Managing your digital information

Dr Beatrice Gini, Training coordinator, Office of Scholarly Communication
bg403@cam.ac.uk

With thanks to Dr Lauren Cadwallader, Dr Rhys Morgan and Dr Sacha Jones
Session agenda

1. Types of data
2. Keeping data safe
3. Organising data
4. Sensitive data
5. Open and FAIR data
6. Data management plans
Types of data
What types of data can you think of?
Data can mean a lot…

<table>
<thead>
<tr>
<th>Documents</th>
<th>Experiment results</th>
<th>Field Notes</th>
<th>Historical documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR data</td>
<td>Images</td>
<td>Interviews (transcripts)</td>
<td>Lab books</td>
</tr>
<tr>
<td>Materials</td>
<td>Physical samples</td>
<td>Protocols</td>
<td>Social media</td>
</tr>
<tr>
<td></td>
<td>Surveys</td>
<td>Videos</td>
<td></td>
</tr>
</tbody>
</table>
Why manage data?

- Funder requirements
- Data Management Plans
- Data sharing
- Increases efficiency
- Good academic practice
- Data preservation
2 Keeping data safe
Could this be you?

****Please SHARE****Lost bag containing all my research notes on laptop and paper. I can’t complete my research masters degree without this - eighteen months worth of daily research. Left it on a train arriving at Charing Cross Saturday 8th July 08:50 ish which became the 09:06 to Sevenoaks. Black and grey with red "Swiss" features. I'm ruined if I don't get it back, please share as widely as possible! Thank you.
andrewpenson@hotmail.co.uk

Credit: Peter Murray-Rust,
These things really do happen...

Nottingham university fire destroys new multimillion-pound chemistry building

Police investigating cause of blaze in state-of-the-art centre that was due to be completed next year

Police are investigating the cause of a "significant" fire that destroyed a new multimillion-pound chemistry building at the University of Nottingham.

https://www.theguardian.com/uk-news/2014/sep/13/nottingham-university-fire-police-investigate-significant-blaze
Backup strategies

At least 2 backups, at 2 different locations

- Memory stick at the office
- Departmental server
- Cloud (cautiously)
- Hard drive at home
3 Organising data
Is this you?

Vincent Gaggioli; Jeffrey Beall
Which one is better and why?

Which one looks like your computer?

Be honest…
Data organisation should:

- Be consistent
- Be meaningful to you & your colleagues
- Allow you to find files easily
File naming conventions – do they matter?

In 3 years time would you know what these are?
An example:

TILS Document Naming Convention

Document naming for the TILS Division should follow this convention:

GDL_TILSDocNaming_V1_20090612.docx

- A prefix shows the document type
- The document title describes the content
- The version number
- The date in the format yyyymmdd

http://www.data.cam.ac.uk/files/gdl_tilsdocnaming_v1_20090612.pdf
OSC shared drive: organised!

http://www.data.cam.ac.uk/files/gdl_tilsdocnaming_v1_20090612.pdf
What makes a good file name?

ARCHIVED  This content was archived in November 2014

https://www.jisc.ac.uk/guides/managing-information/good-file-name
4 Personal and sensitive data
University Data Security Classification Levels

0  Unclassified or public information (majority)

1  Cambridge only  Only for staff and students. E.g. minutes of meetings, licensed software

2  Confidential  General personal information, can be dealt by any staff with delegated responsibility

3  Personal and strictly confidential  Highly sensitive, delegated authority is not appropriate.

Anything that identifies individuals (in aggregation). E.g. name, email address, location.

Ethnicity, political opinions, union membership, religion, genetic/biometric data, health, sexual orientation, offences

informationmanagement@uis.cam.ac.uk
General Data Protection Regulation (GDPR)

Protects EU citizens anywhere in the world
Covers any personal data
Tight rules on consent
Require privacy risk assessment
Require data breach notification
Right to be forgotten
Exemptions for research
When asking for consent, provide:

Who you *may* share the data with

Any plans to combine datasets

Retention period

A link to the University’s privacy notice for research subjects

Other relevant information (e.g. risks, withdrawal procedures, contact details)

Legal basis
Legal basis for collection

Normally “necessary for the performance of a task carried out in the public interest”

**Not** normally consent

Sensitive data only collected where proportionate to the risk (ethical approval usually required)

Sensitive data requires additional legal basis “necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes”
Think before you collect!

What may **identify** people?

What data do you **really need**?

Get informed consent for **anything** you may need

Anonymise **ASAP**

Is pseudonymised, keep **key** secure

https://www.research-integrity.admin.cam.ac.uk/
Sharing with colleagues

Do they really need the data?
Where will they store it?
Do you have consent?
Share anonymised data if possible
Use secure routes
Responsible file sharing

- Encrypted email
- File Transfer Protocol
- Encrypted hard drive
- Data Storage Services
- UoC solutions
- Email
- Private cloud solutions

Icons: Dropbox for Business, Google Drive, OneDrive for Business
Why the private cloud is not secure

Google services Terms of Use:
“…By submitting, posting or displaying the content you give Google a perpetual, irrevocable, worldwide, royalty-free, and non-exclusive license to reproduce, adapt, modify, translate, publish, publicly perform, publicly display and distribute any Content which you submit, post or display on or through, the Services. This license is for the sole purpose of enabling Google to display, distribute and promote the Services … You agree that this license includes a right for Google to make such Content available to other companies, organizations or individuals with whom Google has relationships for the provision of syndicated services”

5. Open and FAIR data
Publications without supporting data (and code!) are just claims.
Benefits of sharing data

- Improved reproducibility
- New findings from old data
- Releasing social/commercial value
- Educational resource
- Wider collaboration
- More visibility and citations
- Better credibility
- Complies with funders requirements
Funders policy on data

“Publicly funded research data are a public good (...), which should be made openly available with as few restrictions as possible”

www.rcuk.ac.uk/research/datapolicy
Datasets available ‘on request’ are not available

The Availability of Research Data Declines Rapidly with Article Age

Data availability decreases by 17% per year
Chance of email address working decreases by 7% per year
Do you want to keep getting requests?
Where to store data?
Uploading on Apollo

www.data.cam.ac.uk

We will check the data, upload it into the repository and send you a link to it
FAIR data

Findable Accessible Interoperable Reusable

Image credit: Sungya Pundir, Wikimedia Commons CC BY-SA 4.0
Findable

Accessible

Interoperable

Reusable

DOI + Right repository + Good metadata = Found by humans and machines
README files should contain:

- **General information** – title, authors, date of collection
- **File overview** - short description of the data each file contains and date it was created
- **Sharing information** – licence or restrictions placed on the data
- **Methodological information** – description of methods for data collection or generation, description of methods used for data processing
- **Data specific information** – variable list (including definitions) for tabular data, units of measurement, definitions for codes or symbols used to record missing data
Findable

**Accessible**

Interoperable

Reusable

Metadata accessible

Open download

“Request a copy”
Findable
Accessible
Interoperable
Reusable

Format
Vocabularies
Good integration
Data management plans
What is a DMP?

DMP = Data Management Plan

Plan outlining:
- Data to be collected/used during a project
- How it will be managed
- Covers the whole project … and beyond

Usually 2/3 pages in length
DMP includes:

Sources of data
Data input methods
Back up strategies
Plans for data sharing

Ethical/legal restrictions
Who is responsible for the data
Support from the Research Data team

www.data.cam.ac.uk/DMPsupport

www.data.cam.ac.uk/funders
CREATE
Now review each other’s plan

Have they described their data well?

Have they thought about back up and storage?

Have they thought about copyright and permissions issues?

Have they considered funder requirements about Open Access or data sharing?
Need more information?

www.osc.cam.ac.uk

data.cam.ac.uk

info@data.cam.ac.uk

@CamOpenData