Teaching the best research data management practices to PhD students

Edinburgh Open Research 2023

Dr Ishwar Kapoor, Assistant Professor, University of Warwick
16 May 2023
Who am I?

- Assistant Professor (TF) of Materials and Manufacturing Engineering at Warwick.
- Before the above role, the University’s Research Data Officer at Warwick library.

Provide support to university students 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).

Why research data management (RDM) in my current role?

- Teach and supervise UG, PGT and PGR engineering student projects where students are involved in collecting various types of data such as public, sensitive, commercial etc. for their studies.
- I teach and supervise company employees (JLR, Dyson, Airbus etc.) and guide students in ethics application for collecting and analysing sensitive and commercial data.
An hour or so lecture on RDM to 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, questionnaires, lab notebooks etc.)

- **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.)

**Digital** (excel spreadsheets, audiotapes, videotapes etc.)

- **Born digital**
- **Made digital**
Research Records?

Administrative materials and supporting documentation that are produced before, during, and after a research project. Examples include:

- Correspondence
- Ethics applications
- Technical appendices
- Research reports
- Signed consent forms
- Social media communications (blogs, wikis, tweets, etc.)
FAIR data principles

FINDABLE
Data includes rich metadata and unique identifier

ACCESSIBLE
As open as possible, as closed as necessary
Data is easy to download or usable via standard protocols

INTEROPERABLE
Use of metadata in accessible and standard language

REUSABLE
Data is well-described and licensed for reuse by others
Research Data Management (RDM)?

- Research data management (RDM) means the storage, curation, preservation and provision of continuing access to analogue and digital research data.

RDM includes activities such as:

- Creating backups of your work and controlling who has access to them.
- Choosing file formats that can be opened easily in the future.
- Describing methodology and keeping track of versions of files.
Why should I invest time in RDM?

- Data can have a longer lifespan than that of the research project that creates or collects it
- Data can be re-used by other researchers in future for different projects
- Data may also be valuable or sensitive, and so require careful handling

Saves one's time and problems, for example

- Helping you to work more efficiently and effectively
- Saving frustration during the project
- Allowing you to see the data more clearly
- Validation, Stem Cell Research Fabrication
How much data would you lose if...?

- Your laptop was stolen
- Your lab burnt down
- You lost your USB stick
- Your portable hard drive corrupted
- Your stuff on third party cloud services disappeared
Why should I invest time in RDM?

- To meet the **University’s Research Data Management Policy** requirements

- To meet the **Funder’s Research Data Management Policy** requirements
Research Data Lifecycle

Plan starts here!

Data created/collected/observed

Data reused

Data processed

Data archived

Data analysed

Data published

Credits Data management plans, The Library, The University of Warwick, UK
Documentation – describing data!

- More detailed equivalent of ‘README’ file for data
- Documentation includes following pieces of information:
  - Who has collected the data?
  - What is the type of data?
  - Why the data has been collected?
  - Description of the data
  - What methodologies were used to create the data?
  - What hardware and software were used to create the data?
  - Are there any assumptions made during data collection, processing and analysing?
  - Why are there anomalies in the data
<table>
<thead>
<tr>
<th>Order by date:</th>
<th>Order by type:</th>
</tr>
</thead>
<tbody>
<tr>
<td>2022-04-12_meeting-recording_PHY.mp3</td>
<td><strong>Meeting-recording</strong>_CHEM_2021-12-15.mp3</td>
</tr>
<tr>
<td>2022-04-12_interview-transcript_PHY.docx</td>
<td><strong>Meeting-recording</strong>_PHY_2022-04-12.mp3</td>
</tr>
<tr>
<td>2021-12-15_meeting-recording_CHEM.mp3</td>
<td><strong>Meeting-transcript</strong>_CHEM_2021-12-15.docx</td>
</tr>
<tr>
<td>2021-12-15_meeting-transcript_CHEM.docx</td>
<td><strong>Meeting-transcript</strong>_PHY_2022-04-12.docx</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Order by subject:</th>
<th>Forced order with numbering:</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHEM_meeting-recording_2021-12-15.mp3</td>
<td>01_PHY_meeting-recording_2022-04-12.mp3</td>
</tr>
<tr>
<td>CHEM_meeting-transcript_2021-12-15.docx</td>
<td>02_PHY_meeting-transcript_2022-04-12.docx</td>
</tr>
<tr>
<td>PHY_meeting-recording_2022-04-12.mp3</td>
<td>03_CHEM_meeting-recording_2021-12-15.mp3</td>
</tr>
<tr>
<td>PHY_meeting-transcript_2022-04-12.docx</td>
<td>04_CHEM_meeting-transcript_2021-12-15.docx</td>
</tr>
</tbody>
</table>
Example folder structure

Main folder

2022
2021
2020
2019

Aeroplane

Train

Car

Car1

Car2

Car3

Raw data

Processed data

Analysed data

List of files
- File1.txt
- File2.mp3
- File3.docx
Data Classification

Public  Protected  Restricted
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

File Transfers & Backup: Example scenario

Encrypted hard disk drive
Sharing data after a project completes can...

- encourage further research branching from the original project
- can lead to new collaborations
- encourages the transparency and the improvement of research practice
- can reduce the cost of further data collection
- can increase your profile as a research output in its own right
Two hours case study on RDM to students...
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?

How will your data be structured and stored?

How will the data be documented and described?

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

Case Study to PhD Students
Feedback received from students

How would you rate the skills session?
Feedback received from students

- Was RDM lecture useful to your project?
- Was RDM case study useful to understand your project lifecycle?
Feedback received from students

Would you like to have a similar skills-based session related to your project?

- yes
- maybe
- no

student response (%)
References

- Kapoor, Ishwar (2022) *FAIR principles for research data management*. In: EUTOPIA Webinar series on Research Data Management, The EUTOPIA European University ; Online, 8-10 Nov 2022, [https://wrap.warwick.ac.uk/171217/](https://wrap.warwick.ac.uk/171217/)

- Kapoor, Ishwar (2022) *Manage PhD research data according to the FAIR principles*. In: Fair Research Data [Fair Raziskovalni Podatki], Doctoral School of the University of Ljubljana, University of Ljubljana, Slovenia, 12 May 2022, [http://wrap.warwick.ac.uk/165546/](http://wrap.warwick.ac.uk/165546/)


- Managing your research data, [https://warwick.ac.uk/services/library/staff/research-data/](https://warwick.ac.uk/services/library/staff/research-data/)

- MANTRA Research Data Management Training, [https://mantra.ed.ac.uk](https://mantra.ed.ac.uk)
Any questions?

Dr Ishwar Kapoor
Assistant Professor
University of Warwick, UK

ishwar.kapoor@warwick.ac.uk

https://www.linkedin.com/in/ishwar-kapoor-phd-27b90b73/

https://twitter.com/ishwar_kapoor