## FAIR in (Biological) Practice

https://edcarp.github.io/2022-02-15\_ed-dash\_fair-bio-practice/

### Day 4

### List of attendees

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

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

### 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>

1. What elements make it FAIR?

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

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

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

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

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

3. Give +1 to the statement that the best describes your latest public dataset or a current project folder:

1. It does not have readme like description:
2. It description has only one/two paragraph(s), I have not thought of adding so many details:
3. My data set description is similar in details:
4. My data /project description is much richer or formalised:

DONE:

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

#### 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:
* It is clear what the content of the data set is:
* It is clear why (what for) the data could be used:
* They are well described:

DONE:

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

#### Exercise 3: Domain specific repositories.

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:
* 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:

-

-

-

* What domain specific features can you see:

-

-

-

* 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:
* 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

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>)

- [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:

b) Your favourite/chosen data type and a recommended repo for it:

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

DONE:

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

#### Exercise 5: Wrap up discussion

Discuss the following questions:

* Why is choosing a domain specific repositories over zenodo more FAIR?
* 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?
* What’s your favourite research data repository? Why?

DONE:

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

### 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:
* Give credit through citations:
* Use open source software:
* Attach PID to your data:
* Attach descriptive metadata:
* Backup your data:
* Create figures and plots in python/R:
* Organize your files in folders:
* Select data repository:
* Add open licence:
* Link publications, data and methods:
* Create a template for assay description:
* Use institutional repositories:
* Use controlled vocabularies:
* Convert numerical data to csv:
* Track versions of files:
* Performing statistical analysis:
* Deposit datasets to Zenodo/Dryad:
* Record experiment details in Electronic Lab Notebook:
* Use github for your code:
* Ask someone to revise your project structure:
* Reformat and clean data tables:
* Use a Minimal Information Standard:
* Use PID in data description:

DONE:

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

**(30 minute exercise)**

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..

-

-

-

**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

-

-

-

**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:**

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)

**Green Room <-> Blue Room**

**Red Room <-> Yellow 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.

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

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

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

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

6. Your metadata should be standardised and descriptive.

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.

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

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

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

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:

• include license with code / scripts:

• use git as version control:

• store code in github:

• create DOI for datasets / code:

• add date availability section to a manuscript:

• use minimal information standards:

• use ontology terms:

• use generic data repository:

• use domain specific data repository:

• have description templates for various techniques in the lab:

• store data in a shared, network drive:

• have an automatic backup solution for files:

• follow a file naming conventions:

• create standard project folder structure:

• use Electronic Lab Notebooks:

• create figures and plots in python/R:

• select data repository:

• know non-restrictive licenses:

• create readme for each dataset:

• use institutional repositories:

• use controlled vocabularies:

• have ORCID

• have dedicated folder / database for protocols / SOP

• have a way to reference different versions of a protocol

• convert numerical data to csv:

• follow conventions for tidy data tables:

• use jupyter notebooks or R-markdown:

• use metadata format / standards:

• use PID from repositories (eg UniProt, GenBank) in data description:

• use database for bio-samples / strains etc:

• use pipelines for data analysis:

• can access all group data from your own PC:

• discuss with team how particular data type should be described and stored:

• use tools / resources you organization offers for data management:

• use support you organization offers for data management:

•

DONE:

**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:

-

-

-

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

#### 

## 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 :)

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

#### 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:

•       My knowledge has not changed much:

2.      How was the pace of the lesson:

•       Too fast:

•       About right:

•       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:

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

#### Feedback For The Whole Course:

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

How good was this course:

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

How useful was this course:

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

How likely are you to recommend this course:

What other topics would you add:

What topics can we shorten: