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 (circadian) Practice

   
   
   
 

### List of attendees

#### You and your research

Organism you work with (type +1 or add new):  
-          Arabidopsis thaliana +1m+1  
-          Drosophila melanogaster +1  
-          Danio rerio +1+1+1  
-          Mus musculus+1  
-          Human+1+1 +1+1+1+1+1+1+1+1+1+1+1  
-          Fragilariopsis cylindrus (polar diatom)+1  
-           Triboilum castaneum+1 parasitoid wasps  
-             
-             
   
Type +1 if you:  
-          Use BioDare2 +1+1+1+1+1  
-          Use R or Python+1+1+1 +1+1+1+1+1+1+1+1+1+1+1+1+1  
-          Use Jupyter notebooks/RMarkdown+1+1+1+1+1+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:+1+1+1+1+1+1+1+1  
   
- I have made my data available only as Supporting Information for a paper:+1+1+1+1+1+1+1  
   
- I have made my data available as both Supporting Information and as a dataset in a repository:+1+1+1  
   
- I have made my data/code available without having it published in a paper: +1  
   
-  I share my code in github or another code repository:+1+1  
   
- I make my code available on demand:+1+1+1+1+1+1+1  
   
- I have used a dataset from a public repository:+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+1+1+1+1+1+1+1+1+1+1+1  
   
-----------------------------------------------------------------------------------------------------------------  
 13:51 - start of OS lesson

### Lesson 2: Open ScienceExercise 1. Benefits of openness 14:02 ()

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:  
-making it easier to gather information  
-increasing sample sizes  
- easier access for marginalised groups - broader collaboration  
- data available from different methodologies  
- meta analyses  
- can be difficult to access everything  
difficult to conclude from many ideas  
   
(Room 2) Open Software:  
-open software typically means it is free - increasing accessibility and increasing usability (updates)  
-increases collaborative opportunities  
- open software get updated faster than comercial software (e.g. R)  
- learning by example  
Often less user-friendly compared to comercial software   
  
  
  
   
(Room 3) Open Notebooks:  
- Reliability of information- Benefits  
- Quality check the process and analysis- Benefits  
-Need for a clearer structure of database- Benefits  
- Potential for collaboration- Benefits  
- Need for a lot of documentation to explain and contextualise data- Problems  
- Need for extra measures and elaborate data management protocols (potentially an ethical concern)- Problems  
  
   
(Room 4)   
Open Educational materials:  
- Easily updated with new information  
-enhances scientific debate, allows everyone to learn the prerequisite knowledge to engage in current and up to date scientific discussion.  
- Anyone can access  
- Accuracy and peer-review of the materials - Problem  
- faster dissemination of knowledge  
DONE:  
   
-----------------------------------------------------------------------------------------------------------------

#### Exercise 2. Personal benefits of being “open” 14:15

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++11+1+1+1+1+1+1+1+1+1+1+1+1+1+1  
·         complying with funders’ policies:+1+1+1+1+1+1  
·         receive higher citations:+1+1+1+1+1+1+1+1  
·         demonstrate research impact:+1+1+1+1+1+1+1+1  
·         save own time (reproducibility but also communication overhead):+1+1+1+1+0  
·         become pioneers:0+00+00  
·         distinguish yourself from the crowd:+1000+0+00+0+000+00+0+0  
·         plan successful research proposals:+1+1+1+1+1+1  
·         gain valuable experience:0  
·         form community:  
·         increased speed and/or ease of writing papers:00000+00+0+00+0+0+0+1+0  
·         speed up and help with peer review:+1+0  
·         build reputation and presence in the science community:+1+1+1  
·         evidence of your scientific rigour and work ethic:+1+1+1+1+1+1+1++11+1+1+1+1+1+1+1+0+1+0  
·         avoid embarrassment/disaster when you cannot reproduce your results:  
   
 DONE: +1+1+1+1+1+1+1+1+1+1+1+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 14:22

Open Science barriers, type bellow the reasons for not being open:  
- competition within fields (before publication)  
-invested time in models produced (e.g long time spent producing transgenic organisms) with others requesting the line; potentially a fear of not producing publications due to competition  
-Ethical issues, publication bias,   
People might not know how to do it (People need to have that specific background in Github or a certain respository),   
Errors in data that might be picked up by others,   
Data security, concerns around IP and commercialisation of data  
Some kind of data is difficult to share (eg, genome data)  
- confidential information  
- lack of confidence in unfinished datasets  
- sensitive data types can't be shared openly  
- difficulty in finding tools to handle certain types of data to allow sharing  
- inaccessibility of coding platforms to unfamiliar users, not all platforms are user friendly  
deposition of data (e.g. genomic) is complicated and time consuming  
scientists are not rewarded (carrer-progression wise) for sharing data.  - that is changing!  
  
   
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 14:40

Which of the following statements about the OS movement are true/false? T or F  
   
·         Open Science relies strongly on the internet:FFFFTTTFTTTTTTTTTTTT  
·         Open Access eliminates publishing costs:FFFFFFFFFFFFFFFFFFFF  
·         Open Data facilitates re-use:TTTTTTTTTTTTTTTTTTTT  
·         Open Data increases confidence in research findings:TTTTTTTTTTTTTTTTTTTT  
·         In Open Peer Review, readers vote on publication acceptance:TTTTFTTTTTTTTTTTTTT NOPEF  
·         Open Notebooks improve reproducibility:TTTTTTTTTTTTTTTTTTTT  
·         Open Notebooks can create patenting issues:TTTTTTFFTTFFTFFTFTTT  
·         Open Access permits the whole society to benefit from scientific findings:TTTTTTTTTTTTTTTTTTTT  
·         Citizen Science engages public in the research process:TTTTTTTTTTTTTTTTTTT  
·         Citizen Science can help getting  ecological data quickly:TTTTTTTTTTTTTTTTTTT  
   
DONE: +1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1  
   
Back  here 14:50  
14:50   
------------------------------------------------------------------------------------------------------------------

### Lesson 3: Being FAIR 14:53

#### Exercise 1a. Protocol (Room 1, 2)

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?  
 the paper cited another and protocol in the other paper was not detailed (not quantitative)

#### Exercise 1b. Average content (Room 3, 4)

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.  
 5.61 -G4  
 data was in a pdf, not csv tsv or other table file.  
   
DONE:  
   
-----------------------------------------------------------------------------------------------------------------

#### Exercise 2. FAIR Example 15:25

   
Zenodo is general data repository.   
Have a look at the dataset record with COVID-19 data:  
<https://doi.org/10.5281/zenodo.6339631>   
   
Hint: navigate to linked github record to easily access the README file  
   
   
**Identify elements that make it FAIR**  
Findable:  
-DOI/Identifier number  
-  
-  
   
Accessible  
- They have put it on GitHub using CSV format which is accessible by any machine/Readme file for accessible data  
- Creative commons license   
-  
   
Interoperable  
-File descriptions  
-Easily described names of variables  
-  
   
Reusable  
- Data is on Github which is open access  
-  
-  
   
   
DONE:  
   
-----------------------------------------------------------------------------------------------------------------

#### Exercise 3. FAIR and You

The FAIR acronym is sometimes accompanied with the following labels:  
·         Findable - Citable  
·         Accessible - Trackable and countable  
·         Interoperable - Intelligible  
·         Reusable - Reproducible  
   
Using those labels as hints discuss how FAIR principles directly benefit you as the data creators.  
Get recognition  
 more citations  
 More engagement with your publication  
 High impact   
  
More collaborations - more grants  
  
interoperable - reduce errors and loss of information   
  
Resusable - better science? Quicker discoveries  
DONE:+1+1+1+1+1  
   
-----------------------------------------------------------------------------------------------------------------

#### Exercise 4. FAIR Quiz  3:40

Which of the following statements is true/false (T or F).  
·         F in FAIR stands for free. FFFFFFFFFFFFFFFFFFFFF  
·         Only figures presenting results of statistical analysis need underlying numerical data. FFFFFFFFFFFFFFFFFFFF  
·         Sharing numerical data as a .pdf in Zenodo is FAIR. FFFFFFFFFFFFFFFFFFFF  
·         Sharing numerical data as an Excel file via Github is not FAIR. FFFFFFFFFFFFFFFFFFF  
·         Group website is a good place to share your data. FFFFFFFFFFFFFFFFFFF  
·         Data from failed experiments are not re-usable. FFFFFFFFFFFFFFFFFFFF  
·         Data should always be converted to Excel or .csv files in order to be FAIR. TTTTTTTTTTTTTTTTTTFT  
·         A DOI of a dataset helps in getting credit. TTTTTTTTTTTTTTTTTTTT  
·         FAIR data are peer reviewed. FFFFFFFFFFFFFFFFFFFFF  
·         FAIR data accompany a publication. FFFFFFFFFFFFFFFFFFFF  
   
DONE:+1+1+1+1+1+1+1+1+1++11+1+1+1+1+1+1+1  
  
------------------------------------------------------------------------------------------------------------------  
 BACK 3:50

### Lesson 5: Intro to metadata 15:52

#### 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  
 contact / funder / author names  
 exp DOI  
 title/funder/contact/study date range  
 title funder  
•Descriptive/citation: assists with discovery/identity   
  e.g. Authors, persistent identifier  
measurement protocols and growth protocols include DOIs/ Kegg IDs  
growth & measurement protocols/authors  
   
•Structural: how the data came about & is structured   
  e.g. Collection method, folder structures  
 Column titles - row 8  
 acronym descriptions  
Strain ID  
 KEGG ID  
 plant IDs  
   
DONE:  
   
-----------------------------------------------------------------------------------------------------------------  
**Minimal 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)  
  
They  are not exclusive  
You can just give number,letter(s) without having to type  
Scroll to **Reporting requirement** and decide which of the points 1-8 are:  
**a)** important for understanding and reuse of data: General features/Study subject  
5 / 7  
7  
   
**b)** important for technical replication:2. Study Subject  3. Recording Location, 4. Task, 5. Stimulus, 8(b) Sampling Rate  
 4Task  5stimulus  
7. Recording 8.Time series data - sampling rate  
 7. Electrode  
**c)** could be applied to other experiments in neuroscience: 1. General Features/2. Study Subject  
1(a) , 1(b) , 1(c) , 2(i), 2(j)4. Task/5. Stimulus/6. Behavioural Event  
DONE:  
   
-----------------------------------------------------------------------------------------------------------------

#### Exercise 3. What to include 16:23 10minutes!

Think of the data you generate in your projects and imagine you are going to share them.  
What information would another researcher need to understand or reproduce your data (the structural metadata)?  
   
**Think as a consumer** of your data not the producer!   
For example, we believe that any dataset should have:  
·         A name/title  
·         **Experiment purpose or experimental hypothesis**  
   
Write down your proposals:  
   
Room 1 – focus on behavioral assays (e.g. activity monitoring)   
- time of day ZT0 , what is meaning 0 in the timeseries  
- equipment used  
- measurement interval  
- conditions in the room (light/dark), free runing or not,   
- age of organisms  
- name of researcher(s)  
- basic description of experiment (hypothesis, etc.)  
- external conditions, e.g. stress factors  
- strain/genotype/type of organism  
- sleep hours/insomnia - standard chronotype of subject  
- what this experiment defines as 'sleep' - key definitions  
- description of samples (number)  
- funding bodies, PIs, etc.  
- types of behaviour being monitored  
- social/economic data on subjects  
- ethical issues/agreements associated with the experiment  
- source of the organisms (e.g. where mutants are from)  
- objective or aim of this research  
- time points at which data was taken (every 5 minutes/every hour/etc.)  
- licences needed for experiment (e.g. for working with protected organisms)  
- experimental intervention or treatment  
- defining the control and testing groups  
- challenges expected/faced during experiment  
- key words  
- public engagement associated with experiment  
- location of data in databases/repositories  
Room 2 – focus on timeseries data - obtained from bioluminescence  
- Identify sample (strain, sex, age, in vivo/in vitro, genotype, tissue,)  
experimental conditions (light regime, temprature, humidity, O2)  
identify recording platform;  
chemistry involved. Description of the reporter.  
experimental design (sampling interval, time of the measurement, number of days)  
normalisation of data +1 ; trend removal   
data processing (time series algorithm)   
-  
-  
-  
Room 3 – focus on molecular biology experiments (expression, translations, modifications, (co)localization)  
-how to deliver DNA or RNA and which agent are you using (or virus)  
-Type of cells system - HEK293, COS, ect  
-Method of extraction  
-Time of manipulation, type of manipulation eg, pharmacological  
-plasmid maps  
-Focus of analysis  
-Type of detection for target you want to analyse- Output of the experiment  
- Product numbers of reagants  
- Description of equipment   
-Unique identifiers   
- Number of replicants   
- Concentration/dilusion factors- what kind of concentration of pharmacology.  
  
Room 4 – focus on “intervention-based” experiments (drug treatments, activating/disactivating gene/protein, phase response, light pulses)  
 mediums used for drugs/doses used/how it is applied / time of drug application (square wave / pulse)

* - is the inverention every day / continuous. is it spiked - experimental schedule (did / how did you offset your sample interventions)

- external parameters of the experiment - light sources / temperature / CO2  
-subject - species, healthy/not, mutant/not  
- gene tagging / how it was tagged  
- info about gene /protein that you're looking at  - openly available  
- administration method of intervention  
- devices used for measuring & calibration calcs e.g. for LOQ/LOD. Bioluminescence cameras / plate readers etc. Syringe pumps...  
- devices used for applying the intervention  
protocol/procedure  
- purpose of intervention  
- what is the intervention  
-  replicates  
sampling rate/frequency  
- where does the intervention (drug, protein...) come from  
- populations (e.g. control group)  
  
   
DONE:  
------------------------------------------------------------------------------------------------------------------  
   
See  you all 4:55

### Lesson 6: Being precise

   
If you have not done it yet, register yourself on ORCID (<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.  
-  
-  
-  
-  
   
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?  
 KEGG ID - doI  
 TAIR  
 STRAIN ID  
·         Can you find the meaning behind those IDs?   
-  
-  
   
DONE:   
   
 SHOULD I STOP or 10 minutes more:  
 STOP: +` +1+1+1+1+!+11+1+1  
 GO: +1+1+1+1+1+1  
   
   
-----------------------------------------------------------------------------------------------------------------

### Lesson 7: Meta(data) in Excel 7:10

Exercise 1: What can go wrong with data in Excel  
Have a look at the example excel data-file:  
<https://carpentries-incubator.github.io/fair-bio-practice/fig/bad-metadata.png>  
<https://github.com/carpentries-incubator/fair-bio-practice/raw/gh-pages/files/04-bad-metadata.xlsx>  
   
   
**Questions:**

* - What do you find confusing?
* different units
* uneven column spacing
* colours
* No information about genotypes
* color code
* no conventions

 poor labelling  
 Bolded row  
 -1 in period data for elf4  
 chloro. abbreviation (chlophyll/chloroplast?)  
 starch fro were ? What chloro ?   
 denomination for samples/cell  
- What would you try to clarify with the author before doing anything with the file?  
 why is one row in red  
 acronyms  
 colors  
 "updated" ?  
 only number in data column  
  
- What will be the issues with calculation of: average biomas, biomas per genotype?  
 Different units - mg vs g  
 Missing values  
 significant values  
 error   
 header not on the same row  
   
- Typically, more advance data analysis is done programmatically, which requires e.g. conversion to a text format as csv, tsv format. Or using a library that reads Excel file and "kind of makes this conversion on the fly". Save this file in a text format, close Excel and reopen the saved files. What has changed?  
   
**Answers:**  
-  
-  
-  
   
Have you seen similar tables? Do you believe this example is realistic? (add +1)+1+1+1+10+1+1+1+1+1+1+1+1  
   
   
DONE:+1  
   
-----------------------------------------------------------------------------------------------------------------

#### Q&A:

Do you have any questions about the topics discussed 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++11+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  
•       Too slow:+1+1+1+1  
   
3. If the lesson could be 5 minutes longer, what would you add or spend more time on:  
Different analysis methods of circadian data+1+1+1+1+1+1+1  
Trying out some of the repositories+1+1+1  
   
4. What could be improved:slightly longer breakout rooms, +1  
Practical examples of data analysis+1  
   
5. What did you like: Breakout rooms+1, +1+1+1+1+1+1  
Enthusiasm for improving science  
engaging questions+1+1