WARWICK

Manage PhD research data according to the FAIR principles

University of Ljubljana, Slovenia

Dr Ishwar Kapoor, Research Data Officer, The Library
12/05/2022

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
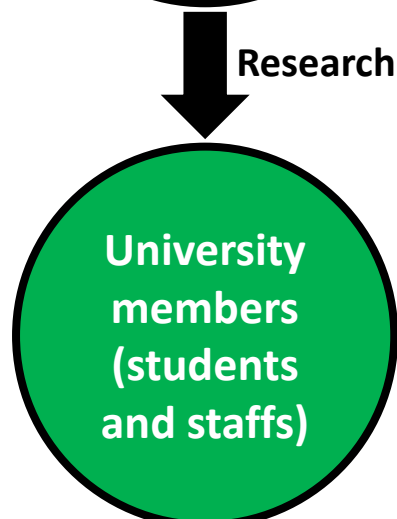
WARWICK

The Library, photo taken in February 2018 @ishwarkapoor

Outline

- The basic introduction to research data, metadata and research records

- What is Research Data Management (RDM)?

- What is Data Management Plan (DMP) and Research Data Lifecycle?

- University of Warwick RDM support to university members

Please make this session as interactive as you wish.

What is produced by Research?

- Research Data

- Metadata

- Research Records

- Data documentation

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.)

A bit more on types of Research Data…

- Documents (text, Word, PDF), spreadsheets
- Laboratory notebooks, field notebooks, diaries
- Questionnaires, transcripts, codebooks
- Audiotapes, videotapes, photographs, films
- Test responses
- Slides, artefacts, specimens, samples
- Collection of digital objects acquired and generated during the process of research (including digitised archive material)
- Models, algorithms, scripts

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

# 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

Image: https://book.fosteropenscience.eu/, CC0 1.0 license
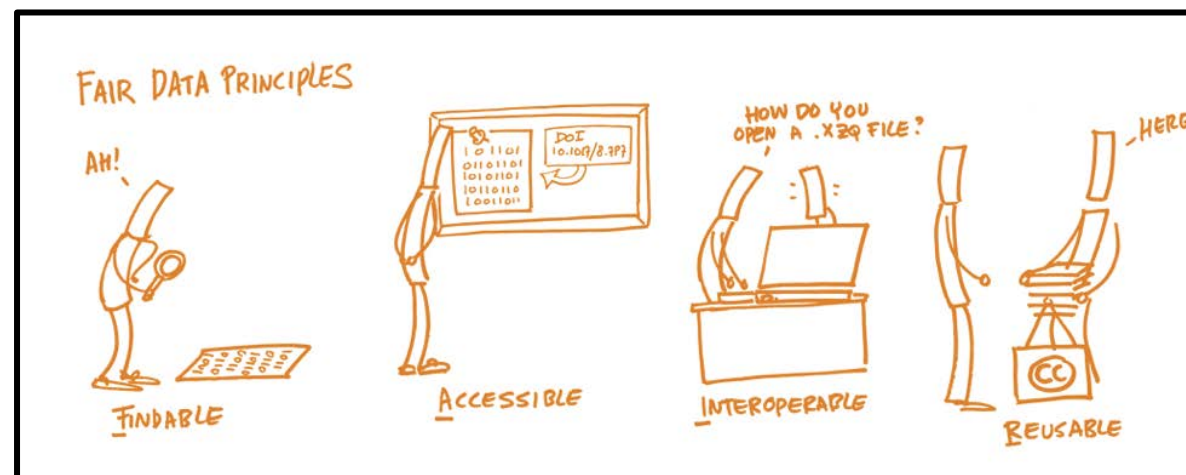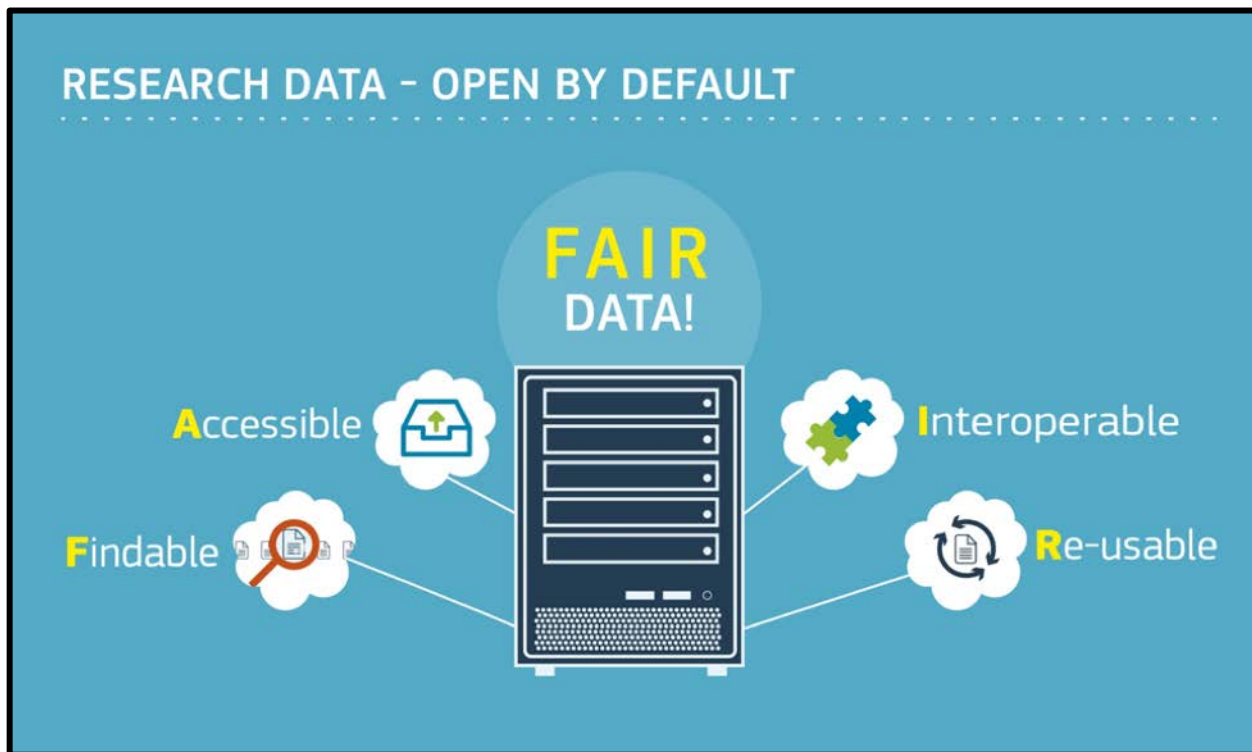
Let's understand Research Data Management (RDM)....

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
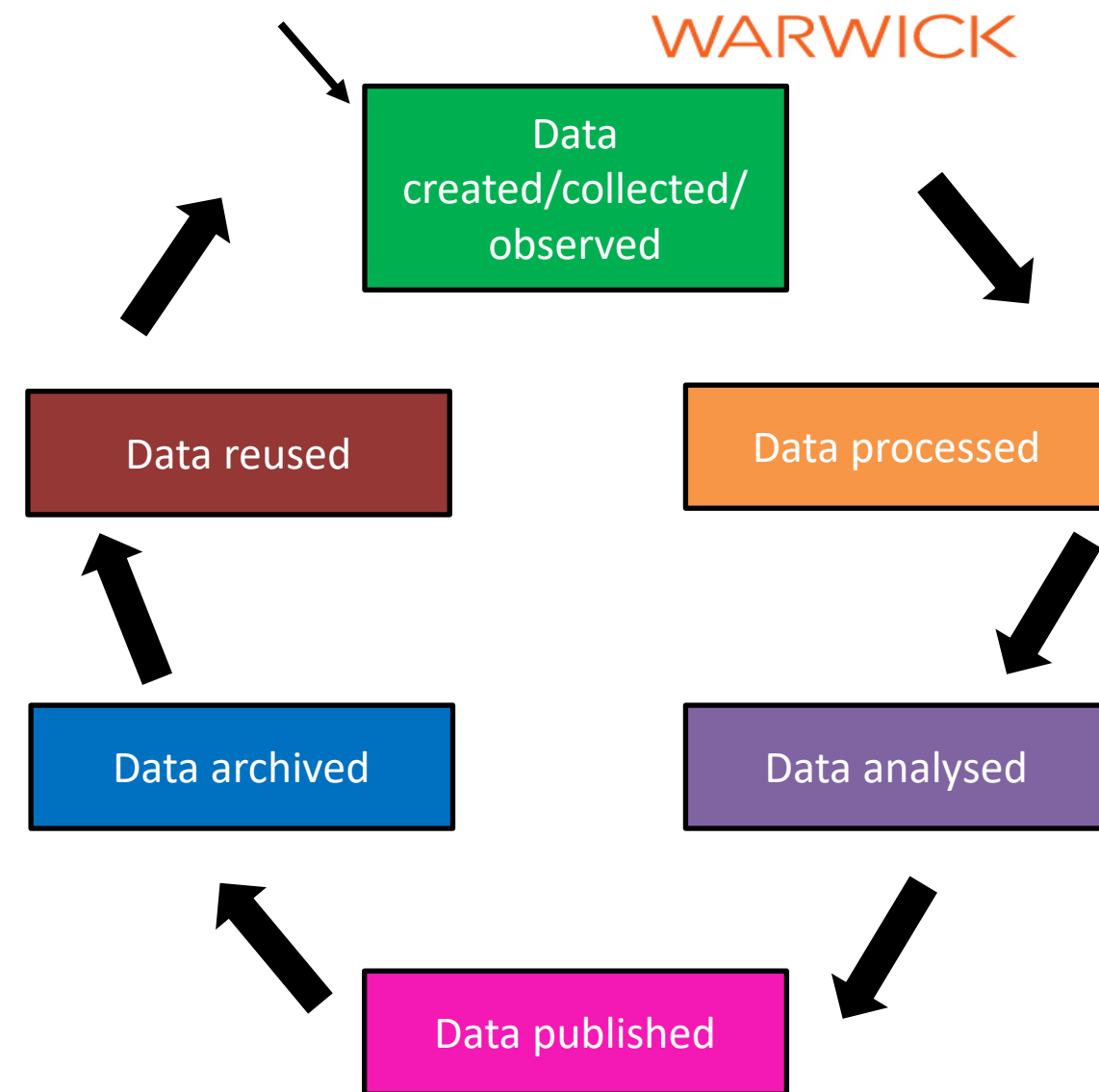
Why should I invest time in RDM?

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

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

Let's understand Research Data Lifecycle and Data Management Plan….

Research Data Lifecycle

- Where are you?
- What questions need to be thought about at each stage?
- Data creation - What data will you produce?
- Data processing and analysis - How will you look after your data once it has been created/gathered?
- Data preservation and access - Can you/others understand the data?
- Data reuse - Who owns the data? Where will the final data be stored?

Plan starts here!

Data created/collected/observed

Data processed

Data analysed

Data published

Data archived

Data reused

Data Management Plan (DMP)

- DMPs are **living document**

- Useful for checking you've considered all aspects of your data management

- Covers each aspect of the **lifecycle of your data**

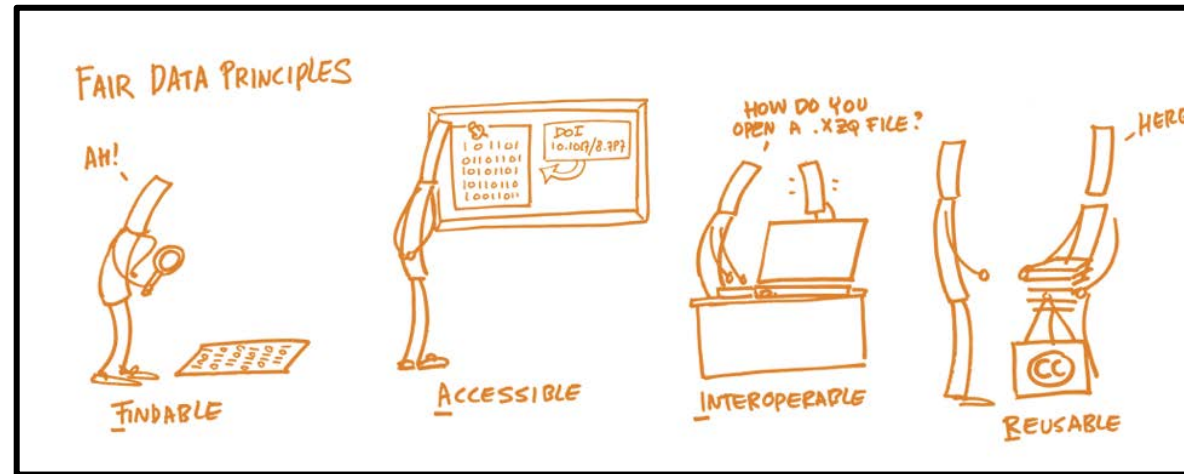- Often required by funders
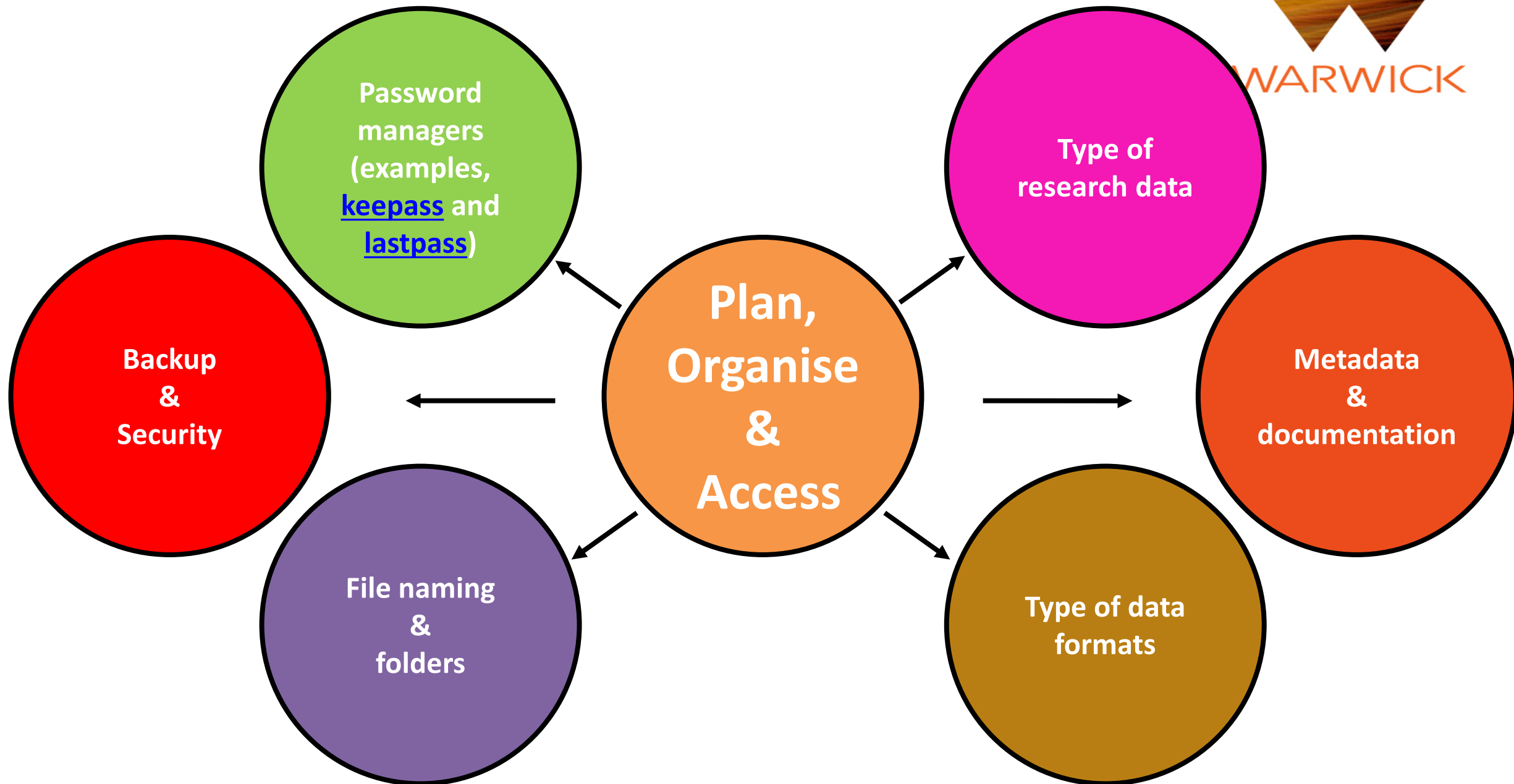
- Valuable for PGRs

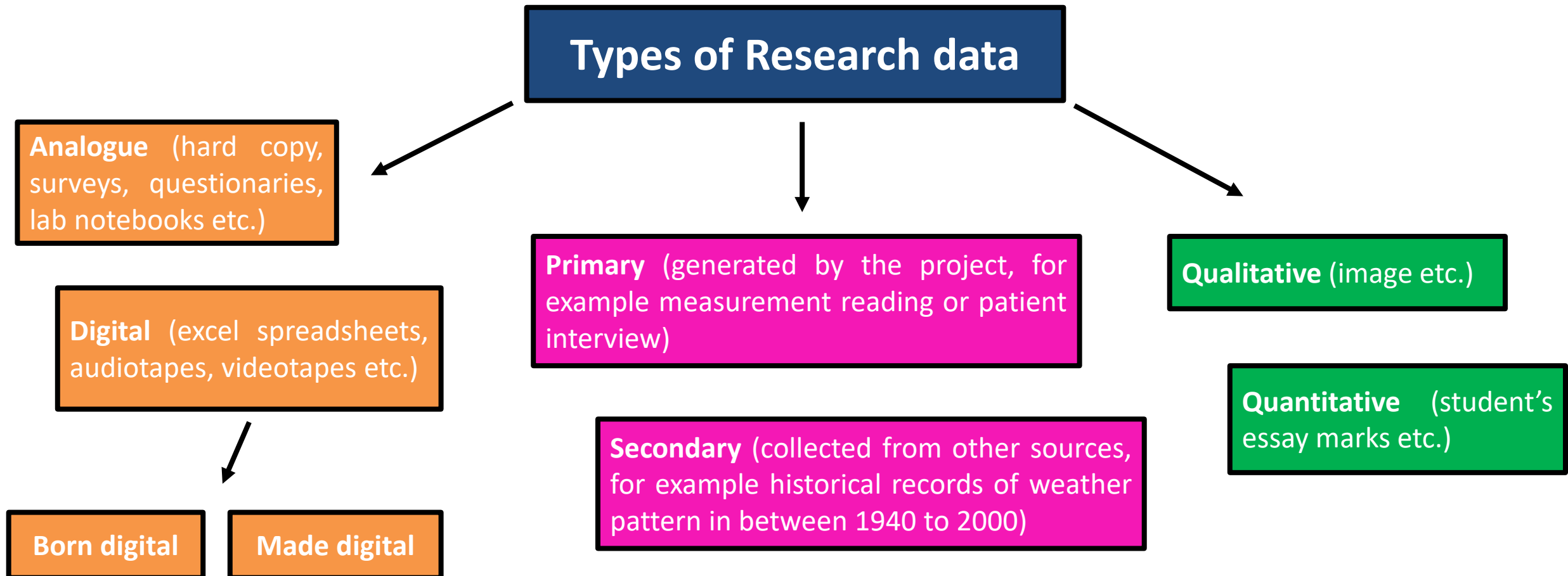Image: https://book.fosteropenscience.eu/, CC0 1.0 license

Let's understand different stages (plan, organise and access) **of Research data lifecycle and explore research support available for University members...**

What data will you produce?

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.)

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 and convention

- A good file name should be **objective**, **meaningful**, **concise** and **standardised**

- Including version information if relevant

- **BE CONSISTENT!** Pick a system and stick to it

- Think about the ordering of elements within a filename (e.g., starting YYYY-MM-DD dates allow chronological sorting)

- Advice on Warwick records management

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
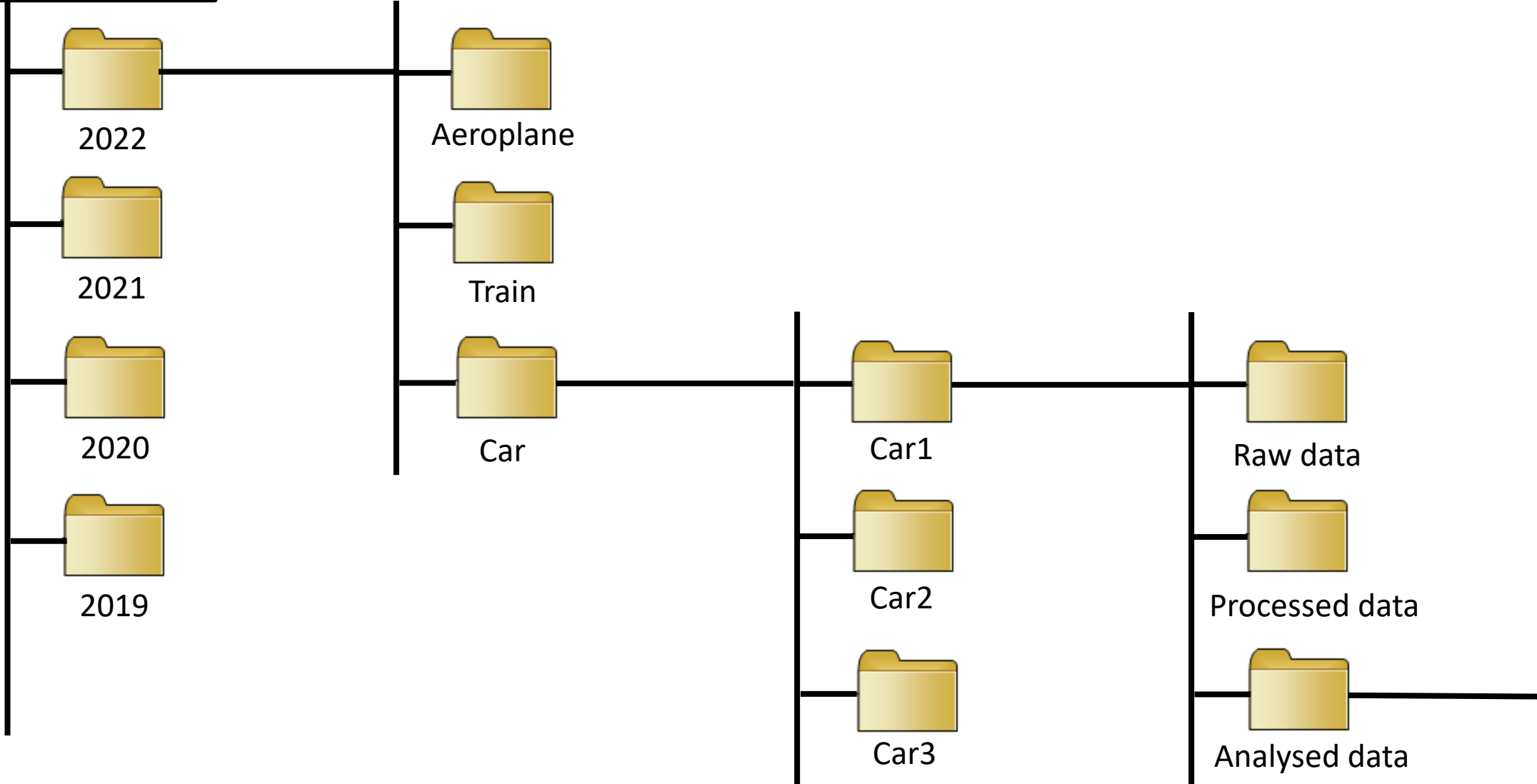
PHY_meeting-transcript_2022-04-12.docx

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 storage and security

- Apply [data classification](#) and handling rules.

- What are the risks to data security (e.g., fire, theft, hardware failure)

- If collecting data offsite, how will you safely transfer it onto the University network storage?

Data Classification

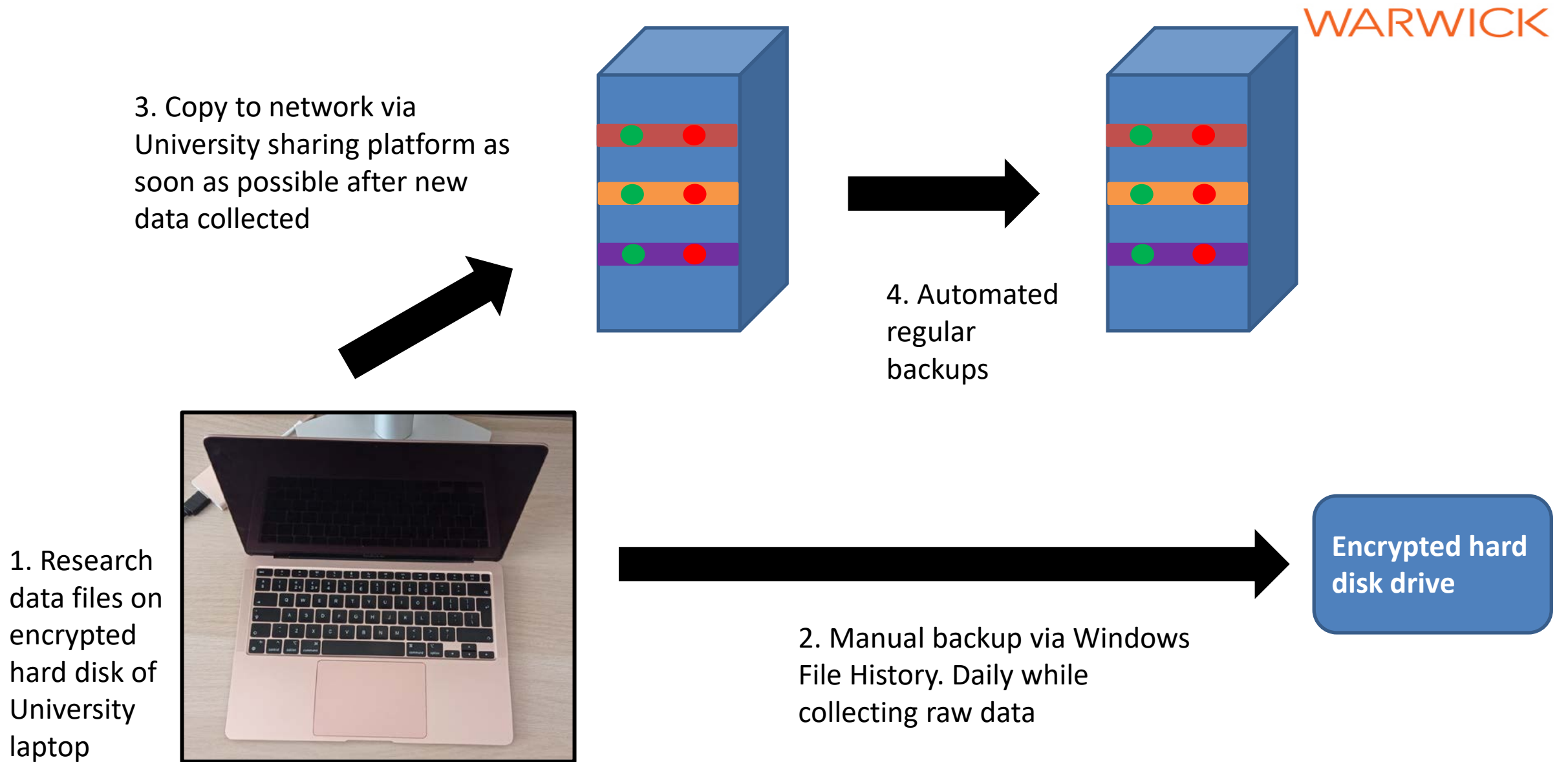Public Protected Restricted

IG05: Information Classification Policy

File Transfers & Backup: Example scenario

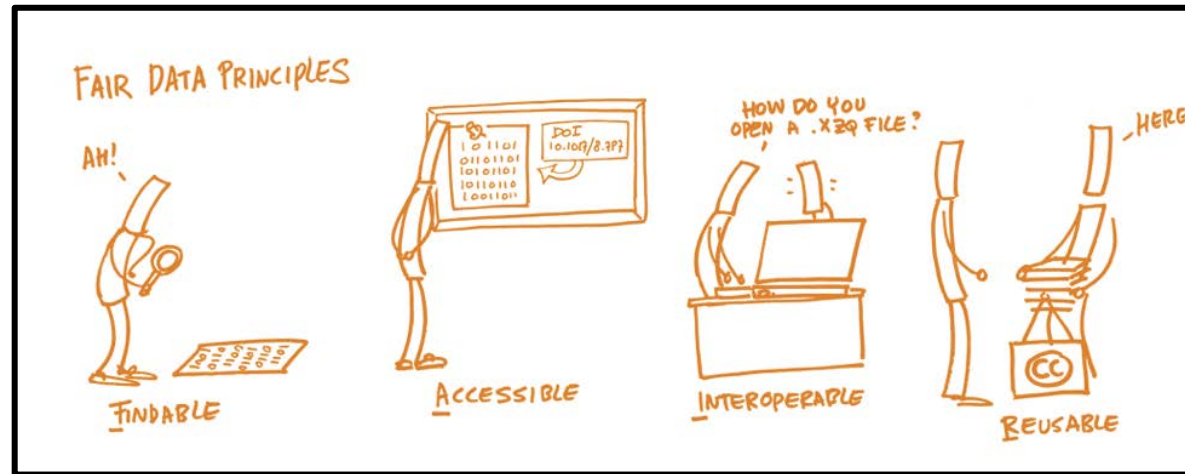3. Copy to network via University sharing platform as soon as possible after new data collected

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

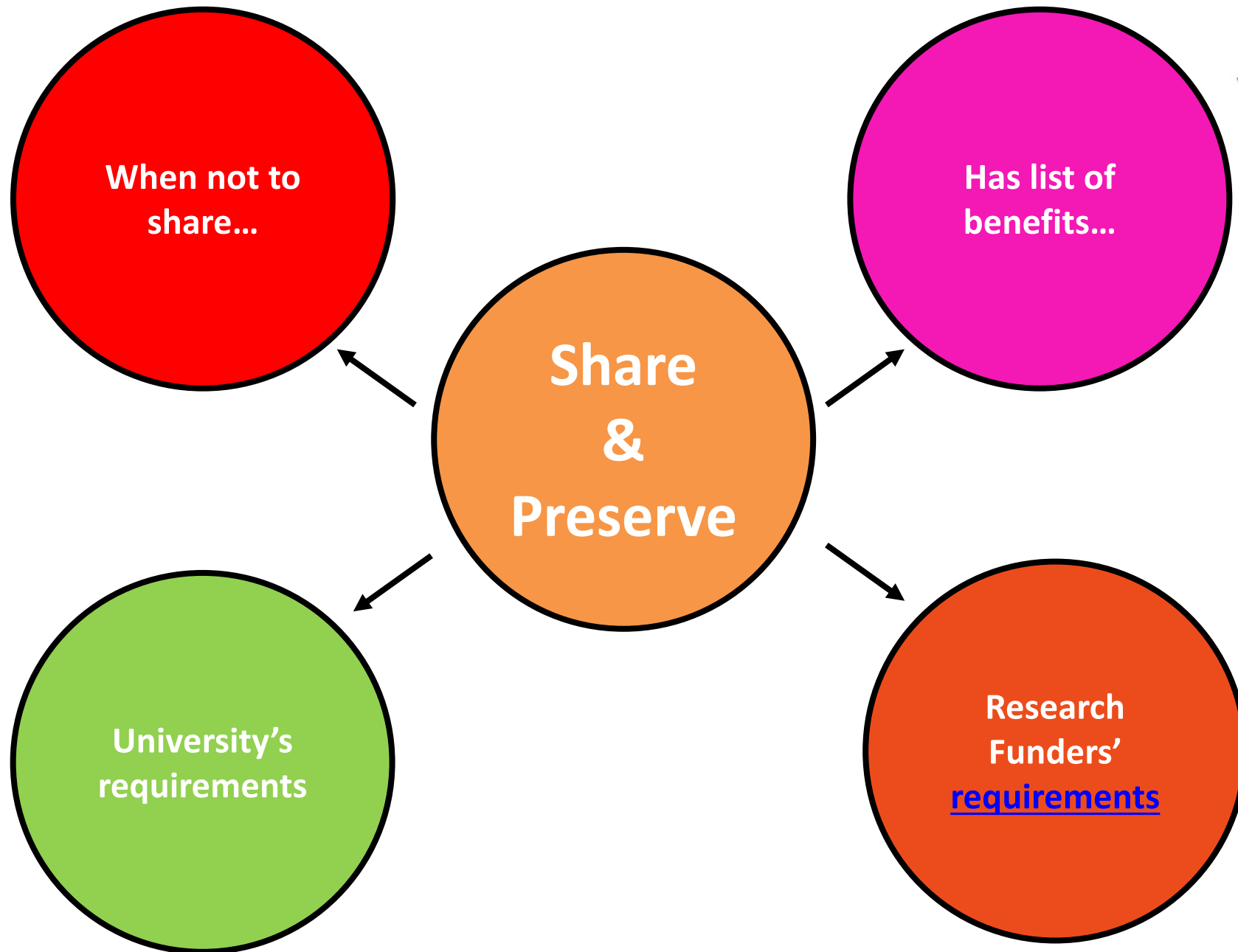**Encrypted hard disk drive**

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

Let's understand last two stages (share and preserve) of **Research data lifecycle and explore research support available for University members...**

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

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

Why should I share data after project completes?

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

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

When not to share…

- data could be of **financial value** or is the basis for **potentially valuable patents** that could be exploited by the University

- data contain **sensitive**, **personal information about human subjects** that could violate Data Protection Act, ethics codes, or your own written consent forms to share it, even with other researchers

- **Anonymising** the data either during or after a project can allow researchers to share and more easily store in the long term

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/

Acknowledgments

- MANTRA Research Data Management Training

https://mantra.ed.ac.uk

- Data Management Rollout at Oxford (DaMaRO) Project

http://damaro.oucs.ox.ac.uk/index.xml

- Managing your research data

https://warwick.ac.uk/services/library/staff/research-data/