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 dataWorkshop led by BioRDM team on 15th November 2024, 9:30-12:30.**

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

Please also register attendance here (for Grad School) <https://forms.office.com/e/Pb0Dx6H4YL>

28 participants

**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+1+1

·         complying with funders’ policies: 000000000+1000

·         receive higher citations: +1 +1 +1

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

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

·         become pioneers: +1+100000000

·         distinguish yourself from the crowd: +1000000

·         plan successful research proposals:000+1

·         gain valuable experience: +1+1 +1+1+1+1

·         form community:+1+1+1+1+1+1+1

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

·         speed up and help with peer review:+1000000000000

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

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

·         avoid embarrassment/disaster when you cannot reproduce your results:0000000+1

 DONE: +1+1+1+1+1++11+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:

-Muscle biopsies were homogenized in reducing sample buffer containing 4% sodium dodecyl sulfate (SDS) and 10% β-mercaptoethanol, and heated for 5 minutes at 100°C. SDS–polyacrylamide gel electrophoresis and Western blotting were performed according to standard methods. Two different primary antibodies raised against titin M10, rabbit polyclonal antibody M10-15 and mouse monoclonal antibody 11-4-3, were used. In addition, mouse monoclonal calpain 3 antibody clones 2C4 and 12A2 (Leica Biosystems, Newcastle Upon Tyne, UK) were used, followed by horseradish peroxidase–conjugated secondary antibodies (Dako, Glostrup, Denmark) and enhanced chemiluminescence detection using the Pierce SuperSignal West Femto substrate (Thermo Fisher Scientific, Waltham, MA).

-The original paper referenced another paper in the methods section but that paper is not open access. Even with an institue login you can only access the abstract, not the whole paper. If you do get access to the paper (by paying) it just says it does a Western Blot using a standard protocol

-It was only possible to find the abstract of the refenced paper, as the 'full paper' linked to a collection of abstracts, even with a university login. more digging may have found the full paper.

**Exercise 2b. Impossible average (Room3, Room4, Room5)**

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:

- Table legends to understand column abbreviations was in a separate part of the journal article

-we could not understand which one of the parameters on the data seet should we choose, difficult to undesrtand

 - Unsure about which data column to use, not clearly described and difficult to find without probably going through the entire paper in detail-agree

 The data is not well annotated - difficult to understand

**Exercise 3 10:15**

(5+3 min)

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

Identify elements that make this dataset FAIR

- **Findable**:

- clear doi

- clear, because it is related to the DOI

-Files are named and organised sensibly

-clear link to the project website

- scripts are put in sensible places in clearly understandable folders

easy

clear links to the paper

open access

-**Accessible**:

-data files are all presented in a clear format for download

- Labelled well

- Most of the data available in spreadsheets and easy to download from the website

- reasonable file size in a zip file

- Clear organisation

Easy to locate files, and link to text body

nice

publically available

easy navigation on files -> well organized/named/categorized

-**Interoperable**:

-Linked protocols

-README file

- Well titled files

-csv and R files

- well documented and easy manipulated throught spreadsheets

nice

-**Reusable**:

-Beginning paragraph(s) to introduce topic make sure you have reached the right source, and readme files and explanations throughout make it easy to follow

Linked papers from where the paper was published so methods can be reproduced

-README file

method clearly described

- There is a licence

-R code for data transformations is provided

good

well described --> linked papers --> can have access --> reprodive dataset

-

DONE:

+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1

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

**Metadata**

**Exercise 4**

  (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:

- magnification/scale+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1

-Figure legend +1+1+1+1+1+1+1

-What is the flo labelled protein+1

-species/stain/organism type+1+1+1

-fluorescence indicator used+1+1+1+1

-stain used+1+1

- scale used/ color scheme/camera +1

figure legend of what we are actually looking at+1+1+1

type. of assay used+1

type of sample used+1

- Microscope specifications+1+1+1+1+1+1+1+1+1

-normalization of fluoresence / controls+1+1

observer

- Specific information about the sample type (e.g. organism, what part of the organism, etc.) +1+1

Fluorescence laser emmitance / detector absorbance +1

 DONE: +1+1++11+1+1+1+1+1+1+1+1+1+1+1+1+1+1

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

**Record keeping**

**Exercise 5:**

(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 & Room3**

Advantages of traditional analog records

- Can write it down in the lab (no laptop required)+1+1+1

-Quick and easy way to write down ideas or chnages to the protocol +1+1

- Easier to reinforce / remember ideas or protocols+1

-can easily/quickly sketch figures+1

quicker and more versatile

Data protection

Advantages of digital records

- Stored on cloud so can't be damaged or lost as easily

-Easier to read and share with others+1

-have a look after years --> saved on clouds cannot be damaged+1

-better organized in categories

Easily accessible from anywhere+1+1

 - always backed up on the cloud - can't be damaged or lost, can revist old versions +1

 - saves physical space, paper etc.

 - when in the could, is easily accessible from different dispositives

 - easier to make changes

**Room4 & Room5**

Disadvantages of traditional analog records

- The notes can potentially get damaged (things can be spilled, pens can smudge, etc.)

- Handwriting may be difficult to read+1

- Laborious and time-consuming, especially for writing longer protocols

- Hard to understand as several handwritings overwritten make the steps harder to follow in order

- Not at all compliant to cGMP standards... (if you're working in pharma you'll get into big trouble)

- Storage of originals in collections/institutions might mean differing loans policies depending on institution, hard to access/even physical travel simply to see originals should researchers need that.

-cannot reuse data - cannot copy/paste

Disadvantages of digital records

-Files can be corrupted+1

- Difficulty taking laptop into certain lab spaces

Laptop can break

- Difficulty of maintaining hygiene and sterility when using laptop both in and out of the lab+

- Data repository can change its policies and turn into a paywall (like protocols.io)

- Free for use platforms might not comply with data security guidelines

- If the laptop cannot be taken into the lab as some other people have already mentioned, you may need to take some handwritten notes while working and if you don't copy over your notes right away, you might end up forgetting to add things

 - Needs to be backed up consistently

**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

b) rna-levels-by-site.v002.xlsx+1+1+1+1+1+1+1+!++11+1+1+1+1+1+1+1+1+1

c) analysis of rna levels from 5Aug2021.xlsx

2.

a) 20210906-birds-count-EDI.csv+1+1+1+1+1++11+1+1+1+1+1+1+1+1+1+1+1+1++11

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+11+1+1++11+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+1++1+11+1+1+1

**Exercise 7 11:45**

**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)

A) - analysing all the data together

- collecting all the data at once

- sharing the data between peers/ co-authors

B)  - keeping individual experiments separate to analyse separately

- experiments carried out at different times

-

**Room3 & Room4 & Room5:**

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

C)

- If you're doing many experiments, pharmacology related?

- when designing experiments/interested in the results of multiple experiments on certain individuals?

When you already collecting data about each individual seperately

D)

To give importance to timing of drug infusions, have separate folders

- to follow the results of multiple sequential experiments

 A small thing would be to include numbers before folders in this project, as otherwise the "pre-dose" folder appears after every other data folder and even after the methods. So perhaps labelling folders as "1\_pre-dose", "2\_2h-post-dose", etc.

DONE: +1+1+1+1+1

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

**Resources for Data Management**

Course materials:

<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. FFFFFFFFFFFFFFFFFFFFF

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

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

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

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

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

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

·         FAIR data are peer reviewed. TTTTTTTTTTTTTTFF

·         Open Science relies strongly on the internet TTTTTTTTTTTTTTTTTT

·         Good record keeping ensures transparency TTTTTTTTTTTTTTTTTTTTT

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

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

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

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

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

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

2.      How was the pace of the lesson:

•       Too fast:

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

•       Too slow:

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

The second part looking at data from real studies

4. What could be improved:

5. What did you like:

Good resources mentioned

Lots of good examples and opportunities to practice some of the topics discussed

Lots of resources mentioned

Good explanations and resources

6. Would you have preferred that this workshop was in person rather than online?

* Yes +1+1+1+1+1+1+1+1+1
* No+1+1+1+1+1+1+1+1+1