We will be using an electronic whiteboard (Etherpad).
Please log onto your computer and access:
https://public.etherpad-mozilla.org/p/2019-2-26-cruk-ci-myrd

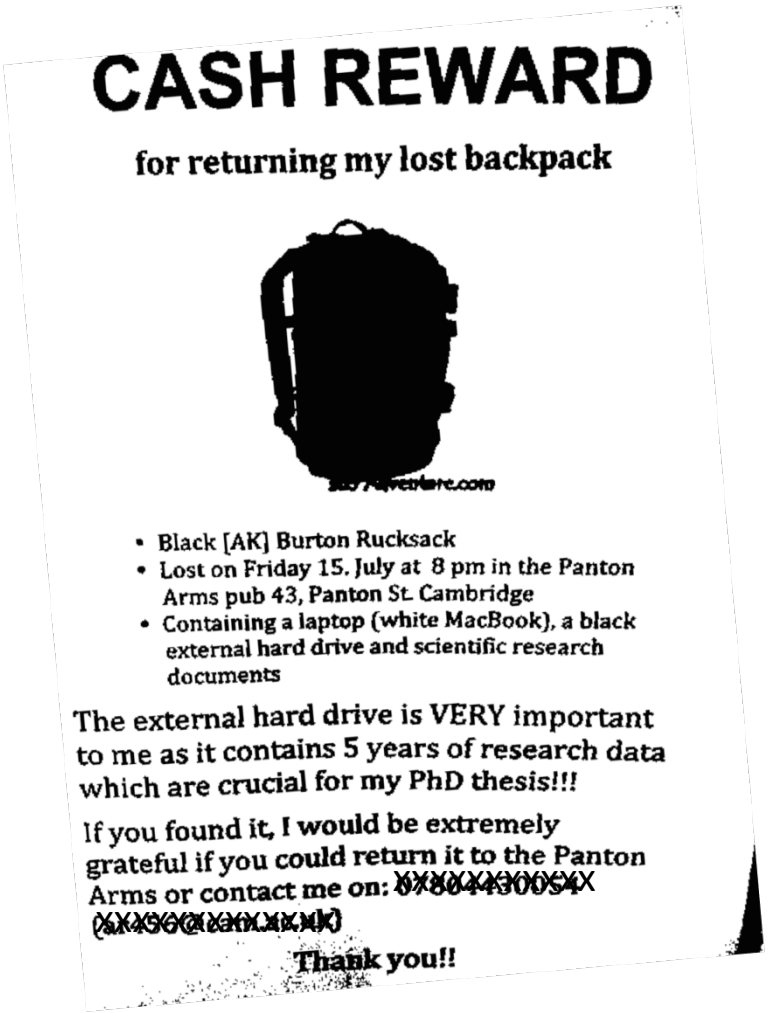
Data sharing and Backup

Anne Pajon

26th February 2019

To avoid data ***disasters*** ...

What would you do if you'd lose your data tomorrow?



CASH REWARD

for returning my lost backpack

- Black [AK] Burton Rucksack
- Lost on Friday 15. July at 8 pm in the Panton Arms pub 43, Panton St. Cambridge
- Containing a laptop (white MacBook), a black external hard drive and scientific research documents

The external hard drive is VERY important to me as it contains 5 years of research data which are crucial for my PhD thesis!!!

If you found it, I would be extremely grateful if you could return it to the Panton Arms or contact me on: 07804430054 (xxxxxxxxxxxxxx)

Thank you!!

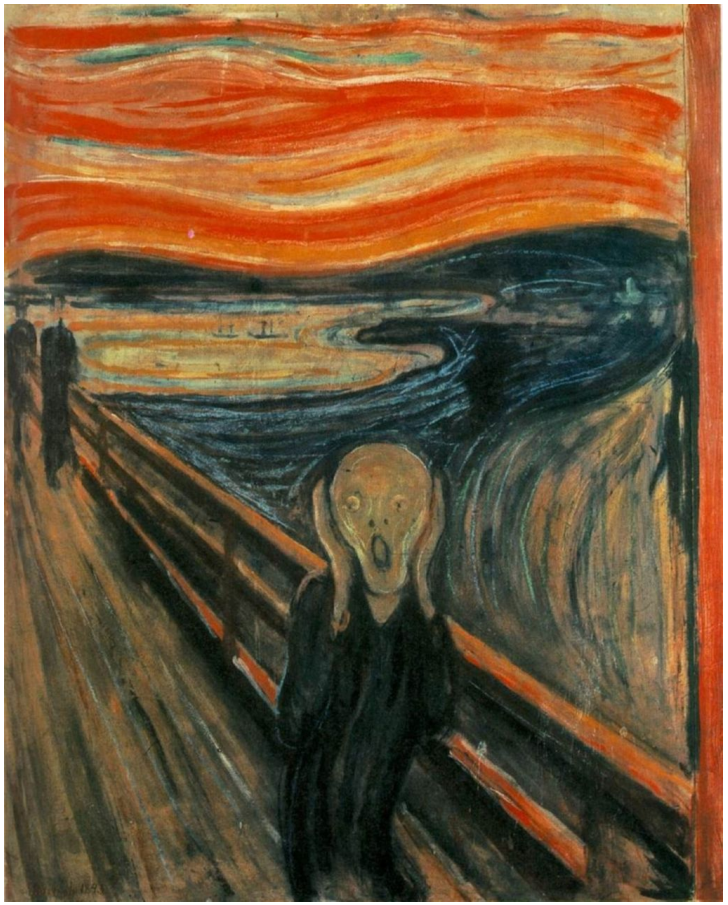
https://blogs.ch.cam.ac.uk/pmr/2011/08/01/why-you-need-a-data-management-plan/

What if?

Cancer Research UK – University of Manchester – 27 April 2017

What would you do if you'd lose your data tomorrow?

- Your laptop got stolen
- Your office/house burnt
- Your USB stick is lost
- Your portable hard disk is damaged
- Your data in Dropbox disappeared



https://en.wikipedia.org/wiki/The_Scream

Never work directly on the raw data

Leave it intact

Always **make a copy**, and work on the copy

Data backup

*At least **2** backups at **2** different locations*

External disks

Online backup

Dropbox Google Drive OneDrive

Servers

Department
College
IT

	External disks	Online backup	Servers
🙂	Cheap £10-15 / TB (1024GB)	Accessibility Free (limit)	Managed by experts
☹	Failure rate 1.5%/year	Personal data Hacking	Moving between institutions

Data backup

Manual

Copying files to relevant folders

Automated

- Install software e.g. Time machine (Mac users)
- RAID technology
- Checksums

Dropbox Google Drive OneDrive

Copying files to relevant folders

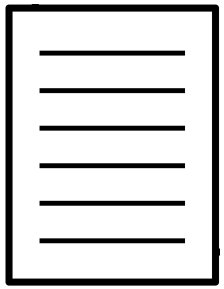
Automatically upload files to the cloud when any changes are saved

If manual … how often?

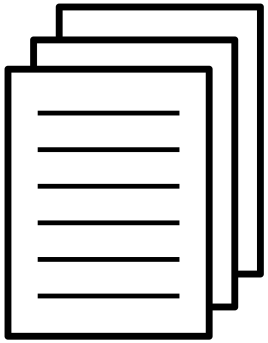
↓

How much would you be willing to lose?

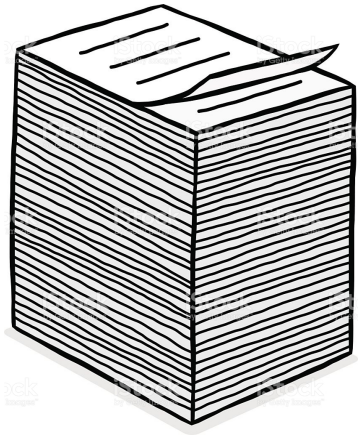
1 day



1 week



1 month-year



*Software allows you to set up **backup time** automatically*

Data backup and file sharing

	Dropbox	Google Drive	OneDrive
Space/price	2 GB (free) 1TB (£96/year)	15 GB (free) 1 TB (£80/year)	5 GB (free) 1 TB (£56)
File history and recovery	Yes, unlimited	Yes	Last 90 days
File size limit	None	5 GB	15 GB
Support	UIS	UIS	UIS
OS	Windows, Mac, Linux, Android, iOS	Windows, Mac, Android, iOS	Windows, Mac, Android, iOS
Accessibility	Sync anywhere on any devices	Live editing	Integration with Microsoft Office

More … file sharing

Email

Website

FTP

Why data sharing is important?

Data should be shared to move our knowledge forward

What data to share?

- Raw data and associated metadata
- Software and scripts
- Methods used
- Processed data
- Papers
  - results and figures

DOAJ DIRECTORY OF OPEN ACCESS JOURNALS

https://doaj.org/

- What about non-positive results?
  - RIO - Research Ideas and Outcomes

https://riojournal.com/

  - Would you consider Pre-Print for your draft manuscripts?

bioRxiv

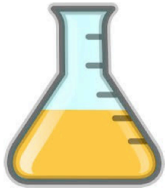
THE PREPRINT SERVER FOR BIOLOGY

https://www.biorxiv.org/

When to share?

Close data

Open data



Initial experiments

Method optimisation

Answering biological questions

Write up story

Ideas

Paper

Tests

First scripts

Robust analysis pipeline

Figures Tables

Document computational method

Private code

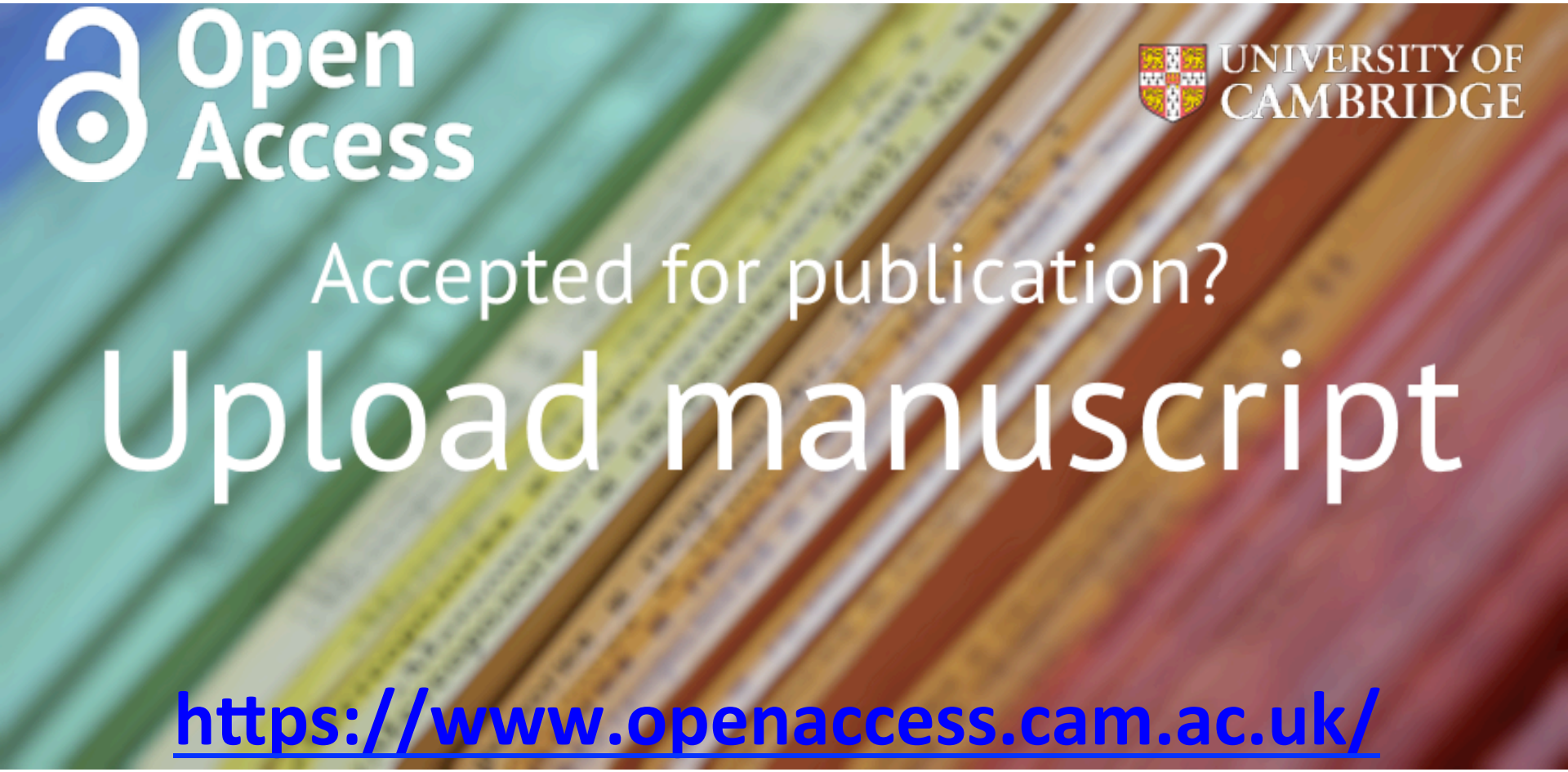
Public code

Gather information early

- Think of data submission from the start
- Collect metadata before submission
- Keep a **ReadMe** file about your project
- Consider using an electronic notebook

Making your publication Open Access

The Open Access Team will check your funder and journal policies and advise on how to comply with Open Access requirements.



How to share your data?

Store, describe and deposit your data in suitable and trusted public data repositories and add a link to your data in your publication.

Repositories for datasets

- Discipline specific
  - Registry of Research Data Repositories http://www.re3data.org/
  - EMBL-EBI services https://www.ebi.ac.uk/services
- General purpose
  - Zenodo https://zenodo.org/

Repositories for software

- GitHub https://github.com
- GitLab https://gitlab.com
- Bitbucket https://bitbucket.org

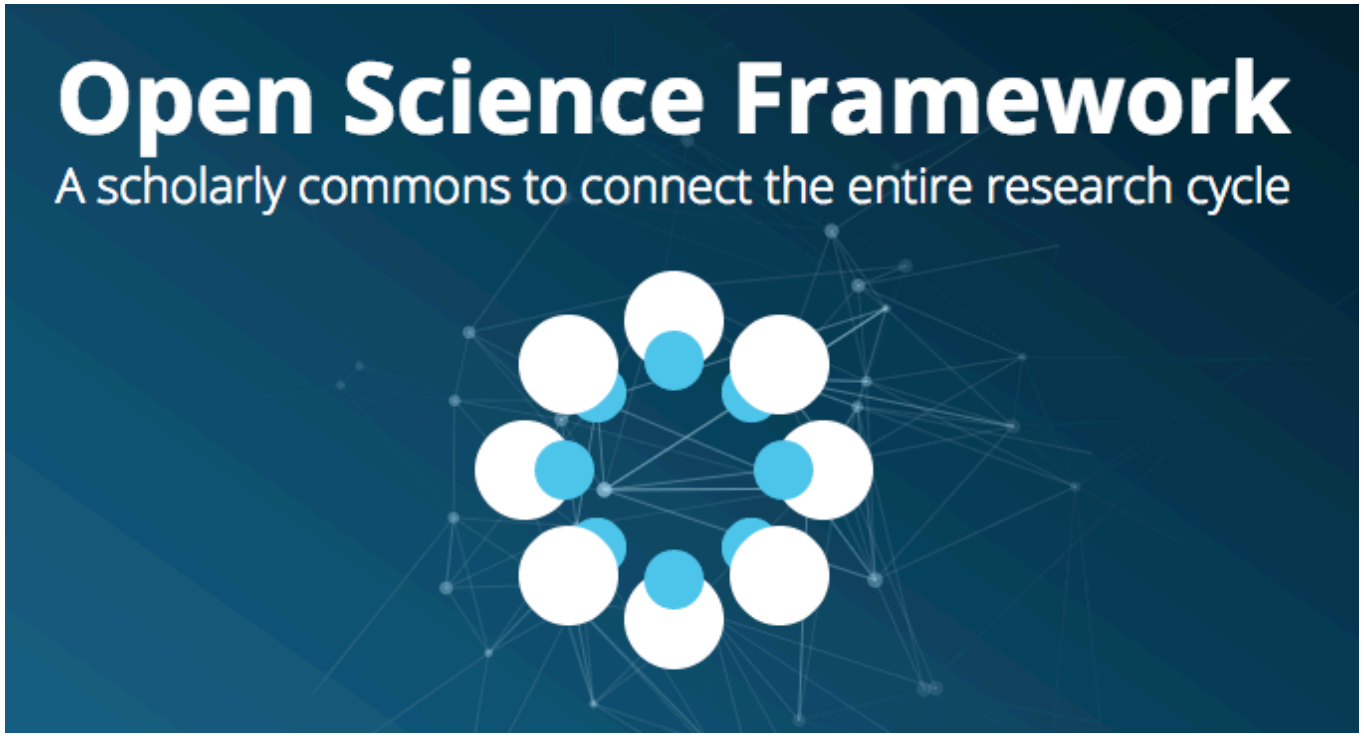
re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

zenodo

Zenodo assigns a Digital Object Identifier (DOI) to make the upload easily and uniquely citeable, with GitHub integration to enable tracking of each release.

Open Science Framework

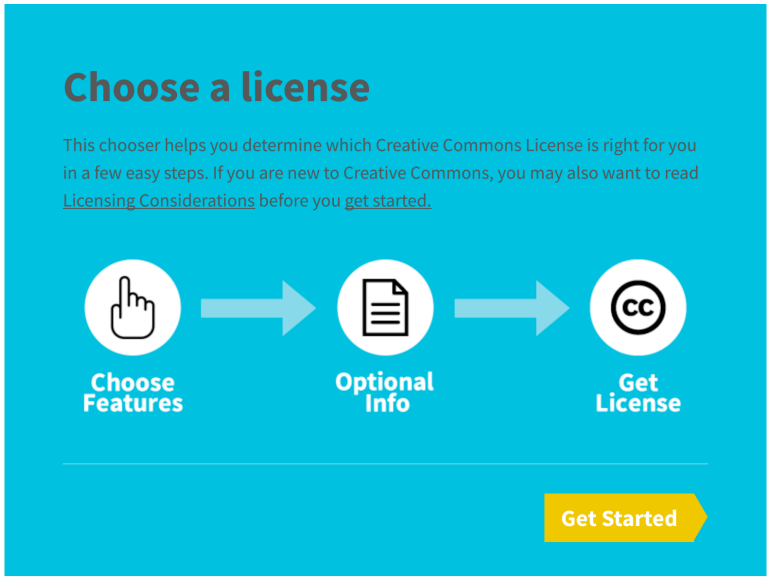
Cloud-based management for your projects @ https://osf.io/



Under which license?

Share your work with one of the Creative Commons licenses

https://creativecommons.org/



Software licenses

https://choosealicense.com

Conclusion

- Always **make a copy** of your raw data
- **Backup** your data at least **twice** at two different locations
- Document your process using a **ReadMe** file
- Ideally most data should be shared
  - Sharing is essential for all publicly funded research
  - Share as early as possible
  - Using suitable repositories and DOI
- Under Creative Commons or Open Source license