Welcome to The Carpentries Etherpad!

This pad is synchronized as you type, so that everyone viewing this page sees the same text. This allows you to collaborate seamlessly on documents.

Use of this service is restricted to members of The Carpentries community; this is not for general purpose use (for that, try [https://etherpad.wikimedia.org](https://etherpad.wikimedia.org/)).

Users are expected to follow our code of conduct: <https://docs.carpentries.org/topic_folders/policies/code-of-conduct.html>

All content is publicly available under the Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>

 ----------------------------------------------------------------------------

**FAIR in (Biological) Practice**

DAY SECOND PAD at <https://pad.carpentries.org/fair-bio-2022-11-23>

[**https://edcarp.github.io/2022-11-22\_ed-dash\_fair-bio-practice/**](https://edcarp.github.io/2022-11-22_ed-dash_fair-bio-practice/)

Online, 22 - 25 November 2022, 13:00 - 17:00

**Day 1 - Tuesday 22 November**

**List of attendees**

-Stephan Nylinder

- Tomasz Zielinski

- Cass

-Marina Vabistsevits

- Loreto Pino

-Chloë Thimonier

-Stephanie MacMaster

-Emma Ramsey

-

-Oscar Jackson

-Livia Scorza

- Pauline Ward

-Ugne Baronaite

**1. You and data sharing**

Thinking of how you make your data or code available to others and how you use others data, write +1 next to the statments that matches your own experience:

- I do not really share data, I only publish the results as a part of a publication:

- I have made my data available only as Supporting Information for a paper: +1

- I have made my data available as both Supporting Information and as a dataset in a repository: +1+1+1+1+1

- I have made my data/code available without having it published in a paper: +1+1+1+1+1

-  I share my code in github or another code repository:+1+1+1+1+1+1

- I make my code available on demand:+1+1+1

- I have used a dataset from a public repository: +1+1+1+1+1+1

- I have used others code from github or such: +1 +1+1+1+1+1+1

 DONE: +1+1+1+1+1+1+1+1+1+1

-----------------------------------------------------------------------------------------------------------------

**Lesson 2: Open Science**

**Exercise 1. Benefits of openness**

Being open has other benefits beyond giving free access to information.

For example “Open access”:

·         speed of work and knowledge distribution

·         new metrics of impact: views, downloads, tweets etc

Discuss in your group additional benefits, or addressed problems for the selected open practices, type them bellow:

(Room 1) Open Data:

- Reproducibility

-scientific transparency

- improved analysis as technology develops

- encourages good record keeping

- encourages collaborations

(Room 2) Open Software:

-other researchers can view/check your code if it's on github (to check reproducibility or reuse in their analysis)

- Free licence.

- Compatible with different operative systems (Mac/Windows).

- suggesting/adding on specific functions that you need and can benefit other users as well

(Red Room) Open Notebooks:

-

-

(Yellow Room) Open Peer Review:

-

-

-

(Orange) Open Educational materials:

-

-

DONE:

-----------------------------------------------------------------------------------------------------------------

**Exercise 2. Personal benefits of being “open”**

Below are some personal benefits to adopting Open Science practices. Read through them, select the 3 most important/attractive for you and mark them with +1, select two least important for you and mark them with 0

·         get extra value from your work (e.g. collaborators, reuse by modellers, ML specialists): +1+1+1+1+1+1+1+1

·         complying with funders’ policies:

·         receive higher citations:+100

·         demonstrate research impact:+1+1+1+1+1+1+1

·         save own time (reproducibility but also communication overhead):

·         become pioneers:0000

·         distinguish yourself from the crowd:00000

·         plan successful research proposals:

·         gain valuable experience:+1

·         form community:+100

·         increased speed and/or ease of writing papers:0000

·         speed up and help with peer review:

·         build reputation and presence in the science community: +1+1+1+1+1+1

·         evidence of your scientific rigour and work ethic:+1+1+10

·         avoid embarrassment/disaster when you cannot reproduce your results: 00000

 DONE: +1+1+1+1++11+1+1+1

Can you think of other benefits? How do personal benefits of Open Science compare to the benefits for the (scientific) society.

**OA links**

Details of funding bodies and their involvement and requirements can be found at

Plan S/cOAlition S: <https://www.coalition-s.org/plan-s-funders-implementation/>

There is also a cOAlition S journal checker tool (<https://www.coalition-s.org/blog/unboxing-the-journal-checker-tool/>) to assess compliance being developed. The Directory of Open Access Journals (DOAJ - <https://doaj.org/>) is a tool to find which journals are Open Access.

-----------------------------------------------------------------------------------------------------------------

**Exercise 3. Why we are not doing Open Science already**

Discuss Open Science barriers, type bellow the reasons for not being open:

- Historically, no demands for being open. Science was a personal venture that only the researcher was concerned with

-they don't know how to do it +1

- don't understand the benefits of doing it(i.e. gatekeeping data over getting more citations if dat a is open)

- mistakes and errors can be amplified unintentionally  +1+1

Difficult to host large datasets and share before advances

- Data sensitive

- Data are of commercial value

Where to next links

•  Challenges & benefits of OS: <https://doi.org/10.1371/journal.pbio.3000246>

•  Centre for Open Science: <https://www.cos.io/>

•  Ted talk supporting OS: <https://youtu.be/c-bemNZ-IqA>

-----------------------------------------------------------------------------------------------------------------

**Exercise 4. Open Science Quiz**

Which of the following statements about the OS movement are true/false? T or F

·         Open Science relies strongly on the internet:TTFTFTT

·         Open Access eliminates publishing costs:FFFFFFFFF

·         Open Data facilitates re-use:TTTTTTTT

·         Open Data increases confidence in research findings:TTTTTTTT

·         In Open Peer Review, readers vote on publication acceptance:FFFFFFFF

·         Open Notebooks improve reproducibility:TTTTTTTT

·         Open Notebooks can create patenting issues:TTTTTTTT

·         Open Access permits the whole society to benefit from scientific findings:TTTTTTTT

·         Citizen Science engages public in the research process:TTTTTTTT

·         Citizen Science can help getting  ecological data quickly:TTTTTTTT

DONE:+1+1+1+1+1+1+1

------------------------------------------------------------------------------------------------------------------

**Lesson 3: Being FAIR**

**Exercise 1a. Protocol (green or 1)**

You need to do a western blot of the protein Titin, the largest protein in the body with a molecular weight of 3,800 kDa. You found an antibody sold by Sigma Aldrich that has been validated in western blots and immunofluorescence. Sigma Aldrich lists the publication by Yu et al 2019 (<https://doi.org/10.1002/acn3.50831>) which uses their antibody.

Can you find a complete protocol for separation and transfer of this large protein?

·         Hint 1: Find the Western blot in the methods section.

·         Hint 2: Follow the references

How easy was it? Not at all, problem is twofold - Not described in paper, then either not described properly in referred paper, or no access to secondarily referred paper.

 Found it , but jumping though papers is annoying

**Exercise 1b. Average content (red or 2)**

The Ikram 2014 (<https://doi.org/10.1093/jxb/err244>) paper contains data about various metabolites in different accessions (genotypes) of *Arabidopsis plant.* You would like to calculate the average nitrogen content in plants grown under normal and nitrogen limited conditions.

Please calculate the average (across genotypes) nitrogen content for both experimental conditions.

·         Hint 1. Data are in Supplementary data (Experiment 2 - <https://tinyurl.com/hjkdzsd4>) link does not work!

·         Hint 2. Search for nitrogen in paper text to identify the correct data column.

DONE: Found it, but pdf is not the best for data (neither is excel as it tends to assign data types sometimes)

-----------------------------------------------------------------------------------------------------------------

**Exercise 2. FAIR Example**

Check a dataset record in the Zenodo repository (a general repo)

<https://doi.org/10.5281/zenodo.6339631>

Identify elements that make it FAIR

(try to access content of the readme file via github)

Findable:

 -Yes, the filename is descriptive

 - Repository description is rich

 -

 Yes, the multiple links can be used to access the same dataset, e.g. Zenodo, GitHub, DOI

- deposit on zenodo insures that the code version at publication is preserved

- DOI is provided.

Accessible

 - Data are provided as csv format.

 -The links are easily accessible

 - Data are described in a README file.

 - Rich metadata desription of repository post

 - Rich parameter descriptions

 - Good and well described file/folder structure

Interoperable

 -Data license.

 - Data formats in csv to alllow cross platform usage

 -using R - open source language

 - markdown documentation

Reusable

 -Naming of the files is descriptive

 -project website - easy entry point

DONE: Red Room.

-----------------------------------------------------------------------------------------------------------------

**Exercise 3. FAIR and You**

The FAIR acronym is sometimes accompanied with the following labels:

·         Findable - Citable

Reference to in years after the data has been collected

·         Accessible - Trackable and countable

Any researcher elsewhere in the world has easy access to the data

·         Interoperable - Intelligible

collaboration - others can reuse the data in different ways

·         Reusable - Reproducible

results are persistent over time, making my research trustworthy

 Can different researchers arrive to the same conclusions (credibility)

 Prove robustness

Using those labels as hints discuss how FAIR principles directly benefit you as the data creators.

DONE:

-----------------------------------------------------------------------------------------------------------------

**Exercise 4. FAIR Quiz**

Which of the following statements is true/false (T or F).

·         F in FAIR stands for free. FFFFFFFF

·         Only figures presenting results of statistical analysis need underlying numerical data. FFFFFFF

·         Sharing numerical data as a .pdf in Zenodo is FAIR. FFFFFFFF

·         Sharing numerical data as an Excel file via Github is not FAIR. FFTFTFFF

·         Group website is a good place to share your data. FFFFFF

·         Data from failed experiments are not re-usable. FFFFFFFF

·         Data should always be converted to Excel or .csv files in order to be FAIR. FFFFTFFT

·         A DOI of a dataset helps in getting credit. TTTTTTT

·         FAIR data are peer reviewed. FFFFFFFF

·         FAIR data accompany a publication. FTTFTFF

DONE:

------------------------------------------------------------------------------------------------------------------

**Lesson 4: Intellectual property, Licencing and Openness**

**Exercise 1. Checking common licence:**

1. Open CC BY licence summary <https://creativecommons.org/licenses/by/4.0/>

is it clear how you can use the data under this licence and why it is popular in academia?

 Must cite

Yes

2. Check the MIT licence wording: <https://opensource.org/licenses/MIT>

is it clear what you can do with software code under this licence?

promote and protect open source software and communities

share the software in modified and unmodified form

Yes

3. Compare the full wording of CC BY

<https://creativecommons.org/licenses/by/4.0/legalcode>

can you guess why the MIT licence is currently the most popular for open source code?

A much more simple version, says the same thing but in fewer words and more clearly

DONE:

------------------------------------------------------------------------------------------------------------------

**Lesson 5: Intro to metadata**

 The microscopy example:

<https://publicomero.bio.ed.ac.uk/webclient/?show=dataset-231>

**Exercise 1. Identify types of metadata**

Here we have an excel spreadsheet that contains project metadata for a made-up experiment of plant metabolites: <https://carpentries-incubator.github.io/fair-bio-practice/fig/04-metadatafull_spreadsheet.png>

In groups, identify different types of metadata (administrative, descriptive, structural) present in this example.

Just as a reminder:

•Administrative: relevant to managing it

  e.g. Experimental code, PI

 Authors,

 Funding

•Descriptive/citation: assists with discovery/identity

  e.g. Authors, persistent identifier

 Authors, DOI number

 Contact

•Structural: how the data came about & is structured

  e.g. Collection method, folder structures

 protocols; line 7

 column A row 7

DONE:

 Any idea of reusing of the example metadata as data:

 - the protocol

 -meta data analysis from multiple experiments conducted through different timepoints/conditions

 -Genotypes

-----------------------------------------------------------------------------------------------------------------

**Minima Information Standards**

<https://fairsharing.org/collection/MIBBI>

<https://fairsharing.org/standards/>

**Exercise 2. Minimal Information Standard**

Look at Minimum Information about Neuroscience Investigation (MINI) Electrophysiology

<https://www.nature.com/articles/npre.2008.1720.1.pdf>

which contains recommendations for reporting the use of electrophysiology in a neuroscience study. (Neuroscience, or neurobiology, is the scientific study of the nervous system)

Scroll to **Reporting requirement** and decide which of the points 1-8 are:

**a)** important for understanding and reuse of data: 1 123417654

**b)** important for technical replication: 2,2332,4,7,8,765

**c)** could be applied to other experiments in neuroscience:2,3,4,4,53,7654

DONE:

-----------------------------------------------------------------------------------------------------------------

Questions

:

:

**Feedback:**

1.      How do you feel about the presented topics after this session (type +1 next to the statement that best describes your feeling):

•       I am more confused:

•       I have a better understanding of them now:+1+1+1+1+1

•       My knowledge has not changed much:+1

2.      How was the pace of the lesson:

•       Too fast:+1+1

•       About right:+1+1+1+1+1

•       Too slow:

3. If the lesson could be 5 minutes longer, what would you add or spend more time on:

The exercises and discussion parts+1+1+1

4. What could be improved:

 Introduction to course leaders and fellow participants+1+1+1+1 can do it on here? JK:) yes I meant that! too late!+1

 More time for discussion/exercises in BO rooms+1+1

5. What did you like:

Great contents!+1+1

-----------------------------------------------------------------------------------------------------------------

 Ask tomorrow about more time.

 Ask about the external users

-----------------------------------------------------------------------------------------------------------------

**Q&A:**

Do you have any questions about the topics dicussed today? Please write them down here. Use +1 to upvote the ones you are interested in if someone already asked it. We will briefly discuss them before the following set of lessons.

-

-

-