Research Data Management

Dr. Sara El-Gebali
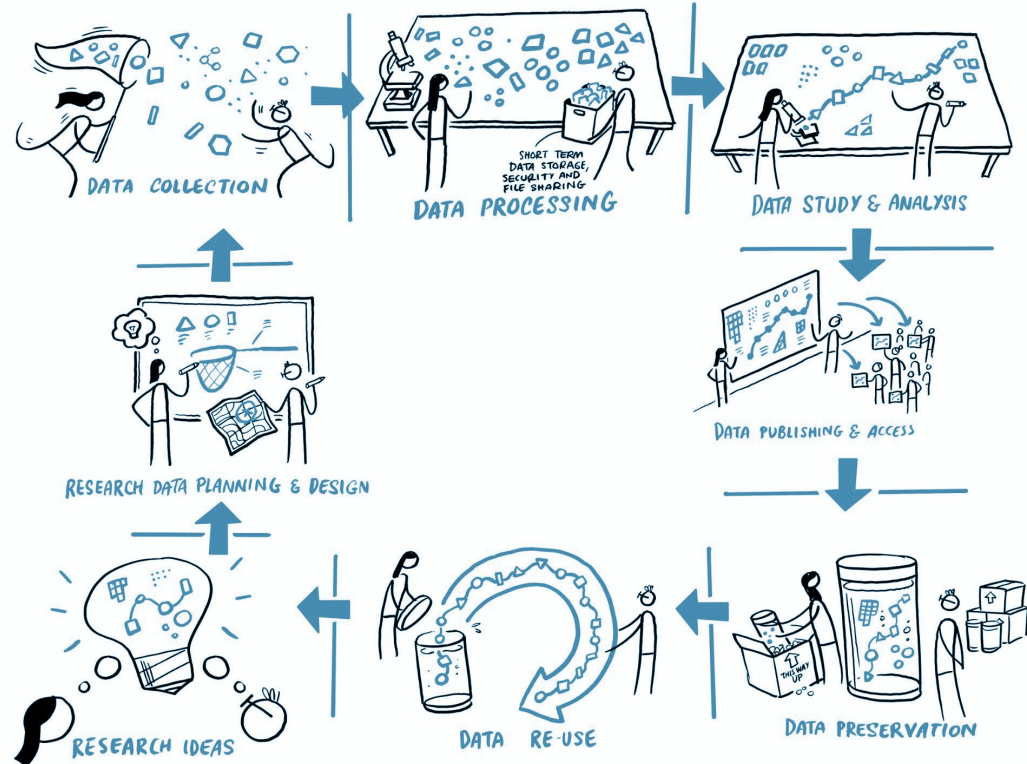
iD 0000-0003-1378-5495
@yalahowy

https://doi.org/10.5281/zenodo.5579455

Open Life Science

SciLifeLab

Research Data management

> *Everyday management of research data during the lifetime of a research project to preserve and share it beyond the project completion.*

Open Life Science

FAIR principles

FAIR is a set of principles to define the best practices for data and software to facilitate discovery, access and reuse by humans and machines.

FAIR is not rules and not a standard, it is an evolving process and a vision.

What does FAIR stand for?

Findable, Accessible, Interoperable and Reusable.

https://doi.org/10.5281/zenodo.1212496

Open Data

Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike.

The Open Data Handbook- Open Knowledge Foundation

Open Life Science

Open Data

FAIR vs Open

"FAIR means thinking about the people who could benefit from your data,"

"When we're talking about open data, we're generally referring to data that can be downloaded freely from the internet."

"A love letter to your future self":
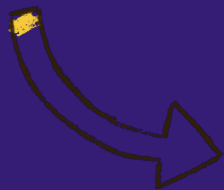What scientists need to know
about FAIR data

Open Life Science

FAIR ≠ Open

'FAIR is not the equivalent of open, but open needs to be FAIR to be useful'

Making your data/software freely and openly available does not translate to it being reusable!

To do so, we need clear, detailed contextual information and data description.

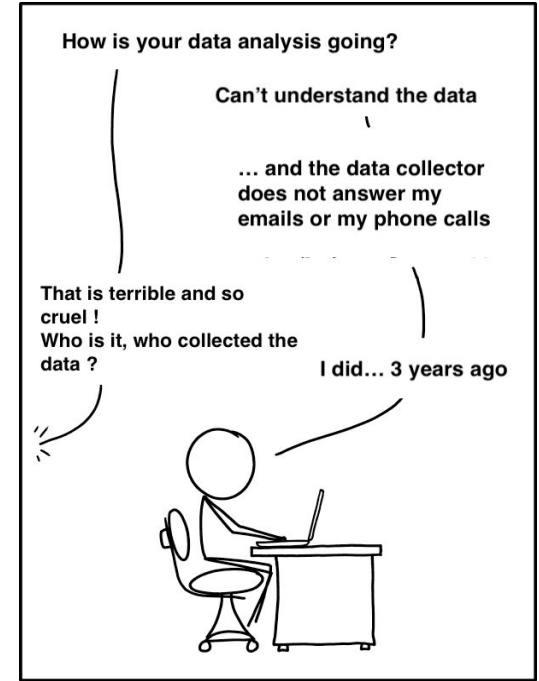Data can be FAIR but not Open! FAIR data motto "as open as possible, as closed as necessary"

Ideally you want FAIR data/software shared openly!

Why should make it FAIR?

You are the first one to reuse your data. Do you understand what you did a year ago?

You are not alone! Research relies on collaboration, can your collaborators understand what you did?

How is your data analysis going?

Can't understand the data

... and the data collector does not answer my emails or my phone calls

That is terrible and so cruel !
Who is it, who collected the data ?

I did... 3 years ago

Your first collaborators are your future selves, be nice to them !

your future self, by Julien Colomb, CC-BY-NC, derived from .NORM Normal File Format, CC-BY-NC, by Randall Munroe

Open Life Science

Why should make it Open?

Ginny Barbour
@GinnyBarbour

Today's reminder of how ~~insane~~ it is that academic research is not #openaccess The 1953 Watson/Crick paper requires a subscription to read. 😠

Content ˅   About ˅   Publish ˅

Published: 25 April 1953

Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid

J. D. WATSON ✉ & F. H. C. CRICK

Access options

Subscribe to Journal
Get full journal access for 1 year

185,98 €
only 3,65 € per issue

Subscribe

All prices are NET prices.
VAT will be added later in the checkout.
Tax calculation will be finalised during checkout.

Rent or Buy article
Get time limited or full article access on ReadCube.

Why should make it Open?

Why should make it Open?

Why should I make it open?

The cornerstone of the open-source philosophy is that the recipients of technology should have access to all its building blocks, such as software code, schematics for electronics and mechanical designs, in order to study it, modify it and redistribute it to others. Ever since releasing the World Wide Web software under an open-source model in 1994, CERN has continuously been a pioneer in this field, supporting open-source hardware (with the CERN Open Hardware Licence), open access (with the Sponsoring Consortium for Open Access Publishing in Particle Physics - SCOAP³) and open data (with the Open Data Portal for the LHC experiments).

The CERN Open Data portal is a testimony to CERN's policy of Open Access and Open Data. The portal allows the LHC experiments to share their data with a double focus: for the scientific community, including researchers outside the CERN experimental teams, as well as citizen scientists, and for the purposes of training and education through specially curated resources. The first papers based on data from the CERN Open Data portal have been published.

What can I do?

BE FAIR!

1. Deposit your data where others can find it, keep in mind where your peers can find it, i.e. field specific repository and give it a stable unique identifier (PID).
2. Make your data & metadata accessible via standard means such as http/API.
3. Create metadata and explain in detail what this data is about, never assume people know!
4. Deposit metadata with PID and make it available with/out data i.e. in case data itself is heavily protected.
5. Include information on ownership and provenance.
6. Outline what the reusers of your data are/not allowed to do, use clear license. Commonly used licenses like MIT or Creative Commons (keep in mind funders requirements).
7. Specify access conditions, if authentication or authorization is required.
8. Describe your data in a standardized fashion using agreed terminology and vocabulary.
9. Share the data in preferred & open file formats.

10.   Start the process early on!

BE OPEN!

——

No license = No access!

Data:

Add an open license (e.g. CC-by or public domain):

- ○ https://chooser-beta.creativecommons.org/

Make sure it is inline with your f institutional and funders requirements, intellectual property rights!!

Software:

Add a license, Apache-2.0 and MIT are permissive licenses with few restrictions, allowing reuse:

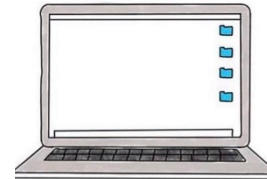- ○ https://choosealicense.com/
- ○ https://tldrlegal.com/

Build new habits

- Organize your data
- Name your files appropriately
- Choose file formats wisely
- Use versioning strategies
- Outline quality control strategies

THERE ARE 2 TYPES OF PEOPLE IN THIS WORLD:

#1   #2

@KAYDENHINES

Open Life Science

Document everything!

- **Who** created and owns this data?
- **What** are the contents?
- **What** output and results?
- **When** was this data created and last updated?
- **Where** is it stored and published?
- **Which** methods were used?
- **Which** instruments were used?
- **How** was the data created, controlled and analysed?
- **How** can I use this data i.e. license?

Remember kids, the only difference between screwing around and science is writing it down.

Thank you!

Get in touch
Email: sara.elgebali@scilifelab.uu.se
Twitter: @yalahowy

Resources

Resources- Research data management:

- [Research Data Management 1 day workshop](#)
- [Making the Case for Research Data Management](#)
- [What is Research Data?](#)
- [New England Collaborative Data Management Curriculum](#)
- [Science Europe- RDM guide](#)
- [Roberts Lab Handbook- Data management in life sciences](#)
- [Research data management (RDM) open training materials](#)

Resources- What is Data & FAIR data:

- Research Libraries
- Zenodo-FAIR principles
- "A love letter to your future self": What scientists need to know about FAIR data
- Invest 5% of research funds in ensuring data are reusable
- FAIRaware- FAIR assessment
- H2020 Programme Guidelines on FAIR Data Management in Horizon 2020
- FAIRsFAIR Europe
- How to FAIR
- Go FAIR
- FAIR sharing

Resources- Open Data & reuse, reproducibility:

- Mozilla open science- challenges to open data and how to respond
- Ten arguments against Open Science that you can win
- 'I ain't afraid of no myth' – busting the myths on data sharing
- OpenAIRE Research Data Management Hand Book
- Open Data Hand Book
- Open Science Hand Book
- Papers Without Code - submission
- Open Data and FAIR Data: differences and similarities
- Open Science Mooc
- Rein in the four horsemen of irreproducibility
- Making experimental data tables in the life sciences more FAIR: a pragmatic approach
- Open Scientist Handbook
- The Turing Way

Resources- Data Organization:

- [Towards a Standardized Research Folder Structure](#)
- [Swedish National Data service](#)
- [DataOne research data management modules](#)
- [Imperial College Research data management guides](#)
- [King's college- Managing your data](#)
- [UK Data Services](#)

Resources- Software

- Reproducible analysis and Research Transparency
- Data Science for the Biomedical Sciences
- From FAIR research data toward FAIR and open research software
- Towards FAIR principles for research software
- Software vs. data in the context of citation
- Assessment report on 'FAIRness of software'
- Sharing interoperable workflow provenance: A review of best practices and their practical application in CWLProv
- FAIR Software

- Python package to analyze a GitHub or GitLab repository's compliance with the fair-software.eu recommendations.
- Checklist for a Software Management Plan
- Top 10 FAIR Data & Software Things
- Research Software Alliance
- CodeMeta
- Open source for open science- CERN
- Software Reproducibility - The Nuts and Bolts
- Is software reproducibility possible and practical?
- Make a README
- README awesome list