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 Biomedical Practice**

**Day 1 - Tuesday 28 March 2023**

**List of attendees**

- Tomasz Zieliński

-Caroline Wise, Jax

-Camille Liedtka, JAX

-Abigail Miller, JAX

-Sue McClatchy, JAX

-Jenn Stauffer, JAX

-Cat Witmeyer. JAX

- Robin Shaw

-Ian Powley

-Fiona Ballantyne

-Gianluca Giusti

-Anna Salvian, University of Surrey

-Jayson Felty

-Livia Scorza - University of Edinburgh

-Rocky Onda (JAX)

-

**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:+1

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

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

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

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

- I make my code available on demand:

- 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

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

**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:

(Green Room 1) Open Data:

-Reuse, more efficient use of resources

-Reproduce; studies can be reproduced with greater precision and elegance

Better information decreases liklihood of fruitless re-work.

(Blue Room 2) Open Software:

- removes duplication of effort

-quicker improvements

 -reproducability

 -Reusable data/frameworks

 -Source of shared learning

 More eyes is better!

(Red Room 3) Open Notebooks:

-reproducibility

-archiving

sharing with documentation

collaboration

change tracking

(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+1:+1+1+100

·         receive higher citations:00000000

·         demonstrate research impact:

·         save own time (reproducibility but also communication overhead):+1+1+1+1+1+1+1+1+11+1

·         become pioneers:000

·         distinguish yourself from the crowd:000000

·         plan successful research proposals:0

·         gain valuable experience:+1

·         form community:1+1+1

·         increased speed and/or ease of writing papers:+1+1

·         speed up and help with peer review:0

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

·         evidence of your scientific rigour and work ethic+1:+11+1+1

·         avoid embarrassment/disaster when you cannot reproduce your results:00+10

 DONE: +1+1+1+1+1+1+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:

-This can be a totally different process which can be hard to implement initially

No support for implementation

- Don't have access to tools or training to use them

-Fear of "giving away your data"

 -Unequal sharing of data between private industry vs academic (the private industry can use your work, but you can't use theirs)

 -"a fear of giving away all your data, eg you might have wanted to publish some other smaller papers based on the same dataset"

 -Privacy laws

Not knowing how to go about doing it

Pulling disparate kinds of data together is a big challenge and time-consuming

-Your work is niche - why go through the effort of sharing?

Discomfort of fear of public finding an error after sharing

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>

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

 SEE YOU 15 past 2 in UK and 10:15 USA

**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: TTTTTTTTTT

·         Open Access eliminates publishing costs: FFFFFFFFFF

·         Open Data facilitates re-use: TTTTTFTTTT

·         Open Data increases confidence in research findings: TTTTTTTTTT

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

·         Open Notebooks improve reproducibility: TTTTTTTTTT

·         Open Notebooks can create patenting issues: TTTTTTTTTT

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

·         Citizen Science engages public in the research process: TTTTTTTTTT

·         Citizen Science can help getting  ecological data quickly: TTTTTTTTTT

DONE: +1+1+1+1+1+1+1+1+1+1

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

**Lesson 3: Being FAIR**

**Exercise 1a. Protocol (green 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?

-some of us hit a paywall on second reference. Interlibrary loan is an option, but a slower one.

What about those who did not have pay wall?Short paragraph.

**Exercise 1b. Average content (red 2, yellow 3)**

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://github.com/carpentries-incubator/fair-bio-practice/raw/gh-pages/files/err244_Supplementary_Data-2023-03-28.zip> )

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

How easy was it?

The data are in a PDF! Well a lot of data is :) pdf are favourites for many :) Aaaarggh! I guess it looks pretty at least. You can still get the number though. Okay, if I have to. That is what the exercise is about :)

All the data seem to be provided, but very hard to work out the thing we need

DONE:

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

**Exercise 2. FAIR Example**

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

**Identify elements that make it FAIR**

Findable:

 -DOI+1+1

 versioned+1

 keywords+1

 indexed in OpenAire database and Github

Accessible

Download via standard browser+1

Public repository

Data as CSv files+1

Available in Github

Deadlink in readme.md

Interoperable

common data formats+1+1

 Zenodo

 Files simply named+1

R-code shared

Reusable

 README file+1+1

 common file types

license information+1

 Creative Commons Attribution 4.0 International license+1+1+1

descriptive file names and a well-organized file and directory structure+1+1

link to journal article

File description in README

DONE:

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

**Exercise 3. FAIR and You**

The FAIR acronym is sometimes accompanied with the following labels:

·         Findable - Citable

Less time expended searching for information.

·         Accessible - Trackable and countable

Analytics

·         Interoperable - Intelligible

vendor specific softare not needed

·         Reusable - Reproducible

Someone taking over from you can recreate experiments

Your analyses can be reproduced by others (or by yourself).

It is legal for others to reproduce your work without worry about violating copyright or other laws.

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

DONE:

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

 I will see you 16:15 or 11:15

**Exercise 4. FAIR Quiz**

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

·         F in FAIR stands for free. FFFFFFFfFFF

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

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

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

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

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

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

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

·         FAIR data are peer reviewed. FFFFFFFfFF

·         FAIR data accompany a publication. FFFFFfFFF

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

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

**Lesson 4: Being precise**

*If you have not done it yet, register yourself on ORCID (*[*https://orcid.org/*](https://orcid.org/)*)*

**Exercise 1. Public ID in action 1**

The Wellcome Open Research journal uses ORCID to identify authors.

Open one of our papers <https://doi.org/10.12688/wellcomeopenres.15341.2> and check how public IDs such as ORCID can be used to interlink information.

-ORCID link contact card with roles and email address, references other publications for the author+1

-Open reviews

-

-

DONE:

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

**Registries**

•species e.g. NCBI taxonomy

<https://www.ncbi.nlm.nih.gov/Taxonomy>

•chemicals e.g. ChEBI

<https://www.ebi.ac.uk/chebi>

•proteins e.g. UniProt

<https://www.uniprot.org/>

•genes e.g. GenBank

<https://www.ncbi.nlm.nih.gov/genbank/>

•metabolic reactions, enzymes e.g KEGG

<https://www.genome.jp/kegg/>

**Exercise 2. Public ID in action 2**

The second metadata example (the Excel table) contains two other types of public IDs.

<https://carpentries-incubator.github.io/fair-bio-practice/fig/04-metadatafull_spreadsheet.png>

·         Can you find them?

 TAIR ID

 KEGG ID - can search for these in KEGG

 Strain ID would be specific for this study?

 Genes searchable in Genbank

·         Can you find the meaning behind those IDs?

-TAIR IDs are Arabidopsis genes, e.g. phytochrome B for AT2G18790

-<https://www.arabidopsis.org/index.jsp>

<https://www.genome.jp/kegg/>

DONE:

**Ontologies**

Finding ontologies:

<https://bioportal.bioontology.org/>

<https://www.ebi.ac.uk/ols/index>

List of recommended ontologies:

<http://www.obofoundry.org/>

•<https://bioportal.bioontology.org/ontologies/SO?p=classes&conceptid=http://purl.obolibrary.org/obo/SO_0000167#details>

•

•<https://bioportal.bioontology.org/ontologies/ZFA/?p=classes&conceptid=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FZFA_0001109#details>

•

•<https://bioportal.bioontology.org/ontologies/CL?p=classes&conceptid=CL:0000129#details>

Exercise

1.       The prefix CL stands for:

a)       Class ontology:

b)      Cell ontology:+1+1 +1 +1+1+1+1+1+1+1+1

c)       Cell line ontology +1

2.       The recommended ontology for chemical compounds is:

a)       cheminf

b)      chmo

c)       chebi +1+1+!+1+1+1+1+1+1

3.       Which terms captures both Alzheimer's and Huntington's diseases

a)       DOID\_680

b)      DOID\_1289++1+1+1+1+1+1+1+1+1

c)       DOID\_0060090

DONE: +1+1+1

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

 SEE YOU 17:10 12:10

**Lesson 5: 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?

It's very clear and simple to understand (it's not in legalese). You can share the data with others and adapt it as you see fit as long as you give credit where credit is due. It's likely popular in academia because it is not restrictive.

Data can be adapted and reused which is appealing for research or industry contexts

it's what researchers have always done, reuse something but give credit by citation

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

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

It is clear! You can " use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software"

 You can do pretty much anything with this license. There's no attribution requirement though, but there is a disclaimer.

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?

Long and complicated. Does it say anything substantively different from the Summary?

 MIT seems popular because it recuses the provider of the software from any risk/liability - essentially users of it can't try to sue the creators for something going wrong+1

 Short enough that it can be included in source code repositories

CCBY is in dense legalese.

DONE:+1

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

**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.

-

-

-

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

**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+1+1+1 +1

•       My knowledge has not changed much:+1+1

2.      How was the pace of the lesson:

•       Too fast:

•       About right:+1+1+1+1+1+1+1 ++11

•       Too slow:+1+1+1

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

 -a bit more practice with ontologies because they seem straightforward, however, understanding how they work together is not so straightforward

 -upstream organisation of different kinds of data (but maybe outside the scope of this course)

 -more about the repositories such as zenodo and figshare+1+1+1

 -demonstration on some tools such as OpenNotebook+1+1

4. What could be improved:

 -remove ontology section+1+1+1

 -

 -

5. What did you like:

 -the practical application - especially the example at the beginning of non-FAIR work+1

 -discussion of how FAIR can be applied in practice/applies to "me"+1

 - the overview of open science and FAIR principles

 -clarified for me the differences between interoperable and reusable+1

 the examples where we had to find why certain tasks were or were not FAIR

 -good mix of content and exercises