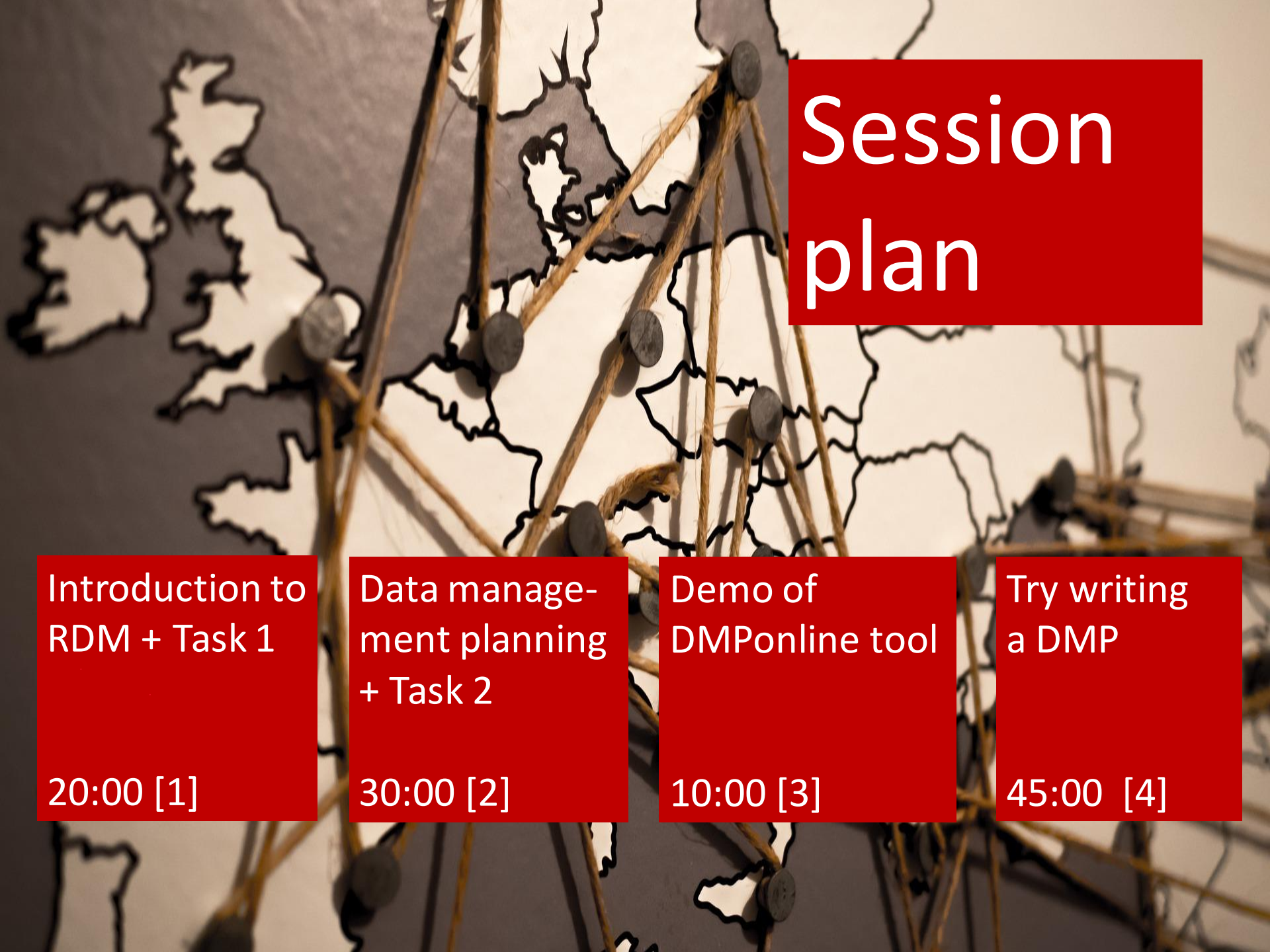
Writing a good Data Management Plan (DMP): A workshop

Nicholas Syrotiuk
Research Data Manager
Durham University

15 November 2023



Session plan

Introduction to RDM + Task 1

20:00 [1]

Data manage-ment planning + Task 2

30:00 [2]

Demo of DMPonline tool

10:00 [3]

Try writing a DMP

45:00 [4]

Part one:

INTRODUCTION TO RDM

What do we mean by RDM?

RESEARCH

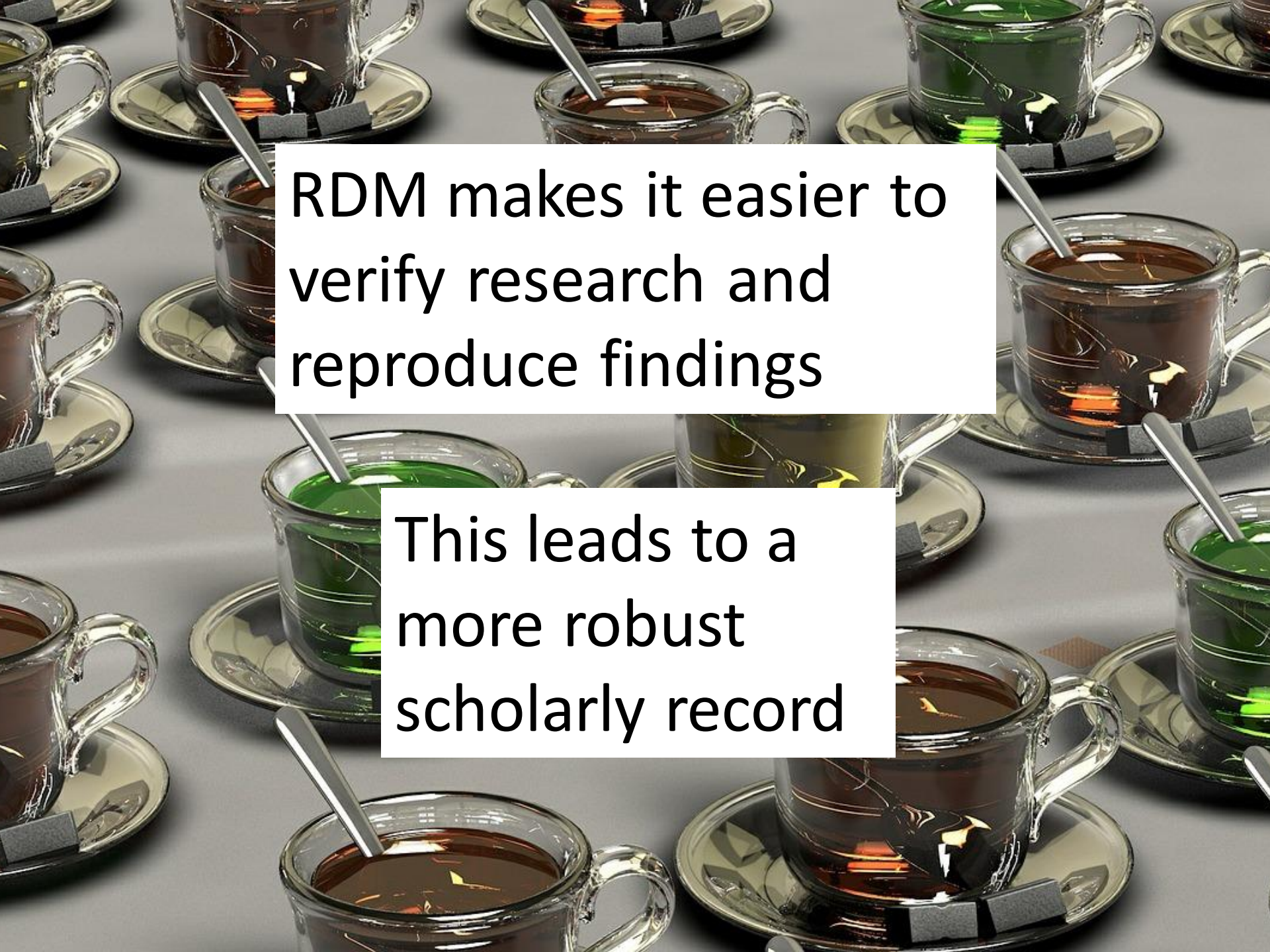
DATA

MANAGEMENT

RDM helps to preserve, protect and proliferate the data behind scientific (research) discoveries and claims

The aim is high quality RDM


RDM leads to increased transparency of the research process


RDM makes it easier to verify research and reproduce findings
This leads to a more robust scholarly record

RDM and sharing (anonymised) research data can lead to making more progress as a research community collectively

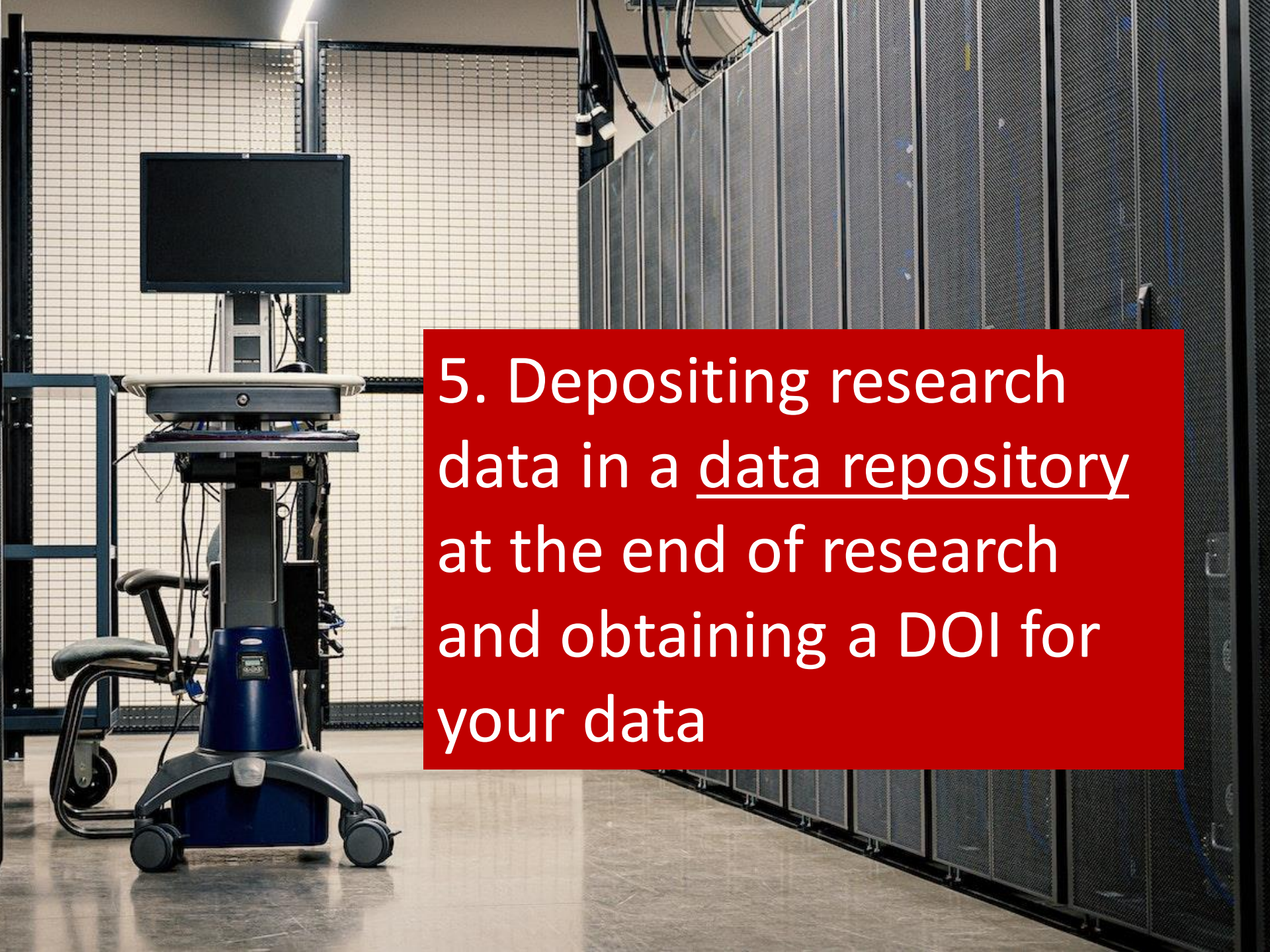
What are core RDM activities?


1. Planning and describing data-related work before it takes place
BED ROOM
13'x15'
KITCHEN
6'11"x7'
BALCONY
9'x3'6"
D2
D3
W1
W2
W3

2. Documenting your data, your data processing and your workflows


3. Choosing open or "standardised" file formats where possible


4. Storing research data
safely during a project


5. Depositing research data in a data repository at the end of research and obtaining a DOI for your data

6. Linking publications to the datasets that underpin them and increasingly code/scripts too

Why bother with data management?

2005/10/30:
Fire destroys Southampton research centre

2019/01/28:

Reward for Royal Oak wreck data (laptop and backup discs) stolen from flat

BBC NEWS

150,000 arrest records wiped in tech blunder

Offenders may go free after software bug deletes fingerprint and DNA files on police computer

Fiona Hamilton, Crime and Security Editor

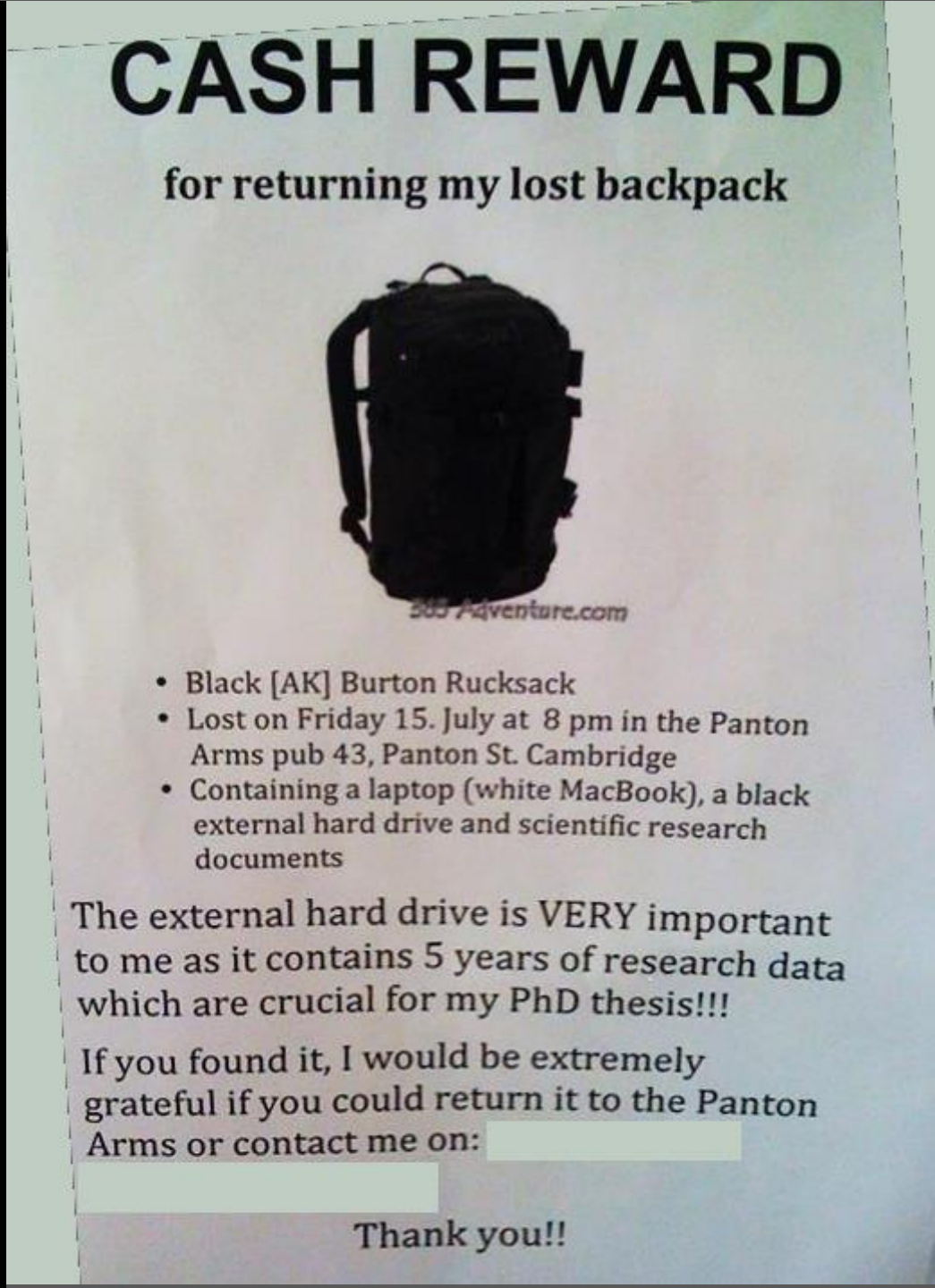
Friday January 15 2021, 12.01am, The Times

UK politics | Politics

The error may allow offenders to go free because biometric evidence left at crime scenes will not be flagged up
GETTY IMAGES

Source: The Times online (15 Jan 2021)

Lost five years of research data



DOES ANYONE HAVE A HORROR STORY ABOUT DATA LOSS?

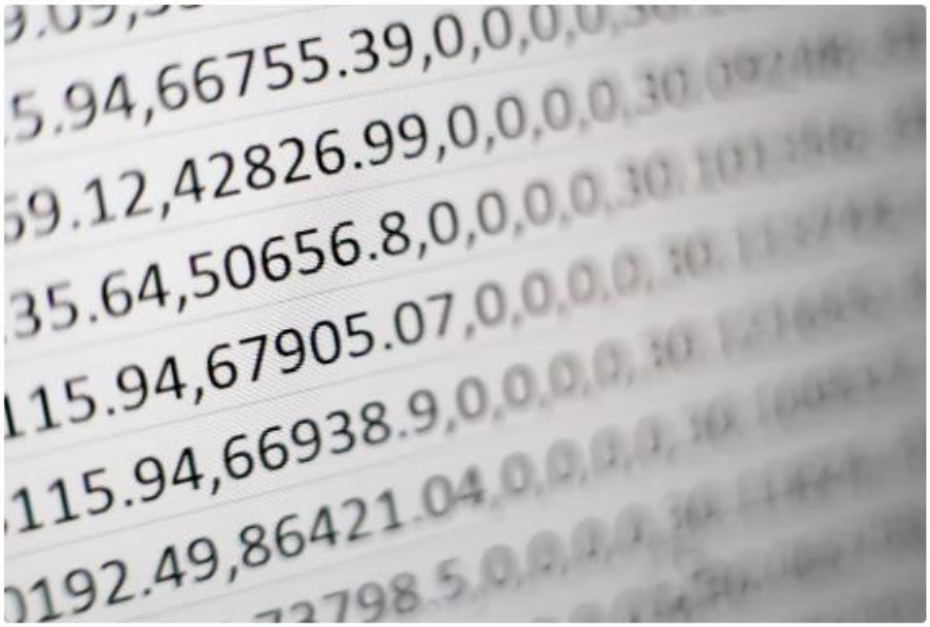
Types of research data

- Survey data
- Recordings
- Text documents
- Tabular data
- Photographs
- Image data
- Big, new, novel or voluminous data
- Scientific measurements
- MRI data in BIDS format
- NVivo data
- International macrodata
- Census data
- Code / scripts
- Project web site

Task one:

WHAT TYPE(S) OF RESEARCH DATA DO YOU CREATE?

Mentimeter



Name one or two types of research data you collect or create.

Enter a word 25

Enter another word 25

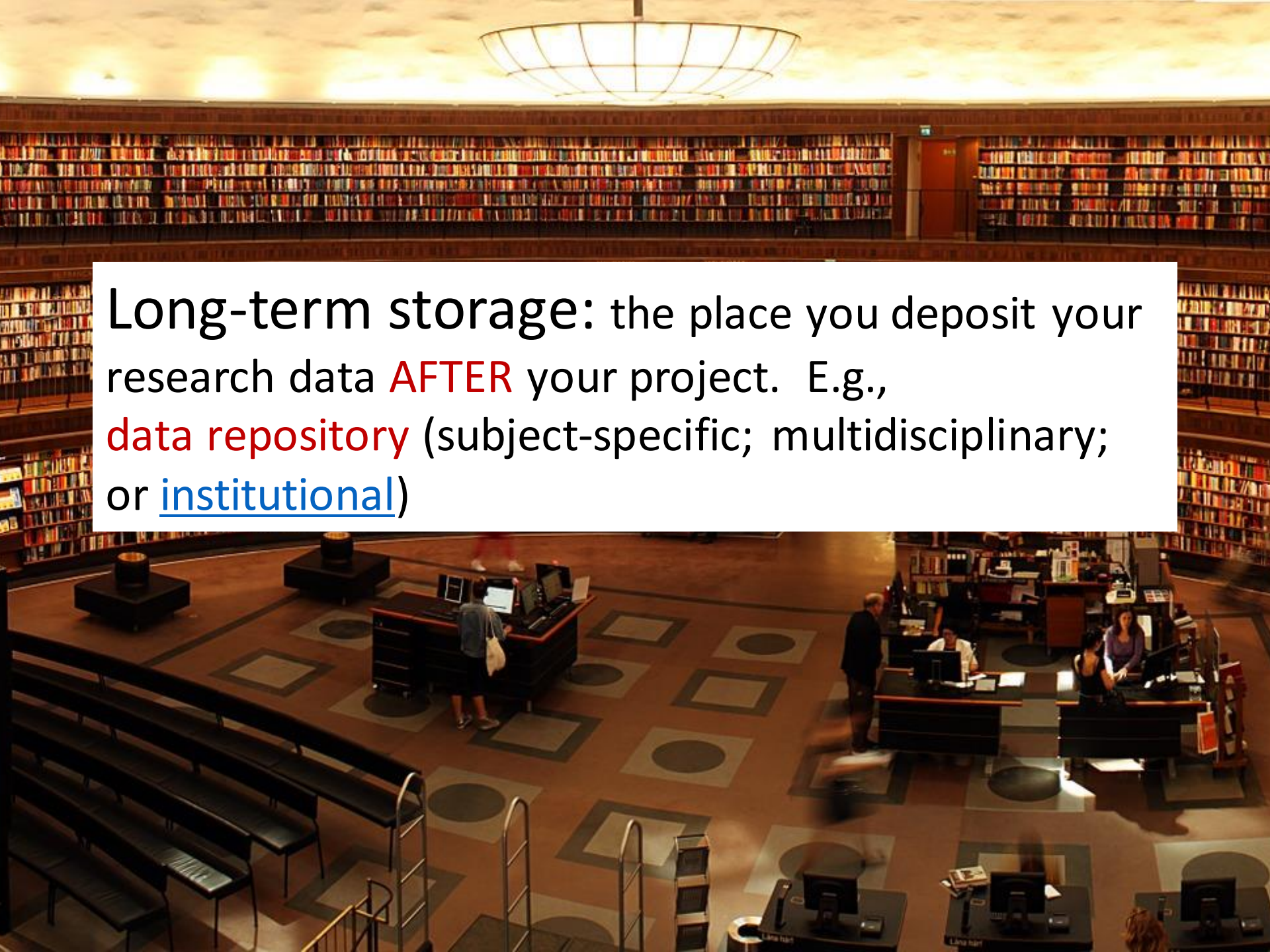
Submit

Name one or two types of research data you collect or create.

39 responses


TERMINOLOGY


Short-term storage: the place you keep your research data DURING your project. E.g., OneDrive for Business or Shared Research Storage (SRS)



Long-term storage: the place you deposit your research data AFTER your project. E.g., data repository (subject-specific; multidisciplinary; or institutional)

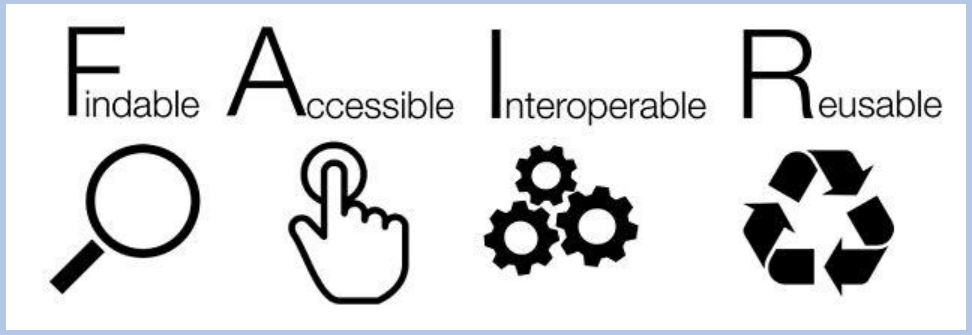
Data sharing: often

means publishing anonymised research data in a data repository** AFTER your research. Other meanings also.

** A repository typically has three levels of access: open access; access for registered users; and restricted access (i.e., access may be granted upon request).

Data sharing agreement: a legal agreement between two organisations for sharing personal data. Read *Data sharing code of practice* from ICO. Seek legal advice.

F.A.I.R. data is ...



The aim is to create **F.A.I.R. data**

as opposed to unFAIR data

Excellent talk by Henry Rzepa (Imperial College) to Durham researchers. DOI: http://doi.org/cppz

Part two:

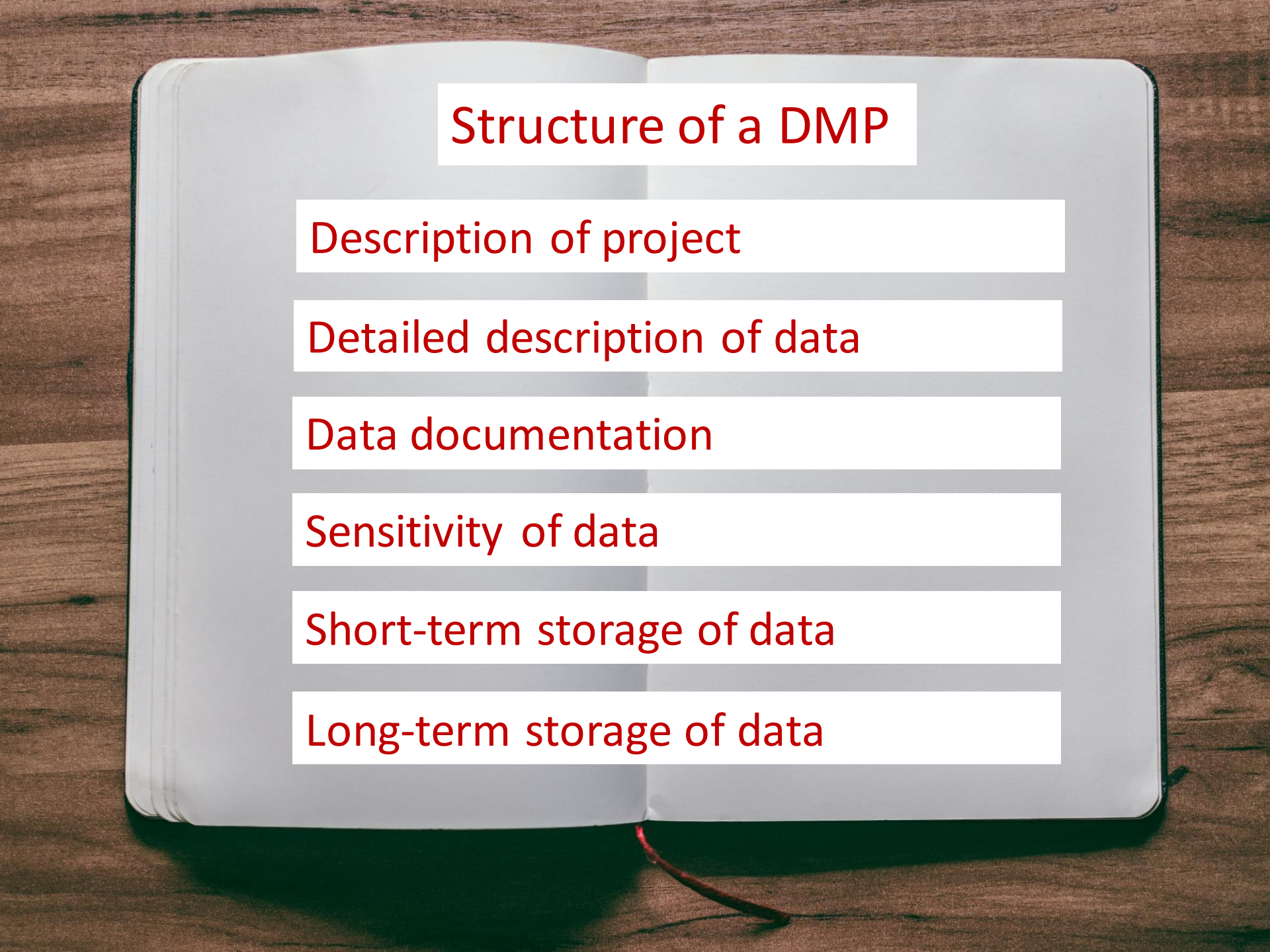
DATA MANAGEMENT PLANNING

UKRI provides guidance on best practice in the management of research data


Plan to keep your sensitive data safe
ABUS
65/50
ABUS


Plan to maximise the re-use potential of your data


Plan to preserve your data for the long term


Structure of a DMP
Description of project
Detailed description of data
Data documentation
Sensitivity of data
Short-term storage of data
Long-term storage of data

Describe research data in a table

Data created or collected	Data type	Data format	Volume or duration	Planned storage or access
Raw ethnographic field notes	Notebooks	Paper	n/a	No shared access
Photographs	Digital images	JPEG	100 photographs * 4 Mb each = ca. 400 Mb	OneDrive for Business
Interviews	Sound recordings	MP3	20 interviews, ca. 30 minutes each	Destroyed after transcription
Transcriptions	Document	Word	20 documents, 5 Mb each. Total: approx. 100 Mb	Anonymised transcripts to be published in the Durham research data repository
Magma flows	X-ray images	Tiff	50 Tb storage per year for five years = 250 Tb total storage	Shared Research Storage

Example row in table

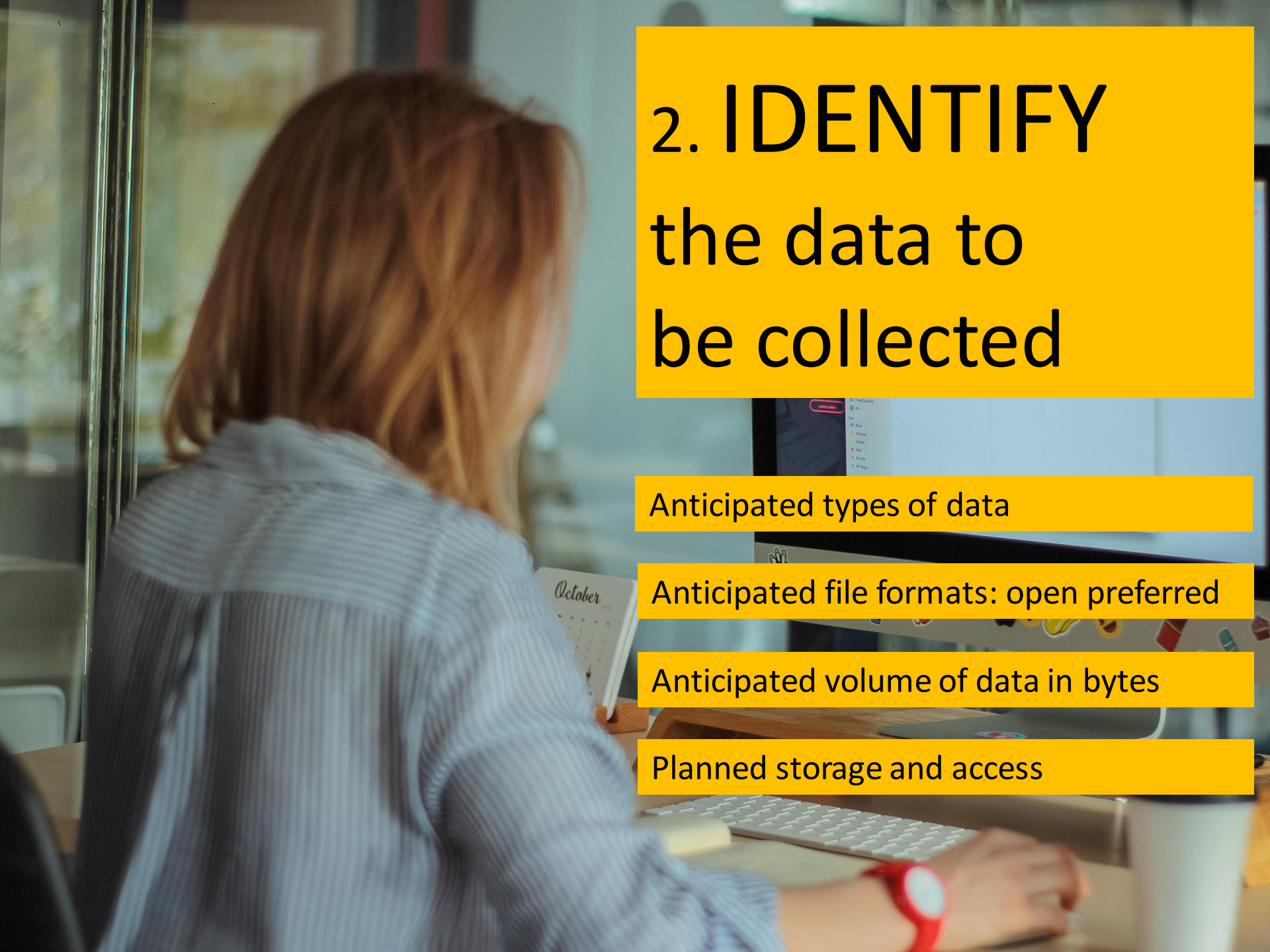
Data created	Data type	Data format	Volume or duration	Planned access
Magma flows	X-ray images	Tiff format	50 Tb storage per year for five years = 250 Tb total storage	Shared Research Storage

Task two:

DESCRIBE YOUR RESEARCH DATA IN A TABLE

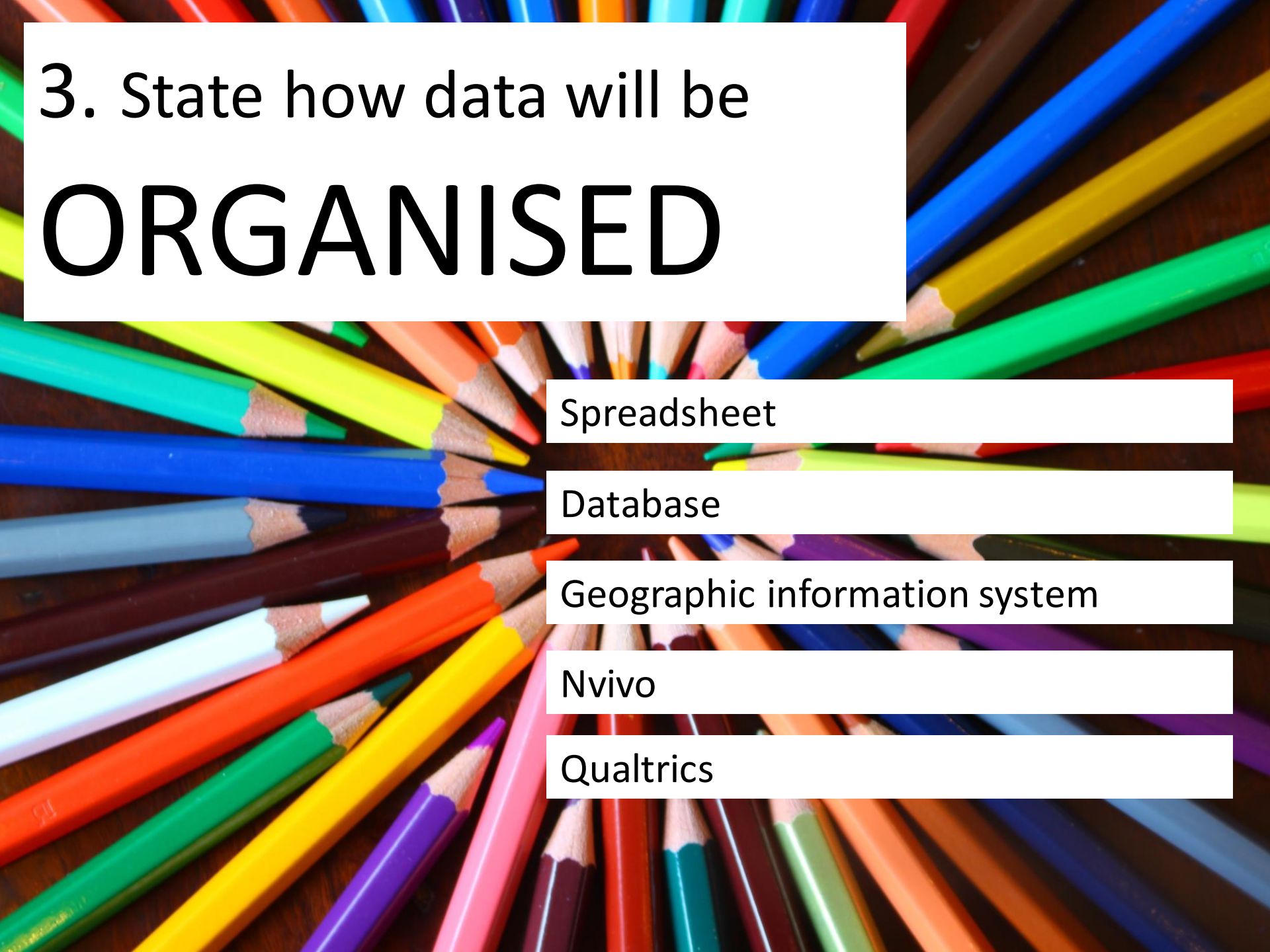

TEN RULES
FOR A GOOD
DMP


1. Determine
FUNDER
requirements



2. IDENTIFY the data to be collected

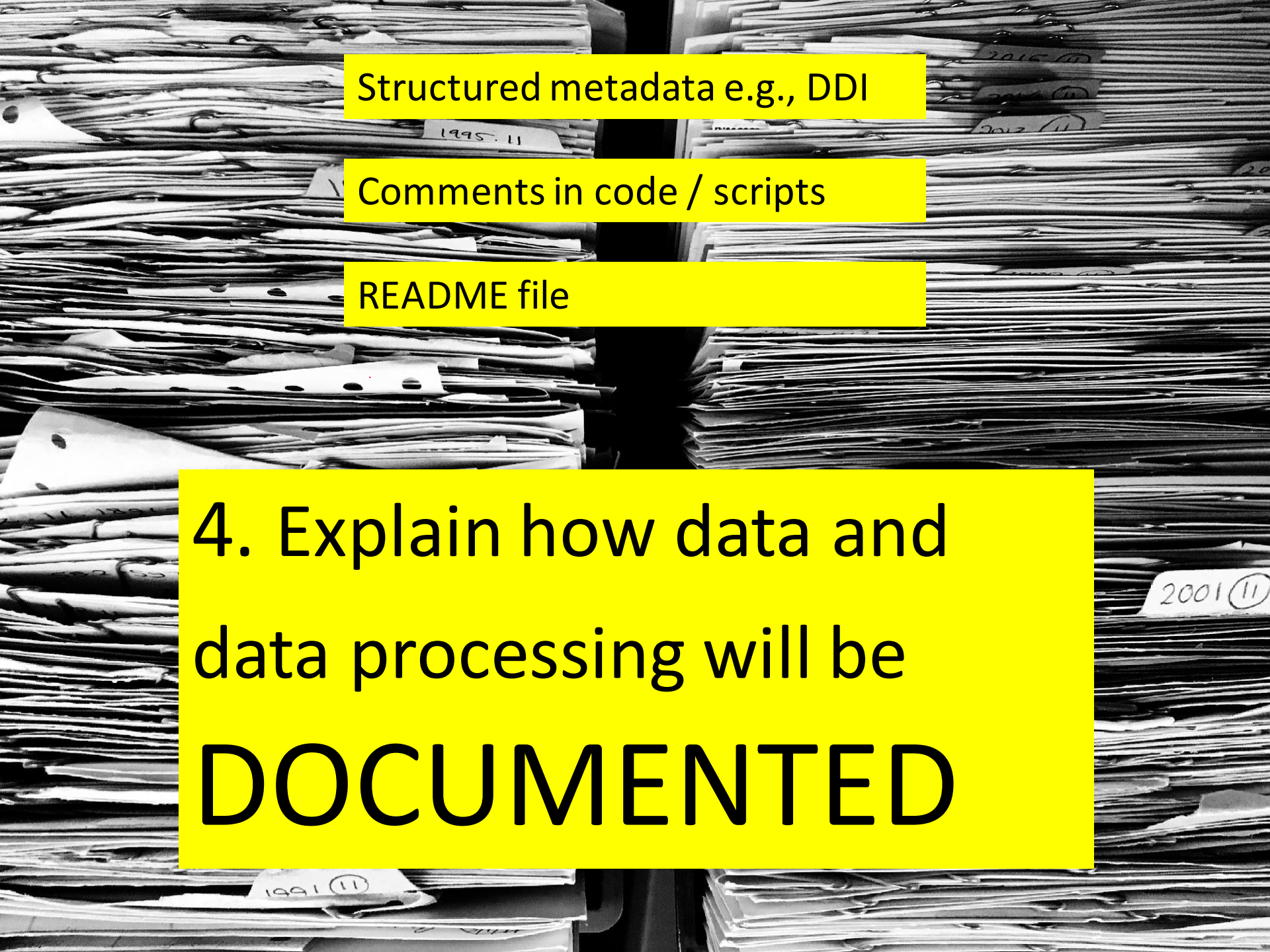
- Anticipated types of data
- Anticipated file formats: open preferred
- Anticipated volume of data in bytes
- Planned storage and access



3. State how data will be

ORGANISED

- Spreadsheet
- Database
- Geographic information system
- Nvivo
- Qualtrics


Structured metadata e.g., DDI
Comments in code / scripts
README file
4. Explain how data and data processing will be
DOCUMENTED

QA during data collection phase

QA during data entry phase

QA during data analysis phase

5. Describe how data

QUALITY

will be assured


6. Provide a solid
STRATEGY for
short-term data
storage and long-
term data storage

7. Define the project's data

POLICIES

Data (and software) licencing

Anticipated plans for data sharing

Management of personal data


Deposit data in a research data repository (perhaps with access restrictions)
Submit data to a journal
Publish a data paper
8. Describe how data will be
DISSEMINATED

9. Assign ROLES and RESPONSIBILITIES

Guidance on authorship & acknowledgements in Library guides:

CRediT (Contributor Roles Taxonomy)

- Data collector
- Quality control
- Data analyst
- Data visualisation
- Data deposit
- DMP editor

10. Prepare a realistic

B U D G E T

Software

Hardware

Short-term storage

Tiers: Silver | Gold

Long-term storage

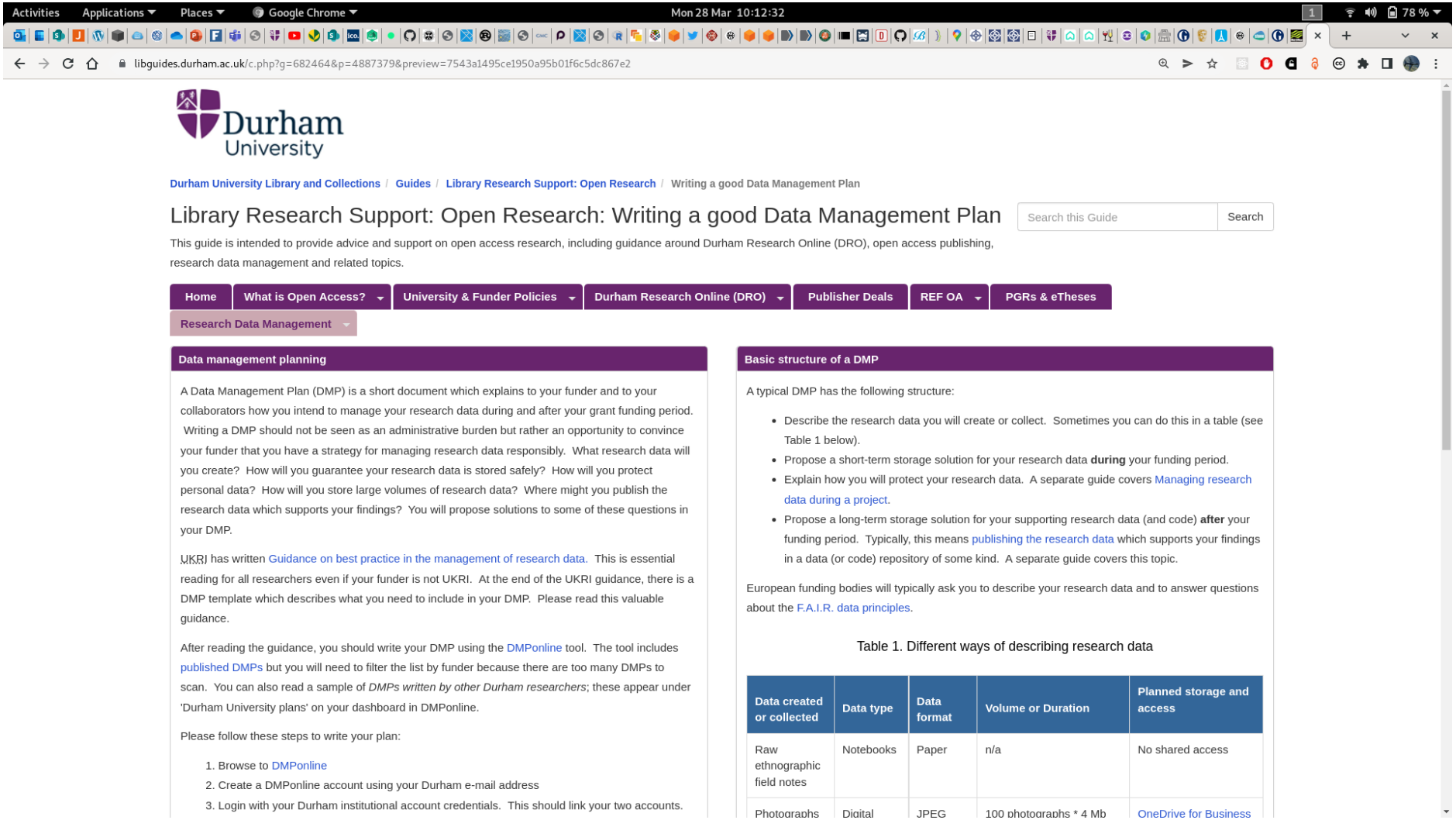
Part three:

DEMONSTRATION OF DMPonline


Live
Demonstration

Part four: **Try writing a**

DATA MANAGEMENT PLAN



Guidance on writing a good DMP

https://bit.ly/3ykxwrU

DMPONLINE | Home | Public DMPs | Funder requirements | Help | Language

Plan to make data work for you

dmponline.dcc.ac.uk

Sign in | Create account

Forgot password?

Remember email

Sign in

- or -

Sign in with your institutional credentials

DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It is provided by the Digital Curation Centre (DCC).

59,972 Users

314 Organisations

65,212 Plans

89 Countries


Thank you
Nicholas Syrotiuk
@DurhamRdm

References:

Wilkinson, M. et al. (2016): The F.A.I.R. Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. DOI: http://doi.org/10.1038/sdata.2016.18

UK Research and Innovation (2018): "Guidance on best practice in the management of research data." UKRI web site.

CESSDA Training Team (2017 - 2020). CESSDA Data Management Expert Guide. Bergen, Norway: CESSDA ERIC. DOI: http://doi.org/10.5281/zenodo.3820472

Wilson G. et al. (2017): Good enough practices in scientific computing. PLoS Comput Biol 13(6): e1005510. DOI: http://doi.org/10.1371/journal.pcbi.1005510

Rzepa, Henry (2018): F.A.I.R. data as a first class citizen in scientific publishing. Imperial College. DOI: http://doi.org/cppz