# List of attendees

-

-

-

-

-

-

# Episode 4. Notebooks to speed up work

## Exercise 1.

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

## Exercise 2.

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

## Exercise 3.

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

# Episode 5. Public repositories

## Exercise 1: Public general record

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

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

Let’s check which elements make it FAIR?

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):

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 2: Dataset discovery

Try to find:

- similar data sets 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 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

Using Fairsharing (<https://fairsharing.org/>) or Registry of research data repositories re3data (<https://www.re3data.org/> )

a) Find and type a repo for flow cytometry data:

b) Find a recommended repo for Your favourite/chosen data type (write both repo and data type):

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

d) What’s your favourite research data repository? Why?

DONE:

Our own curated repository list:

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

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

## Exercise 5: Using repositories

Discuss the following questions:

• Why is choosing a domain specific repository(ies) over Zenodo more FAIR?

• 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?

• What is the most surprising thing you’ve learned today?

DONE:

# 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 do not know 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:
* user 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:
* know what tools / resources you organization offers for data management:

DONE:

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

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 asspects:

* 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