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>).

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/>

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

**List of attendees**

**Episode 1. Why are we here? Why should you know it?**

**Exercise 1. Get to know each other (4 - 5) 10:06**

Introduce yourselves telling each other why you have joined this course.

Then, try to find one professional/academic thing that your group has in common.

For example:

- we all had our latest grant proposals accepted by MRC

- we are all desperately searching for an experienced lab technician

(1 Room)

We all:

(2 Room)

We all: we all care about data and how to manage it well

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

**Exercise 2. You and data sharing (3)**

Thinking of how you and your group make data or code available to others and how your group uses others’ data, write “+1” next to any statements that match your own experience:

- We do not really share data, we only publish the results as part of a publication: +1+1

- We have made our data available only as Supporting Information for a paper: +1++11+1

- We have made our data available as both Supporting Information and as a dataset in a repository:+1+1+1+1+1+1+1+1

- We have made our data/code available without having it published in a paper:+1+1+1+1

- We share the code in GitHub or another code repository:+1+1+1

- We make the code available on demand:+1

- We have used a dataset from a public repository:+1+1+1+1+1

DONE: +1+1+1+1+1+1+1+1

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

**Exercise 3. Why we are not doing Open Science / Data Sharing already 10:34**

Discuss Open Science barriers, and type below the reasons for not being open:

- Great effort of collecting data and no direct reward - relative to the effort and time of collection - for data sharing (having a valuable dataset is also a researcher's "life insurance") BTW: We do share our data :)

-Time and effort to curate data, institutional infrastructure to support;  institutional policies (IP, ethics) may inhibit. Fear of reprocusions of being truly open if errors are discovered - need for culture change in fields.

-Slow work, risk being scooped, try share as much possible still though, lack of good structures

-Not knowing how to share it or where, it would take a lot of time to prepare data for sharing (e.g.annotating it that others would know exactly what was done

-Lack of knowldege and expertise in open science practices

difficulties with uploading and dealing with large datasets in difficult formats

stories are in progress and still evolving - publish when interpretation is solid

Some research groups are still uber-competitive and not keen to share!

we share data as we are required by the journal when a publication was accepted

 Primary data is machine readable and needd espensive software to decode

 Curation and access is difficult; formatting of data to be shared should be common and understood

 Human data - ensure compliance with GDPR and privacy

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

**Exercise 4. Your presence**

Write +1 next to any statements that match your own experience:

- I currently supervise at least 2 postdocs:+1+1+1+1

- I have promoted at least 3 PhDs:+1+1+1+1+1+1

- I review at least 4 articles a year:+1+1+1+1+1+1+1+1

- I have been a member of a grant funding panel:+1+1+1+1+1+1

- I have been a member of a school/college/university committee:+1+1+1+1+1+1+1+1

- I have contributed to development of an institutional/community policy:+1+1+1+1+1+1

- I have been involved in the selection process for fellows / lecturers / readers +1:+1+1+1+1+1

- I am a member of a Research Council’s Council:+1

- I am editor of a journal:+!+1+1+1

- Any other activities through which you exert academic influence in the research community?+1+1+1+1

DONE: +1+1+1+1+1+1+1+1

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

**Exercise 5. Your minions**

Write +1 if your soon-to-be leaving postdoc:

- Has released software:+1+1

- Has made any dataset(s) available under an open licence:+1+1+1+1+1+1+1

- Can demonstrate outreach activities: +1+1+1+1+1+1

- Is active in some scientific community group:

 (e.g. journal club, carpentries, ReproducibiliTea)+1+1+1+1+1+1

DONE:+1+1+1+1+1+1+1+1+1+1

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

**Exercise 6. Lottery winner 11:07**

Imagine a situation in which you suddenly lose a postdoc because she/he has won the National Lottery and won’t be coming to work any more (or more realistically, they were hit by a bus). Write +1 next to any scenarios to which you can relate:

- everything should be recorded in their notebook, which you hope is in the office:+1+1+1+1+1

- (But frankly, you have never checked how good their lab notes are):+1+1

- everything should be in the team’s Electronic Lab Notebook, and you can quickly check if that is the case:+1+1+1

- all data, excel, presentations and paper drafts are in a shared network drive:+1+1+1+1+1+1+1+1

- some data and documents may only be in the postdoc’s PC/laptop:+1+1+1

- every now and then, you check people’s data and notes, so you are fairly confident they follow good practices and you know where you can find what is needed:+1+1+1+1+1+1+1

- your group has a “data management” policy/plan to which all members are introduced as part of their induction, so at least in principle all should be fine:+1+1+1+1+1+1+1

- you have left it to your group to organise such trivial matters and you hope they did it well:

- your lab manager should know it all:

- there was that old postdoc who kept track of things but they left last year:

- you are getting nervous:+1+1+1+1

DONE:+1+1+1+1+1+1+1+1

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

**Feedback: Intro episode4**

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

How good was this lesson:

-4

4

 3

 4

 3

 4

 3

 4

 3

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

Was it worth your time:

-43

 33

 4

 2.5

4

4

3

2

 See you 11:20

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

**Episode 2. Being FAIR Exercise 1 (5+3) 11:20**

Data from publications

**Exercise 1a. Impossible protocol (Room 1)**

You need to do a western blot of the protein Titin, the largest protein in the body with a molecular weight of 3,800 kDa. You found a Titin-specific antibody sold by Sigma Aldrich (‘SAB1400284’) that has been validated in western blots and immunofluorescence. The Sigma SAB1400284 webpage lists the publication by Yu et al 2019 (<https://doi.org/10.1002/acn3.50831>) which uses the antibody.

**Can you find a complete protocol for separation and transfer of this large protein?**

·         Hint 1: Find the Western blot in the methods section.

·         Hint 2: Follow the references

How easy was it?

Answers: No

Replication of the protocol would not be possible on the basis of the information provided."Western blotting were performed according to standard methods"

-

**Exercise 1b. Impossible average (Room 2)**

The Ikram 2014 (<https://doi.org/10.1093/jxb/err244>) paper contains data about various metabolites in different accessions (genotypes) of *Arabidopsis plants.* You would like to calculate the average nitrogen content in plants grown under normal and nitrogen limited conditions.

**Please calculate the average (across genotypes) nitrogen content for both experimental conditions.**

* Hint 1. Data are in     Supplementary data (Experiment 2 - <https://academic.oup.com/jxb/article/63/1/91/552676#supplementary-data> )
* Hint 2. Search for     nitrogen in paper text to identify the correct data column.

Answers:

-           Supplementary table 2, check SN%??? and Nutrition conditions

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

**Exercise 2- Public general record – FAIR elements 11:50**

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?

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

- doi

- refers to the original article

-keywords

-

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

- creative commons licence

- has readme file, so it is easy to understand what the data is about

-

-

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

- licesce in txt format

- cleaned images in tif format

-raw images in nd2 format (?!) for use in Fiji and they provide a link

-

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):

- has definitions of the data (read me)

-

We meet at 13:00

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

**Exercise 3. FAIR and You (3+2) (13:00)**

The FAIR acronym is sometimes accompanied with the following labels:

·         Findable – Citable

·         Accessible - Trackable and countable

·         Interoperable – Intelligible

·         Reusable – Reproducible

Using those labels as hints describe how each of FAIR principles directly benefit you and your team as the data creators.

Findable – Citable

- If people can find your data, they will be able to use it for their research and cite creating a greater impact of your research

-

 enhance collaborations

 visibility

Accessible - Trackable and countable

- more people can use your data if they know what it is about

-

Interoperable – Intelligible

-Requirement for the benefits  of Re-use/reproduction

- researchers would not be limited to the software/tools that you use

-fewer queries

Reusable – Reproducible

- Allow for maximising value from publicly funded work

-verification of results by others

 -remine to find novel aspects, meta studies etc

 can be combined with other datasets to make more powerful analyses

DONE:

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

**Exercise 4. FAIR Quiz (3+2)**

Which of the following statements is true/false (T or F)?

·         F in FAIR stands for free. FFfFFFF

·         Only figures presenting results of statistical analysis need underlying numerical data. FFFFFFF

·         Sharing numerical data as a .pdf in Zenodo is FAIR. FFFFFFFF

·         Sharing numerical data as an Excel file via GitHub is not FAIR.F FFFFFF

·         Your group website is a good place to share your data. TFFFFFF

·         Data from failed experiments are not re-usable. FFFFTFF

·         Data should always be converted to Excel or .csv files in order to be FAIR. T TFTTTFF

·         A DOI of a dataset helps in getting credit. TTTTTTT

·         FAIR data are peer reviewed. FFFFFFF

·         FAIR data accompany a publication. TT/F T&FT/FT/F

·         Open Science relies strongly on the internet: TTTTTTT

·         Open Access eliminates publishing costs FFFFFFF

DONE:+1+1+1

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

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

**Feedback: FAIR episode**

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

How good was this lesson:

-4444444

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

Was it worth your time:

-4453344

**Feedback: Open Science and FAIR**

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:+1

2.      How was the pace of the lessons:

•       Too fast: +1exercises

•       About right:+1+1+1+1

•       Too slow: explanations

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

- Breakout discussions yes please - longer in the rooms, please

- Seeing more examples of good data repositories and examples of how to prepare the data

4. What could be improved:

-More invovlement of audience in discussions after exercises

-Consideration of data that is not quant lab-based

exercise 1-4 in pre-work rather than in session

5. What did you like:

- Practical examples, breakout sessions +1

-Practical examples

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

**LUNCH BREAK**

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

**Episode 3. Tools for Overlords**

**Jupyter notebooks for reusable data analysis**

**Exercise 1: Basics of Jupyter Notebooks**

Open this Jupyter Notebook: <http://mango.bio.ed.ac.uk/jupyter>

We will first show you how to duplicate a notebook, save it and run code.

1.     Select the notebook titled **'student\_notebook\_light\_conditions.ipynb**' and click 'Duplicate'.

2.    Confirm with **Duplicate** when asked if you are certain that you want to duplicate the notebook.

3.    A copy of the notebook has appeared with the suffix '-Copy' and a number.

4.    Rename the notebook from -copy number to your initials e.g. “student\_notebook\_light\_conditions\_TZ” , open it by clicking

Explore the anatomy of the notebook (<https://github.com/carpentries-incubator/fair-bio-practice/blob/gh-pages/fig/10-02-jupyter_anatomy.png> )

5.    Save the notebook: Click on the disk symbol in the toolbar

6.    Run the notebook: Select the top cell of the notebook with the title (this is likely pre-selected already and will show with a light-blue bar to its left), and click “Run” in the tool bar.

7.    Click two times. What can you see?

-

-

8.    We want to run ALL the code: In the top tool bar click Cell > Run All. What can you see?

-

-

 DONE:

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

**Exercise 2: How to add and remove content**

You have duplicated a notebook and saved it under your own name. Now we will add text, remove cells and change code.

1.     Change the author name of the document to your name: Double click on the cell containing the author name and change the name.

2.     Press Run again.

3.     You can experiment using different formatting, dbl click on the cell,

add # or ## or ### before text, surround text with \*text\* \*\*text\*\* ; Run the cell again

4.     Change colours of your graph: Where the code of the graph reads the comment “# change colour of groups” you can replace the HEX codes, # followed by 6-symbol code, with names of colours (e.g. blue, green…) or other HEX codes if you are familiar with them.

8.     Save the graph under new name: Add your initials to the file name under which the image is saved. Press Run. Your image should be visible in overall file hierarchy.

DONE:+1+1

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

**Exercise 3: Sharing of your Jupyter Notebook**

You have now generated your own analysis and want to share this with your colleagues.

1.       Download your Notebook (ensure all code has been run) as .html

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

**Exercise 4: Accessibility of Jupyter Notebooks**

On a scale from -2 to 2, based on these examples how do you feel about the following statements (R is interchangeable with Python),

where -2 (strongly disagree), 0 no opinion to +2 strongly agree:

Type a number between -2 to 2 behind the statement:

•             it is easier to generate a series of plots with similar layout in R than Excel: +2+2+2+1+200

•             using notebooks does not require any programming knowledge:-1-2-1-1+1-2-2-2

•             notebooks give you a better overview of your data analysis than Excel: +2+2+1+1+2+2+1+1

•             notebooks replicate laboratory style records for data analysis: +2+2+1+2+1+1+1

•             you need to learn R to do any data processing in notebooks: +20+1+1+2+1+2+2

•             notebooks assure reproducible computing: +2+1+2+2+1+1+1+1

DONE: +1+1+1+1+1+1+1

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

GitHub

<https://github.com/ewallace/pseudonuclease_evolution_2020/commits/master>

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

Benchling 14:00

<https://benchling.com/s/etr-8u10BD9KalyiIh9ASHXA?m=slm-IOyleZlGh3xMpJXoWRrO>

<https://benchling.com/covid_ww/f/lib_tpnkCXlW-archived-experiment-notes/etr_92p2OLL5-pcr004-test-of-existing-rt-qpcr-probes-and-primers-on-cww-samples/edit>

BACK  14:10

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

**Exercise 5. Tools and the management**

Imagine your team has successfully adopted ELN , Jupyter notebooks and version control with GIT for your data. You are facing one of the scenarios bellow. Based on what you learnt from the short demonstrations, as well as your “breakout room” experience, how could these tools help you to act in the assigned situation? What actions would you take?

For example:

SCENARIO: Your building has been flooded, all electronic equipment has been damaged, you won’t have access to your office nor lab for the next 3 months.

REACTION: You mostly worried about the delay and loss of biological samples. Your data and methods are safe on the servers.

ACTIONS:

- you ask group members to list experiments which need repeating

- You divide group into those who have data to analyse and can work on publications, and those who will spend the next couple of months on training and learning new programming skills

Room 1:

You suddenly lost a postdoc because she/he has won the National Lottery and won’t be coming to work any more. They will stay two more weeks with you, what actions could you take:

-documentation (reagents, analyses)

- sort out reagents (physical location, labels) - electronic records should be checked to ensure inventory is up to date and includes locations

- Plan ahead: teach others skills - succession planning!+1

-apply clinical trials procedures

-

-

Room 2

Your new postdoc has been trying to set up RT-qPCR for strain detection of Staphylococcus aureus for the last 3 months. He says he has no success and looks super stressed and depressed. Your group previously successfully used RT-qPCR for strain detection of E. coli. What actions could you take:

- Try to work together with another colleague from the lab

- If there is no colleague, refer them back to the digital lab books to check the protocols there

- keep shared folder of lab protocols

**Exercise 6. Learning curve 14:28**

What is the minimum knowledge that you as a PI should know about each of these tools to have them adopted in your team? Or what should be the basic skill set for the PI with those tools.

For example:

ELN – know which ELN is being used by your group and why they chose this one

Answers

ELN

- How to access it and how to edit it, how to share with others+1+1+1

-If it covers all the information for a paper to be written+1+1+1

-how to access, duplicate and use it; add comments / queries+1

-how long we are allow to access it if people move to other labs/countries, etc+1

how safe is the data and how is it backed up+1+1

Git/GitHub

- where the data is, where to find description for it+1

-Access

-basic structure of GitHub+1

Jupyter Notebooks / R-markdown

- how to access the information, edit it, even replot the graphs if ever needed in the future (depends on how big the lab group is, if small, PI should have knowledge about it all)

-

-how to access, read and comment - basic proficiency with R+1+1

- A certain command of analysis functions

**Feedback: Tools lesson**

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

How good was this lesson:

-444443

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

Was it worth your time:

-454444

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

**Episode 4. Public repositories 14:36**

**Exercise 1a: Public general record description**

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

We have discussed which elements of the record make it FAIR.

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: 55555

•             It is clear why the data could be used (i.e., what for): 55555

•             It is well described: 55555

•             How confident will you be to work with this data set:55555

•             How easy it is to access the data set content:55555

•             Your team datasets are equally well described (or better):44424

DONE: +1+1+1+1+1

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

**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: 5555

•             It is clear what the content of the other data sets are:5444

•             It is clear why the data could be used (ie what for):53 (depends on descriptions)not all3

•             They are well described:44some

DONE:+1+1

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

**Exercise 2: Domain specific repositories (5 min) 14:55**

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?:

 link with similar expression data

 very specific to this type of technology

•             What domain specific features can you see?:

link to related databases

•             Searching:

as soon as you start typing, there are suggestions from a dropdown menu

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?: Have easy to understand summary on the side of the page with species, tissue used, etc.

•             What domain specific features can you see?:

 Raw data files provided making the instrument specific parameters available, this is key for method tranfer to other makes of instruments

•             Searching:

 Within the project file names are not self explanatory.

 When looking for the datasets, you can an automatic dropdown lists of relevant search terms.

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

•             What makes it better than Zenodo?:

 It is specific for metabolites

•             What domain specific features can you see?:

 Descriptors, protocols, sample info with clear headers, assays, metabolites, files sub menues

•             Searching:

 Many possibilities to browse cross studies, by compounds, species etc

DONE:+1

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

**Exercise 3: 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:

 MIFlowCyt, FCS

-FlowRepository(FlowRepository)

-

-<http://flowrepository.org/>

once done, search for repository for genomics data

-FlowRepository - Immunophenotypic Landscape in Three Relapsed Acute Myeloid Leukemia Patients

-genomics england

-<https://doi.org/10.25504/FAIRsharing.mya1ff> The European Genome-phenome Archive

<https://doi.org/10.25504/FAIRsharing.88v2k0> Database of Genotypes and Phenotypes(dbGaP)

DONE: +1+1+1

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

**Exercise 4: Using repositories (5) 15:16**

Describe/Type:

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

 only used GEO

 Arabase and GEnBank - only ones used

Did use Genebank, UniProt (updates), ArrayExpress, GEO

 I used GEO

 Zenodo - easy to find data

•             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?

 make sure you collate all the required metadata and see what kind of processed data and format may be requied

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

 See if they can all go into the same repository as a linked project

DONE:

 Complete excercise and we are back 15:25

**Feedback: Repositories lesson**

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

How good was this lesson:

-4.5444

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

Was it worth your time:

-4444

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

**Episode 5. It is all about planning**

**Exercise 1: OS and Data Management practices**

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

0 if your team/group do not perform it

? if you are not completely sure what it stands for

1 if your team/group adheres / practices it

•                     include license with datasets: 1???

•                     include license with code / scripts:100

•                     use git as version control:0000

•                     store code in github:0000

•                     create DOI for datasets / code:000

•                     add date availability section to a manuscript:11111

•                     use minimal information standards:11111

•                     use ontology terms:11111

•                     use generic data repository:0100

•                     use domain specific data repository:111

•                     have description templates for various techniques in the lab:0010

•                     store data in a shared, network drive:11111

•                     have an automatic backup solution for files:11111

•                     follow a file naming conventions:1?00

•                     create standard project folder structure:0000

•                     use Electronic Lab Notebooks:00000

•                     create figures and plots in python/R:00010

•                     select data repository:?111

•                     create readme for each dataset:0000

•                     use institutional repositories:1111

•                     use controlled vocabularies:1???

•                     have ORCID11111

•                     have dedicated folder (database/protal) for protocols (SOP):11111

•                     have a way to reference different versions of a protocol1000

•                     follow conventions for tidy data tables:?????

•                     use jupyter notebooks or R-markdown:00010

•                     use metadata format / standards:00100

•                     use PID from repositories (eg UniProt, GenBank) in data description:11???

•                     use database for bio-samples / strains etc:111

•                     use pipelines for data analysis:?111

•                     can access all group data from your own PC:10001

•                     use tools / resources you organization offers for data management:11111

•                     use support you organization offers for data management:??11

DONE: 1+1+1

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

**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

 Group 1 = DMP\_1

 Group 2 = DMP\_2

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

**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. Practical Adoption**

**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+1+1+1+1+1

Introduce common folder structures for the projects+1

Develop templates for experiment types+1+1

Encourage data plotting/analysis in R/Python+1

Create catalogue of group research outputs+1+1

Use shared network (or cloud) storage for group data+1+1

Introduce ELN+1+1+1

Collect group protocols in shared/online platform

Introduce version control (git) for scripts and analysed data

Adopt Jupyter notebook (R-markdown) for data analysis+1

Require DMP for each project+1

Develop strain / samples data base+1

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

**Exercise 2: Policy enforcement without micromanagement**

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

-

-

-

-

-

-

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

**Materials**

Slides and excercises available are on github:

<https://github.com/carpentries-incubator/fair-for-leaders/tree/gh-pages/instructors>

The rendered version of the course is:

<https://carpentries-incubator.github.io/fair-for-leaders/>

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

**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:+1+1

•       My knowledge has not changed much:

2.      How was the pace of the lessons:

•       Too fast:

•       About right:+1+1

•       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:

 -44

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

How useful was this course

44

 -

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:

-45

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 follows a similar format to this one:

-45

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

Was the workshop worth your time:

 - 44

 -

Would you like this course to be 1.5h longer and include practical exercises with ELN and Version Control? (T F)

-TT

What other topics would you add:

 -n/a

 -

What topics could we shorten:

-. topics  in episode 1

-