# Paste this template content into the Etherpad at:

<https://pad.carpentries.org/fair-4-leaders-begins-20YY-MM-DD>

# List of attendees

-

-

-

-

-

-

# Episode 3. Tools for Oracles and Overlords

## Exercise 1.

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

## Exercise 2.

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

## Exercise 3.

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

# Episode 4. Public repositories

## Exercise 1: Public general record (demo 4 mins + 4 mins evaluate + 4 mins finding with instructor)

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

Boehm et al. (2020). Confocal micrographs and complete dataset of neuromuscular junction morphology of pelvic limb muscles of the pig (Sus scrofa) [Data set]. In Journal of Anatomy (Vol. 237, Number 5, pp. 827–836). Zenodo.

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

Let’s check which elements make it FAIR? (instructor should explain the elements)

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

ACCESSIBLE (The (meta)data retrievable by their identifier using 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, and any conditions on reuse stated clearly with a licence):

Exercise 1a (4min)

Now, skim through the data set description (HINT there is also a README), try to judge the following, and indicate your evaluation using marks from 0 to 5 (5 best) as to whether:

• It is clear what the content of the data set is:

• It is clear why the data could be used (ie what for):

• 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:

• Your team datasets are equally well described (or better):

DONE:

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

## Exercise 1b: Dataset discovery

Try to find:

- data sets related to neuromuscular junction in Zenodo

Judge the following, indicating your assessment using marks from 0 to 5 (5 best)

• how easy it is to find similar or interesting data sets:

• It is clear what the content of the other data sets are:

• It is clear why the data could be used (ie what for):

• They are well described:

DONE:

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

## Exercise 2: Domain specific repositories (5 min)

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 (3 min +3)

Our own curated repository list:

<https://www.wiki.ed.ac.uk/display/RDMS/Suggested+data+repositories>

Using Fairsharing (<https://fairsharing.org/>) find a repo for flow cytometry data and type the name below:

-

-

-

once done, search for repository for genomics data

-

-

-

DONE:

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

## Exercise 5: Using repositories (5)

Describe/Type:

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

• How can selecting a repository for your data as soon as an experiment is performed (or even before!) benefits your team research and helps data become FAIR?

• What to do if your publication contains multiple data types?

DONE:

## Feedback: Repositories lesson

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

How good was this lesson:

-

On the scale 0 - 5 (zero not at all, 5 yes it was productive way of spending my time)

Was it worth your time:

-

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

# Episode 5. It is all about planning

## Exercise 1: OS and Data Management practices

Read through the following activities / practices, type next to each

-1 if your team/group do not perform it

0 if you are not completely sure what it stands for

+1 if your team/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:

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

## DMP example and reusable paragraphs

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

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

## Exercise 2: Data Management Plan

In your group evaluate the example data management plan. Mark / Comment any section which should be improved.

As a “grant reviewer” focus on following aspects:

* does it look as if the author(s) thought what they would do with their data or is it some ‘whatever’ copy pasted text
* does it look as if the data will be stored in a secured way
* does it look as if the data will be shared in a FAIR way
* are there any standards that should be followed
* does it look as if the data will be made accessible / findable to others for re-use
* are the usage restrictions justifiable

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

## Feedback: DMP

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

How good was this lesson:

-

On the scale 0 - 5 (zero not at all, 5 yes it was productive way of spending my time)

Was it worth your time:

-

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

# Episode 6. Wrapping it up

## Exercise 1: Adoption of Data Management

Select 3 actions that you believe have the biggest impact in improving productivity and data management of your group, type +1 next to them

Check (meta)data standards for the experimental data in your group

Investigate domain specific repositories for your domain(s)

Mandate (FAIR) data management training for your group

Introduce common folder structures for the projects

Develop templates for experiment types

Encourage data plotting/analysis in R/Python

Create catalogue of group research outputs

Use shared network (or cloud) storage for group data

Introduce ELN

Collect group protocols in shared/online platform

Introduce version control (git) for scripts and analysed data

Adopt Jupyther notebook (R-markdown) for data analysis

Require DMP for each project

Develop strain / samples data base

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

## Exercise 2: Policy enforcement without micromanagement

Propose solution how you can enforce / propagate good data management practices without excessive micromanagement and policing.

-

-

-

-

-

-

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

## Feedback: 2nd Session, tools, repositories and DMP

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 lessons:

• Too fast:

• About right:

• Too slow:

## 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 to other PIs:

-

On the scale 0 - 5 (not chance, definitely will send)

How likely are you to send your minions to the 2-day full course “FAIR in bio practice”, which follow similar format to this one:

-

On the scale 0 - 5 (zero not at all, 5 yes it was productive way of spending my time)

Was it worth your time:

-

-

What other topics would you add:

-

-

What topics could we shorten:

-.

-

Would you like this course to be 1.5 longer but have practical excercises with ELN and Version Control? (T F)

-