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

 

## Day 4

 

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

### Lesson 13: Public repositories Exercise 1: Public general record

Have a look at the following record for data set in Zenodo repository:  
<https://doi.org/10.5281/zenodo.5045374>

* What elements make it FAIR?

FINDABLE (persistent identifiers, easy to find data and metadata):  
 DOI+1+1+1+1+1+1+1+1+1+1+1+1+1  
   
ACCESSIBLE (The (meta)data retrievable by their identifier using a standard web protocols)  
 Open access+1+1+1+1+1+1+1+1+1+1  
 Repository  
 README file+1+1+1  
INTEROPERABLE (The format of the data should be open and interpretable for various tools):  
 Comon file extensions+1+1+1+1+1+1+1+1+1++11+1+1+1  
 Methods are easily accessible+1  
REUSABLE (data should be well-described so that they can be replicated and/or combined in different settings, reuse states with a clear licence):  
 Meta data+1+1+1+1+1  
 License\_text file+1+1+1  
Proper formating+1  
   
2. Skim through the data set description (HINT there is also a README), try to judge using marks from 0 to 5 (5 best) if  
·         It is clear what the content of the data set is:222443333322  
·         It is clear why (what for) the data could be used:5445444444  
·         It is well described:4555444445445  
·         How confident will you be to work with this data set:422432232322  
·         How easy it is to access the data set content: 555555555555  
   
3. Give +1 to the statement that the best describes your latest public dataset or a current project folder:  
-It does not have readme like description:  
-It description has only one/two paragraph(s), I have not thought     of adding so many details: +1+1+1+1+1+1+1+1+1+1+1+1+1  
-My data set description is similar in details:  
-My data /project description is much richer or formalised:  
   
DONE:+1+1+1+1+1+1+1+1+1+1+1+1+1  
------------------------------------------------------------------------------------------------------------------

#### Exercise 2: Dataset discovery  (13:25)

   
Try to find either:  
- similar data sets in Zenodo  
- data sets of interest for you    
   
Judge using marks from 0 to 5 (5 best)  
·         how easy is to find similar or interesting dataset:4434434333  
·         It is clear what the content of the data set is:44333333333  
·         It is clear why (what for) the data could be used:333343333  
·         They are well described:1323222222  
   
DONE:1+1+1+1++11+1+1+1+1+1  
 ------------------------------------------------------------------------------------------------------------------

#### Exercise 3: Domain specific repositories. 13:32

Select one of the following repositories based on your expertise/interests:    
    
Have a look at mRNAseq accession 'E-MTAB-7933' in [ArrayExpress]   
(<https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-7933/>)   
   
·         What makes it better than Zenodo:   
 More information/clearer   
 Information more easily readable and accessible  
·         What domain specific features can you see:  
 Protocols section+1  
·         Searching:  
 Direct links more easily accessible  
   
   
Have a look at microscopy 'project-1101' in [IDR]  
(<https://idr.openmicroscopy.org/webclient/?show=project-1101>)    
   
·         What makes it better than Zenodo:   
- Can open the file directly without downloading+1+1+1+1+1  
-  
-  
·         What domain specific features can you see:  
categorized more  
-  
-  
-  
·         Searching:  
-  
-  
-  
   
   
Have a look at the synthethic part record 'SubtilinReceiver\_spaRK\_separated' within the 'bsu' collection in [SynBioHub](<https://synbiohub.org/public/bsu/SubtilinReceiver_spaRK_separated/1>)    
·         What makes it better than Zenodo:   
-  
-  
-  
·         What domain specific features can you see:  
-  
-  
-  
·         Searching:  
-  
-  
-  
   
Have a look at the proteomics record 'PXD013039' in [PRIDE]  
(<https://www.ebi.ac.uk/pride/archive/projects/PXD013039>)    
·         What makes it better than Zenodo:   
 A very clear summary at the top of the project and easy access to the protocol  
   
·         What domain specific features can you see:  
 Modifications, Instruments and quantifications etc clearly identifiable  
   
·         Searching:  
 Links to similar studies attached  
 Can search for specific proteins / AA sequences. Or target the search through specific instruments / modifications / organisms etc  
   
Have a look at the metabolomics record 'MTBLS2289' in [Metabolights](<https://www.ebi.ac.uk/metabolights/MTBLS2289/descriptors>)  
·         What makes it better than Zenodo:   
 In-detail information about components and data  
   
·         What domain specific features can you see:  
 Search based on organism, compound and so on  
   
·         Searching:  
 Targeted search based on identifier/ email/ keywords ...  
   
DONE:+1+1+1+1  
   
 ------------------------------------------------------------------------------------------------------------------

#### Exercise 4: Finding a repository 13:48

Firstly, check the publisher's / funder' recommended list of repositories, some of which can be found below:  
- [BioMed Central / Springer Nature](<https://www.springernature.com/gp/authors/research-data-policy/recommended-repositories>)  
- [eLife](<https://submit.elifesciences.org/html/elife_author_instructions.html#policies>)  
- [Elsevier](<https://www.elsevier.com/about/policies/research-data>)  
- [EMBO Press](<https://www.embopress.org/page/journal/14602075/authorguide#datadeposition>)  
- [F1000 Research](<https://f1000research.com/for-authors/data-guidelines>)  
- [GIGAscience - OUP](<https://academic.oup.com/gigascience/pages/instructions_to_authors>)  
- [PLoS](<https://journals.plos.org/plosbiology/s/recommended-repositories>)  
- [Scientific Data - Nature](<https://www.nature.com/sdata/policies/repositories>)  
- [Taylor and Francis](<https://authorservices.taylorandfrancis.com/data-sharing-policies/repositories/>)  
- [BBSRC](<https://bbsrc.ukri.org/research/resources/>)  
- [NERC](<https://nerc.ukri.org/research/sites/environmental-data-service-eds/policy/>)  
- [Royal Society](<https://royalsociety.org/journals/ethics-policies/data-sharing-mining/>)  
- [Wellcome Open Research](<https://wellcomeopenresearch.org/for-authors/data-guidelines>)    
     
Secondly, check [Fairsharing recommendations](<https://fairsharing.org/recommendations/?q=>)  
- alternatively, check the [Registry of research data repositories - re3data](<https://www.re3data.org/>)  
   
BioRDM's curated list of repos: <https://www.wiki.ed.ac.uk/display/RDMS/Suggested+data+repositories>  
   
a) Find and type a repo for genomics data:   
dbGAP:  database of Genotypes and Phenotypes  
 The European Genome-phenome Archive (EGA)+1+1  
 cBioPortal - cancer genomics  
 ClinVar: aggregates information about genomic variation and its relationship to human health (no abbreviation)  
Gene Expression Omnibus (GEO)  
b) Your favourite/chosen data type and a recommended repo for it:  
 Protein data - UniProt  
 Organism-focused resource - Zebrafish Model Organism Database (ZFIN)  
 OpenNeuro (EEG, iEEG)  
 BioDare  
 GenBank  
c)   List repositories you have used (either to download or to share your data): Figshare, +1  
GWAS catalog,   
ClinVar  
The Gene Expression Omnibus (GEO)  
TAIR- The Arabidopsis Information Center  
ZFIN  
 GitHub+1  
   
DONE:+1+1+1. +1+1+1  
 ------------------------------------------------------------------------------------------------------------------

#### Exercise 5: Wrap up discussion 14:02

Discuss the following questions:  
·         Why is choosing a domain specific repositories over zenodo more FAIR?  
 It can make your data more easily accessible to the right people. Also likely has requirements for specific information you should include that's unique for your data+1+1  
more specific search options+1  
 Visualisation of data  
 more accessible  
·         How can selecting a repository for your data as soon as you do an experiment (or even before!) can benefit your research and help your data become FAIR?  
 You can find the list of metadata items   
 Ensure you're recording / collecting all the information you need as you go  
 Provides a safe backup +1+1  
·         What’s your favourite research data repository? Why?  
   
   
DONE:+1+1  
  
Come back at 14:15.   
  
README template: <https://github.com/carpentries-incubator/fair-bio-practice/blob/gh-pages/files/readme_template_BioRDM.txt>  
------------------------------------------------------------------------------------------------------------------

### Lesson 14: It's all about planning 14:20

### Exercise 1: Action plan challenge

Where would you say the following actions belong in the Research Data Life Cycle? How do they help in achieving FAIR principles? (type the corresponding number of the research data Life Cycle after each corresponding action)  
   
**Research Data Life Cycle**

* Creating data
* Processing data
* Analysing data
* Preserving data
* Sharing data
* Reusing data

   
**Actions:**  
·         Clarify usage rights:56555565555555  
·         Give credit through citations:56 56555556656  
·         Use open source software:5366616233623653666  
·         Attach PID to your data:555555545555  
·         Attach descriptive metadata:453554455464646  
·         Backup your data:44444444444446  
·         Create figures and plots in python/R:3313131233231333  
·         Organize your files in folders:444444441234414  
·         Select data repository:2255555256555  
·         Add open licence:55555555555  
·         Link publications, data and methods:5544445555455  
·         Create a template for assay description:1122111112452  
·         Use institutional repositories:555544555545455  
·         Use controlled vocabularies:666566645656  
·         Convert numerical data to csv:3323332422453  
·         Track versions of files:444444423444  
·         Performing statistical analysis:333333333333  
·         Deposit datasets to Zenodo/Dryad:5555455555555  
·         Record experiment details in Electronic Lab Notebook:11/4144411441111  
·         Use github for your code:546556665545556  
·         Ask someone to revise your project structure:22422222222  
·         Reformat and clean data tables:22222222222  
·         Use a Minimal Information Standard:1211112111145  
·         Use PID in data description:555555555255  
   
DONE:+1+1+1+1+1+1+1+1+1+1+1+1  
   
   
Is there an actions above which is not clear for you or you do not know how to perform it?  
 List them:  
 -  
  ------------------------------------------------------------------------------------------------------------------

#### Exercise 2: Challenge We gave 20 for plan 10 for feedback

**(30 minute exercise)**  
   
 <https://drive.google.com/drive/folders/1AHxyKHoJ2ixVWJnfvPCqRHN2BxHqM-lu?usp=sharing>  
   
Working in groups, think of your last papers (or projects). Pretend that you have a joined project that combines the outputs of at least two your papers/projects (so more than one data type and a collaborative project).  
   
You can look at the example of DMP and resuable paragraphs:  
<https://www.wiki.ed.ac.uk/display/RDMS/Short+paragraphs+that+you+might+find+useful+when+preparing+your+DMP>  
   
Our list of suggested Data Repositories can be found here: <https://www.wiki.ed.ac.uk/display/RDMS/Suggested+data+repositories>  
   
For finding standards and repositories:  
<https://fairsharing.org/>  
   
For ontologies: <http://www.obofoundry.org/>, <https://bioportal.bioontology.org/>     
   
Write a short DMP for this **joined project**.   
   
   
Your DMP should contain the following three sections:  
   
**1. What data will you acquire during the project:** Please describe the type of data you will generate (for example ‘flow cytometry data’) as well as file formats and data volume..  
-thermal imaging over time (time series) (images converted to temprature readings in spreadsheet)  
-Write in the document on google drive :)  
<https://drive.google.com/drive/folders/1AHxyKHoJ2ixVWJnfvPCqRHN2BxHqM-lu?usp=sharing>  
-  
   
**2. How will you store and organize the data:** Please describe how you will store and organize your data, what metadata will you capture in what form. Explain how you will document the data during the duration of the project  
- <https://drive.google.com/drive/folders/1AHxyKHoJ2ixVWJnfvPCqRHN2BxHqM-lu?usp=sharing>  
- please use google drive  
-  
   
**3. How will you share the data:** Please describe the strategies for data sharing, licensing and access information.  
-  
-  
-  
   
**Remember: it is a joined project**  
   
**Drop the DMP document at:**    
<https://drive.google.com/drive/folders/1AHxyKHoJ2ixVWJnfvPCqRHN2BxHqM-lu?usp=sharing>  
   
HINT: You can drop a document there and start to collaborate on it online doing simultanous edits!  
   
   
   
   
   
 

#### Exercise 2, part 2: check each other DMPs

(10 minute exercise)  
   
Now, take a look at the **other group's DMP** and make comments/suggestions on how to improve it (at the end of the DMP)  
   
**Room 1 <-> Room 3**  
  
   
  
We are back 15:25  
   
   ------------------------------------------------------------------------------------------------------------------

#### Exercise 3: Quiz

Answer the following questions with true or false (T or F):  
   
1. The best time to do data management is at the end of a project, when you've collected all the data you're managing.FFFFFFFFFF  
2. Data management plans (DMPs) detail what will happen to data before collection begins.TTTTTTTTTT  
3. The best storage method for data is multiple backups to USBs.FFFFFFFFF  
4. There is a single best way to manage, organise, and share data.FFFFFFFFF  
5. For grant applications, DMPs should mention data preservation, longevity, sharing, discover, and reuse.TTTTTTTTTT  
6. Your metadata should be standardised and descriptive.TTTTTTTTTT  
7. Taking the time to plan out what's needed in metadata and your DMP will save you time in the long run and make your data more FAIR.TTTTTTTTTT  
8. DMP online is a tool which constructs DMPs for researchers.FFFFFFFF  
F  
9. Data addressed in a DMP can be freely shared regardless of confidentiality.FFFFFFFFFF  
10. Data can be given creative commons licenses to dictate how others can and cannot use it.TTTTTTTTT  
   
   
DONE:+1+1+1+1+1+1+1+1+1+1  
-----------------------------------------------------------------------------------------------------------------

### 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?  
-It is clear +1+1- the ability to copy and share is good for publications  
-it is popular because all academic want is just credit  
-  
   
2. Check the MIT licence wording: <https://opensource.org/licenses/MIT>  
is it clear what you can do with software code under this licence?  
 -Yes,  to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software  
 -  
 -  
   
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 one of the most popular for open source code?  
   
   
DONE:   
------------------------------------------------------------------------------------------------------------------  
 

### Your journey to be FAIRproductive

   
   
**Exercise 1**  
   
Read through the following activities / practices, type next to each  
-1 if you (your group) do not perform it  
? if you are not completely sure what it stands for  
0 if you only learnt abou it at this workshop  
+1 if you (your group) adheres / practices it  
   
•                     include license with datasets: +1-1+100-10  
•                     include license with code / scripts:+100-100-10  
•                     use git as version control:+1-10-1-1+10-1+1  
•                     store code in github:+1+1+1-1-1+10-1-1  
•                     create DOI for datasets / code:+1+100000000  
•                     add date availability section to a manuscript:+1+1+1-1+1+1+1+1+1  
•                     use minimal information standards:+10000+0000  
•                     use ontology terms:?????+1???  
•                     use generic data repository:+10+1-10+0+1+10  
•                     use domain specific data repository:+1+1+1+1+1+1+1+1+1  
•                     have description templates for various techniques in the lab:+1+1+1-1+1+1+1+1  
•                     store data in a shared, network drive:+1+1+1+1+1+1+1+1+1  
•                     have an automatic backup solution for files:0+1+1-1-1-1-1-1-1  
•                     follow a file naming conventions:+1+10-1+1+1-1+1+1  
•                     create standard project folder structure:+1+10-1+1+1-1+1+1  
•                     use Electronic Lab Notebooks:\_1+1-1+1+10-1-10  
•                     create figures and plots in python/R:+1+1+1+1-1-1-1-1-1  
•                     select data repository:+1+10+1+1-1+1+10  
•                     know non-restrictive licenses:-10-100+0+00  
•                     create readme for each dataset:+10-10+0+00  
•                     use institutional repositories:+1-1-1-10+100  
•                     use controlled vocabularies:+1-10-1+1+1+100  
•                     have ORCID+1+1+1+1+1+1+1+1+1+1  
•                     have dedicated folder / database for protocols / SOP+1+1+1-1+1+1+11+1+1+1  
•                     have a way to reference different versions of a protocol+10-1+1+10+1+1+1  
•                     convert numerical data to csv:+1+1-1+1+1-1-1+1+1  
•                     follow conventions for tidy data tables:+1-1-1+1?-1-1+1+1  
•                     use jupyter notebooks or R-markdown:-1-1+1-1-1+10+000+1  
•                     use metadata format / standards:+10+1+1-1+0+1+10  
•                     use PID from repositories (eg UniProt, GenBank) in data description:-1-1?+1+1+1-1-10  
•                     use database for bio-samples / strains etc:-1+10-1-1+1-1-1-1  
•                     use pipelines for data analysis:+1+10+1-1-1-1-1-1+1  
•                     can access all group data from your own PC:+1+1+1+1-1+1-1-1+1+1  
•                     discuss with team how particular data type should be described and stored:+1+10+1+1+1+1+1+1  
•                     use tools / resources you organization offers for data management:+1+10+1-1+1+1+1+1+1  
•                     use support you organization offers for data management:+1+1+1+1-1-1-1+1+1+1  
•                        
   
DONE: +1+1+1+1 +1+1+1+1+1+1  
   
   
**Exercise 2:**  
   
Type below the things you are going to change in your work habits or actions you are going to take after this course:  
- deposit data in depositories as soon as they are generated.  
-better file / data organisation - instead of dumping things in one place+1+1  
- Keep track of metadata for all my experiments - ensure the information is ready to use / understand+1  
-DMP+1+1+1+1  
Put in more effort in studying R+1+1+1  
Store my data in safe places and have backups+1+1+1  
- Will try electronic lab notebook, and Jupyter+1  
- Better naming of file and storage stucture+1  
   
   
-----------------------------------------------------------------------------------------------------------------

#### WHERE are all of our materials

   
Our course website is:  
<https://biordm.github.io/fair-in-circadian-practice/>  
   
It contains the text for self learnings, the exercises and the links.  
   
There is a github repository that contains the lessons text as well as the powerpoint slides (  
 <https://github.com/BioRDM/fair-in-circadian-practice/tree/gh-pages/instructors>  
   
-----------------------------------------------------------------------------------------------------------------

#### 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+1+1  
•       My knowledge has not changed much:  
   
2.      How was the pace of the lesson:  
•       Too fast:   
•       About right: +1+1+1+1+1+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:  
- longer breaks  
- different licences  
   
4. What could be improved:  
- Should ne more interactive  
-Less biological examples - not my background  
-It could be shorter (people could do some excerises in own time maybe? I would!)  
   
5. What did you like: DMP +1, file naming and folder organisation+1  
   
-----------------------------------------------------------------------------------------------------------------

#### Feedback For The Whole Course:

   
On the scale 0 - 5 (zero a terrible course, 5 a fantastic course)  
How good was this course: 544444344+43  
   
On the scale 0 - 5 (zero useless, 5 useful)  
How useful was this course:54443344344  
   
On the scale 0 - 5 (zero only for my worst enemy, 5 highly recommended)  
How likely are you to recommend this course:5553334334  
   
What other topics would you add: other kinds of database rather than biological ones, like dtat from sensors  
 More data analysis exercises+1  
Deeper explanation of Biodare- i mean the analysis methods  
What topics can we shorten:  
   
 16:20 Period anlaysis  
   
 