Teaching the best research data management practices to PhD students

WARWICK

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)

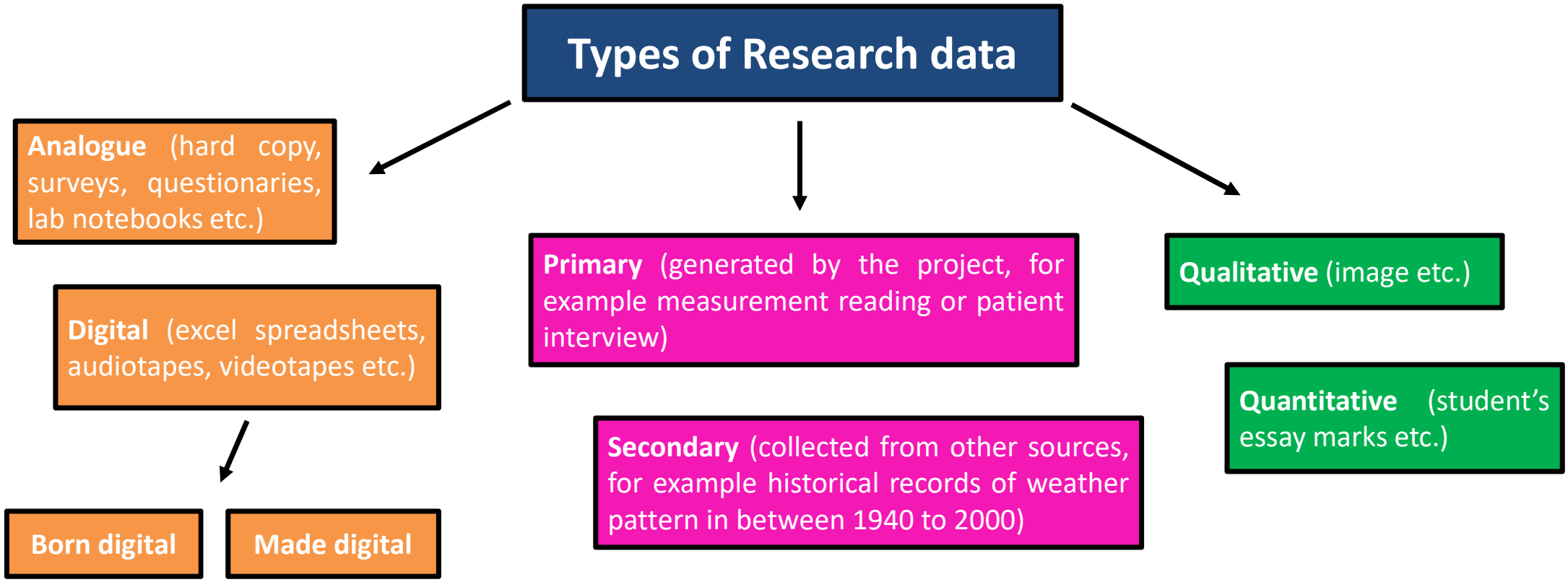
Why research data management (RDM) in my current role?

- 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



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

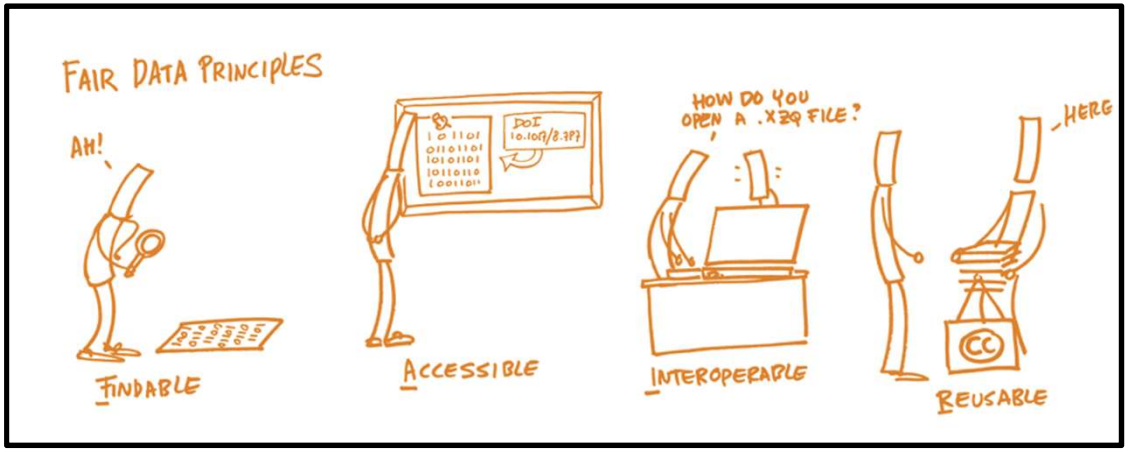
FINDABLE

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

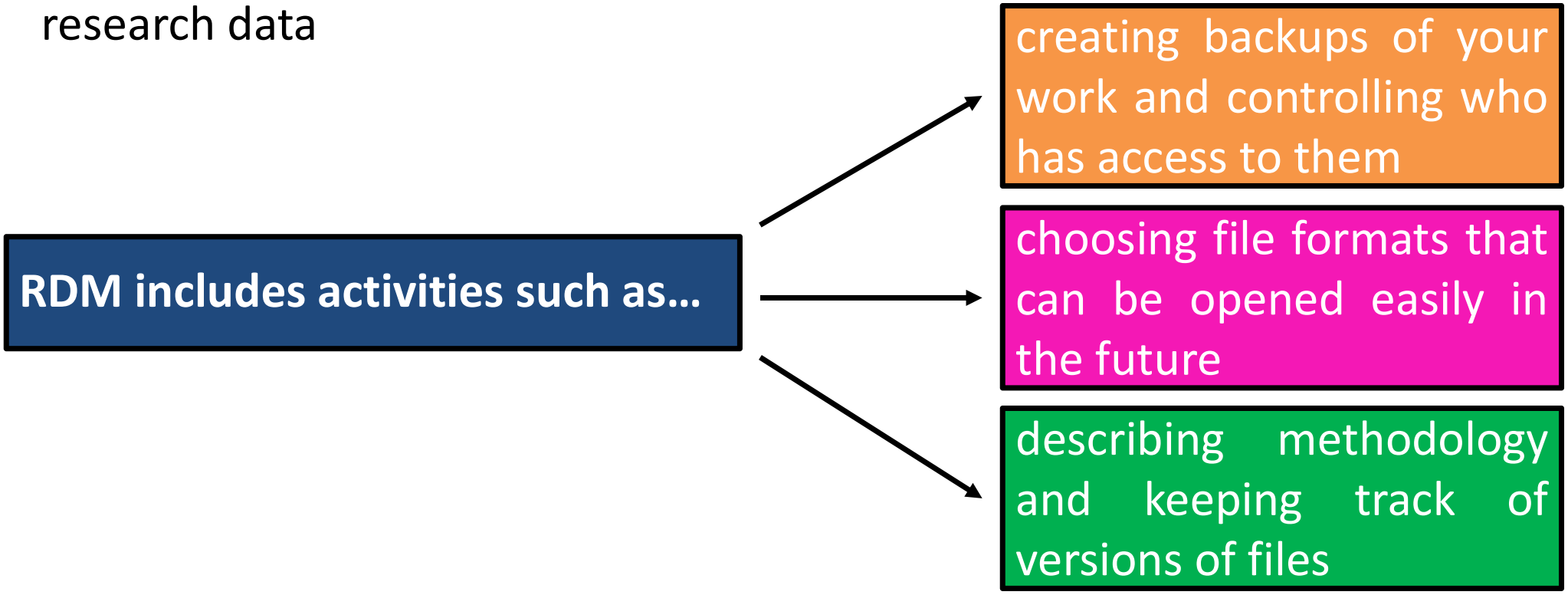
REUSABLE

Data is well-described and licensed for reuse by others

Image: https://book.fosteropenscience.eu/, CC0 1.0 license

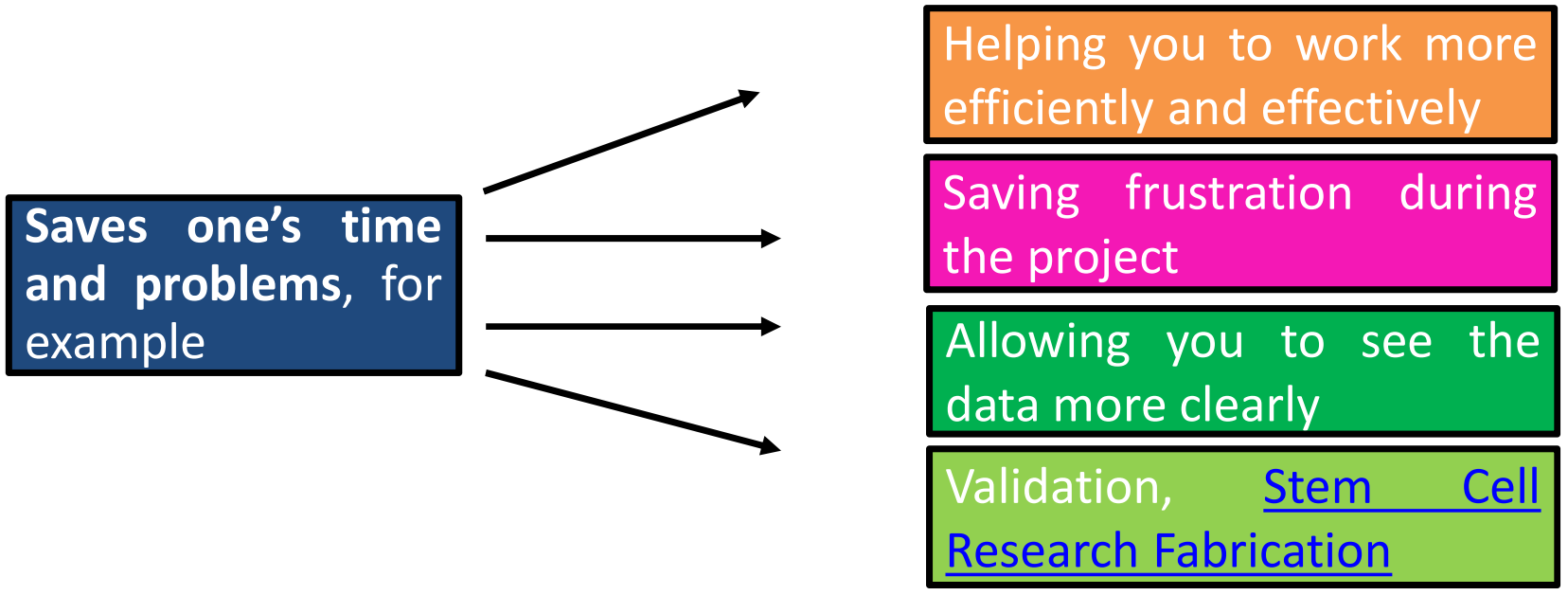
Research Data Management (RDM)?

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



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



How much data would you lose if… ?

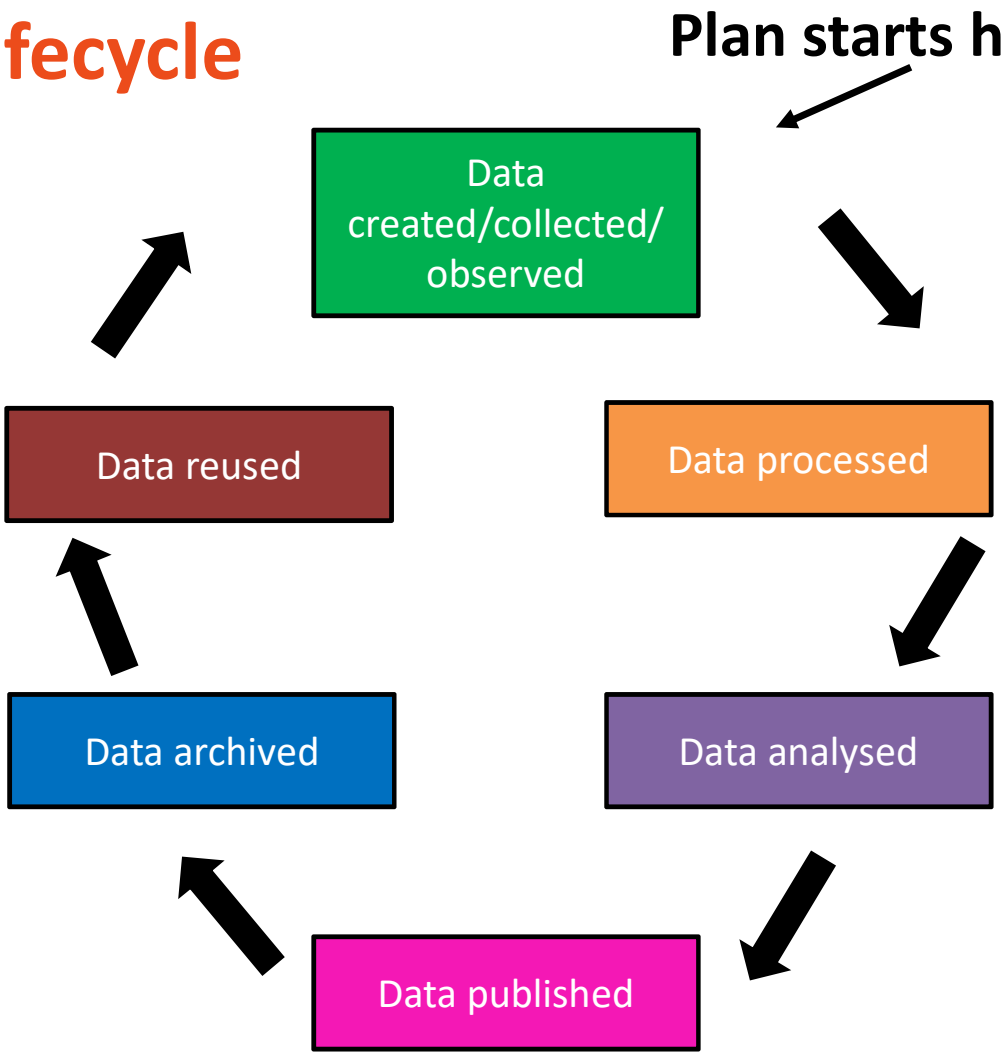
- Your laptop was stolen
- Your lab burnt down
- You lost your USB stick
- Your portable hard drive corrupted
- Your stuff on third party cloud services disappeared

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



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

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:

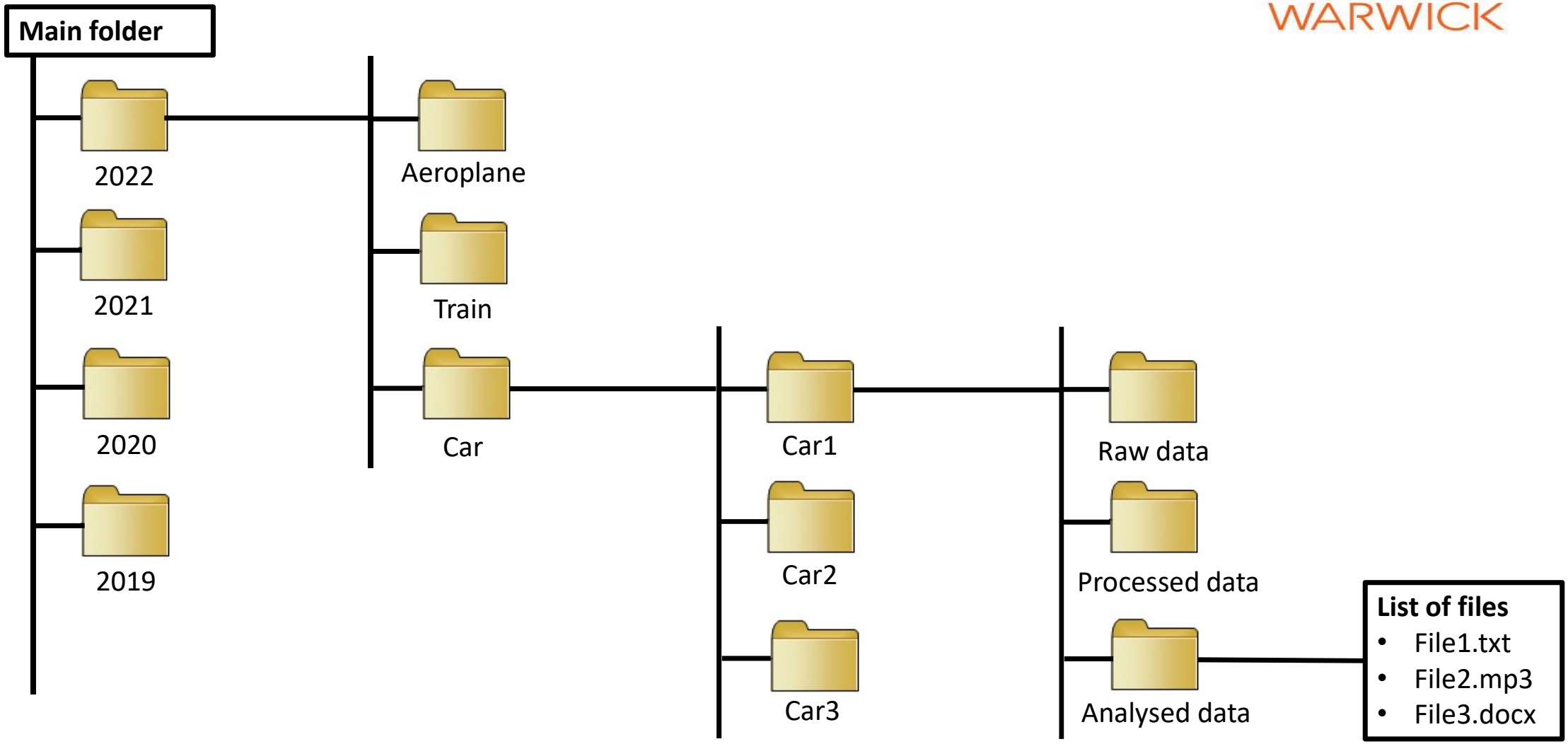
01_PHY_meeting-recording_2022-04-12.mp3

02_PHY_meeting-transcript_2022-04-12.docx

03_CHEM_meeting-recording_2021-12-15.mp3

04_CHEM_meeting-transcript_2021-12-15.docx

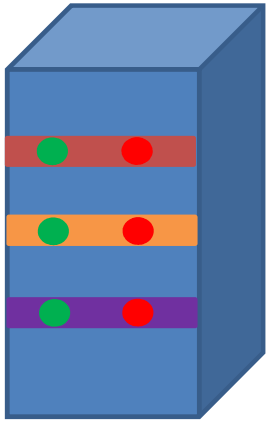
Example folder structure

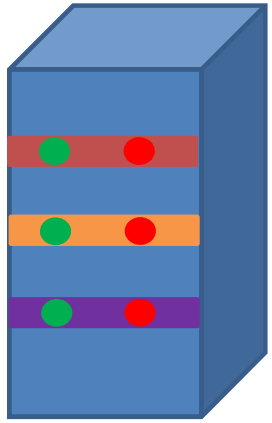


Data Classification

File Transfers & Backup: Example scenario

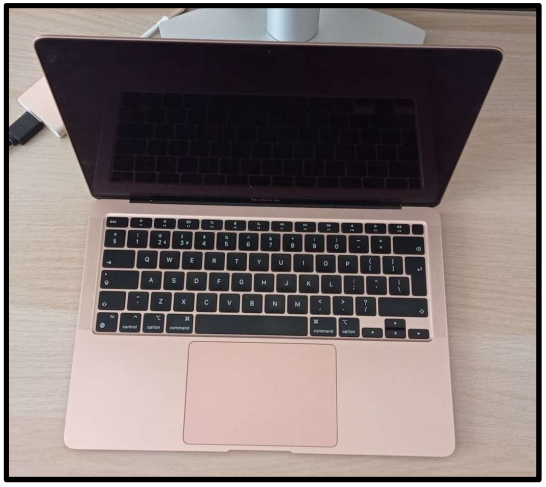
3. Copy to network via University sharing platform as soon as possible after new data collected





4. Automated regular backups

1. Research data files on encrypted hard disk of University laptop



2. Manual backup via Windows File History. Daily while collecting raw data

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…

WARWICK

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

Case Study to PhD Students

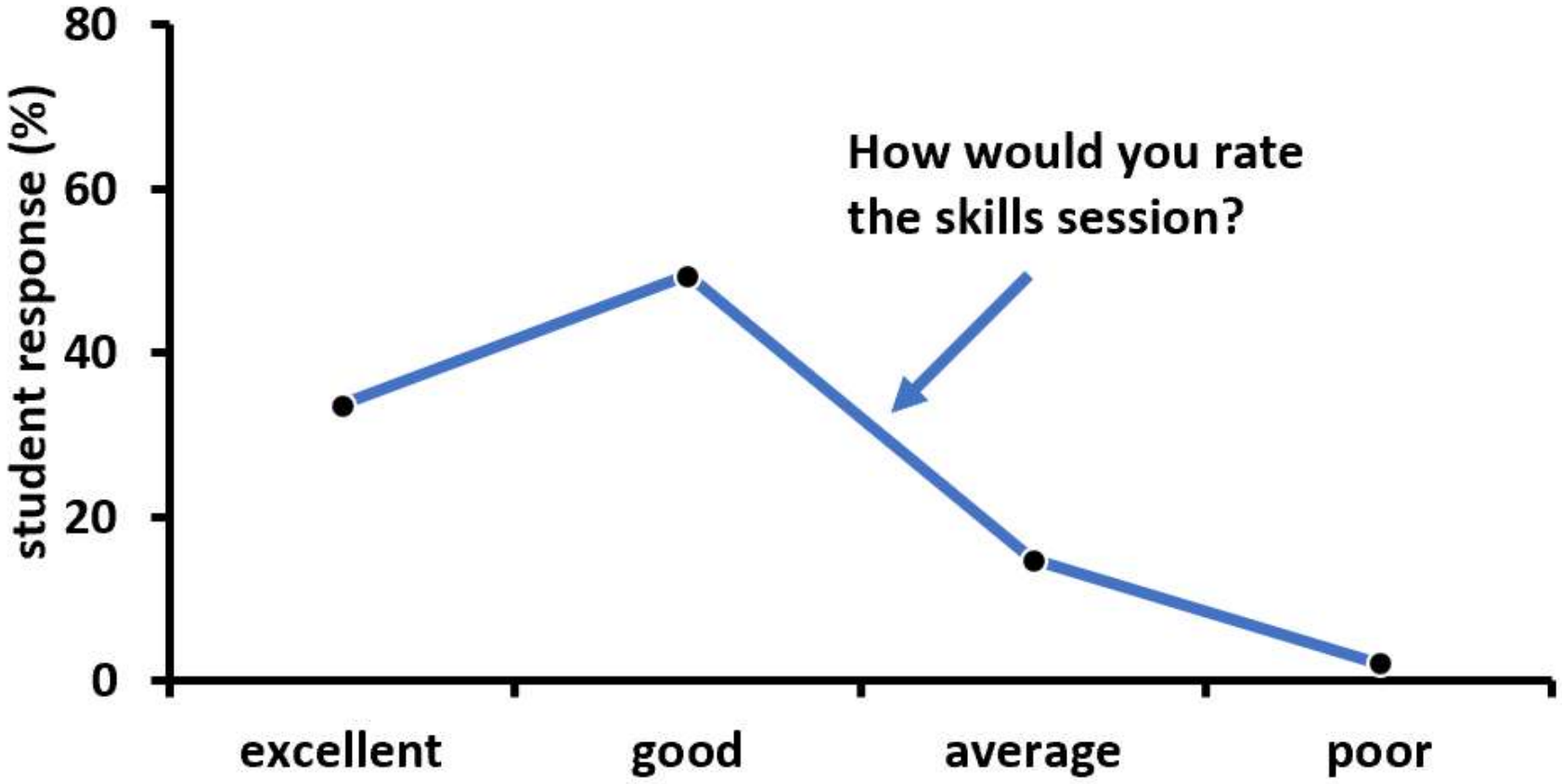
What will be the format of data?

How will your data be structured and stored?

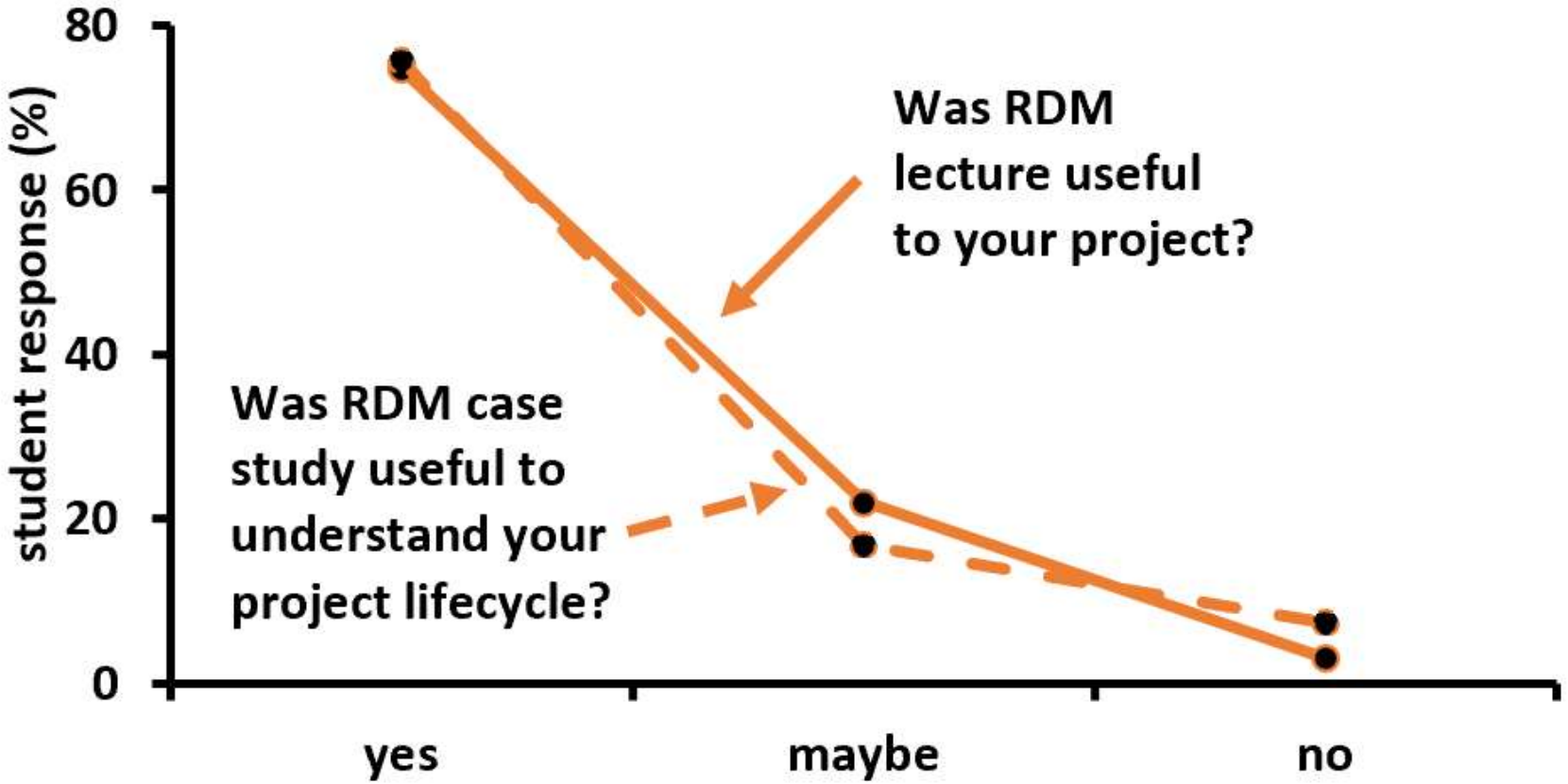
How will the data be documented and described?

What will be the size of data?

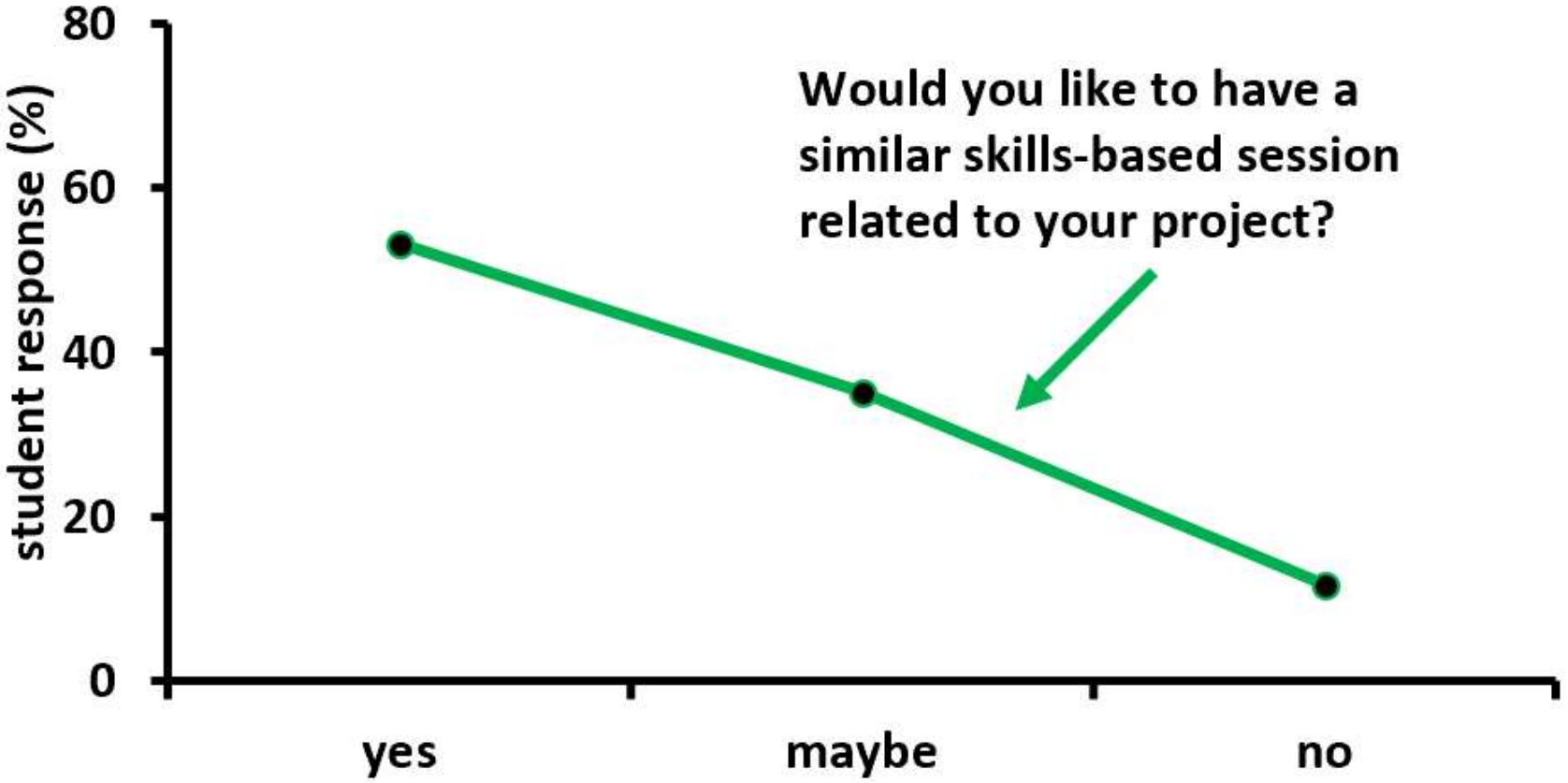
Feedback received from students



Feedback received from students



Feedback received from students



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/
- 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/
- Kapoor, Ishwar (2022) *Teaching the best data management practices to PhD students for long-term preservation.* In: International Conference on Digital Preservation (ipres2022), Glasgow, 12-16 Sep 2022, http://wrap.warwick.ac.uk/169363/
- Kapoor, Ishwar (2022) *Research Data Action Group at Warwick.* In: Redefining Engagement – The Mercian Collaboration Conference, Virtual conference, 6-7 Sep 2022, http://wrap.warwick.ac.uk/169364/
- Managing your research data, https://warwick.ac.uk/services/library/staff/research-data/
- MANTRA Research Data Management Training, 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