

# OPEN SCIENCE, FAIR and OPEN DATA

Czech-Bavarian MINI-SCHOOL 2020  
on large scale facilities and open data



OPEN SCIENCE  
SUPPORT CENTRE  
Charles University

22.10.2020, online lecture  
Dagmar Hanzlíková & Milan Janíček  
[researchdata@cuni.cz](mailto:researchdata@cuni.cz)

This presentation was created under the project **Development of capacities for research and development II**, reg. no. CZ.02.2.69/0.0/0.0/18\_054/0015222.



CHARLES  
UNIVERSITY



EUROPEAN UNION  
European Structural and Investment Funds  
Operational Programme Research,  
Development and Education





# Open Science

# Open Science Definition

- **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.

<https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>





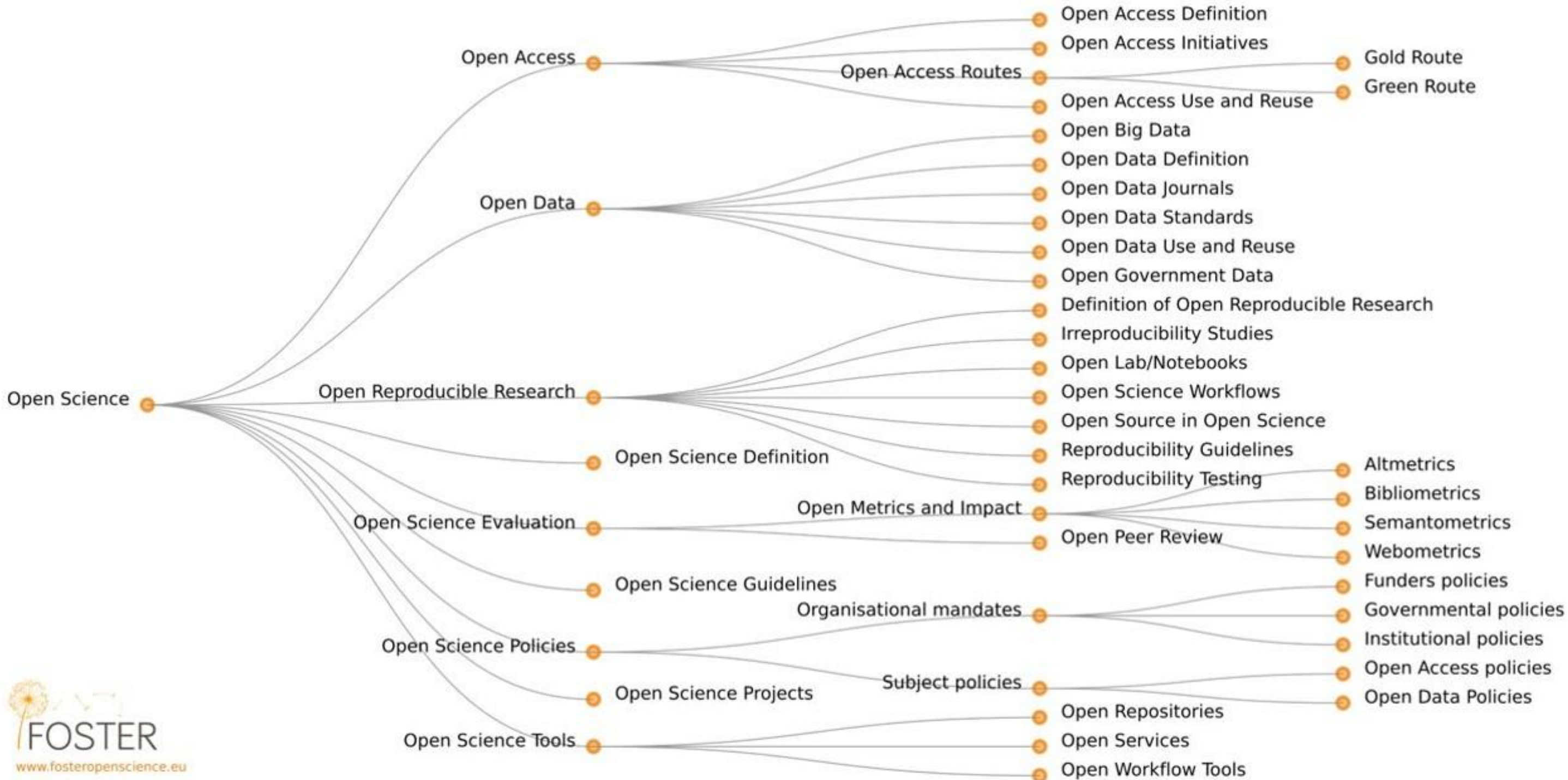
# Open Science Definition

- **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**.

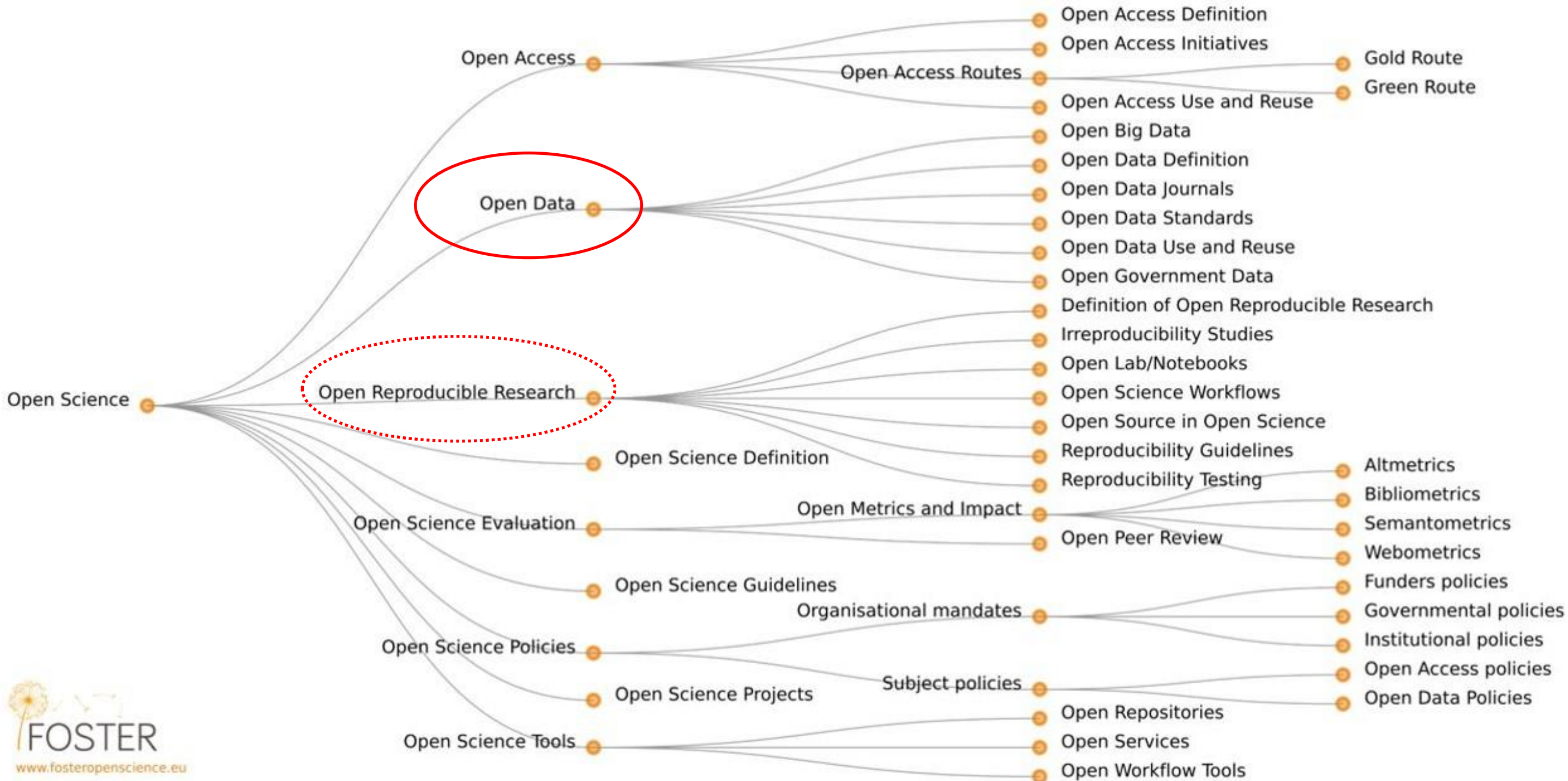
<https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition>



# Open Science Taxonomy



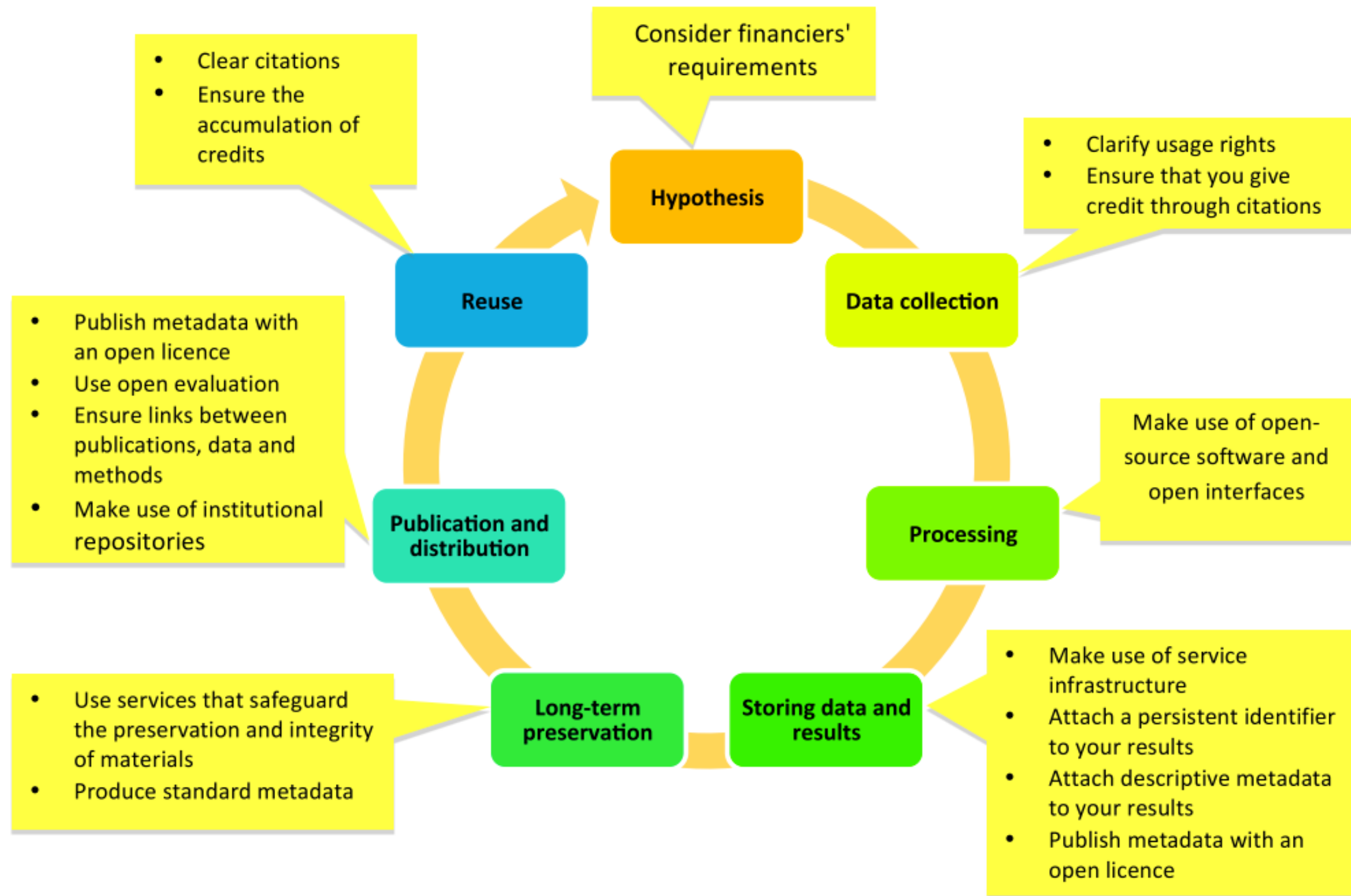
# Open Science Taxonomy



# OS influence on the research process

- Open Science is not only about publishing results – there are changes possible in different phases of research process
  - Hypothesis
  - Data collection
  - Processing
  - Storing data and results
  - Long-term preservation
  - Publication and distribution
  - Reuse







# Reproducibility



# Reproducibility and Replicability

- Reproducibility
  - research data and code are made available
  - ➔ others are able to reach same results (as are claimed in scientific outputs)
- Replicability
  - repeating a scientific methodology
  - ➔ reach similar conclusions

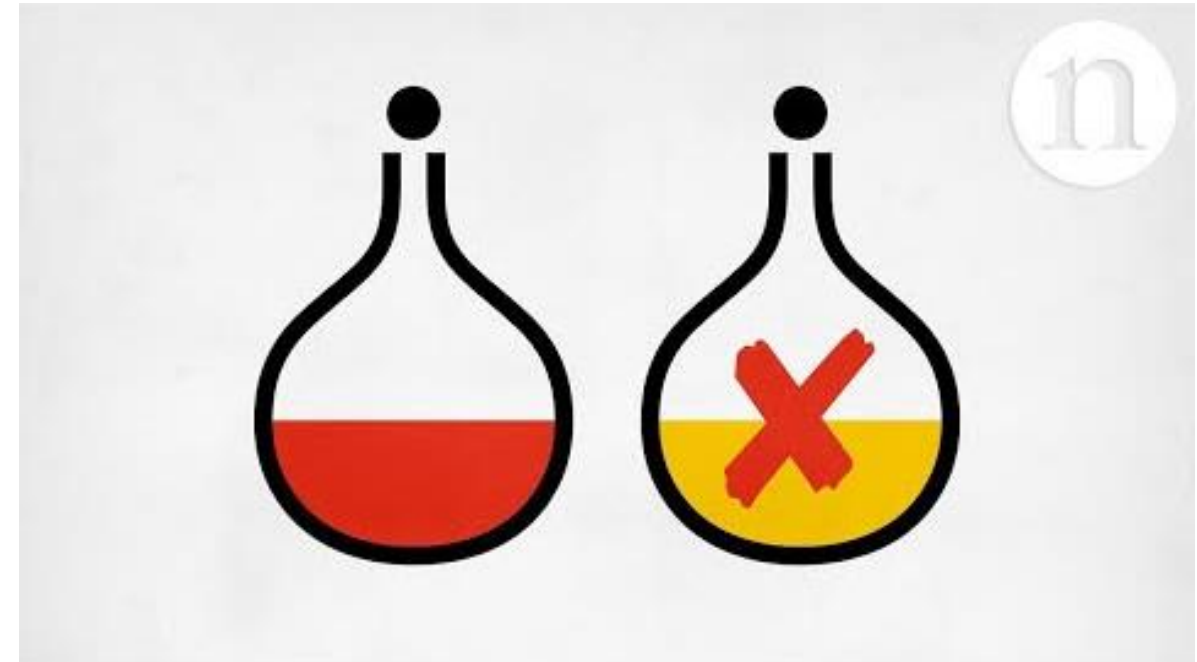
# Reproducibility Crisis?

Nature Survey

*1,500 scientists lift the lid on reproducibility*

<https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>

- Is there a reproducibility crisis?
- How much published work in your field is reproducible?
- Have you failed to reproduce an experiment?
- Have you ever tried to publish a reproduction attempt?
- Have you established procedures for reproducibility?



# Reproducibility Tips (I)

1. Plan for reproducibility before you start
  - create a study plan or protocol
  - choose reproducible tools and materials
  - set-up a reproducible project
2. Keep track of things
  - Registration (preregister important study design and analysis information..)
  - Version Control
  - Documentation

# Reproducibility Tips (2)

3. Share and license your research
  - data
  - materials
  - software, notebooks and containers
4. Report your research transparently
  - Report and publish methods and interventions explicitly
  - Post results to registration platform

<https://book.fosteropenscience.eu/en/02OpenScienceBasics/04ReproducibleResearchAndDataAnalysis.html>



# FAIR Data

# FAIR Data

- Findable
- Accessible
- Interoperable
- Reusable



# FAIR Data - Findable

- Findable
  - possibility to find data
  - both humans and computers should be able to find the data
- persistent identifiers (PIDs) are important feature for identifying the right dataset
  - eg. DOI, but there may be domain specific identifiers as well
- metadata (data describing the dataset) are important
  - There are right ways how to create metadata descriptions! Ask your librarian!;-)



# FAIR Data - Accessible

- Accessible
  - The (meta)**data** should be **retrievable** by their identifier
    - using a standardized and open communications protocol,
    - authentication and authorization is possible
      - not EVERYTHING has to be open...
  - **metadata** should be available even when the data are no longer available



# FAIR Data - Interoperable

- Interoperable
  - it should be possible to combine data with other data or tools
  - format should be open and interpretable for various tools
- both data and metadata should be interoperable
  - for metadata use vocabularies (e.g., subject headings) to describe the dataset where possible



# FAIR Data - Reusable

- Reusable
  - optimize data for reuse
    - describe data sufficiently to allow replication of results or reuse in different context
  - use license to state that data can be reused



# FAIR Principles

## To be Findable:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

## TO BE ACCESSIBLE:

- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
- A1.1 the protocol is open, free, and universally implementable.
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.

## TO BE INTEROPERABLE:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

## TO BE RE-USABLE:

- R1. meta(data) have a plurality of accurate and relevant attributes.
- R1.1. (meta)data are released with a clear and accessible data usage license.
- R1.2. (meta)data are associated with their provenance.
- R1.3. (meta)data meet domain-relevant community standards.



<https://www.go-fair.org/fair-principles/>



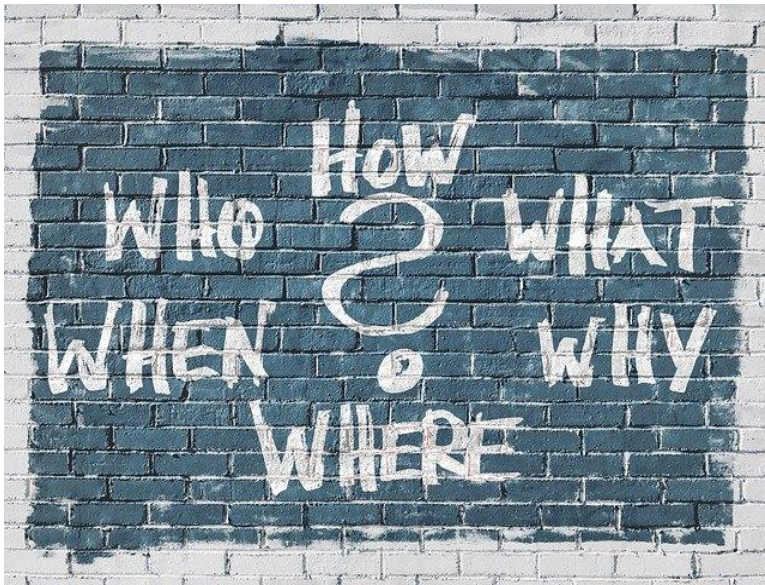
# Data Management Plan





# Data management plan (DMP)

= a document that specifies *what data will be created and how*, and outlines the plans for *sharing and preservation* of the data



## Content of the DMP

1. Administrative data
2. Data collection
3. Documentation & metadata
4. Ethics & legal compliance
5. Storage & backup
6. Selection & preservation
7. Data sharing
8. Responsibilities & resources

# Why should you create a DMP?

- Helps you anticipate **potential issues**
- Decreasing the risk of **duplication, data loss or data security breach**
- Ensuring that the data are **complete, accurate and reliable**
- Ensuring **continuity** and **consistency** in long-term projects
- Saving **time** and **energy** (e.g., when searching the data, writing up papers)
- Helps you prepare your data for **sharing**
- Meeting the requirements set out by **research funders** and **publishers**




← → ↻ 🏠 https://dmponline.dcc.ac.uk ... 🛡️ ☆

**DMP ONLINE** Home Public DMPs Funder requirements Help Language ▾


## Welcome

DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It is provided by the Digital Curation Centre (DCC).


Join the growing international community that have adopted DMPonline:




17,622 Users



203 Organisations



23,083 Plans



89 Countries

Some funders mandate the use of DMPonline, while others point to it as a useful option. You can [download funder templates](#) without logging in, but the tool provides tailored guidance and example answers from the DCC and many research organisations. Why not sign up for an account and try it out?

Sign in Create account

\* **Email**

\* **Password**

[Forgot password?](#)

☐ Remember email

Sign in

- or -

Sign in with your institutional credentials

[dmponline.dcc.ac.uk](https://dmponline.dcc.ac.uk)





# Open Research Data

# Open Research Data (ORD)

*= data that are freely available online to anyone and can be used, modified, and shared for any purpose*

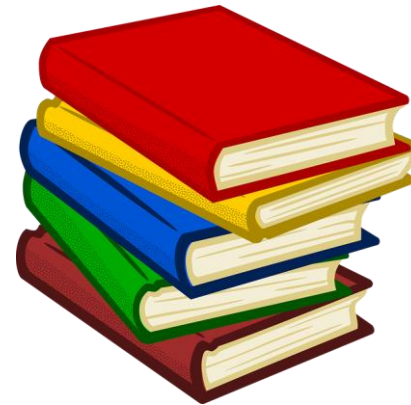
**As open as possible,  
as closed as necessary.**





# ORD: Why should you share data?

- Research funder's policy
- Open data policy of a publisher or a journal ([Nature](#), [PLOS One](#))
- EU [Directive 2019/1024](#) on open data and the re-use of public sector information





# ORD: Why should you share data?

## Benefits for **researchers**

- Ensuring your research is robust
- Enhancing your reputation and greater impact
- Increased citation rates (data & publications)
- Compliance with funders' and publishers' policies

## Benefits for **wider society**

- Effective use of resources
- Speeding up research progress
- Encouraging citizen science
- Reducing academic fraud



# ORD: Which data to share?

Generally, *everything that is needed to replicate your study should be made available*

## Keep in mind!

- Ethics and legal issues
- Who decides whether you can share your research data
- Who do you share your data with



# ORD: How to share data?

- Datasets can be shared via data repositories or data journals
  - **Data repository**
    - **Subject** specific repository, [re3data.org](https://re3data.org)
    - **General-purpose** repository, e.g., [Zenodo](https://zenodo.org) or [figshare](https://figshare.com)
  - **Data journal**: Peer-reviewed journals that publish papers describing datasets that are made available via a repository, e.g., [Scientific Data](https://www.nature.com/scientificdata/)
- Choose an appropriate license for your data, e.g., [Creative Commons](https://creativecommons.org/licenses/by/4.0/) or [Open Data Commons](https://openstax.org/)
- Share **documentation** along with the data

# Data journals: Scientific Data

= *peer-reviewed, open-access journal for descriptions of scientifically valuable datasets*

- **Data descriptors**
  - Detailed descriptions of research datasets
  - Big or small data, from new experiments or value-added aggregations of existing data, from major consortiums and single labs
  - Data published in a public, community-recognised repository (list of recommended repositories)



Q & A



# Použité zdroje

- FOSTER Open Science. Available online at <https://www.fosteropenscience.eu/>
- Open Science and Research Initiative Open Science and Research Initiative. (2014, December). *The Open Science and Research Handbook v1*. Available online at <https://www.fosteropenscience.eu/sites/default/files/pdf/3986.pdf>
- Centrum pro podporu open science. Výzkumná data. <https://openscience.cuni.cz/OSCI-61.html> <accessed 24.8.2020>
- Digital Curation Centre. DMPonline available at <https://dmponline.dcc.ac.uk>
- Journal Information. *Scientific Data*. Nature. Available at <https://www.nature.com/sdata/>





Thank **you** for your attention.

Open Science Support Centre  
Central Library of Charles University  
[openscience@cuni.cz](mailto:openscience@cuni.cz)

