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 (Biological) Practice

<https://edcarp.github.io/2022-02-15_ed-dash_fair-bio-practice/>

**Day 4**

**List of attendees**

1. Amelia Edmondson-Stait
2. Adelaide Young
3. Nneka Nnadi
4. Flávia Fonseca Pezzini
5. Alex Meier
6. Ines Boehm
7. t zhou
8. Caity Ellis
9. Cigdem Selli

Which room where you in yesterday (for the template exercise)?

1. blue room (pcr template)

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

**Q&A For the day 3:**

Do you have any questions about the topics dicussed yesterda (Computing, Version Control, IP, templates)?

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

•       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

•       Too slow:

3. From 0 to 5 how useful was the jupyter demo:5555

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

- solving git merge problems, but maybe not the scope

5. What could be improved:

-Give the big exercise a bit earlier in the day (I was tired in the last 15min!)+1+1

-

-

6. What did you like:overall format+1

the correction of our templates just now was nice

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

**Lesson 13: Public repositories Exercise 1: Public general record (13:25)**

Have a look at the following record for data set in Zenodo repository:

<https://doi.org/10.5281/zenodo.5045374>

1. What elements make it FAIR?

FINDABLE (persistent identifiers, easy to find data and metadata):

 DOI on the link we clicked

 interlinked with ORCID

ACCESSIBLE (The (meta)data retrievable by their identifier using a standard web protocols):

 DOI is available

 open access

 preview before downloading

INTEROPERABLE (The format of the data should be open and interpretable for various tools):

 files in txt can be opened pretty much anywhere (not sure about the images formats)

 It has a readme so you can understand what the data is and the format

 readme tells you image files can be opened with freely available softwareok thanks! (well it says can be opened with FIJI)

REUSABLE (data should be well-described so that they can be replicated and/or combined in different settings, reuse states with a clear licence):

 license text is available

 data reuse section in the README

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: 55555
* It is clear why (what for) the data could be used:55555
* It is well described:55555
* How confident will you be to work with this data set:5555
* How easy it is to access the data set content:55555

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:+1
* It description has only one/two paragraph(s), I have not thought     of adding so many details:+1+1+1
* My data set description is similar in details:
* My data /project description is much richer or formalised:

DONE:

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

**Exercise 2: Dataset discovery  (13:34)**

Try to find either:

- similar data sets in Zenodo

- data sets of interest for you (please type in the url below)

Judge using marks from 0 to 5 (5 best)

·       how easy is to find similar or interesting dataset:5252

·       It is clear what the content of the data set is:452

·       It is clear why (what for) the data could be used:422

·       They are well described:442

DONE:

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

**Exercise 3: Domain specific repositories. (13:45)**

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:

 advanced search is available

better(easier to understand) page layout

release and updated dates are easy to find

contact details are available

 other tools integrated (Blast etc.)

 defined sections that need to be filled in

 allows bulk download

·       What domain specific features can you see:

·       Searching:

Have a look at microscopy 'project-1101' in [IDR]

(<https://idr.openmicroscopy.org/webclient/?show=project-1101>)

·       What makes it better than Zenodo:

- Automatically preview image files

-

-

·       What domain specific features can you see:

-Can visualise metadata and preview images at the same time

-Can zoom/look at separate channels of images on webpage

-

·       Searching:

-Tags don't seem to be used frequently by depositors

-

-

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:

·       What domain specific features can you see:

·       Searching:

Have a look at the metabolomics record 'MTBLS2289' in [Metabolights](<https://www.ebi.ac.uk/metabolights/MTBLS2289/descriptors>)

·       What makes it better than Zenodo:

·       What domain specific features can you see:

·       Searching:

DONE:

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

**Exercise 4: Finding a repository (13:52)**

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](https://submit.elifesciences.org/html/elife_author_instructions.html" \l "policies))

- [Elsevier](<https://www.elsevier.com/about/policies/research-data>)

- [EMBO Press]([https://www.embopress.org/page/journal/14602075/authorguide#datadeposition](https://www.embopress.org/page/journal/14602075/authorguide" \l "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>

**THE task**

a) Find and type a repo for genomics data:

 DDBJ

Gene Expression Omnibus (GEO)+1

b) List one of your data types and a recommended repo for it:

 RNA sequencing data

c)   List repositories you have used (either to download or to share your data):

GEO, ArrayExpress

DONE:+1

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

**Exercise 5: Wrap up discussion (14:02)**

Discuss the following questions:

·       Why is choosing a domain specific repositories over zenodo more FAIR?

 people use them more than zenodo

 Held to a higher standard so data is more likely to be understandable and reusable

 Increased accessibility because people look for specific data in them "standardized communications protocol"+1+1+1

 in general they have data curators

·       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?

 Helps to plan ahead, organisation+1+1

 We know in advance what metadata we will need = how we should structure our files to submit to the repo

·       What’s your favourite research data repository? Why?

 github very clear+1+1

 GEO (by NCBI)

 plasmodb you can search expression level data by stages of the parasite

DONE:

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

 BACK 2:25

**Lesson 14: It's all about planning 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**

1. Creating data
2. Processing data
3. Analysing data
4. Preserving data
5. Sharing data
6. Reusing data

**Actions:**

·       Clarify usage rights: 5551,4,5

·       Give credit through citations:666+1

·       Use open source software:2,3,4,5,65,6All of them+1

·       Attach PID to your data:555

·       Attach descriptive metadata:454

·       Produce standard metadata:24514

·       Backup your data:421,4,6

·       Create figures and plots in python/R:3563

·       Organize your files in folders:511

·       Select data repository:5411

·       Add open licence:55

·       Link publications, data and methods:5

·       Create a template for assay description:451

·       Use institutional repositories:464

·       Use controlled vocabularies:1

·       Convert numerical data to csv:34

·       Track versions of files:2,6

·       Performing statistical analysis:336

·       Deposit datasets to Zenodo/Dryad:55

·       Record experiment details in Electronic Lab Notebook:45123

·       Use github for your code:4536

·       Ask someone to revise your project structure:31

·       Reformat and clean data tables:234

·       Use a Minimal Information Standard:1

·       Use PID in data description:5

·       Download a dataset:6

·       Link to UniProt or GenBank records:5

DONE:

 Is there an action above which is not clear for you or you do not know how to perform it?

 List them:

 -

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

**Exercise 2: Challenge**

**(20 minute exercise)**

Working in groups, think of your last paper (or project). Pretend that you have a joined project that combines the outputs of at least two your papers/projects.

You can look at the example of DMP and resuable paragraphs:

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

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**.

**Drop the DMP document at:** <https://uoe-my.sharepoint.com/:f:/g/personal/tzielins_ed_ac_uk/EtzDrquhA45KlqFPSP3HFL0BpfbwW5ANJKdV00gRqLI2kA?e=QIwXON>

HINT: You can drop a document there and start to collaborate on it online doing simultanous edits!

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. These data will be stored under (include the meta data as well).

-

-

-

**2. How will you store 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

-

-

-

**3. How will you share the data:** Please describe the strategies for data sharing, licensing and access information.

-

-

-

**Remember: it is a joined project**

Exercise 2, part 2: check out each others 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)

**Green Room <-> Blue Room**

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

**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.FFFFFFF

2. Data management plans (DMPs) detail what will happen to data before collection begins.TTTTTT

3. The best storage method for data is multiple backups to USBs.FFFFFF

4. There is a single best way to manage, organise, and share data.FFFFFF

5. For grant applications, DMPs should mention data preservation, longevity, sharing, discover, and reuse.TTTTTT

6. Your metadata should be standardised and descriptive.TTTTTTT

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.TTTTT

8. DMP online is a tool which constructs DMPs for researchers.TFTF

9. Data addressed in a DMP can be freely shared regardless of confidentiality.FFFF

10. Data can be given creative commons licenses to dictate how others can and cannot use it.TT

DONE:

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

**Feedback For The Whole Course:**

On the scale 0 - 5 (zero a terrible course, 5 a fantastic course)

How good was this course:

 555555

On the scale 0 - 5 (zero useless, 5 useful)

How useful was this course:

 555555

On the scale 0 - 5 (zero only for my worst enemy, 5 highly recommended)

How likely are you to recommend this course:

 555555

What other topics would you add:

 -Git

 -More time on templates

What topics can we shorten:

-Jupyter notebook+1-1(I'm not a coder by any stretch and found this interesting)

-

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

**Your journey to be FAIRproductive**

Type below the things you are going to change in your work habits or actions you are going to take after this course:

-use GitHub more and encourage others to use it+1

-naming files better+1+1+1

choose databases to deposit data wisely

-Readme and metadata+1+1+1

Proper data storage and periodically backup+1

-choice of repository before data collection/generation+1

-Version record better

-include a license

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

**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.

- I have heard colleagues saying that data deposited in general repositories such as zenodo often have more details (metadata) because people are free to write whne compared to specific repositories where you have to submit you data in a fixed format. This would be important for interndisciplinary studies. Have oyu ever come across this?

-

-

 Stage in academia:

 - BSc/MSc students:

 - PhD:+1

 - Postdoc:+1

 - Junior PI:+1

 When/Where did you learn to program?

 during the PhD

 Joined my new lab early last year and they had scRNAseq datasets so I self taught along with help from the others in the group

Self-taught on a research assistant job

unrelated question: do you also have courses for biologists who want to improve programming?

<https://edcarp.github.io/>

<https://edcarp.github.io/Ed-DaSH/index.html>

<https://datacarpentry.org/workshops-upcoming/>

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

**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

•       My knowledge has not changed much:

2.      How was the pace of the lesson:

•       Too fast:

•       About right:+1+1+1+1

•       Too slow:

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

-

-

4. What could be improved:

-

-

-

5. What did you like: discussion groups have a good size now

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

**WHERE are all of our materials**

Our course website is:

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

It contains the text for self learnings, the excercises and the links.

There is a github repository that contains the lessons text as well as the powerpoint slides (HINT in instructors folder), but we challenge you to find the repository on your own :)