Offene Forschungsdaten (Open Data)

Johannes Sperling

21.10.2020

tu-dresden.de/kontaktstelle-forschungsdaten/
Topics

• Open Data – Why?

• Obstacles to data publication

• What can you do?
The four pillars of open science

Figure 1 from: Masuzzo & Martens: Do you speak open science? Resources and tips to learn the language. *PeerJ Preprints* (2017) [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0)
Open Science / Open Data

“science carried out and communicated in a manner which allows others to contribute, collaborate and add to the research effort, with all kinds of data, results and protocols made freely available at different stages of the research process.”

Research Information Network, Open Science case studies

“Open research data is data that can be freely accessed, reused, remixed and redistributed, for academic research and teaching purposes and beyond.”

Open Data – Why?

Good scientific practice
• Inhibiting scientific misconduct

Increasing the quality of research
• By increasing re-use and further development in later years

Increasing the efficiency of research
• Preventing duplication of work
Open Data – Why?

Enhancing efficiency for you
• Re-using existing data

Understanding the data (later)
• Helping yourself, colleagues and successors understand the data and the topics

Promoting your work
• paper + data gets more interest than just paper

source: pixabay.com
Policies: EU Horizon

- Horizon 2020: Open Data Pilot
  - „As open as possible, as closed as necessary“
  - obligatory Data Management Plan (DMP)
  - "Opt-out" possible at any time, but must be justified (possible reasons: intellectual property rights, data protection, endangering of the project goal)

- Horizon Europe: no Opt-Out option anymore?
Policies: DFG

- New Codex on Good Research Practice (2019, to be applied till July 2021)
- „Where possible and reasonable, this includes making the research data, materials and information on which the [published] results are based, as well as the methods and software used, available and fully explaining the work processes.” (…) “in recognised archives and repositories in accordance with the FAIR principles (Findable, Accessible, Interoperable, Reusable).” (in guideline 13)

- „Research data are handled in accordance with the requirements of the relevant subject area.” (in guideline 7)
Obstacles to data publication?
Data Curation Continuum

Simple Documentation (Metadata) → Data Curation → Rich Documentation (Metadata)

Data creation „just for yourself“  Data creation for your colleagues  Data creation for (open) publication
Obstacles to data publication

• Requires extra effort / time

• Commercial interests (PPPs, patents etc.)

• Privacy issues (processing of personal data)

• Systemic obstacles (academic culture, reputational systems)

Note: Data publication ≠ Open Data
→ Licenses, embargo, access control
„open“ science vs. „conventional“ science

- openness is already a premise to the very nature of science (actually, you’ve been doing it already to a certain degree all along…)

- open science is an ideal (i.e. not something you can ultimately achieve, and that’s ok)

- open science is not ultimately defined (you may decide what it means for you)

- opening science is a gradual process (you don’t have to do it all at once, you don’t have to be an open science extremist)
What can you do?

1. Know about the ideas and concepts surrounding open data

2. Constantly re-examine your current practices (Is there something (perhaps small) you could do that might be a (baby-)step in the direction of more openness in your research?)
Publishing data

• What data do I publish?
  - Data that might be interesting to others
  - Data that is the basis of published research results (reproducibility)

• Where do I publish my data?
  - institutional repositories
  - subject-specific repositories
  - general repositories
OpARA

OpARA - Open Access Repository and Archive*

• archiving and **publication (DOI)** of research data
• **metadata** need to be added during upload process
• **for static data only**

* for further information please check https://opara.zih.tu-dresden.de/xmlui/?locale-attribute=en
Other repositories

• ZENODO – [https://zenodo.org](https://zenodo.org)
• Figshare – [https://figshare.com](https://figshare.com)
• Dryad – [https://datadryad.org](https://datadryad.org)
• ...

Subject-specific repositories – [Repositorien Recommender](https://www.re3data.org) (by HTW Dresden)

Repository registry – [https://www.re3data.org](https://www.re3data.org)

...and find published data e.g. at:

[DataCite](https://www.datacite.org)
„Open Data ist nice, but requires a lot of work“?

(based on Karl Valentin)
Data management – key principles

Think about who will use your data: just you? Your group? Other partners? Open data publication?

Spending a little time upfront can save a lot of time later on.

Be realistic: strike a balance between doing too much and too little.

There’s no single right way to do it; establish a system that works for you.
Why data hygiene is important

The first person with whom you will share your data is .... you!

„Metadata is a love note to the future“ – Jason Scott
Data hygiene – first steps

- give your folders a logical **structure** (and stick with it)
- Define **rules for naming** files (and stick with them)
- Create **documentation** (e.g. put README file in folder)
Easiest/fastest way to document your data:
• README.txt
• create one for each folder (ideally)

Basic content* of any README.txt:
• Title
• Date
• Author
• Description of folder content

* for further information please check https://data.research.cornell.edu/content/readme
Make your data „FAIR“

- metadata and doi
  
  * at repositories, journals, research information systems*

- decision what data to publish
- license

- repository
  
  * find yours here: re3data.org*

- documentation
- metadata
  
  * (within dataset) standards/vocabularies*
- file formats
Open data

Probably the hardest part

• .... develop a way to monitor and enforce these steps!

Images by Gerd Altmann and kmicican at Pixabay.com
Getting help

Service Center Research Data (TUD / SLUB)

- Data management planning (for funders)

- Support for your Data management
  • consultation
  • workflows and tools
  • technical implementation

- Publishing data (e.g. TUD repository OpARA)

Image: Used with permission of the American Library Association.
Discussion & Questions
Thank you for your attention!

Questions? Need help?
Contact us!

kontaktstelle-forschungsdaten@tu-dresden.de
https://tu-dresden.de/kontaktstelle-forschungsdaten

See also: “What does Open Science imply for me as a researcher?”
https://tu-dresden.de/kontaktstelle-forschungsdaten/wissenswertes/open-data