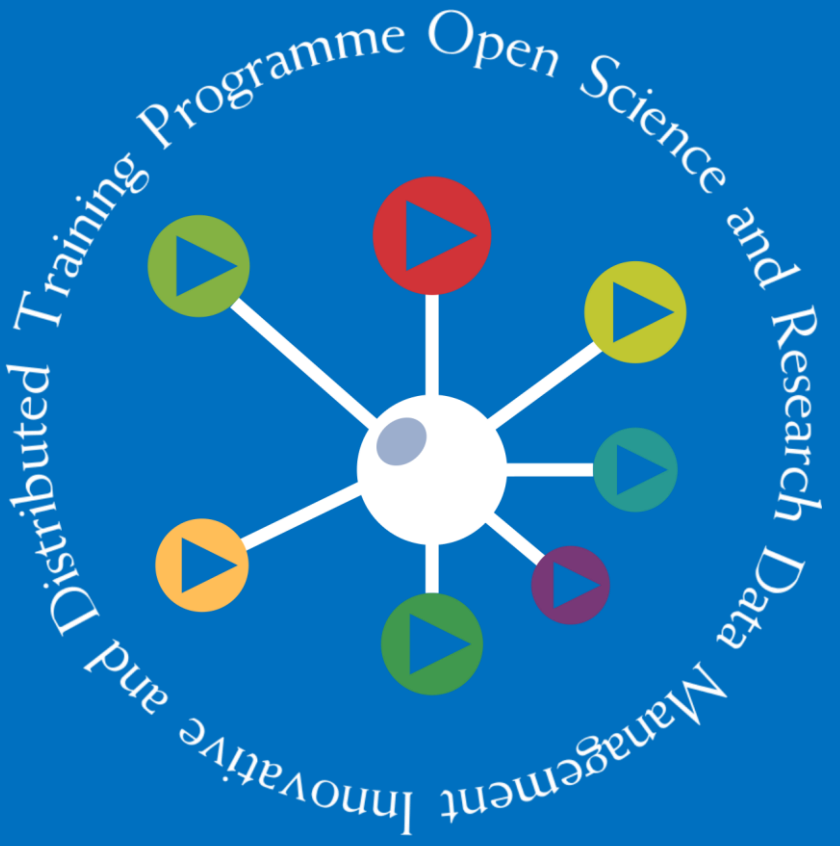
Open Science

Hands-On



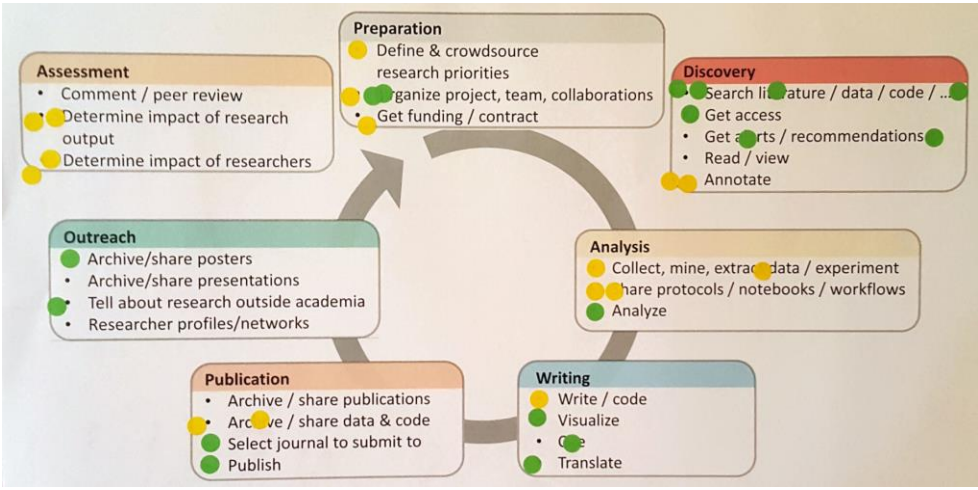
University POLITEHNICA of Bucharest

Based on: https://open-science-training-handbook.github.io/

Open Concepts and Principles

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.

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

Open Science discussion topics (1)

- **Suggestions for discussion topics:**
  - Working in an Open Science manner makes research more fun"
  - "Scooping is a real and existing problem that makes Open Science a hard choice"
  - "APCs (article processing charges) are the main obstacle to publishing more in Open Access"
  - "We need more explicit support for Open Science from funders and the government"
  - "Engaging in open peer review is problematic for young researchers that want to make a career"
  - "We should take citizen scientists more seriously, and also not just see them as data suppliers"
  - "Impact factors are a symptom and not the cause of the publishing rat-race"
  - "There is absolutely no reason we should not publish a paper as a preprint as soon as it is ready"
  - "Just sharing our data is fine, but to speed up science we need to also work on interoperability and reusability of those data"
  - "Sharing ideas and projects through ResearchGate is a good way of doing outreach for our research"
  - "Demands of our PIs are probably the main reason why young researchers do not engage more in Open Science"
  - "We should strive to create a kind of ‘commons’ where we share all our research outcomes/objects to foster collaboration and reuse"

Open Science discussion topics (2)

Discussion topics

Topic 1
Making research more fun

Topic 2
Scooping

Topic 3
APCs

Topic 4
Support from funders & government

Topic 5
Open peer review

Topic 6
Citizen science

Topic 7
Impact factors

Topic 8
Preprints

Topic 9
Data sharing

Topic 10
ResearchGate for outreach

Topic 11
PI demands

Topic 12
Scholarly commons

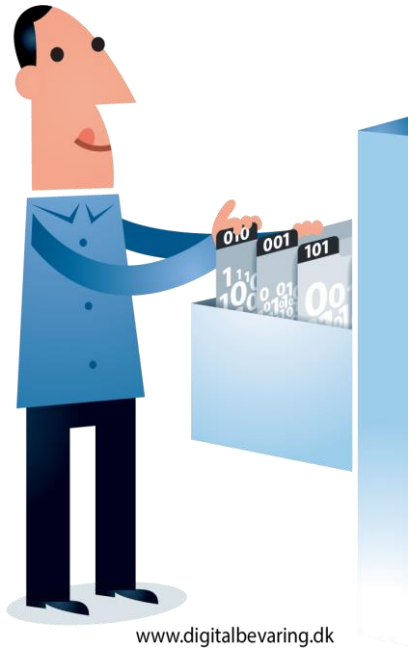
Open Access to Published Research Publications

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
  - 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.

Finding a 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 best repository for your data
- Exercise description
  - Browse www.re3data.org for relevant repositories for your own research project.



Finding a Repository (2)

- Search the Registry of Research Data Repositories (https://www.re3data.org/) for a suitable repository for your research data.

10 minutes

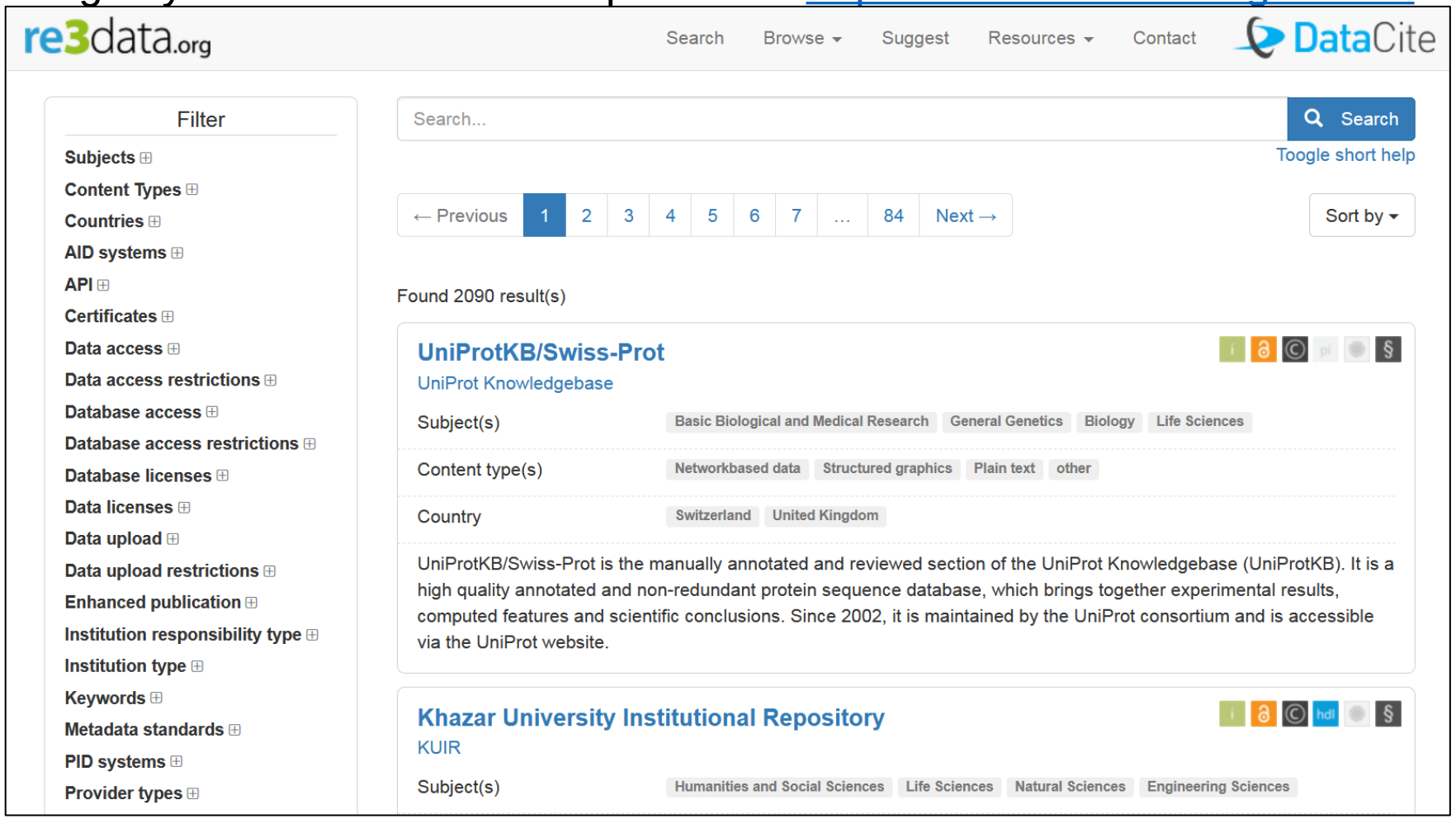
  - How many repositories do you find for your subject?
  - Use the filter criteria (e.g. repository type, country, data upload restrictions) to further limit your result.
  - Take a look at your favorites on the detail page. What do you find regarding metadata standards, data upload and data access? Which persistent identifier system is used?

- Discussion: What did you find? What do you look for in particular when selecting a repository?

10 minutes

How do I find a suitable repository?

Registry of Research Data Repositories https://www.re3data.org/search



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.

Choose the right version for the repository (1)

- Discuss with your neighbor:
  - What criteria would you use to select a repository? What is particularly important to you?
- Consider criteria such as
  - costs,
  - storage location,
  - reliability,
  - visibility,
  - citability, storage duration,
  - support offered, etc.

Choose the right version for the repository (2)

- Do you know of a recognized repository for research data in your discipline?
- Is there an institutional repository in your institution?
- Would publishing on Figshare or Zenodo be an alternative for you?
- Search on re3data.org for possible additional repositories for your research data.
- What criteria are important to you when selecting a repository? Which repository would you select?

Set up OSF project & link to other platforms

- **Format, time needed**
  - Individually or in pairs
- **Topic**
  - Open Research Data and Materials
- **Learning objectives**
  - Learn to use data sharing platforms
- **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.

Set up OSF project & link to other platforms (2)

- Lou is a first year graduate student working on a project in a biomedical research laboratory. He's trying to decipher data left by a former post doc as a start for his thesis project.
- **Download** Lou's files:
  - http://tinyurl.com/hvna4mg
- Create an OSF account
- Create a project called "Lou's project"

Open Research Data and Materials

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

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).

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.

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.

THANK YOU!

Follow us