Teaching the best data management practices to PhD students for long-term preservation

18th International Conference On Digital Preservation

Dr Ishwar Kapoor, Research Data Officer, The Library
13 September 2022
Who am I?

- The University’s Research Data Officer at the Library

- Provide support to university members in planning, managing and preserving their research data in the light of the University and funding body policies and to advise on all aspects of open research data (including its reuse)

- Before the above role, PhD and researcher in Engineering at WMG specialised in lightweight automotive

The Library, photo taken in February 2018 @ishwarkapoor
Research Data Lifecycle

Plan starts here!

- Data created/collected/observed
- Data processed
- Data analysed
- Data archived
- Data reused
- Data published

Credits: Data management plans, The Library, The University of Warwick, UK
What data will be produced?

What will be the format of data?

What will be the size of data?

How will the data be documented and described?

Are there any ‘special’ requirements for your data?

What are the plans for the long-term archiving of the digital data supporting the thesis?

What data will be produced?

What will be the format of data?

What will be the size of data?

What are the plans for data sharing and access during and after submission of the thesis?

How will your data be structured and stored?

What are the plans for the long-term archiving of the digital data supporting the thesis?

How will the data be documented and described?

What will be the format of data?

What will be the size of data?

Case Study to PhD Students
Research Data

The smallest building blocks of research, created, observed or collected for analysis to test a research hypothesis.

Types of Research data

- **Analogue** (hard copy, surveys, questionaries, lab notebooks etc.)
- **Digital** (excel spreadsheets, audiotapes, videotapes etc.)
- **Born digital**
- **Made digital**
- **Primary** (generated by the project, for example measurement reading or patient interview)
- **Secondary** (collected from other sources, for example historical records of weather pattern in between 1940 to 2000)
- **Qualitative** (image etc.)
- **Quantitative** (student’s essay marks etc.)
Metadata

• Structure information about the data

• Includes key pieces of information about the data such as:
  - Title
  - Persistent URL or Digital Object Identifier (DOI)
  - Description of data
  - Subject
  - Creator(s)
  - Funder
  - Language
  - Publication date
  - Publisher
  - Contact email address

Credits: [DataCite Metadata Schema](https://www.datacite.org)
File naming strategies - examples

Order by date:
- 2022-04-12_meeting-recording_PHY.mp3
- 2022-04-12_interview-transcript_PHY.docx
- 2021-12-15_meeting-recording_CHEM.mp3
- 2021-12-15_meeting-transcript_CHEM.docx

Order by type:
- Meeting-recording_CHEM_2021-12-15.mp3
- Meeting-recording_PHY_2022-04-12.mp3
- Meeting-transcript_CHEM_2021-12-15.docx
- Meeting-transcript_PHY_2022-04-12.docx

Order by subject:
- CHEM_meeting-recording_2021-12-15.mp3
- CHEM_meeting-transcript_2021-12-15.docx
- PHY_meeting-recording_2022-04-12.mp3
- PHY_meeting-transcript_2022-04-12.docx

Forced order with numbering:
- 01_PHY_meeting-recording_2022-04-12.mp3
- 02_PHY_meeting-transcript_2022-04-12.docx
- 03_CHEM_meeting-recording_2021-12-15.mp3
- 04_CHEM_meeting-transcript_2021-12-15.docx
Example folder structure

Main folder

- 2022
- 2021
- 2020
- 2019

Aeroplane

- Car

Train

Car

Car1

Car2

Car3

Raw data

Processed data

Analysed data

List of files
- File1.txt
- File2.mp3
- File3.docx
1. Research data files on encrypted hard disk of University laptop
3. Copy to network via University sharing platform as soon as possible after new data collected
4. Automated regular backups

Encrypted hard disk drive
Sharing research data creates secondary data for re-use

• Sources of research data include
  – Re3data.org – great for finding obscure research data
  – https://data.gov.uk/
  – General purpose repositories
    • Figshare, Zenodo, GitHub
    • UK Data Service
  – Specialist repositories
    • http://datacompass.lshtm.ac.uk/
  – Institutional repositories
    • Warwick Research Archive Portal
Feedback received from students

How would you rate the skills session you attended?

- Excellent: 60%
- Good: 40%
- Average: 5%
- Poor: 5%

Do you feel RDM&P Case Study was a useful exercise to understand your project data lifecycle?

- Yes: 80%
- No: 10%
- Maybe: 10%

Do you feel RDM&P lecture (1 hr) was useful to your project?

- Yes: 70%
- No: 10%
- Maybe: 20%

Would you like library to run similar skill-based sessions related to your project?

- Yes: 60%
- No: 10%
- Maybe: 30%
Any questions?

Dr Ishwar Kapoor
Research Data Officer
The Library
The University of Warwick, UK
ishwar.kapoor@warwick.ac.uk
http://warwick.ac.uk/lib-researchers/research-data/