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/>

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

**Dealing with data**

**Workshop led by BioRDM team on 8th November 2024, 9:30-12:30.**

**List of attendees (Please add your names and UUN!)**

- 20 participants

 Also register at <https://forms.office.com/e/Pb0Dx6H4YL>

 (for Grad School)

**Open Science and FAIR principles**

**Exercise 1 9:45**

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+1+1+1+1+1

·         complying with funders’ policies: 000000

·         receive higher citations:0000000+100000

·         demonstrate research impact: +1

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

·         become pioneers:00

·         distinguish yourself from the crowd:+1000000000

·         plan successful research proposals: +1+1

·         gain valuable experience:+1+1+1

·         form community:+1

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

·         speed up and help with peer review: +1+1+1+1+1+1+1+1+1

·         build reputation and presence in the science community:0000+1

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

·         avoid embarrassment/disaster when you cannot reproduce your results: +1+1+1000

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

**Exercise 2 Data from publications**

(5+3 min)

**Exercise 2a. Impossible protocol (Room1, Room2)**

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 a Titin-specific antibody sold by Sigma Aldrich (‘SAB1400284’) that has been validated in western blots and immunofluorescence. The Sigma SAB1400284 webpage lists the publication by Yu et al 2019 (<https://doi.org/10.1002/acn3.50831>) which uses the 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?

Answers:

- the final answer is not very detailed (both papers)

-Both papers use standard method but neither tell you what the standard method is

-The 17th reference is more clear in the methods, in comparison to the description made by the authors of the paper.

**Exercise 2b. Impossible average (Room3, Room4)**

The Ikram 2014 (<https://doi.org/10.1093/jxb/err244>) paper contains data about various metabolites in different accessions (genotypes) of *Arabidopsis plants.* 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://academic.oup.com/jxb/article/63/1/91/552676#supplementary-data> )

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

Answers:

- found experimental setup in methods and raw data in suppl. data but couldn't calculate in the time allocated

-There isn't a clear glossary of terminology/abbreviations

**Exercise 3 10:12**

(5+3 min)

Look at the dataset from Zenodo   <https://doi.org/10.5281/zenodo.6339631>

Identify elements that make this dataset FAIR

**-Findable:**

-It is in a public repository+1+1+1

-It is in an open access repository+1

-The website for data retrieve is mentioned in this paper.

-**Accessible:**

-

-The data has a logical order and everything have a descriptive name

- license is provided in readme

**-Interoperable:**

- The data is in easy-to-read data types (like csv) +1+1+1+1+1+1+1+1+1

**-Reusable:**

- Includes READ ME files +1+1+1+1+1

- README file actually contains proper descriptions of what has been done & how it's presented

- Includes code used to generate figures and results that can be modified+1+1+1

Has DOI link to the dataset so can be found again and cited (full citation provided further down the page)

Creative commons attribution

DONE: +1+1

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

**Metadata**

**Exercise 4 10:36**

  (3min)

What information – metadata would you need to re-use the data like in example picture.

**Think as a consumer** of the data not producer.

Type your proposals:

-legend

-How the image was captured

-staining used

- What the image is of

-species, magnification, microscope settings, colour-description, raw file

 -protocol

 -title

 -goal of the observation

 -if there is modification with photoshop hopefully not lol :-)

 - specific instrument used

 -experiment condition, like solution, temperature, genetic modifications, etc

volume and concentration

-If it is a plant, where was it grown - growth conditions

-where the speciemen come from (origin, lab origin)

-coloration

-Genus

-if it's a code, an example of how to use it+1

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

**Record keeping**

**Exercise 5: 10:49**

(5 +3 min)

**Differences between analog and digital record keeping**

Compare the electronic version of the tea protocol:

<https://www.protocols.io/view/how-to-make-a-cup-of-tea-buhknt4w>

with the paper one from the photo:

<https://github.com/carpentries-incubator/fair-bio-practice/blob/gh-pages/fig/06-handwritten-tea-protocol.jpg>

What are advantages and disadvantages of traditional analog records vs digital records? Try to find at least a handful of advantages and disadvantages for each. With all of these, which system do you think is most advantageous?

**Room1 & Room2**

Advantages of traditional analog records

-can be quicker if you are in the lab/field

-easy to move

easy to annotate

- A good way to make a draft version while you're working in the lab and can write up later +1

-If something catastrophic happens it is more difficult to lose the data.

-They do not require electrical energy (if you're in the middle of nowhere and you need to check them)

Advantages of digital records

-easy to work in a team

-easy to share+1

-easy to not loose becaue you can have several copies

-easy to modify

-easy to track modifications and different versions

-add links+1

-Easier to write up if it's already on a pc

-easier to add command+1

easier to refer back to

easier to distinguish the words (compared with some handwriting)

for a larger amount of data is a must

**Room3 & Room4**

Disadvantages of traditional analog records

-Can be easily lost or damaged

- handwriting can be hard to read+1

- harder to organise (messy)

 -subject to physical conditions (e.g. paper aging)

 can burn

- less standardised

- harder to search for information

Disadvantages of digital records

-hard to use in an actual lab setting

- Easily lost e.g., file is deleted with no back ups, no access via wifi or lack of device to view it on (not fully accessible)

- might not be accessible to everyone if login is lost

 - may rely on proprietary software

 -requires technical skills not every researcher may have

 -may take a lot of memory space and be hard to access if stored in a hard disk

LINKS

Example record:

<https://benchling.com/s/etr-0FdV1H0rpWeHk4H72NOg/edit>

Our ELN resources

<https://www.wiki.ed.ac.uk/x/f0SkGw>

Benchilng tutorial:

<https://www.wiki.ed.ac.uk/display/RDMS/Benchling+%28quick%29+tutorial>

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

**Working with Files**

**Exercise 6**

**A good name**

Select which file options adhere the best to the presented recommendations:

1.

a) analysis-20210906.xlsx +1

b) rna-levels-by-site.v002.xlsx +1+1+1+1+1+1+1+1+1+1+1+1+1

c) analysis of rna levels from 5Aug2021.xlsx +1+1+1

2.

a) 20210906-birds-count-EDI.csv+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+!+1+1

b) birds.csv

c) birds-count&diversity EDI 2021-09-06.csv

3.

a) 2020-7-12\_s2\_phyB\_+\_SD\_t01.raw.xlsx

b) ld\_phyA\_on\_s02-t01\_2020-07-12.norm.xlsx+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1

c) ld\_phya\_ons\_02-01\_2020-07-12.norm.xlsx

DONE: +1+1+1+1+1+1+1+1+1+1+1+1

**Exercise 7**

**Projects structure**

Have a look at the four different folder structures A-D.

<https://github.com/carpentries-incubator/fair-bio-practice/blob/gh-pages/fig/07-file_organisation.png>

The first two” A) B) are recommended for computing, the other two: C) D) are for more wet/biological projects.

**Room1 & Room2:**

When/why would you use A) and when/why B)

depends on the data

A)when you have one species

* All results together
* if you wish on using the same analyses for another species
* For simulations and computational originated data
* If using same proceedure/data analysis throughout the paper
* easy way to present to public

B)when you have several species or several type of data

if you focus on analyse

For experimental data

Able to group based on a set of data

personal structure

**Room3 & Room4**

When/why would you use C) and when/why D)

C)Would not be ideal when there is multiple biological replicates (e.g. many more pigs in this case)

 focus on the different samples/species/strains

 Wider overview, including methods and results

D) good for many time points

 focus on treatment groups/time rather than subject

 both have poor naming, as files have same names in different conditions

DONE:+1+1

 11:50

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

**Resources for Data Management**

<https://biordm.github.io/SBS-PhD-Induction-Dealing_with_data/>

BioRDM wiki

<https://www.wiki.ed.ac.uk/display/RDMS>

RDS page

<https://www.ed.ac.uk/information-services/research-support/research-data-service>

DRS page

<https://digitalresearchservices.ed.ac.uk/>

DataStore

•       <https://www.ed.ac.uk/information-services/computing/desktop-personal/connect-uni-file-storage>

•       <https://www.wiki.ed.ac.uk/x/tet_H>

Sharepoint

[https://uoe-my.sharepoint.com](https://uoe-my.sharepoint.com/)

UoE WIKI

[https://www.wiki.ed.ac.uk](https://www.wiki.ed.ac.uk/)

DMPOnline

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

Jupyter notebook

<https://jupyter.org/>

Benchling

<https://www.benchling.com/>

Protocols.io

<https://www.protocols.io/>

Zenodo

<https://zenodo.org/>

UoE DataShare

<https://datashare.is.ed.ac.uk/>

FAIRification of you project (FAIR cookbook)

<https://faircookbook.elixir-europe.org/>

FAIR in (bio) practice

<https://carpentries-incubator.github.io/fair-bio-practice/>

 Github/ GitLab

<https://github.com/>

<https://git.ecdf.ed.ac.uk/>

Version control with Git - self paced course

<https://swcarpentry.github.io/git-novice/>

**Exercise 8**

**Quiz**

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

·         F in FAIR stands for free. FFFFFFFFFFFffFFFF

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

·         Sharing data as an Excel file is not FAIR. FFFFFFFFfFffFFFF

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

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

·         Data should always be converted to Excel or .csv files in order to be FAIR. TTTTTTTTTtTtf\*TTTF

\* doesnt make sense for images....

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

·         FAIR data are peer reviewed. TTTTTTtTTfFf

·         Open Science relies strongly on the internet TTTTTTTTTTTtTTtT

·         Good record keeping ensures transparencyT TTTTTTTTTttTTtT

·         There are advantages to using analog record keeping when compared to digital record keeping.TTTTTTTTTTTttTTt

·         On balance, digital record keeping is more advantageous than analog record keeping. TTTTTTTTTTTttTt

·         ‘output 3-Aug-2022’ is a good file name FFFFFFFFFFFfFfFf

·         Digital records are easier to search (for and within) than analog records.TTTTTTTTTTttTTt

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

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

**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+1+1+1+1

2.      How was the pace of the lesson:

•       Too fast:+1

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

•       Too slow:+1+1+1+1

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

Or if you could shorter what would you remove? Breakout rooms+1+1

 organizing your folders

4. What could be improved: Perhaps some insights into how to represent data graphically would be nice

 No need for break-out rooms+1

 Maybe split this session into 2 (by topic)+1+1+1

 the lab part is long for people not working in lab

  I think it was too many information for the amount of time+1+1+1+1

  handout with examples for recommended organisation instead of exercise >> enables us to look back on what we might improve+1

5. What did you **like**:

 I think the breakout rooms help make it more interactive instead of it being a wall of info

 orcid presentation & why it's important