶ Derived work
- ▶ Examples: GPL, LGPL, Mozilla, Apache, BSD, MIT

Licensing

Current Funding Cycle for Research Articles

Limited dissemination, economic efficiency and social impact



Optimized Funding Cycle for Research Articles

Maximum dissemination, economic efficiency and social impact



This document by Creative Commons, licensed CC BY

Icons from *The Noun Project* by: CC BY 3.0: Adam Whitcroft (cloud upload) - Anand A Nair (share) - Andrew Forrester (locks) - Diego Naive (speaker, book) - Emma Vila Hopkins (turtle) - José Hernandez (fence) - Henrik Lund Mikkelsen (money) - Sotiros Papavasiliopoulos (light bulb) - Thomas Weber (community) CCB: Antonis Makriyannis (scientist) - Arthur Schmitt (construction) - Dsathiyaraj (folder) - Fission Strategy (download) - Max Hancock (brain) - Mike Wirth (hourglass)

Example: NCN Application

- ▶ NCN supported research must be open access
- ▶ Each proposal must contain information about data management plan
- ▶ Many journals require processing and open access charges
- ▶ Making research available for free can be done also via preprint servers
- ▶ Sometimes it is possible to directly import a preprint when submitting work for review
- ▶ Often journals highly recommend providing free access to data and protocols
- ▶ ... collecting processing and open access charges is often exploited by predatory journals

Example: NCN Application

Data Management Plan

Formal requirements

To be filled out in English.

Before filling out the form, read the Guidelines for completing the data management plan for a research project.

The NCN recognises that some projects will not generate, re-use or analyse research data and similar materials. In these cases, a short explanation is required as an answer to the question: How will new data be collected or produced and/or how will existing data be re-used?

Let's have a look how specific NCN's data management plan is ...

Example: NCN Application

1. Data description and collection or re-use of existing data.

- ▶ How will new data be collected or produced and/or how will existing data be re-used?

How do you generate data? Do you use data generated by other research?

- ▶ What data (for example the types, formats, and volumes) will be collected or produced?

Discipline specific; may be even plain text files or Excel sheets.

Example: NCN Application

2. Documentation and data quality.

- ▶ What metadata and documentation (for example methodology or data collection and way of organising data) will accompany data?
Very important: you have to explain what is in your data and how to use it (e.g., README files, code documentation).
- ▶ What data quality control measures will be used?
Are there any automatic protocols that you can use to ensure that your data is of highest quality? Doing it by hand may be very exhausting and does not exclude human error.

Example: NCN Application

3. Storage and backup during the research process.

- ▶ How will data and metadata be stored and backed up during the research process?

Backups, best to use version control server (Zenodo, GitHub)

- ▶ How will data security and protection of sensitive data be taken care of during the research?

For instance, when dealing with patient records.

Example: NCN Application

4. Legal requirements, codes of conduct.

- ▶ If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?
- ▶ How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

Licences.

Example: NCN Application

5. Data sharing and long-term preservation.

- ▶ How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?
If freely accessible or partially, upon request.
- ▶ How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?
Should be at least 5 years.
- ▶ What methods or software tools will be needed to access and use the data?
- ▶ How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?
Zenodo, arXiv, publications.

Example: NCN Application

6. Data management responsibilities and resources.

- ▶ Who (for example role, position, and institution) will be responsible for data management (i.e the data steward)?
Do you manage your data yourself or is there someone who does it for you?
- ▶ What resources (for example financial and time) will be dedicated to data management and ensuring the data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Exercise

Provide an answer to each point above.

- ▶ For each question up to 3 sentences
- ▶ Must be discipline specific (different requirements for, e.g., mathematics and biology)
- ▶ Write your answers in a **plain text file**
- ▶ Send the file to jr@fizyka.umk.pl