Is your research data management plan

a FAIRy tale

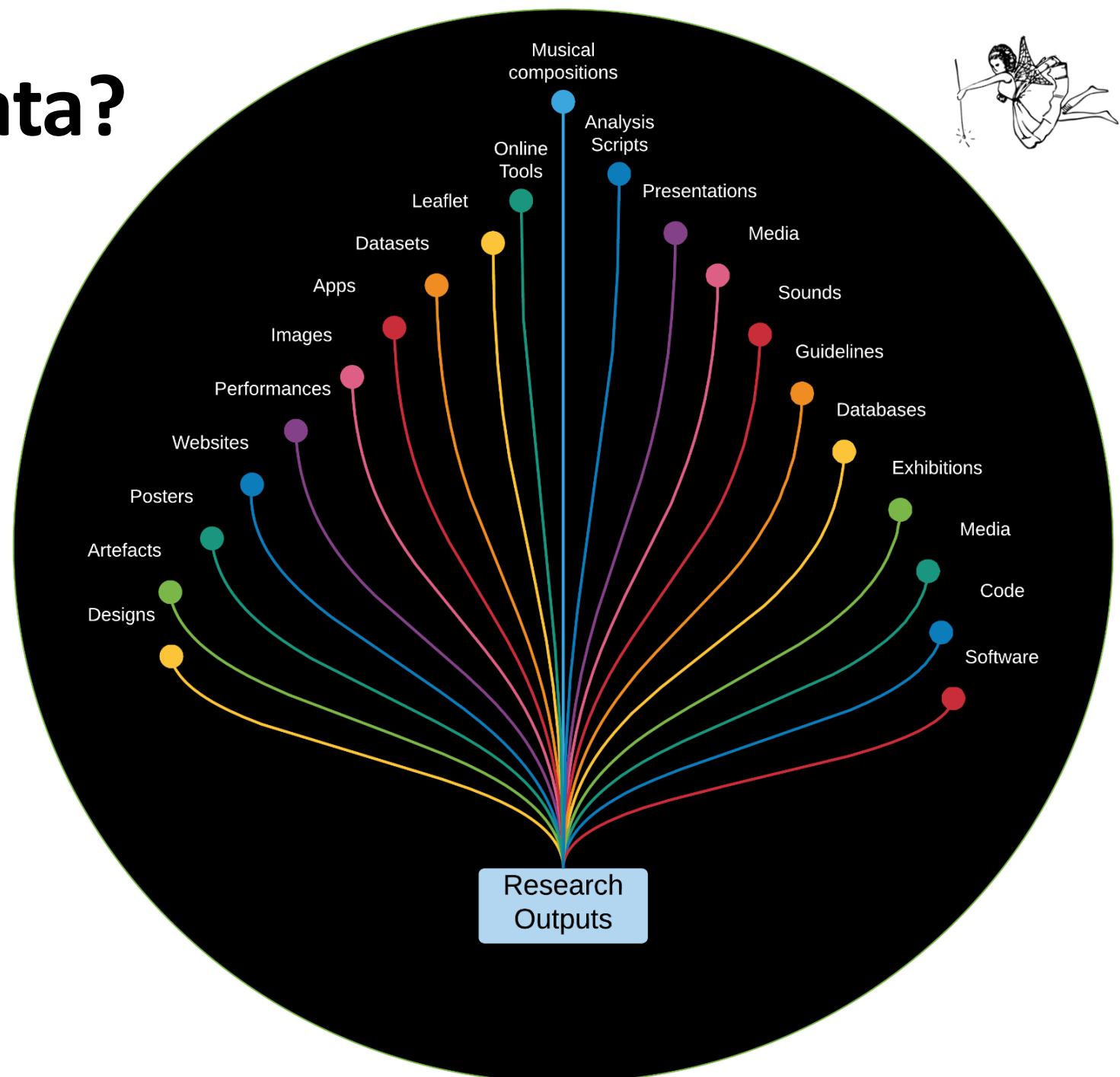Go to www.menti.com and enter **50 78 83 8**
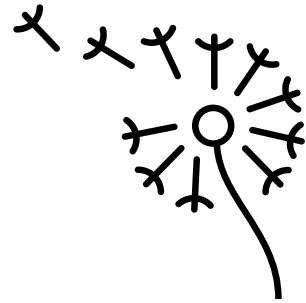
What is Research Data?

"..........information generated, collected or observed during a research project.

It is evidence used to support research conclusions and will go on to form part of the scholarly record.

Go to www.menti.com and enter **50 78 83 8**

What is Open Data?

Open Research Data is data that can be freely used, reused and redistributed by anyone - subject only, at most, to the requirement to credit the curator and share under the same license.

Benefits of Open Data

Researchers
- greater discoverability
- increased efficiency
- attracts funding & support
- new collaborations

Funders
- increased visibility & reuse of funded research
- greater funding impact
- greater ROI

General Public
- self-empowerment
- increased transparency
- greater engagement in science & research

Organisations/ NGOs
- enhanced access to research
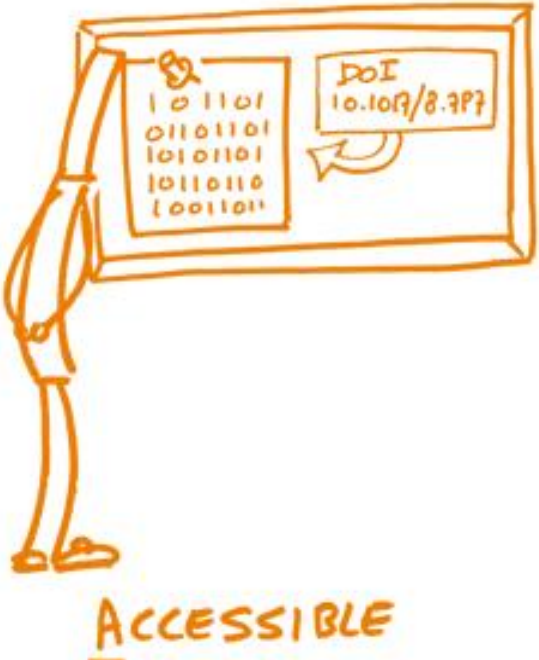- better information-sharing
- more effective advocacy/lobbying

National Governments
- data-driven decision making
- reduced government costs
- more effective & efficient government services

Go to www.menti.com and enter **50 78 83 8**
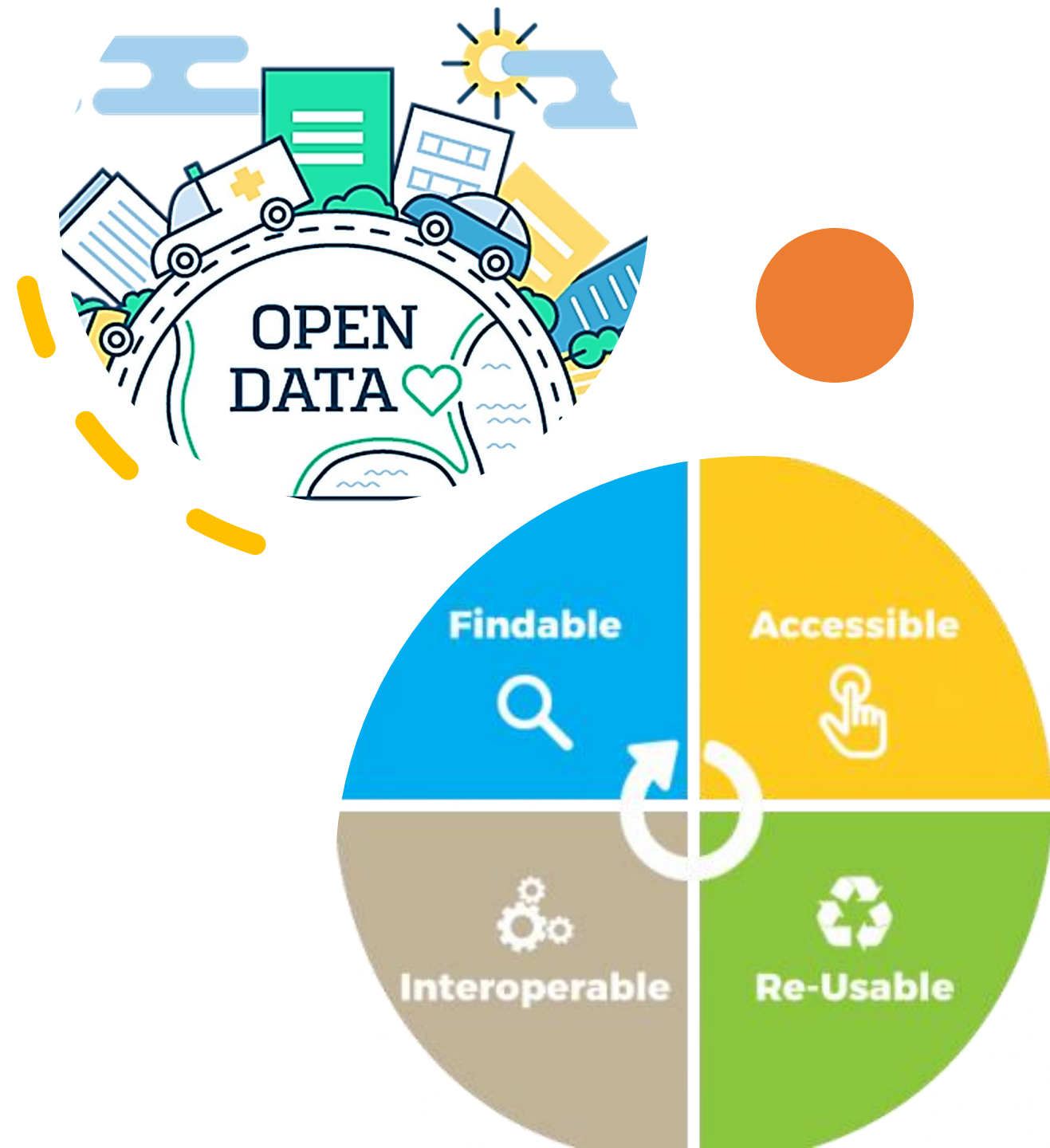
What are the FAIR Principles?

Principles and practice

Some reasons to aim for FAIR and/or Open Data

- Expose your data and research to a wider audience
- Increased transparency
- Reproducibility of results
- Reduce duplication of effort
- Increase longevity
- Share Null results
- Opportunities for new partnerships and collaborations
- Save time and effort in maintaining your own database
- Better value for public money spent on research

Go to www.menti.com and enter **50 78 83 8**

Other reasons to think about FAIR

IRISH RESEARCH COUNCIL
An Chomhairle um Thaighde in Éirinn

Horizon 2020
European Union Funding
for Research & Innovation

An Roinn Talmhaíochta,
Bia agus Mara
Department of Agriculture,
Food and the Marine

epa
Environmental Protection Agency

Science Foundation Ireland
sfi
For what's next

wellcome

HRB
An Bord
Taighde Sláinte
Health Research
Board

IEEE

PLOS

Taylor & Francis
Taylor & Francis Group

NON SOLUS
ELSEVIER

CAMBRIDGE
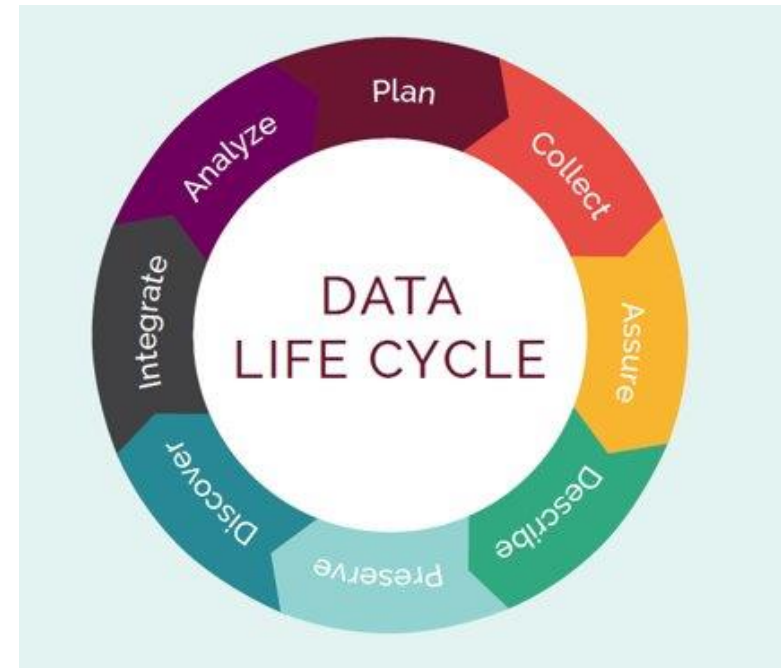UNIVERSITY PRESS

Springer

WILEY

Go to www.menti.com and enter 50 78 83 8

To make data FAIR you must......

plan to do so from early in the research process.

What is Data Management?

DATA LIFE CYCLE
Plan, Collect, Assure, Describe, Preserve, Discover, Integrate, Analyze

Go to www.menti.com and enter **50 78 83 8**

Why do you need a data management plan?

The Only Naming Convention That Works

DOGHOUSEDIARIES

A STORY TOLD IN FILE NAMES:

Location: C:\user\research\data

Filename	Date Modified	Size	Type
data_2010.05.28_test.dat	3:37 PM 5/28/2010	420 KB	DAT file
data_2010.05.28_re-test.dat	4:29 PM 5/28/2010	421 KB	DAT file
data_2010.05.28_re-re-test.dat	5:43 PM 5/28/2010	420 KB	DAT file
data_2010.05.28_calibrate.dat	7:17 PM 5/28/2010	1,256 KB	DAT file
data_2010.05.28_huh??.dat	7:20 PM 5/28/2010	30 KB	DAT file
data_2010.05.28_WTF.dat	9:58 PM 5/28/2010	30 KB	DAT file
data_2010.05.29_aaarrrgh.dat	12:37 AM 5/29/2010	30 KB	DAT file
data_2010.05.29_#$@*&!!.dat	2:40 AM 5/29/2010	0 KB	DAT file
data_2010.05.29_crap.dat	3:22 AM 5/29/2010	437 KB	DAT file
data_2010.05.29_notbad.dat	4:16 AM 5/29/2010	670 KB	DAT file
data_2010.05.29_woohoo!!.dat	4:47 AM 5/29/2010	1,349 KB	DAT file
data_2010.05.29_USETHISONE.dat	5:08 AM 5/29/2010	2,894 KB	DAT file
analysis_graphs.xls	7:13 AM 5/29/2010	455 KB	XLS file
ThesisOutline!.doc	7:26 AM 5/29/2010	38 KB	DOC file
Notes_Meeting_with_ProfSmith.txt	11:38 AM 5/29/2010	1,673 KB	TXT file
JUNK...	2:45 PM 5/29/2010		Folder
data_2010.05.30_startingover.dat	8:37 AM 5/30/2010	420 KB	DAT file

Type: Ph.D Thesis Modified: too many times   Copyright: Jorge Cham   www.phdcomics.com

Go to www.menti.com and enter 50 78 83 8

Why do you need a data management plan?

Data Management ~~Reproducibility~~ is important because the you of 3 months ago is terrible at answering email! - @tracykteal at #2016dssummit

Casey Greene
@GreeneScientist

1:17 PM - 26 Oct 2016 from Manhattan, NY

Go to www.menti.com and enter **50 78 83 8**

Why do you need a data management plan?

- Essential part of project management

- Establishes clear workflows and guidelines surrounding data management

- Sets out transparent data practices

- Plan and allocate resources and cost

- Plan infrastructure to support data

- Day to day you can find and understand your data when you need to use it

- There is continuity if project staff leave or new researchers join

- You can avoid unnecessary duplication e.g. re-collecting or re-working data

- The data underlying publications are maintained, allowing for validation of results

- The data can become another output from a project

Go to www.menti.com and enter **50 78 83 8**

Why do you need a data management plan?

• Essential part of project management

• Establishes clear workflows and guidelines surrounding data management

• Sets out transparent data practices

• Plan and allocate resources and cost

• Plan infrastructure to support data

• Day to day you can find and understand your data when you need to use it

• There is continuity if project staff leave or new researchers join

• You can avoid unnecessary duplication e.g. re-collecting or re-working data

• The data underlying publications are maintained, allowing for validation of results

• The data can become another output from a project

A data management plan is the way to FAIR data.

Go to www.menti.com and enter **50 78 83 8**

Structure of Data Management Plans

1. Data Collection

2. Documentation and Metadata

3. Ethical and Legal Compliance

4. Storage and Backup during active research

5. Data Sharing and Long-term Preservation

6. Requirements and Resourcing

SCIENCE EUROPE
PRACTICAL GUIDE TO
THE INTERNATIONAL ALIGNMENT OF
RESEARCH DATA MANAGEMENT

SCIENCE EUROPE

Go to www.menti.com and enter 50 78 83 8

1. Data Collection

What is the type, format and volume of data?

How will data be collected or created?
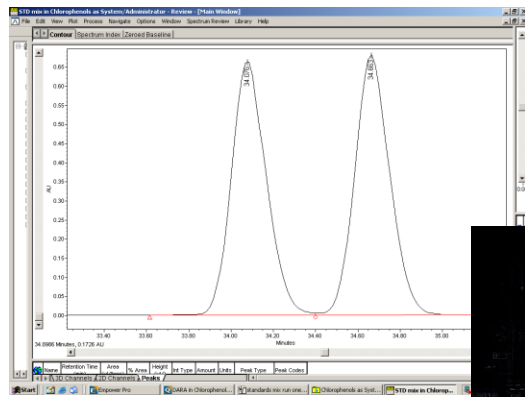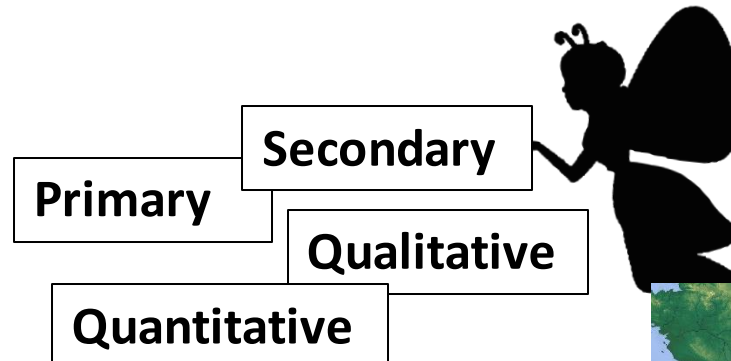
Primary

Secondary

Qualitative

Quantitative

2. Documentation and Metadata and Data Quality

What metadata and documentation will accompany data?
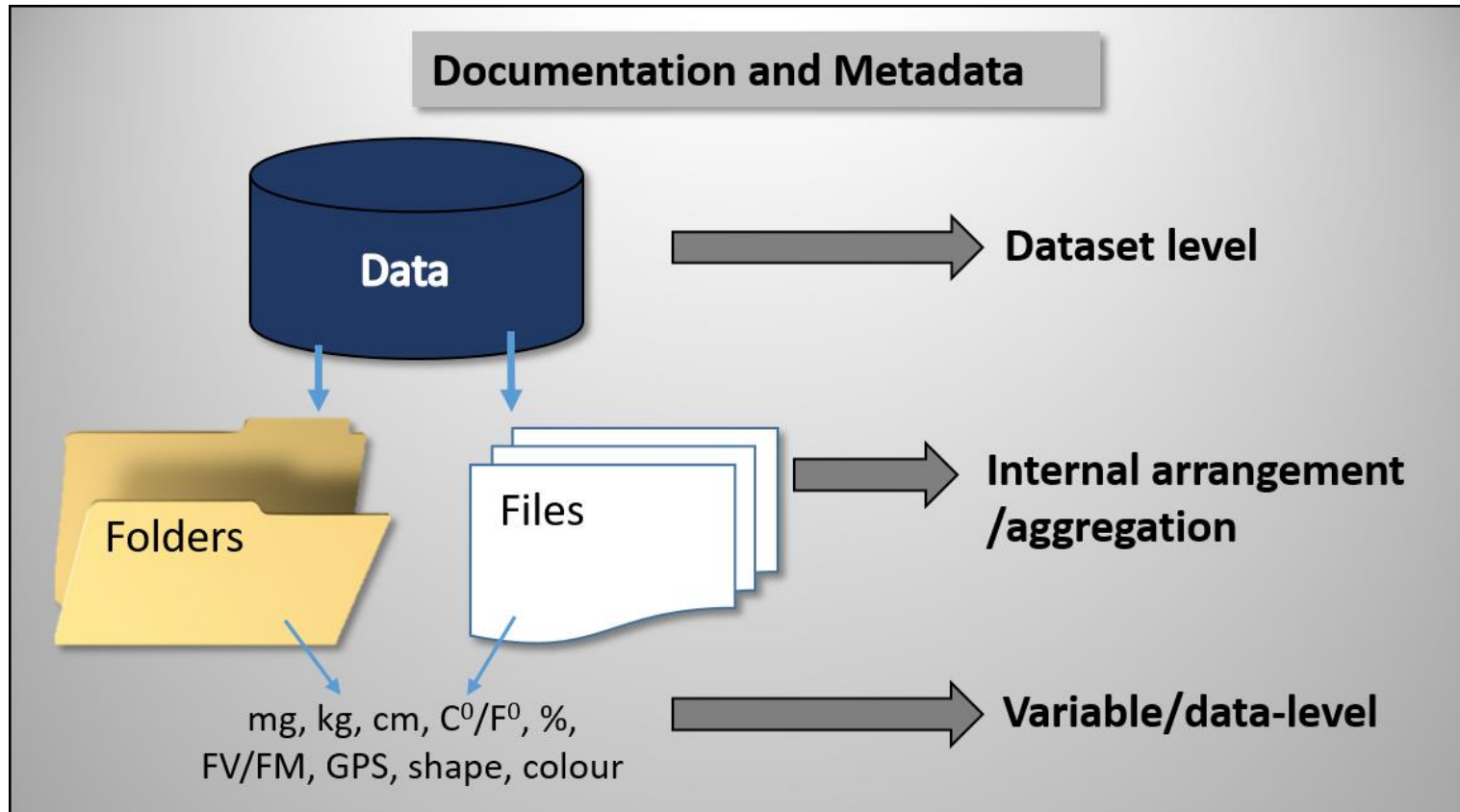
Discovery/Descriptive Metadata
- authors
- funders
- keywords
- abstract

Substantive/Contextual Metadata
- bibliography/citations
- standards used
- file formats
- data dictionary
- analysis scripts
- links to other metadata

Metadata is data about data...it describes and gives information about other data

2. Documentation and Metadata and Data Quality

mg, kg, cm, C^0/F^0, %,
FV/FM, GPS, shape, colour
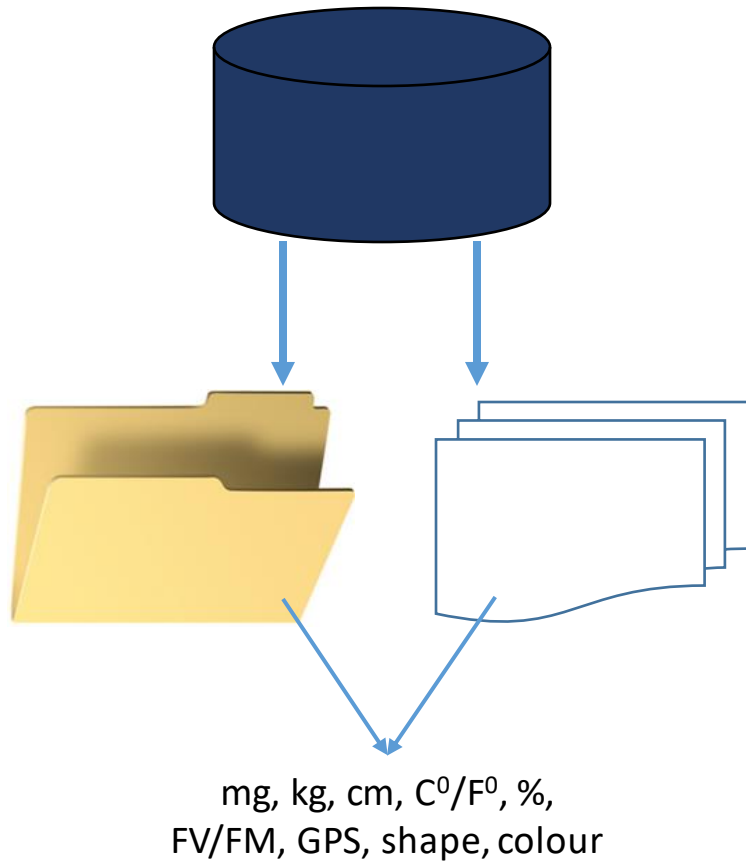
Dataset level:
what is this thing, who made it, when, how...?

Internal arrangement/aggregation:
what files, folders, database tables, components make up this thing, what are they, who made them, how do I open them, how do they relate to each other, how does their naming and arrangement encode meaning?

Variable-level:
where is the label for this variable, what does the label mean, what units were used, what are the acceptable values, how do these variables relate to each other?

Go to www.menti.com and enter **50 78 83 8**

National Water Vole Monitoring Programme survey form

people's trust for endangered species

3

Site Number:		Transect No. (if applicable):		Bank surveyed (N/S/E/W):
Start grid ref:		End grid ref:		Transect length:
Survey date:		Surveyor:		

Water vole signs in each 100m section of the transect
(complete as many as is applicable to the length of your transect)

Field sign	0-100m	100-200m	200-300m	300-400m	400-500m	500-600m
Number of trampled latrines (trodden flat on top)						
Number of untrampled latrines						
Please note the location of the **first** latrine that you encounter (Grid Reference or GPS)						
Please note the location of the **last** latrine that you encounter (Grid Reference or GPS)						
Water vole feeding signs						
Burrows/nests (approximate no.)	☐ None ☐ 1-5 ☐ 6 - 10 ☐ More than 10	☐ None ☐ 1-5 ☐ 6 - 10 ☐ More than 10	☐ None ☐ 1-5 ☐ 6 - 10 ☐ More than 10	☐ None ☐ 1-5 ☐ 6 - 10 ☐ More than 10	☐ None ☐ 1-5 ☐ 6 - 10 ☐ More than 10	☐ None ☐ 1-5 ☐ 6 - 10 ☐ More than 10
Sighting:	☐ No ☐ Yes, if so how many:	☐ No ☐ Yes, if so how many:	☐ No ☐ Yes, if so how many:	☐ No ☐ Yes, if so how many:	☐ No ☐ Yes, if so how many:	☐ No ☐ Yes, if so how many:

ECOLOGICAL MONITORING DATA - COMPREHENSIVE

Savory

© SAVORY INSTITUTE 2017

UCC
Coláiste na hOllscoile Corcaigh, Éire
University College Cork, Ireland

Consent Form

I…………………………………………agree to participate in **YOUR NAME's** research study.

The purpose and nature of the study has been explained to me in writing.

I am participating voluntarily.

I understand that I can withdraw from the study, without repercussions, at any time, whether before it starts or while I am participating.

I understand that I can withdraw permission to use the data within two weeks of the interview, in which case the material will be deleted.

I understand that anonymity will be ensured in the write-up.

Signed: …………………………………….   Date: ………………

PRINT NAME: …………………………………….

Go to www.menti.com and enter 50 78 83 8

https://www.ucc.ie/en/sit/software/endnote/

https://www.zotero.org/

Go to www.menti.com and enter **50 78 83 8**

https://rd-alliance.github.io/metadata-directory/

https://fairsharing.org/

Why control data quality?

- Research Integrity

- You can be confident of your results and analysis

- Less time spent 'cleaning' the data

- More likely that your data will be re-used

What data quality control measures do you use?

- A data management plan

- Defined SOPs for treatment of your data

- Naming conventions

- Version control

- Read only raw data files

- Spreadsheet validation

…….etc.

Go to www.menti.com and enter **50 78 83 8**

3. Ethics and Legal Compliance

How will you manage ethical issues and codes of conduct?

How will you manage copyright and Intellectual Property Rights (IPR) issues?
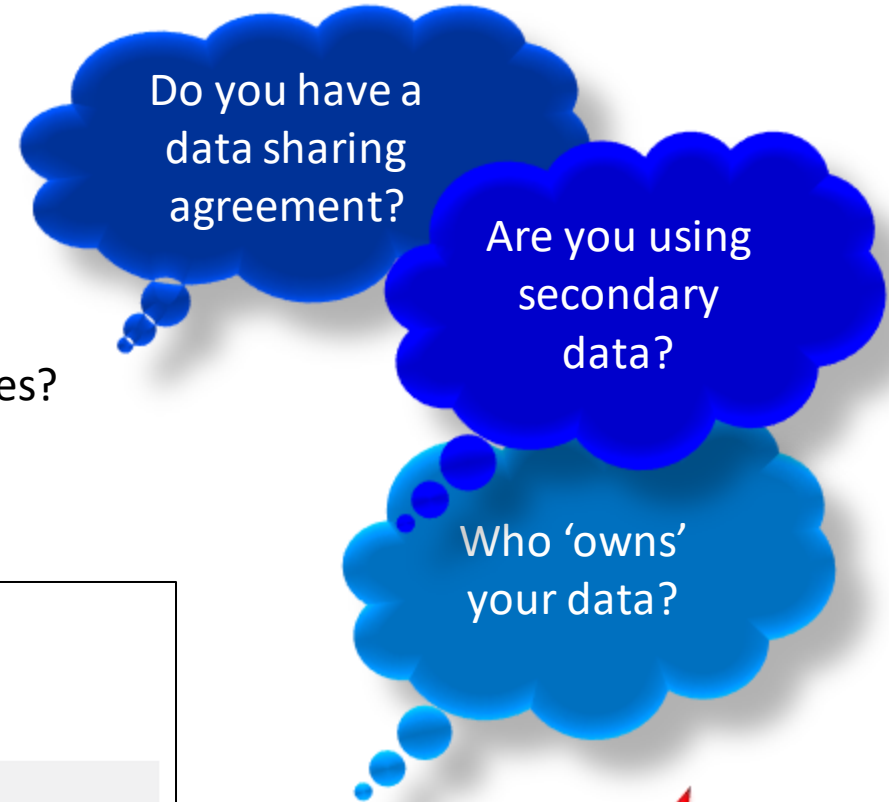
Do you have a data sharing agreement?

Are you using secondary data?

Who 'owns' your data?

GDPR !

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

CODE OF RESEARCH CONDUCT

Research Ethics

- Ethics ● Clinical Research ● Social Research
- Animal Experimentation

UCC | **Office of Technology Transfer**
University College Cork, Ireland | Aistriú Teicneolaíochta
Coláiste na hOllscoile Corcaigh

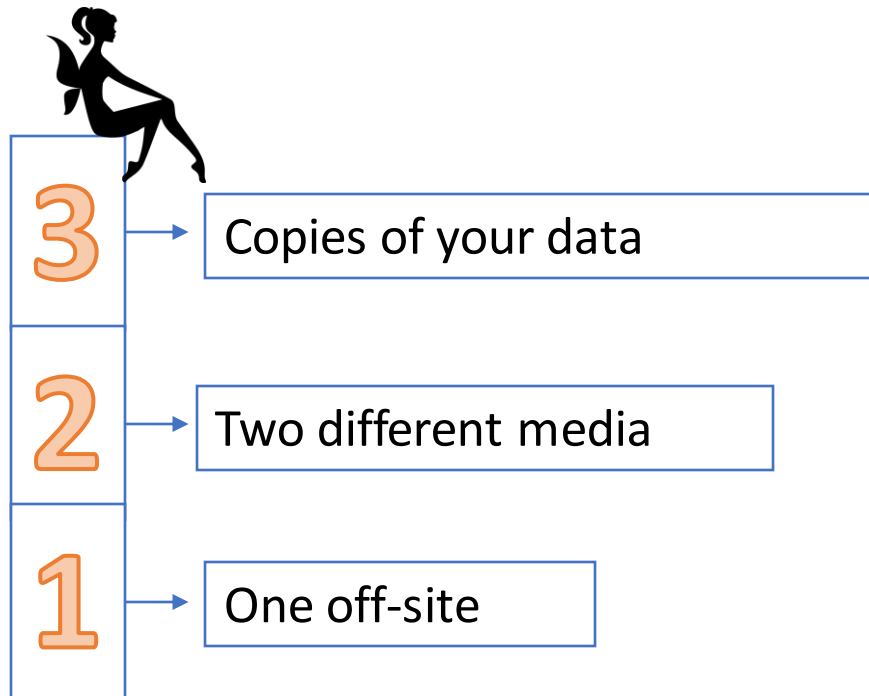Go to www.menti.com and enter **50 78 83 8**

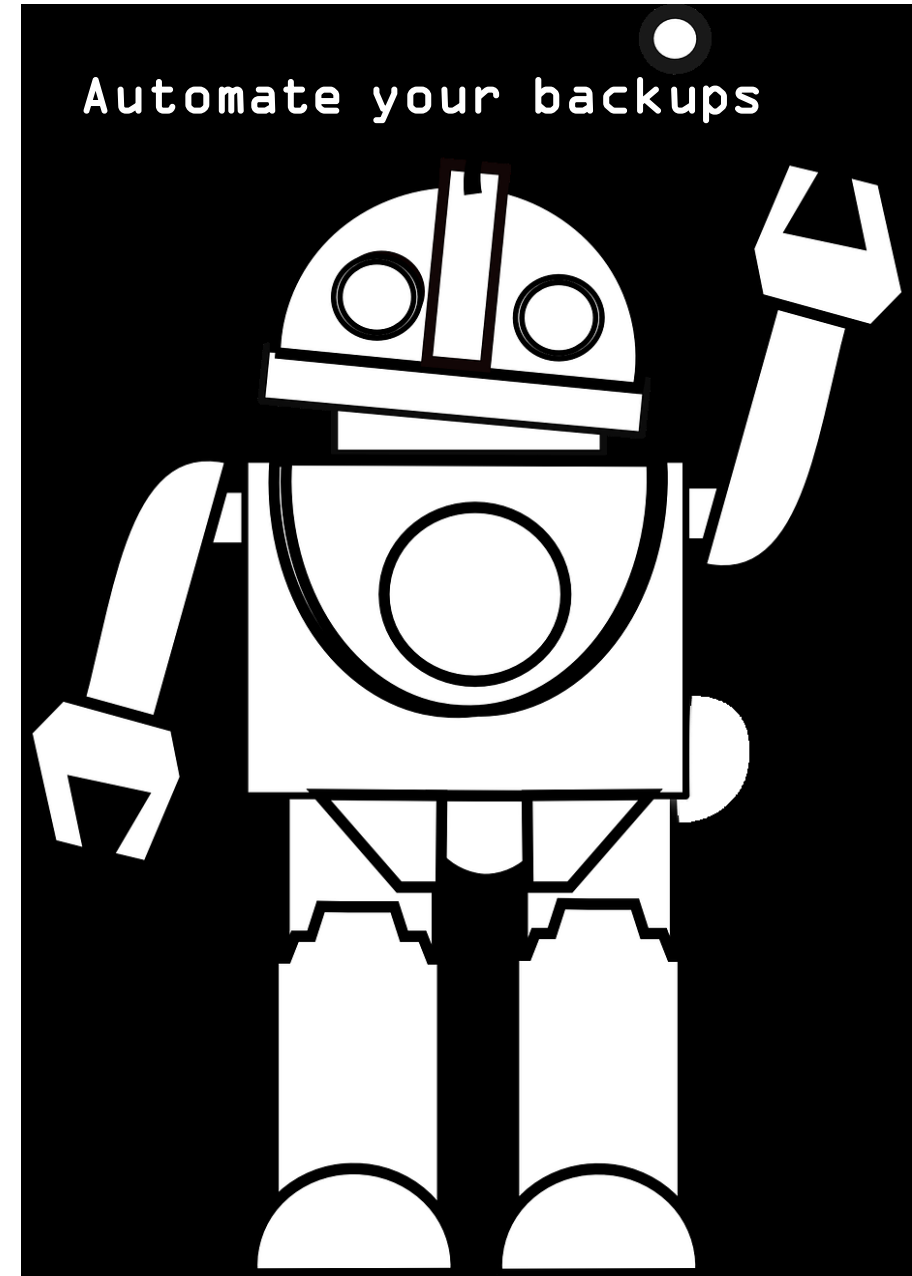# 4. Storage and backup [DURING]

How will data be stored and backed up during the research?

How will you take care of data security and personal data protection?
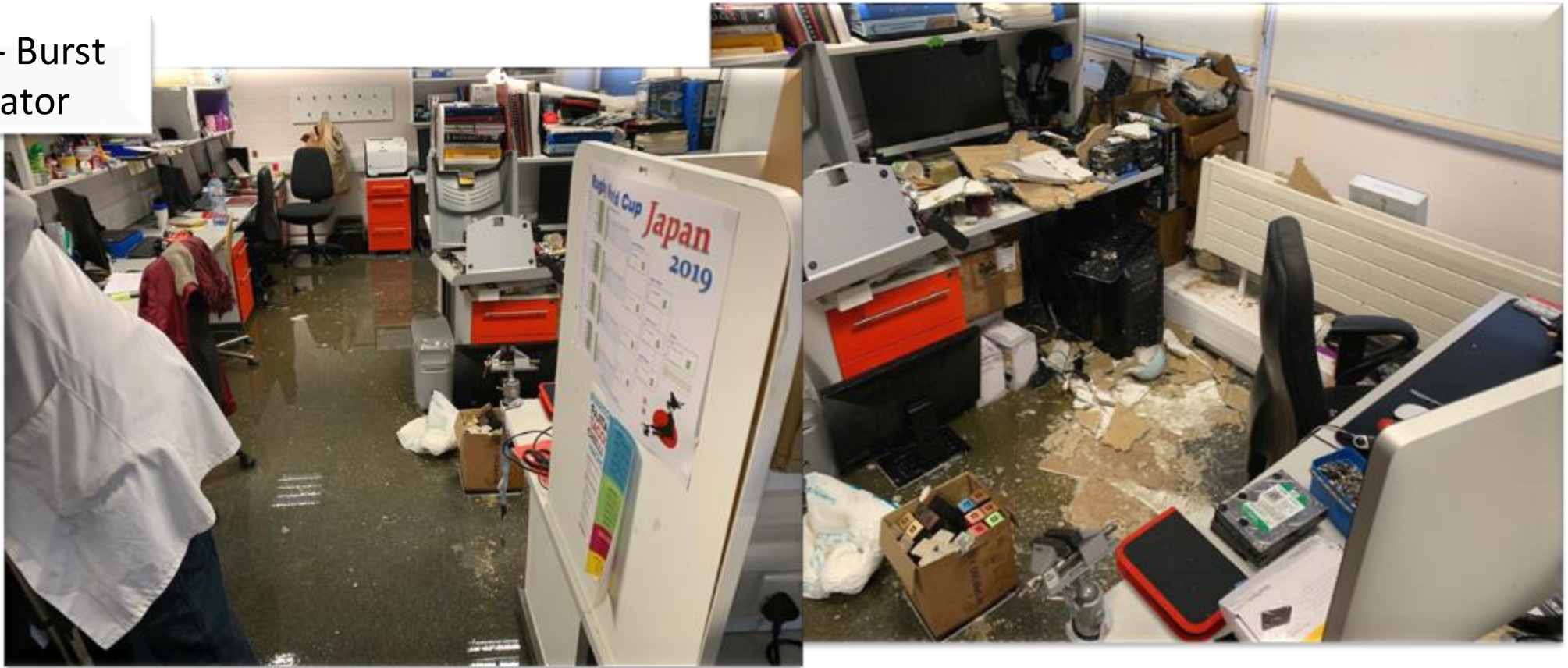
How will you manage access and security?

3 → Copies of your data

2 → Two different media

1 → One off-site

https://www.nature.com/articles/d41586-019-01040-w

Automate your backups

Go to www.menti.com and enter **50 78 83 8**

APC- Burst Radiator

Dear all,

A new colleague has just arrived in Cork and discovered that the hard drive in his laptop is not functioning due to hard drive failure (Unmountable boot volume).

He will get a replacement hard drive from Dell on Thursday, but in the meantime needs to recover the data from the damaged one.

I would be very grateful to receive any advice on how to do this, or details of a local company that could help.

Thanks in advance

4. Storage and backup [DURING]

How will data be stored and backed up during the research?

How will you take care of data security and personal data protection?
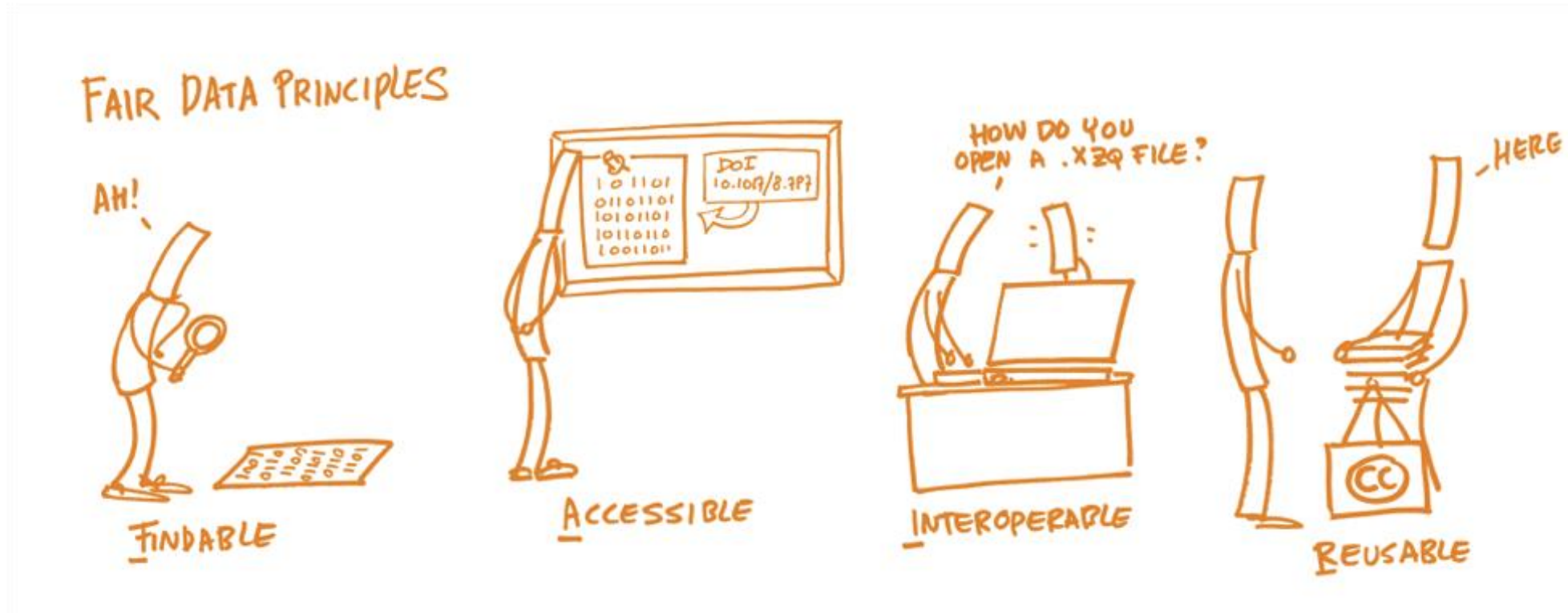
How will you manage access and security?

Research
DataStore

5. Data Sharing and Long-Term Preservation

Heller agrees. However, the true benefit of FAIR, he argues, is in providing a framework for researchers to manage their own data so they can themselves find it, understand it, and reuse it. "As a scientist, you should treat your data like a love letter to your future self," he says.
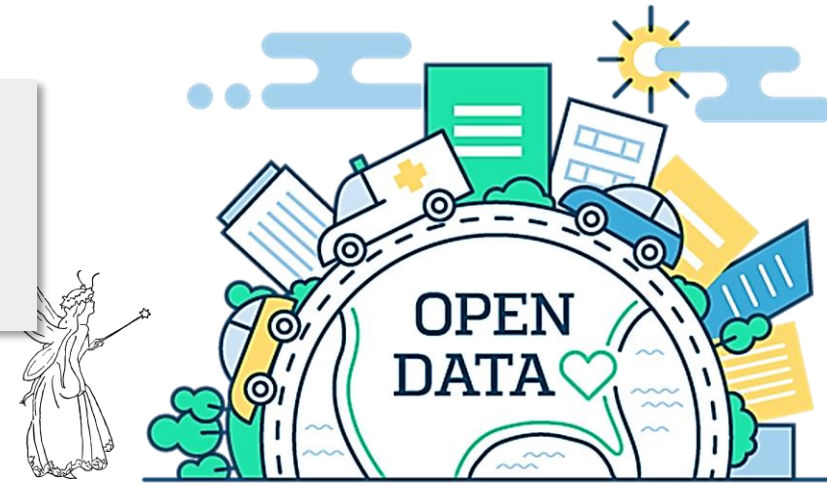
5. Data Sharing and Long-term Preservation

What data? All data?

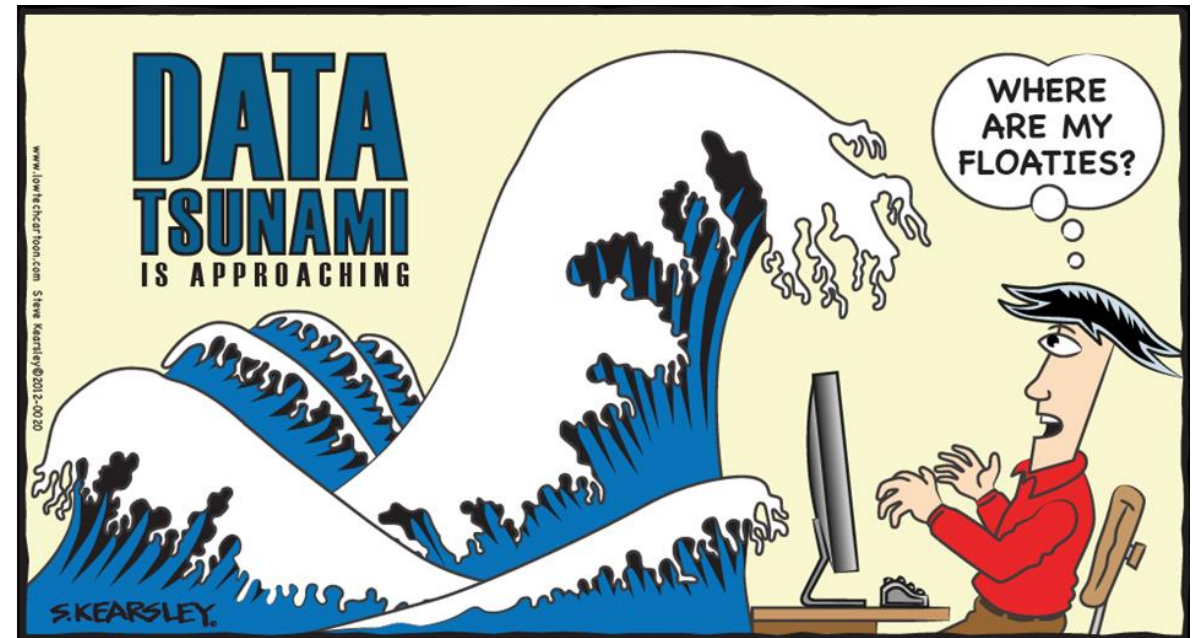Which data should be retained, shared and/or preserved?

What is the long-term preservation plan for the dataset?

H2020

"........underlying data"

"………any other data"

.......as open as possible as closed as necessary"

http://www.lowtechcartoon.com/?p=934

5. Data sharing and Long-term Preservation

How and when will you share data (licences, security, possible embargo reasons)?

Will you make sure unique and persistent identifier is in use (e.g. DOI)?

https://datacite.org/

https://zenodo.org/

https://figshare.com/

https://datadryad.org/stash

https://creativecommons.org/

https://fairsharing.org/

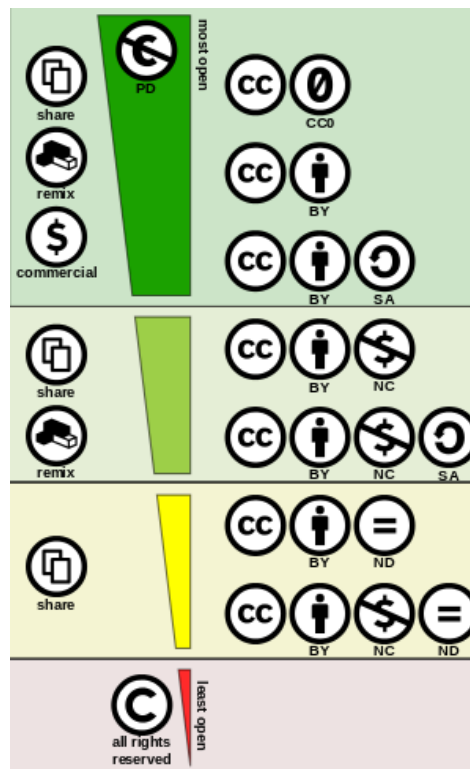https://opensource.org/licenses/MIT

6. Responsibilities and Resources

Who will be responsible of data management?

What resources will you require to deliver your plan?

Take Home Messages

Introduction to UCC Library

Introduction to UCC Library

Finding Library Resources for Your Assignments

Finding Library Resources for Your Assignments

Citing, Referencing & Plagiarism

Citing, Referencing & Plagiarism

Literature Review

Literature Review

Internet for Research

Internet for Research

Referencing Software

Referencing Software

Library Videos & Tutorials

Library Videos & Tutorials
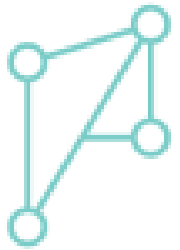
Research Data Management

Research Data Management

The FAIR Principles

The FAIR Principles

Special Collections & Archives

Other useful stuff

https://pubpeer.com/static/about

https://retractionwatch.com/

https://archive.org/web/

Data Horror Escape Room

https://sites.google.com/vu.nl/datahorror/home

https://perma.cc/
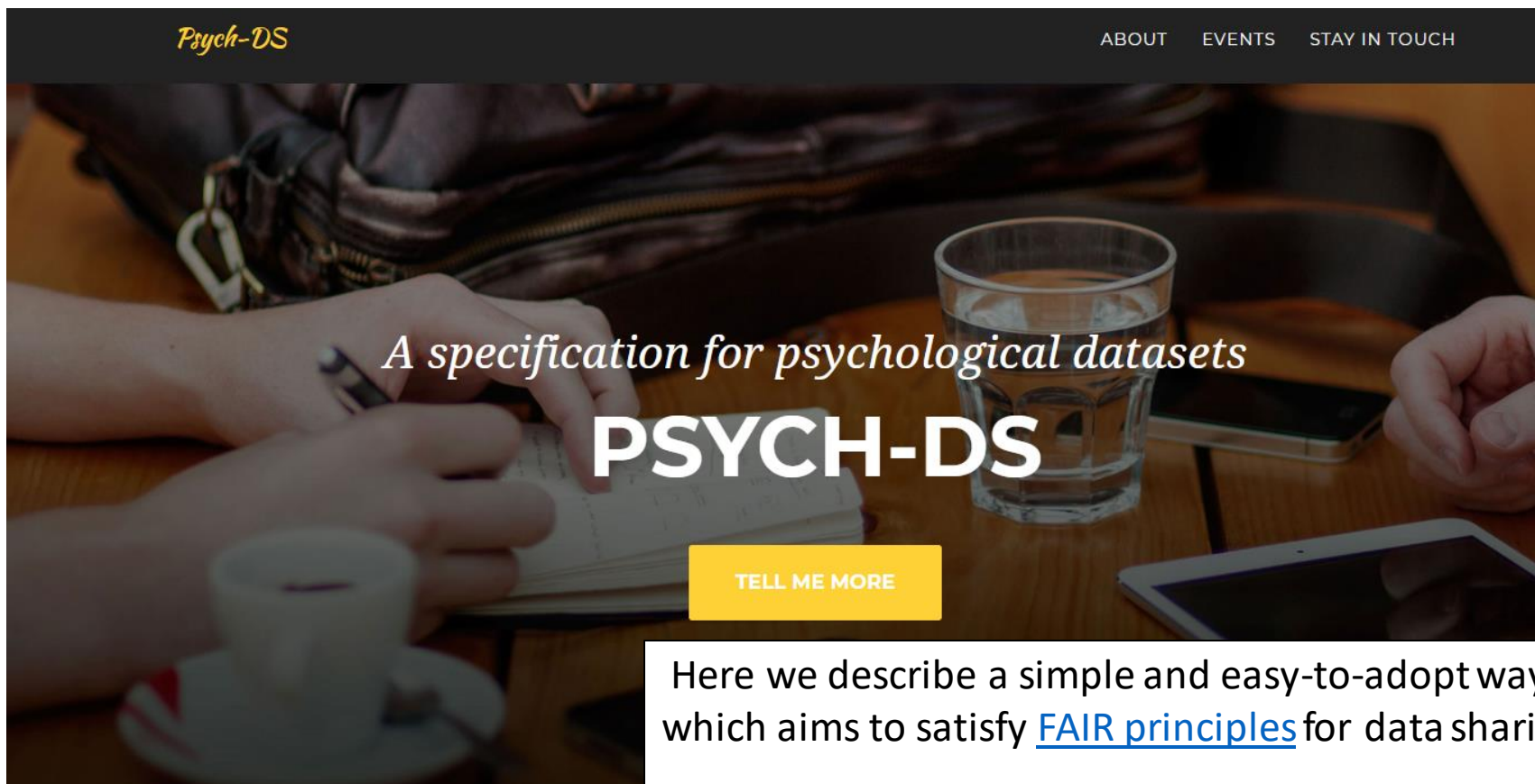
Any Questions

Go to www.menti.com and use the code 82 72 18 6

Specialist metadata standards examples

A specification for psychological datasets

PSYCH-DS

TELL ME MORE

Here we describe a simple and easy-to-adopt way of organizing psychological data, which aims to satisfy FAIR principles for data sharing.

Critically, there are two goals for this standard: (1) we wish to promote the adoption of good practices in the management of scientific data by individual working scientists and (2) we wish to create a machine-readable format for these datasets that can support tools for analysis, discovery, and preparation of datasets in psychology.

https://docs.google.com/document/d/1u8o5jnWk0Iqp_J06PTu5NjBfVsdoPbBhstht6W0fFp0/edit#heading=h.1m6sa5bs9j0e

AUTHOR_DATASET_ReadmeTemplate - Notepad

File   Edit   Format   View   Help

```
-------------------
GENERAL INFORMATION
-------------------



1. Title of Dataset |


2. Author Information


  Principal Investigator Contact Information
        Name:
            Institution:
            Address:
            Email:



  Associate or Co-investigator Contact Information
        Name:
                        n:
```

DataCite - International Data Citation

DataCite Metadata Schema Documentation for the Publication and Citation of Research Data

Citation:

DataCite Metadata Working Group. (2017). DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.1. DataCite e.V. 10.5438/0014.

Document, Discover and Interoperate

The Data Documentation Initiative (DDI) is an international standard for describing the data produced by surveys and other observational methods in the social, behavioral, economic, and health sciences. DDI is a free standard that can document and manage different stages in the research data lifecycle, such as conceptualization, collection, processing, distribution, discovery, and archiving. Documenting data with DDI facilitates understanding, interpretation, and use -- by people, software systems, and computer networks. Use DDI to **D**ocument, **D**iscover, and **I**nteroperate!

https://rd-alliance.github.io/metadata-directory/standards/ddi-data-documentation-initiative.html

Ecological Metadata Language

From Wikipedia, the free encyclopedia

Ecological Metadata Language (EML) is a metadata standard developed by and for the ecology discipline. It is based on prior work done by the Ecological Society of America and others,[1] including the Knowledge Network for Biocomplexity.[2] EML is a set of XML schema documents that allow for the structural expression of metadata. It was developed specifically to allow researchers to document a typical data set in the ecological sciences.

...gned to describe digital resources, however, it may also be used to describe non-digital resources such as paper maps ...al media.

...twork for Biocomplexity project has developed a software client specifically to address this need. Morpho[3] is data ...are intended for generating metadata in EML format. Morpho is part of the DataONE Investigator Toolkit and therefore ...e data sharing and reuse am...

Darwin Core

Darwin Core is a standard maintained by the Darwin Core maintenance group. It includes a glossary of terms intended to **facilitate the sharing of information about biological diversity** by providing identifiers, labels, and definitions. Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, samples, and related information.

| technical specification | current standard | 2009 |

Specialist metadata standards examples

Common semantic features

Version 1.1 *Working specification*

- Introduction
- Definition of terms
- Semantics of data items
  - Data name semantics
    - Namespace
    - Note on handling of units
  - Data value semantics
    - Data typing
    - Data subtyping
    - Special generic values
  - Embedded data semantics
    - CIF conventions for special characters in text
  - Handling of long lines
- Dictionary compliance
- CIF markup conventions
  - Greek letters
  - Accented letters
  - Other characters
  - Typographic style codes
- References

The Crystallographic Information File

Medical Subject Headings
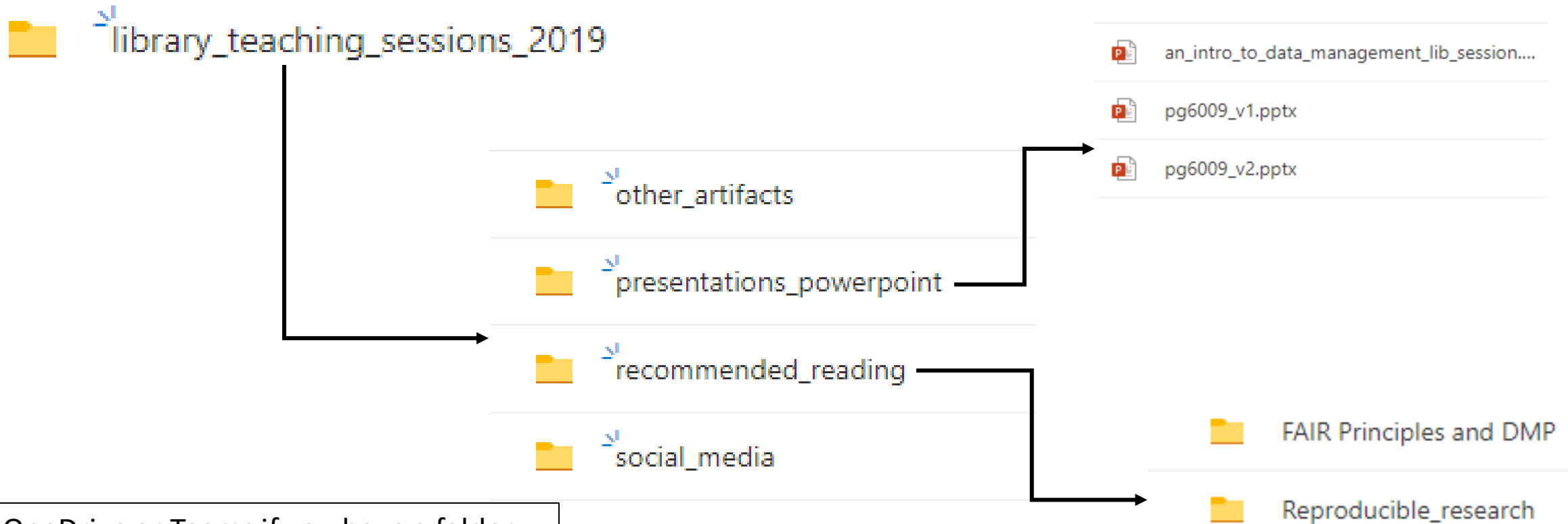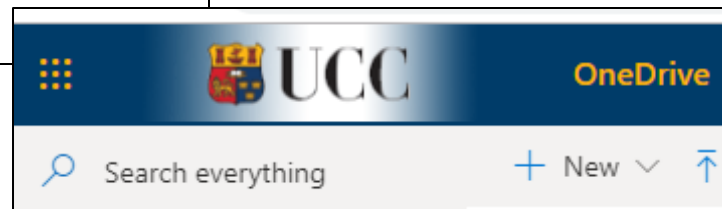
Naming Conventions

- Be Consistent

- Descriptive

- Consensus......on collaborative projects

- Chronological Order YYYYMMDD or YYMMDD

- Short names

- Don't use special characters....@#?~£$€&

- Use leading zeros 0001, 0002 ……..1001

- Don't use spaces …file_name, file-name, filename or FileName

- Include a README.txt file to explain and describe your naming convention

Naming and structuring Folder and Files

📁 library_teaching_sessions_2019

📁 other_artifacts

📁 presentations_powerpoint

📁 recommended_reading

📁 social_media

📄 an_intro_to_data_management_lib_session....

📄 pg6009_v1.pptx

📄 pg6009_v2.pptx

📁 FAIR Principles and DMP

📁 Reproducible_research

In OneDrive or Teams if you have a folder structure more than 3 layers deep you won't be able to search using the search function effectively. This means there is a risk of things getting lost.

UCC OneDrive

🔍 Search everything   + New ∨ ↑

Spreadsheet Design..........be aware of the limitations of excel and take steps to prevent them

1. Be Consistent
2. Decide on naming conventions
3. Use ISO standards
4. Have rules for empty cells
5. One piece of data per cell
6. Make it a rectangle
7. Subjects in rows
8. Variables in columns
9. Create a data dictionary
10. Backup
11. Version Control
12. Use data validation
13. Include a ReadMe file

1. No calculation in raw data files
2. Don't use formatting as data
3. Don't have figures in raw data files
4. Don't use merged cells
5. Don't hide columns
6. Don't use special characters
7. Don't have more than one header row
8. Don't include temporary or dummy variables not of use to researchers

https://www.kristianbrock.com/post/send-me-data/
https://www.youtube.com/watch?v=Ry2xjTBtNFE
https://www.tandfonline.com/doi/full/10.1080/00031305.2017.1375989

Version Control

- Provides a 'way-back' to earlier versions

- Track changes by multiple authors

- Ensures everyone is working on the latest version of the document

- Incorporate into your naming convention

  - v1, v2

  - Draft, final ....be careful with this one!

  - Information about changes made ...'transformed data'

- Manual Version control

  - Don't need a version history

  - Few people working on the files

  - Files are accessed from one locations

- Software for version control

  - Google Drive

  - Word

  - Git