

Research Data Management

Dr. Aoife Coffey, Research Data Coordinator,

Research Services, UCC Library

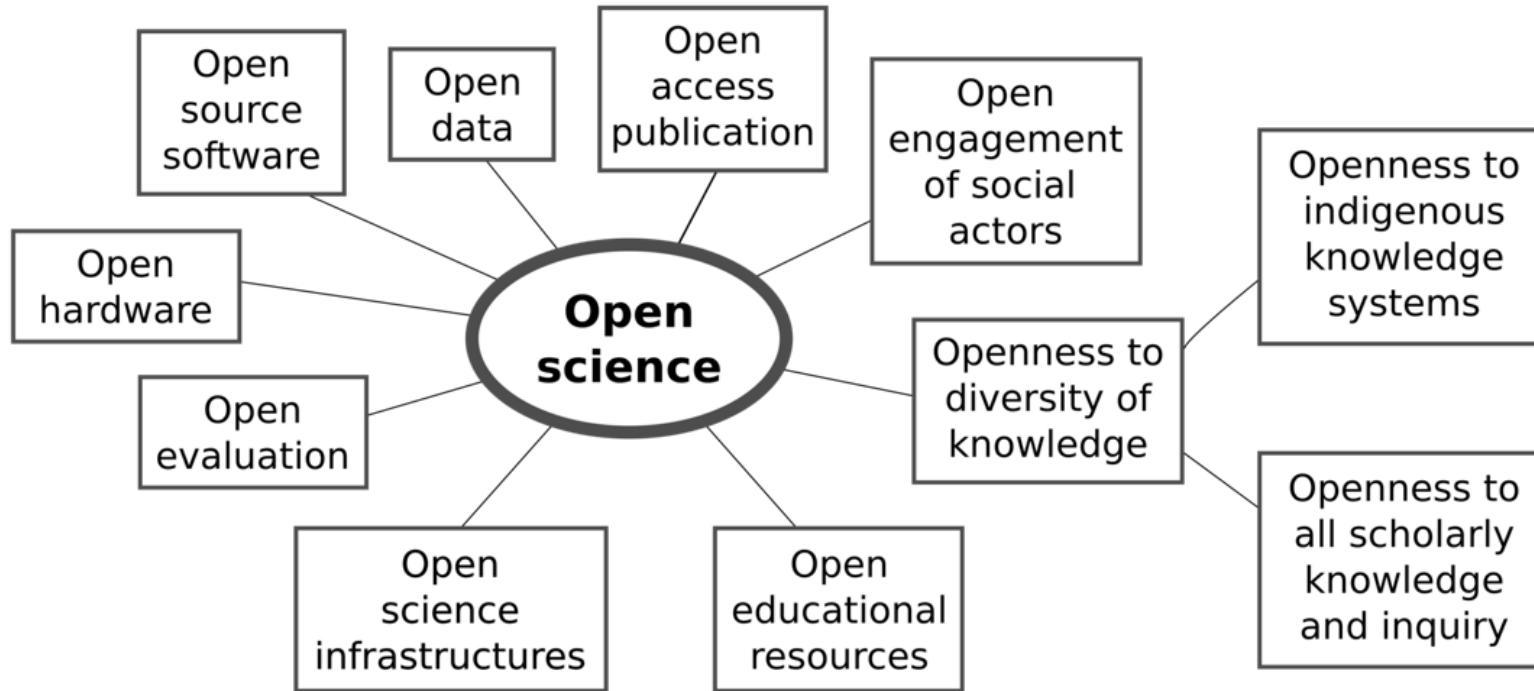
researchdata@ucc.ie



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Library
LEABHARLANN

Open Science /Research



“**Open science** is an approach based on open cooperative work and systematic sharing of knowledge and tools as early and widely as possible in the process. It has the potential to increase the quality and efficiency of research and accelerate the advancement of knowledge and innovation by sharing results, making them more reusable and improving their reproducibility. It entails the involvement of all relevant knowledge actors”

Horizon Europe

Open Research Data

Refers to the data underpinning scientific research results with no restrictions on its access

- Accountability
- Transparency
- Reproducibility
- Building on existing data
- Revisiting existing data
- Better and stronger collaborations
- Value for public money



FAIR Principles

1. Metadata
2. Persistent Identifiers
3. Licensing
4. Repositories



<https://libguides.ucc.ie/researchdataservice/fair>

<https://www.nature.com/articles/sdata201618>

<https://www.go-fair.org/fair-principles/>

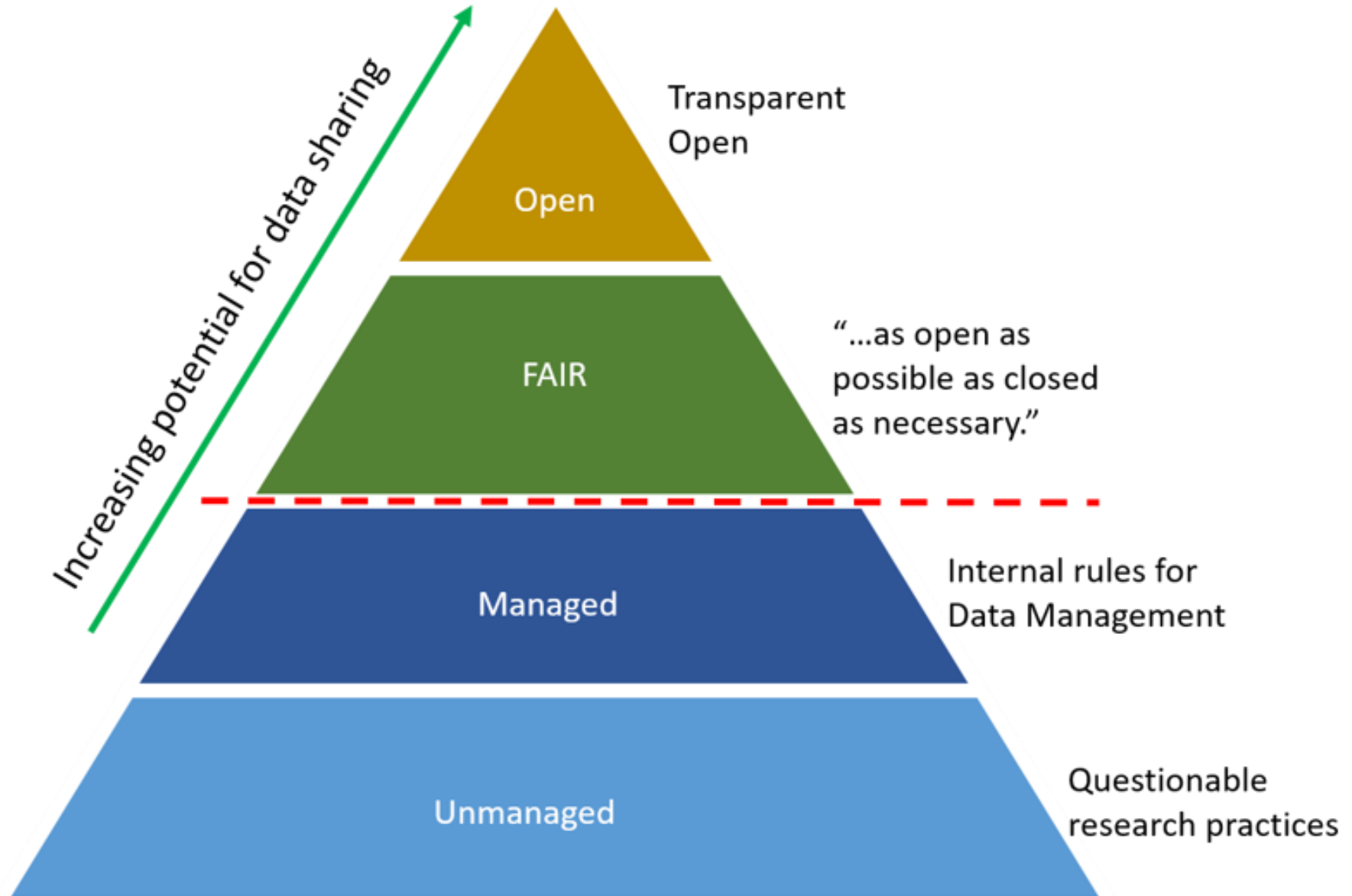
<https://www.andis.org.au/working-with-data/fairdata/training>

National and International Context

- National Funders
 - NORF
 - Impact 2030
 - Research Integrity
 - Publishers
 - Open Data Directive
- EU Funders
 - European Open Science Cloud
 - Research Integrity
 - Publishers
 - COARA –Coalition for Advancing Research Assessment

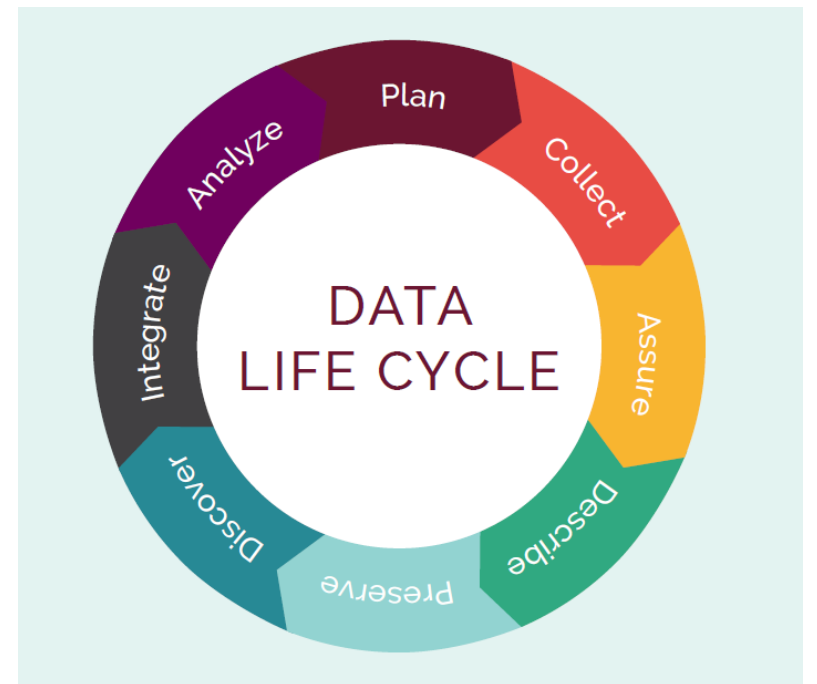
How do
you make
data FAIR?

Degrees of Data Management



What is data management?

- It is the **active** management of research data during the lifecycle of a project and also includes decisions about treatment of the data post-project
- It involves organization, storage, preservation and potential sharing of the data
- It's an integral part of good research practice



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
- Ensures compliance with legislation
- 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

Structure of Data Management Plans



- 1. Data Collection**
- 2. Documentation, metadata and data quality**
- 3. Ethics and Legal Compliance**
- 4. Storage and Back-up during research**
- 5. Data Sharing and Long-term Preservation**
- 6. Resources and Responsibilities**

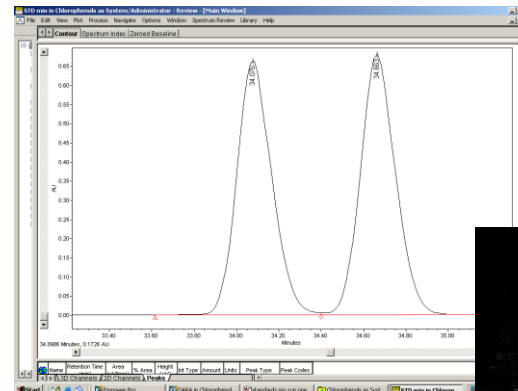
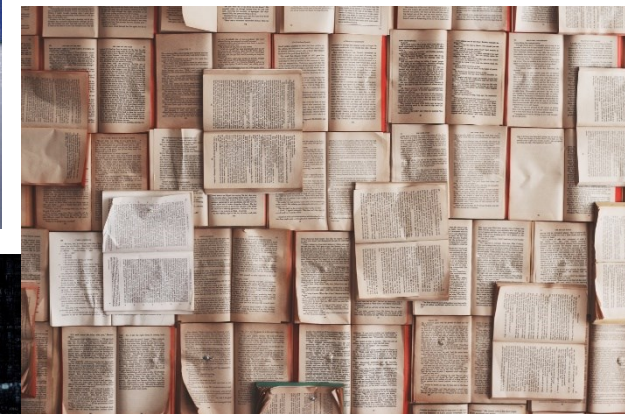
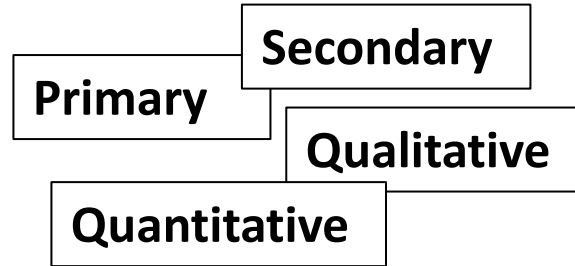
<https://doi.org/10.5281/zenodo.4915861>

[Practical Guide to the International Alignment of Research Data Management - Extended Edition - Science Europe](#)

1. Data Collection

What is the type, format and volume of data?

How will data be collected or created?



https://openi.nlm.nih.gov/detailedresult.php?img=PMC3742216_ijms-14-13763f2&req=4
<https://www.cellsignal.com/contents/resources-protocols/changes-to-recommended-western-blotting-protocols/western-variation>
<http://chem.ch.huji.ac.il/nmr/whatisnmr/whatisnmr.html>

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.

This means it can include;

- datasets
- databases
- images
- musical compositions
- guidelines
- protocols
- websites
- apps
- designs
- bibliographies
- code
- software
- presentations
- performances
- music
- podcasts
- video

.....etc. etc. Ect.

	dataset	Collection method	type	format	location and access	software required
1	Fieldtrial 1 (FT 1)	Small scale fieldtrial. Lettuce grown using 3 UV filters	Flavonoids, Plant growth parameters	Excel, images	FT1 Master File (hyperlink to file location) Only project member have access	Excell, SPSS, ImageJ
2	Field trial 2(FT2)	Arabodopsis Ecotype Trial. Using 3 UV filters	Flavonoids, Plant growthparameters, Fluorescence	Excel Images Fv/Fm	FT3 Master File (hyperlink to filelocation) Project member and 4th year project student have access	Excell, SPSS, ImageJ

Data collection and Description:

Data will be collected from participants using two methods an online form will be used to collect anonymous demographic data and qualitative data will be gathered using focus groups. Transcripts from the focus groups will be anonymised as soon as practical after collection. See Table 1 for further details.

Dataset Name	Type	Content	N	Format	Analysis Software	Associated metadata
Demographic Data	anonymous quantitative	Demographic data collected using an online form	100	Excel spreadsheet	R	Data dictionary, online form, Data SOP
Focus groups	Personal qualitative	Focus groups transcripts	35	Word doc	NVivo	Codebook, Information sheet, recruitment information, SOP for anonymisation

Documentation and Metadata and Data Quality

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

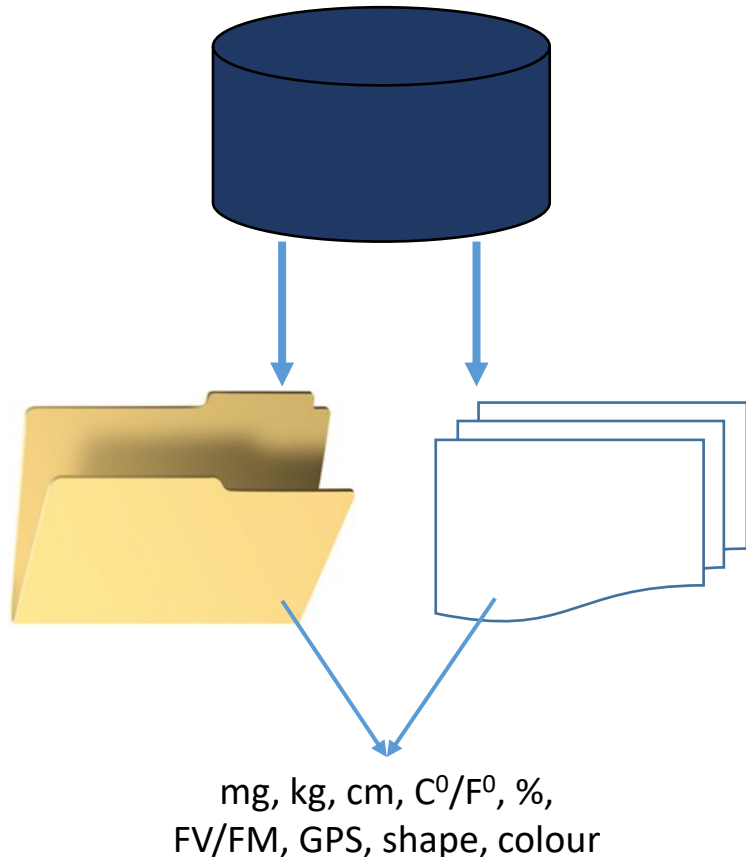
Study level documentation: provides high-level information on the research context and design, the data collection methods used, any data preparations and manipulations and summaries of findings based on the data. E.g. research design; data collection methods; structure of data files; publications etc that help to explain or draw on the data

Data-level documentation: provides information at the level of variables in a database or individual objects such as interview transcripts or pictures. E.g. variable names, labels and descriptions; units of measurement for variables; how do these variables relate to each other

Discovery metadata: can describe the content, context and provenance of datasets in a standardised and structured manner, typically describing the purpose, origin, temporal characteristics, geographic location, authorship, access conditions and terms of use of a dataset.

Creating a ReadMe style metadata (provides information about a data file): [Guide to writing "readme" style metadata | Research Data Management Service Group \(cornell.edu\)](#)

2. Documentation and Metadata and Data Quality



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?

Data



Metadata

Sat 23 Oct 2021 • 11:31



Add description...

Location

[Open in Maps](#)



Cork

Estimated location – [Learn more](#)



Details



20211023_113126.jpg

7.9MP 1960 x 4032 2.4 MB



samsung SM-J415FN

f/1.9 1/279 3.60 mm ISO40


August 2, 2021

Dataset

Open Access

Sampling-event dataset of short-term monitoring in Poblacion and Kadurong Reefs in Liloan, Cebu, Philippines



 Edullantes, Brisneve; Maglangit, Fleurdeliz; Ortiz, Angelito; Casibo, Joana Mie; Vicentuan, Lorraine Louise; Bensig, Eukene

	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	month	day	eventDate	habitat	eventRemarks	locationID	country	countryCo	stateProvi	county	locality	decimalLo	decimalLa	geodeticD	coordinate	georeferen	georeferen
2		3	2015-03	Shallow coral reef	sampled during	S01	Philippine:PH		Cebu	Liloan	Poblacion	124.005	10.40206	WGS84	50	Brisneve E	Hand-held
3		3	2015-03	Shallow coral reef	sampled during	S02	Philippine:PH		Cebu	Liloan	Poblacion	124.0083	10.40275	WGS84	50	Brisneve E	Hand-held
4		3	2015-03	Shallow coral reef	sampled during	S03	Philippine:PH		Cebu	Liloan	Poblacion	124.0055	10.40644	WGS84	50	Brisneve E	Hand-held
5		3	2015-03	Shallow coral reef	sampled during	S04	Philippine:PH		Cebu	Liloan	Poblacion	124.0048	10.40189	WGS84	50	Brisneve E	Hand-held
6		3	2015-03	Shallow coral reef	sampled during	S05	Philippine:PH		Cebu	Liloan	Between k	124.0082	10.39764	WGS84	50	Brisneve E	Hand-held
7		3	2015-03	Shallow coral reef	sampled during	S06	Philippine:PH		Cebu	Liloan	Between k	124.0064	10.39714	WGS84	50	Brisneve E	Hand-held
8		3	2015-03	Shallow coral reef	sampled during	S07	Philippine:PH		Cebu	Liloan	Kadurong	124.0094	10.39264	WGS84	50	Brisneve E	Hand-held
9		3	2015-03	Shallow coral reef	sampled during	S08	Philippine:PH		Cebu	Liloan	Kadurong	124.0104	10.39492	WGS84	50	Brisneve E	Hand-held
0		3	2015-03	Shallow coral reef	sampled during	S09	Philippine:PH		Cebu	Liloan	Kadurong	124.0092	10.39544	WGS84	50	Brisneve E	Hand-held
1		3	2015-03	Shallow coral reef	sampled during	S10	Philippine:PH		Cebu	Liloan	Kadurong	124.0078	10.39546	WGS84	50	Brisneve E	Hand-held

<https://dwc.tdwg.org/terms/>

<https://doi.org/10.5281/zenodo.5154859>

Example that doesn't use a metadata standard

Albertson, Lindsey. (2022). Influence of beaver mimicry restoration on habitat availability for fishes, including Arctic grayling (*Thymallus arcticus*) [Data set].

<https://doi.org/10.5061/dryad.47d7wm3fq>

<https://zenodo.org/record/5839189#.YeANoP7P2Uk>

Even without a metadata standard you can have reusable and reproducible data if you provide the right contextual information, metadata and documentation.

This can include things like;

- readme.txt files
- data dictionaries
- protocols
- analysis scripts
- codebooks
-etc. etc.

Files (98.8 kB)	
Name	Size
AllFishes.csv	1.8 kB
md5:77dc23124b970a05a24be387385bf87f ?	
Climate_data.csv	1.0 kB
md5:9c6654effdd1d5c69f7609db3f8cbc46 ?	
Habitat_Data.csv	2.1 kB
md5:95c042ef6f65ef8cc52fc4057f13087f ?	
July_temp_data.csv	84.0 kB
md5:a93842fa3c1a651ceb6bb2b6a11d30a9 ?	
README_BMS_habitat_for_fishes.txt	6.8 kB
md5:9bbf72670a817c75af29afba7c01b793 ?	
Rifle_Pool_data.csv	3.1 kB
md5:494769fe99ae80602512423608ef5d5b ?	

National Water Vole Monitoring Programme survey form

people's trust for endemism species

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.)	<input type="checkbox"/> None <input type="checkbox"/> 1-5 <input type="checkbox"/> 6-10 <input type="checkbox"/> More than 10	<input type="checkbox"/> None <input type="checkbox"/> 1-5 <input type="checkbox"/> 6-10 <input type="checkbox"/> More than 10	<input type="checkbox"/> None <input type="checkbox"/> 1-5 <input type="checkbox"/> 6-10 <input type="checkbox"/> More than 10	<input type="checkbox"/> None <input type="checkbox"/> 1-5 <input type="checkbox"/> 6-10 <input type="checkbox"/> More than 10	<input type="checkbox"/> None <input type="checkbox"/> 1-5 <input type="checkbox"/> 6-10 <input type="checkbox"/> More than 10	<input type="checkbox"/> None <input type="checkbox"/> 1-5 <input type="checkbox"/> 6-10 <input type="checkbox"/> More than 10
Sighting:	<input type="checkbox"/> No <input type="checkbox"/> Yes, if so how many	<input type="checkbox"/> No <input type="checkbox"/> Yes, if so how many	<input type="checkbox"/> No <input type="checkbox"/> Yes, if so how many	<input type="checkbox"/> No <input type="checkbox"/> Yes, if so how many	<input type="checkbox"/> No <input type="checkbox"/> Yes, if so how many	<input type="checkbox"/> No <input type="checkbox"/> Yes, if so how many

ECOLOGICAL MONITORING DATA - COMPREHENSIVE



Property _____ Transect _____ Photo #s _____ Date _____

Throw Number	AT DIRT ENTRY POINT				6-IN (15-cm) CIRCLE AROUND POINT				DESCRIBE NEAREST PERENNIAL				Examiner(s):
	What Dirt Point Hit (Must Check One)	Yes	Soil Surface (Must Check One)	Evidence of Check if Yes	What It Is (Must Check One)	Its Habitat (Check One)	Its Age (Must Check One)	Its Form (Must Check One)	Its Species (if Known)				
1													
2													
3													
4													
5													
6													
7													
8													
9													
10													
11													
12													
13													
14													
15													
16													
17													
18													
19													
20													
21													
22													
23													
24													
25													
26													
27													
28													
29													
30													
31													
32													
33													
SUBTOTAL													

Other Comments: _____



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:

What data quality control measures do you use?

- Defined SOPs for treatment of your data
- Documenting the calibration of instruments
- Standardise data capture, data entry or recording methods
- Naming conventions
- Version control
- Read only raw data files
- Data entry validation techniques (spreadsheets)
- Methods for transcription
- Peer review of data.



.....there are probably others specific to your discipline

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



Kristian Brock

Principal Biostatistician
UCB
University of Birmingham

<https://www.kristianbrock.com/post/send-me-data/>

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://twitter.com/i/status/1323619993263755264>

<https://www.youtube.com/watch?v=Ry2xjTBtNFE&t=6s>

Article

Data Organization in Spreadsheets

Karl W. Broman & Kara H. Woo

Pages 2-10 | Received 01 Jun 2017, Accepted author version posted online: 29 Sep 2017, Published online: 24 Apr 2018

Download citation

<https://doi.org/10.1080/00031305.2017.1375989>

Check for updates

<https://doi.org/10.1080/00031305.2017.1375989>

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?

[UCC IP Policy](#)

[UCC Research Data Management Policy](#)

[UCC Code of Research Conduct](#)

The DMP should outline how you are going to ensure compliance with any legal or ethical requirements associated with your data.

Do you have a data sharing agreement?

Are you using secondary data?

Who 'owns' your data?

Origin and provenance

Intellectual Property Copyright

4. Storage and backup [DURING]

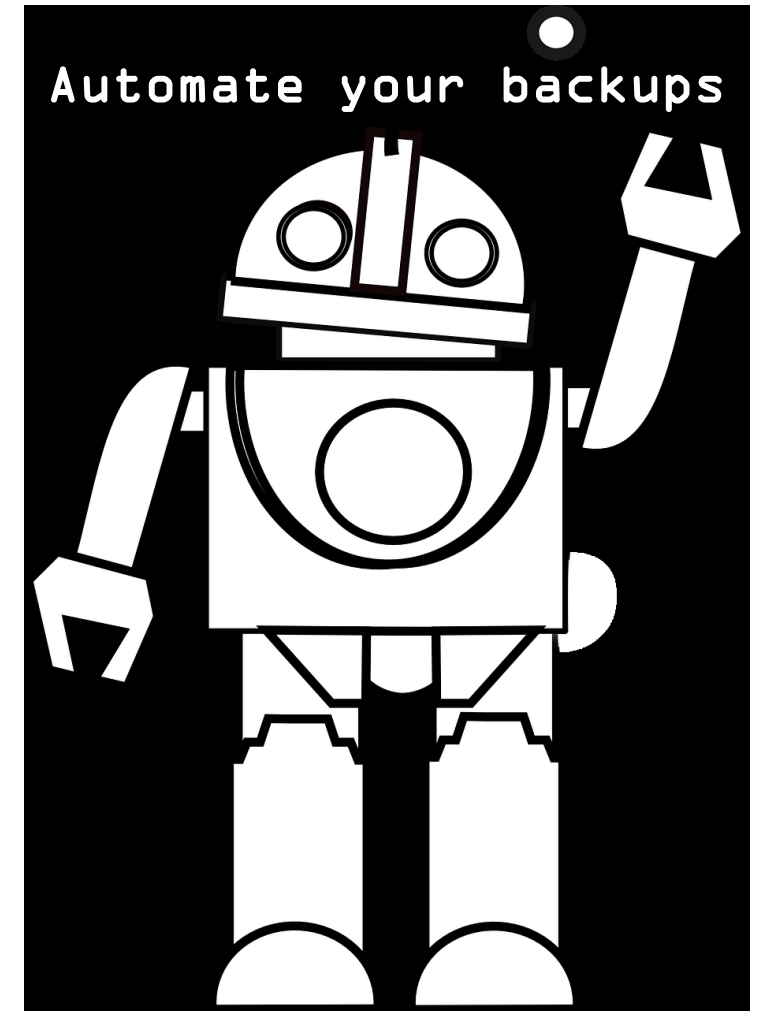
How will data be stored and backed up during the research?

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

How will you manage access and security?

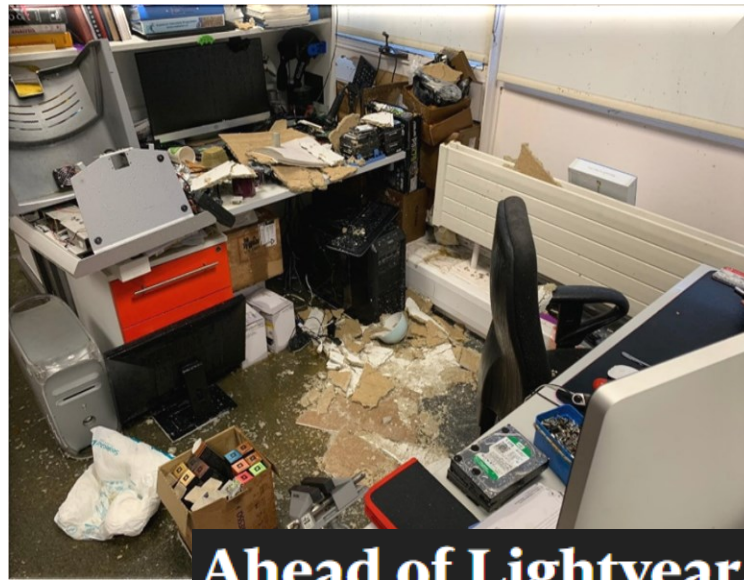
[Research Box](#)

[OneDrive Teams and SharePoint](#)



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

APC- Burst Radiator



Ahead of Lightyear, fans are remembering when Toy Story 2 got deleted – twice

'Backups MATTER, people. Do them. Test them,' warned one commenter



Chris Albon ✓
@chrisalbon

Someone at Pixar deleted all of Toy Story 2 and the backup hadn't worked for a month, and the only reason we saw that movie was b/c someone on maternity leave had a copy of it on her home computer.

Her name is Galyn Susman and she is now the producer for the new Lightyear movie!

8:14 PM · Feb 15, 2022 · Twitter Web App

https://www.independent.co.uk/arts-entertainment/films/news/lightyear-toy-story-deleted-galyn-susman-b2016221.html?utm_content=Echobox&utm_medium=Social&utm_campaign=Main&utm_source=Twitter#Echobox=164500483

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).

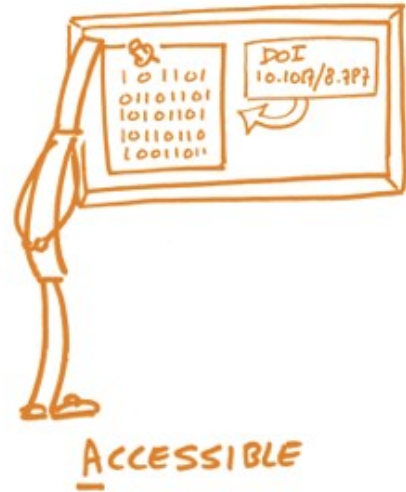
He will get a replacement hard drive from Dell on Thursday, but in the meantime needs to recover the data from the damaged one.

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

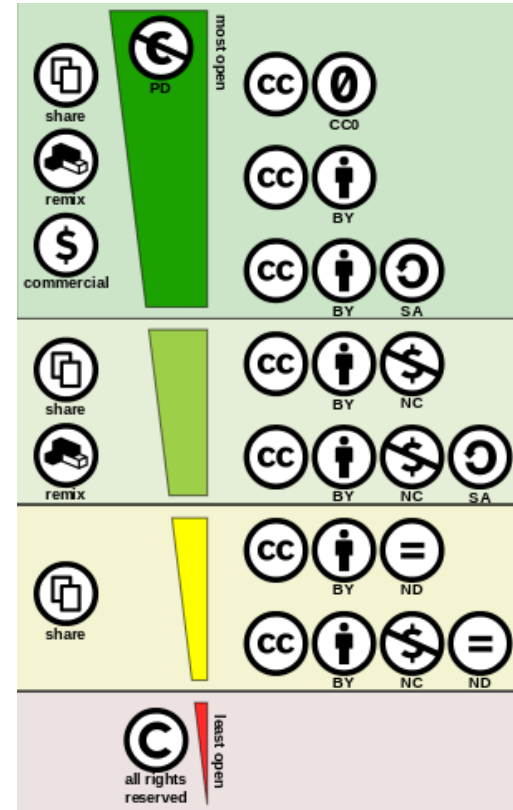
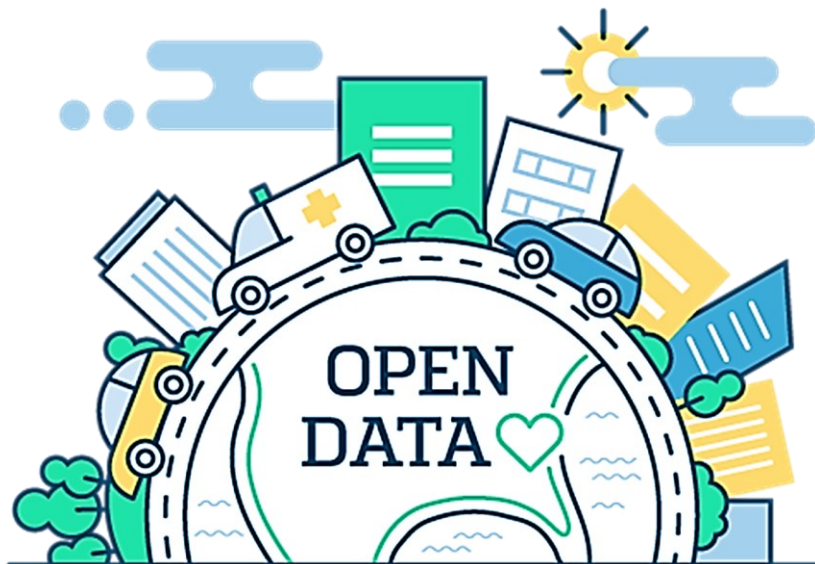
Thanks in advance

5. Data Sharing and Long-Term Preservation

FAIR DATA PRINCIPLES



“.....collect once use many times”



<https://creativecommons.org/>

<https://libguides.ucc.ie/researchdataservice/fair>

<https://www.nature.com/articles/sdata201618>

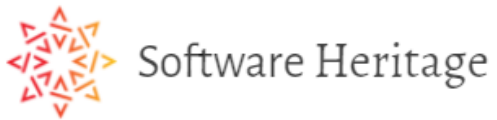
Leveraging repositories to FAIRify your data



<https://dataverse.harvard.edu/>



<https://osf.io/>



<https://www.softwareheritage.org/>



Irish Qualitative Data Archive

<https://www.maynoothuniversity.ie/iqda>



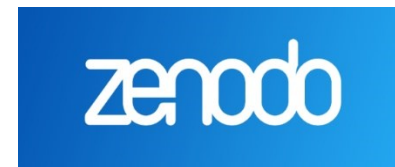
<https://www.dri.ie/>



<https://www.ucd.ie/issda/>



<https://www.scienceeurope.org/our-resources/practical-guide-to-the-international-alignment-of-research-data-management/>



<https://zenodo.org/>



<https://www.ebi.ac.uk/>



<https://fairsharing.org/>



<https://figshare.com/>



<https://www.ncbi.nlm.nih.gov/>

If data cannot be made available, why?

Reasons why you can't share your data

1. Copyright
2. Secondary data
3. Sensitivity -Commercial (embargo)
-Re-identification of anonymised data
4. No consent
5. Endangered species locations/habitat

...but even in projects where the primary raw data cannot be shared there nearly always other outputs which can be shared.

1. Metadata
2. Aggregate data
3. Analysis code
4. SOPs

6. Responsibilities and Resources

Who will be responsible of data management?

What resources will you require to deliver your plan?



Research Data Service

- Canvas Research Essentials [Enroll](#)
- Website [Research Data Service](#)
- One to one appointment booking [Book Now](#)
- Data Management Plan Review
- Advice on active data management
- Advice on FAIR principles, data sharing and long-term preservation
- Scheduled and tailored training in data management and FAIR
- Digital Badge in the Responsible Conduct of Research



Open Access

Being FAIR with your Data

Research Data Management

Metadata for Researchers

Open Access

OpenRefine

Tidy Data

Searching Techniques for Systematic Reviews

**Research
Essentials**



UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

Library
Leabharlann