Writing a good Data Management Plan: A workshop

Nicholas Syrotiuk
Research Data Manager
Durham University

30 March 2022

Session plan

Introduction to RDM and data management planning
60:00 [1]

Questions in chat channel or in person
15:00 [2]

Demo of DMPonline tool
15:00 [3]

Try writing a DMP
30:00 [4]

Part one:
INTRODUCTION TO RDM AND DATA MANAGEMENT PLANNING

What do we mean by

RDM?

RESEARCH

DATA

MANAGEMENT

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

RDM leads to increased transparency of the research process

RDM done well can make it easier to verify research and reproduce findings

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

What are core RDM activities?

Planning and describing data-related work before it takes place***

Documenting your data (and processing/workflows) so that you and others can find and understand it

Choosing open (or at least standardised) file formats where possible

Storing data safely during a project

Depositing data in a trusted repository at the end of research

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

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)

Vines, T. H., et al. (2014): "The availability of research data declines rapidly with article age," Current Biology 24(1): 94-97. DOI: https://doi.org/10.1016/j.cub.2013.11.014

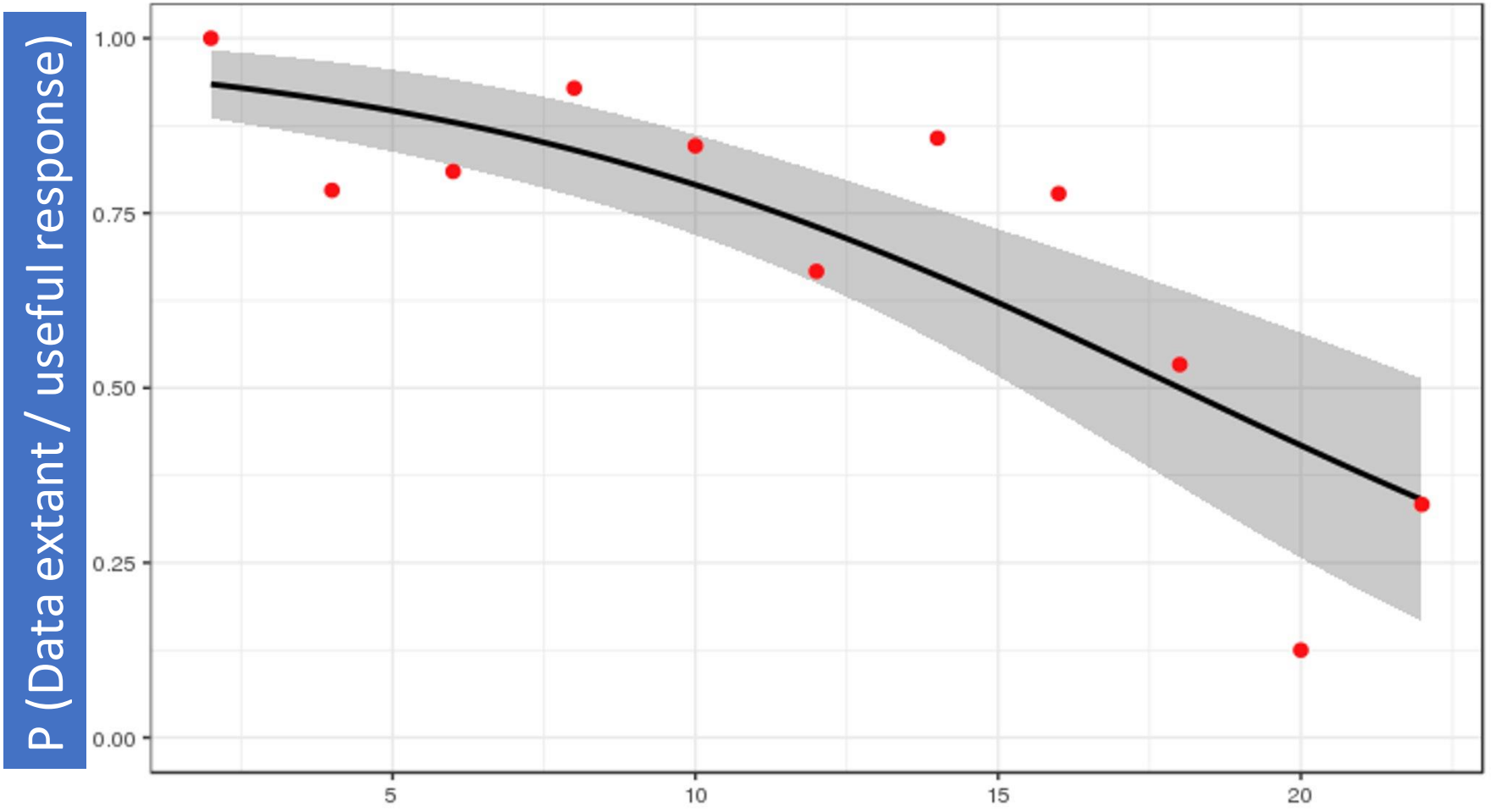
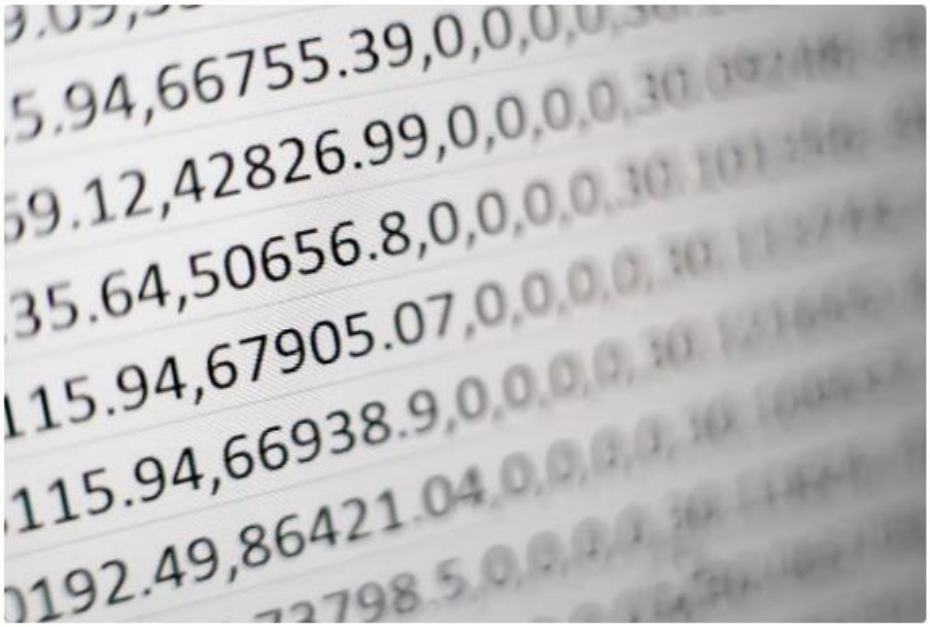


Image created by N Syrotiuk from the original research data and R script.

Types of research data

- Image data
- Survey data
- Recordings
- Transcriptions
- Digital photographs
- Simulation data
- Big, new, novel or voluminous data
- Scientific measurements
- MRI data
- NVivo data
- International macrodata
- Census data
- Code / scripts
- Project web site

Mentimeter



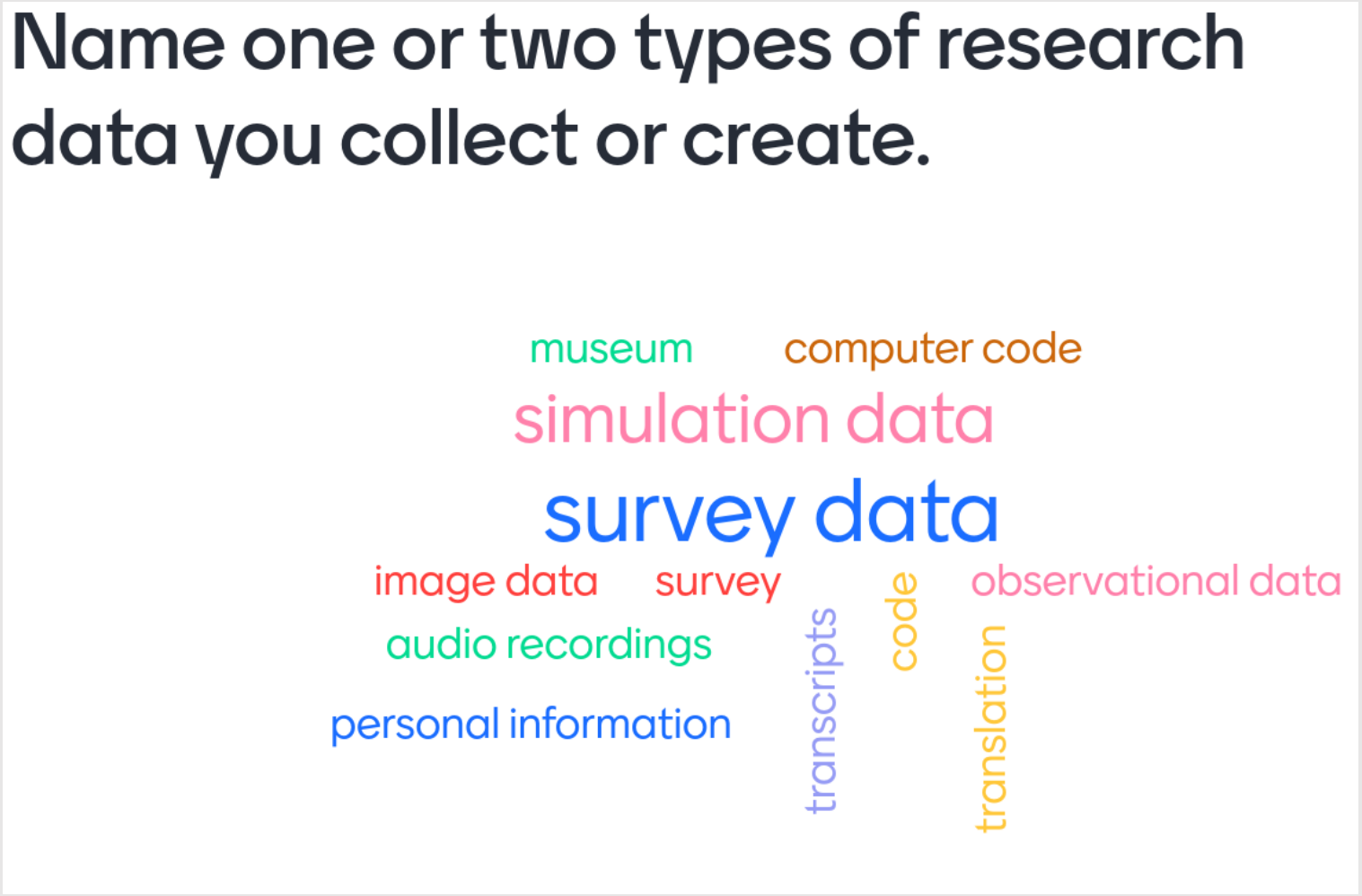
Name one or two types of research data you collect or create.

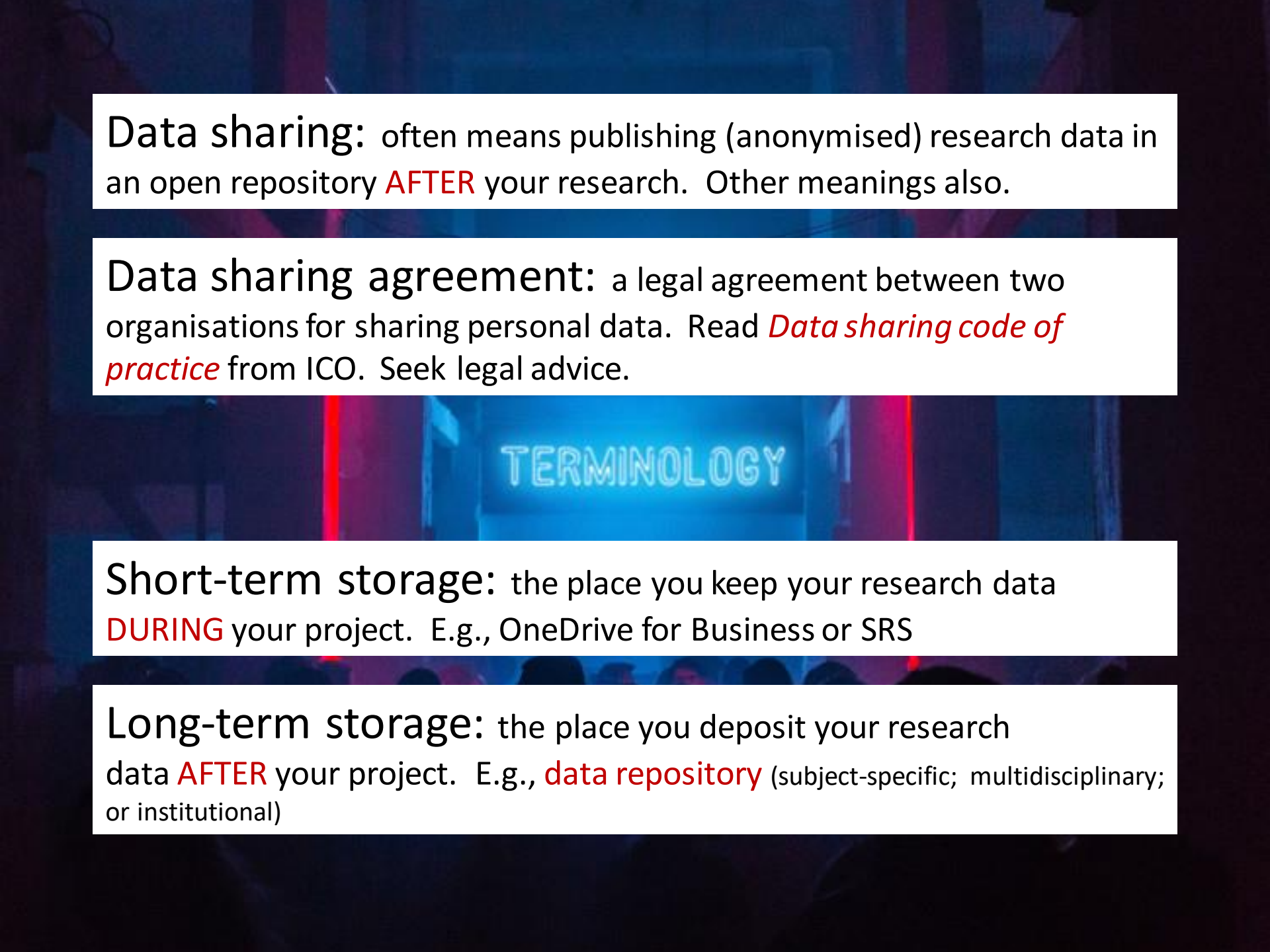
Enter a word 25

Enter another word 25

Submit

Results of Mentimeter





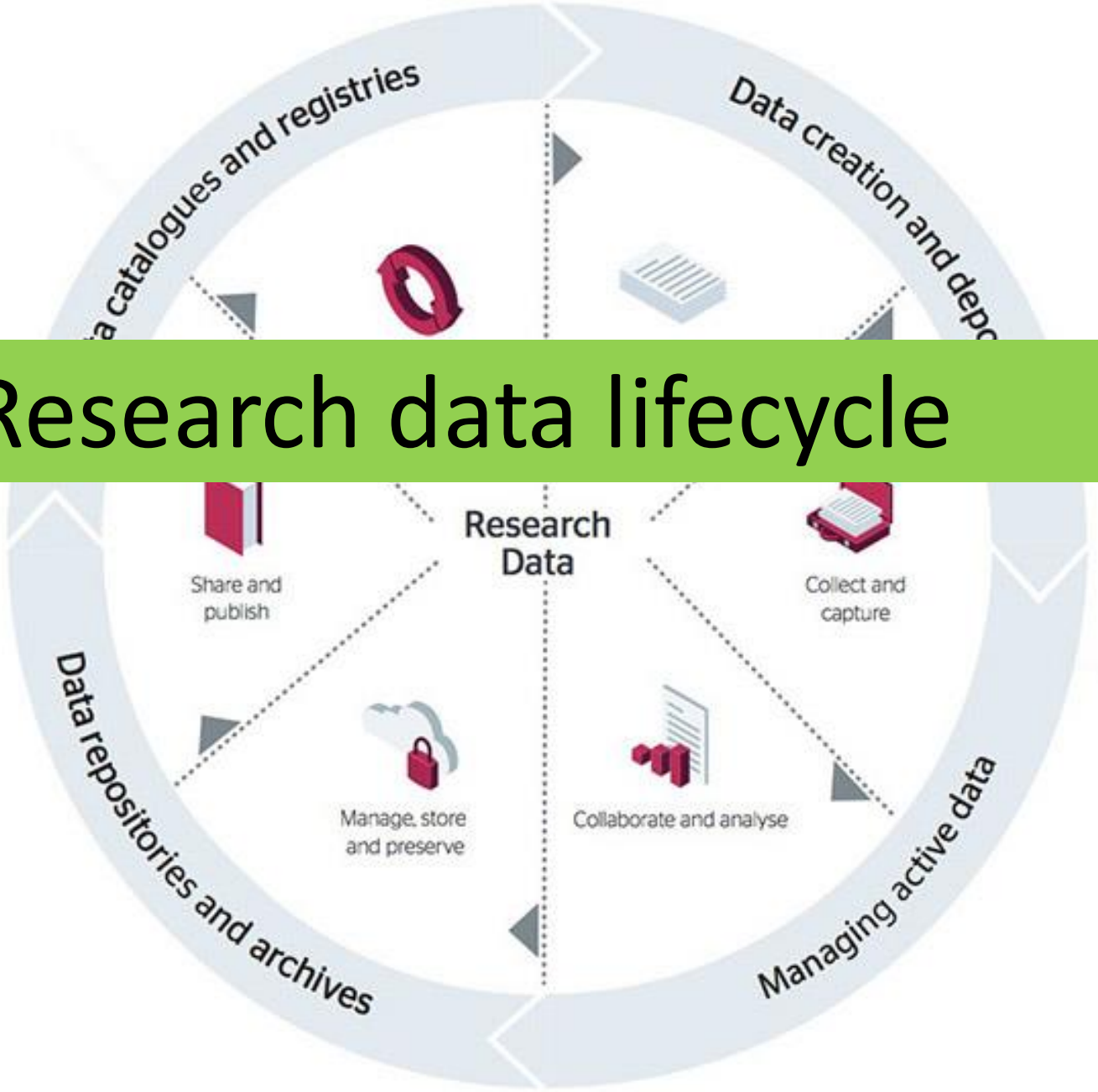
Data sharing: often means publishing (anonymised) research data in an open repository AFTER your research. Other meanings also.

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

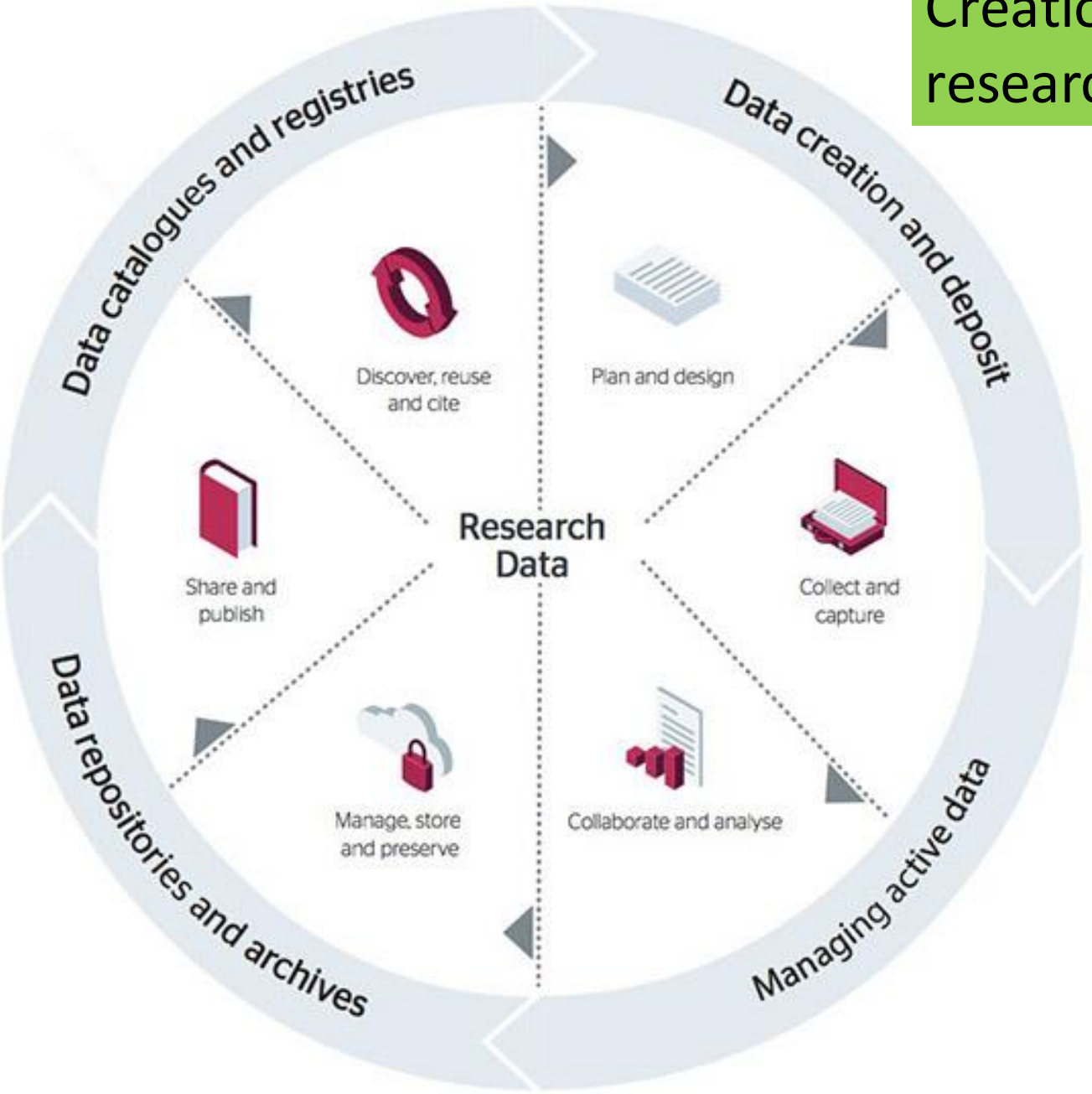
Short-term storage: the place you keep your research data DURING your project. E.g., OneDrive for Business or SRS

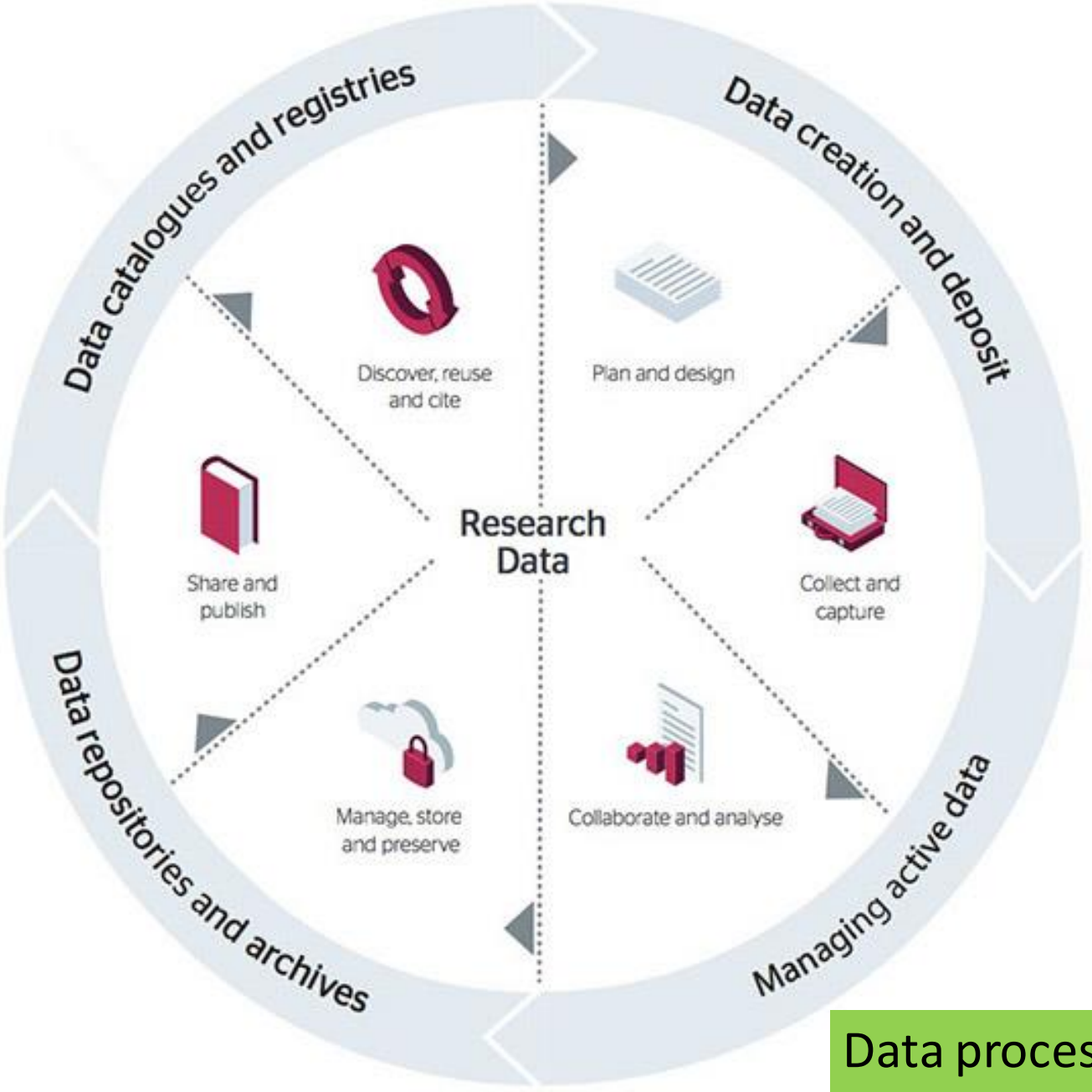
Long-term storage: the place you deposit your research data AFTER your project. E.g., data repository (subject-specific; multidisciplinary; or institutional)

Research data lifecycle

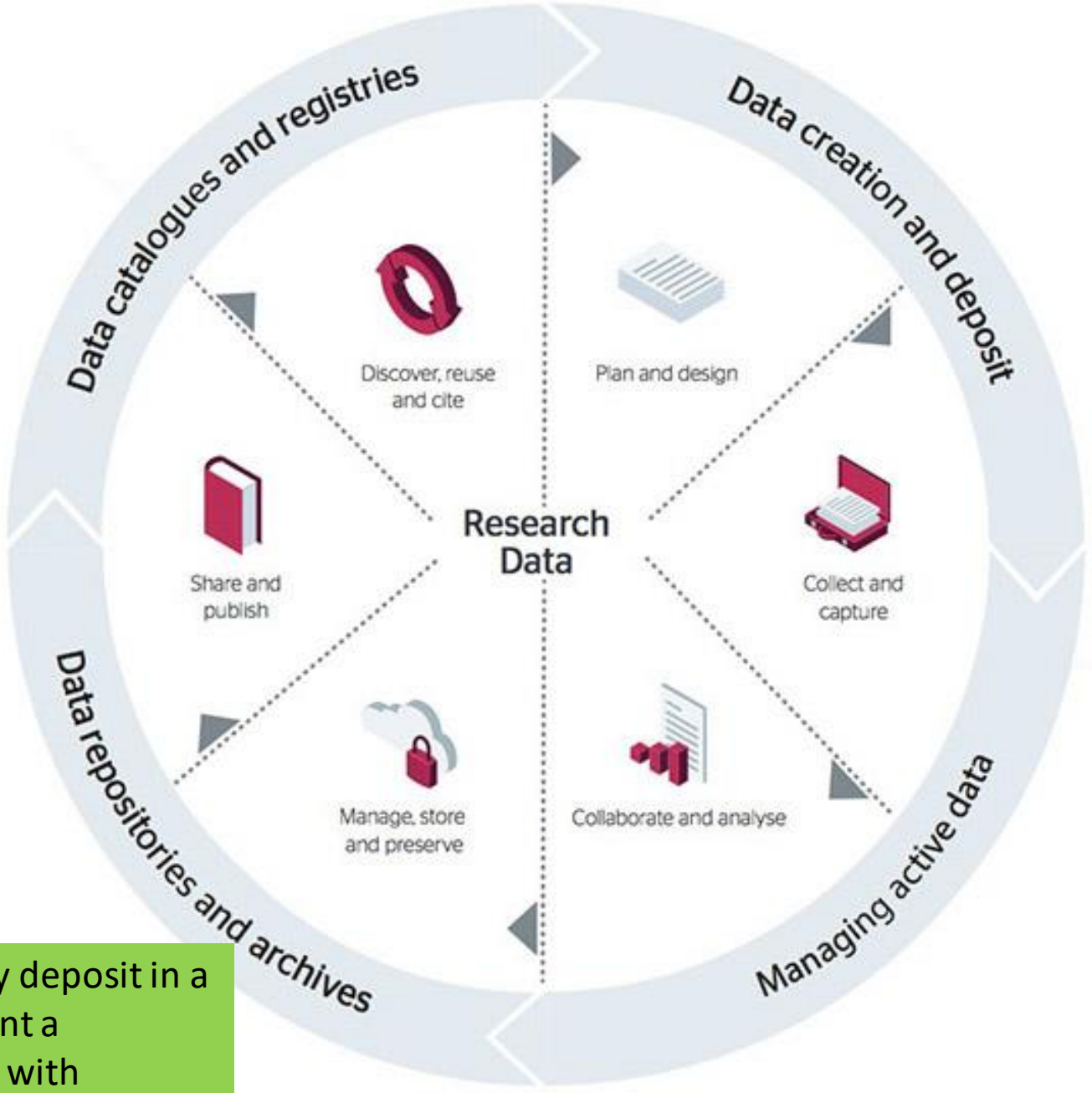


Creation of research data



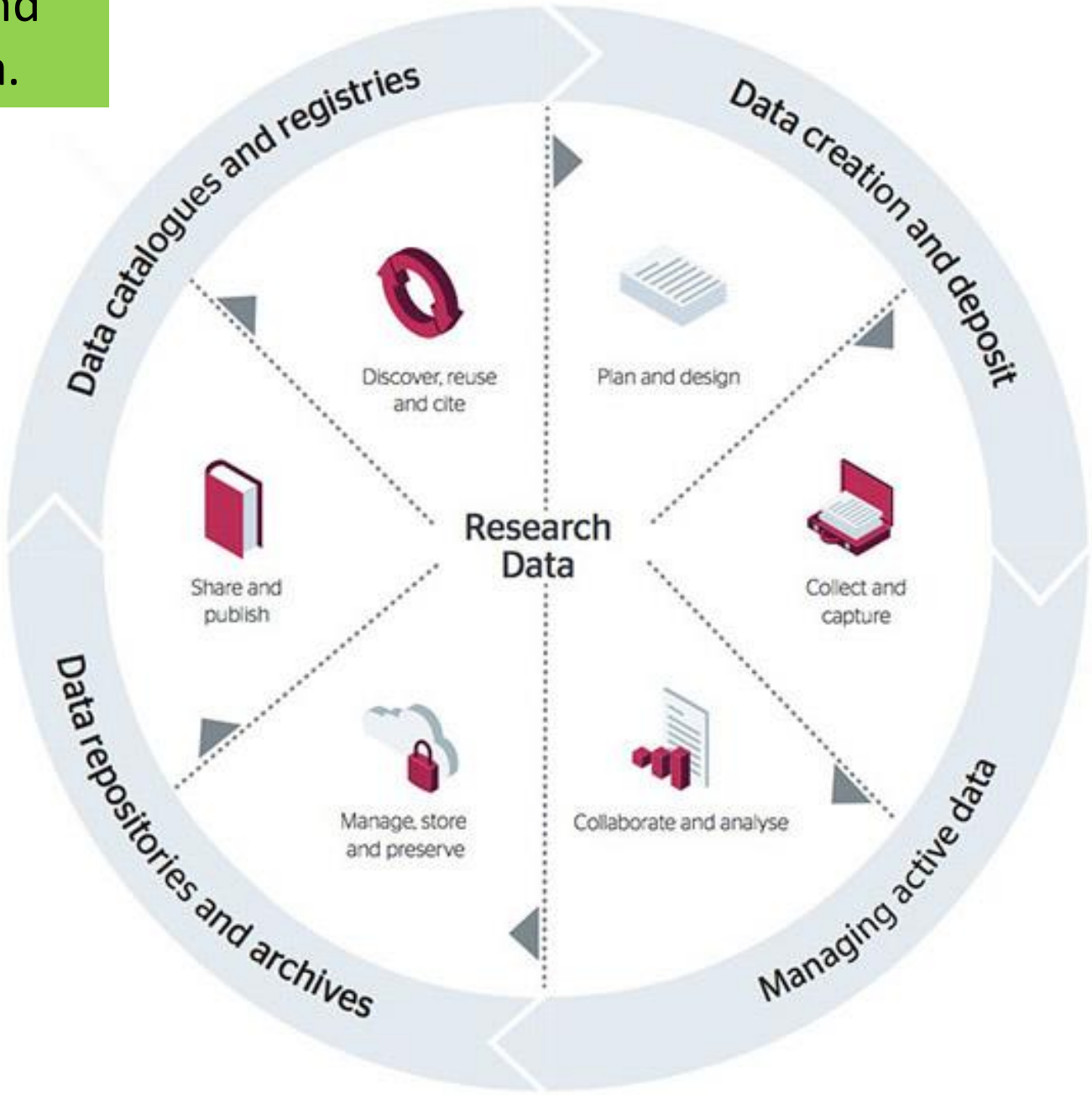


Data processing and analysis



Publish data by deposit in a repository. Mint a DOI. Link data with research paper.

Discover and re-use data.



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

Principle 1.

Publicly funded research data are a public good, produced in the public interest, which should be made **openly available** with as few restrictions as possible in a timely and responsible manner.

Principle 2.

Institutional and project specific data management policies and **plans** should be in accordance with relevant standards and community **best practice**. Data with acknowledged **long-term value** should be **preserved** and remain **accessible** and **usable** for future research.

Principle 3.

To enable research data to be **discoverable** and effectively **re-used** by others, sufficient **metadata** should be recorded and made openly available to enable other researchers to **understand** the research and **re-use** potential of the data. Published results should always include information on how to **access** the supporting data.

Principle 4:

UKRI recognises that there are **legal, ethical and commercial constraints** on release of research data. To ensure that the research process is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these are considered at all stages in the research process.

Principle 5.

To ensure that research teams get appropriate **recognition** for the effort involved in collecting and analysing data, those who undertake Research Council funded work may be entitled to a **limited period of privileged use** of the data they have collected to enable them to publish the results of their research. The length of this period varies by research discipline and, where appropriate, is discussed further in the published policies of individual Research Councils.

Principle 6.

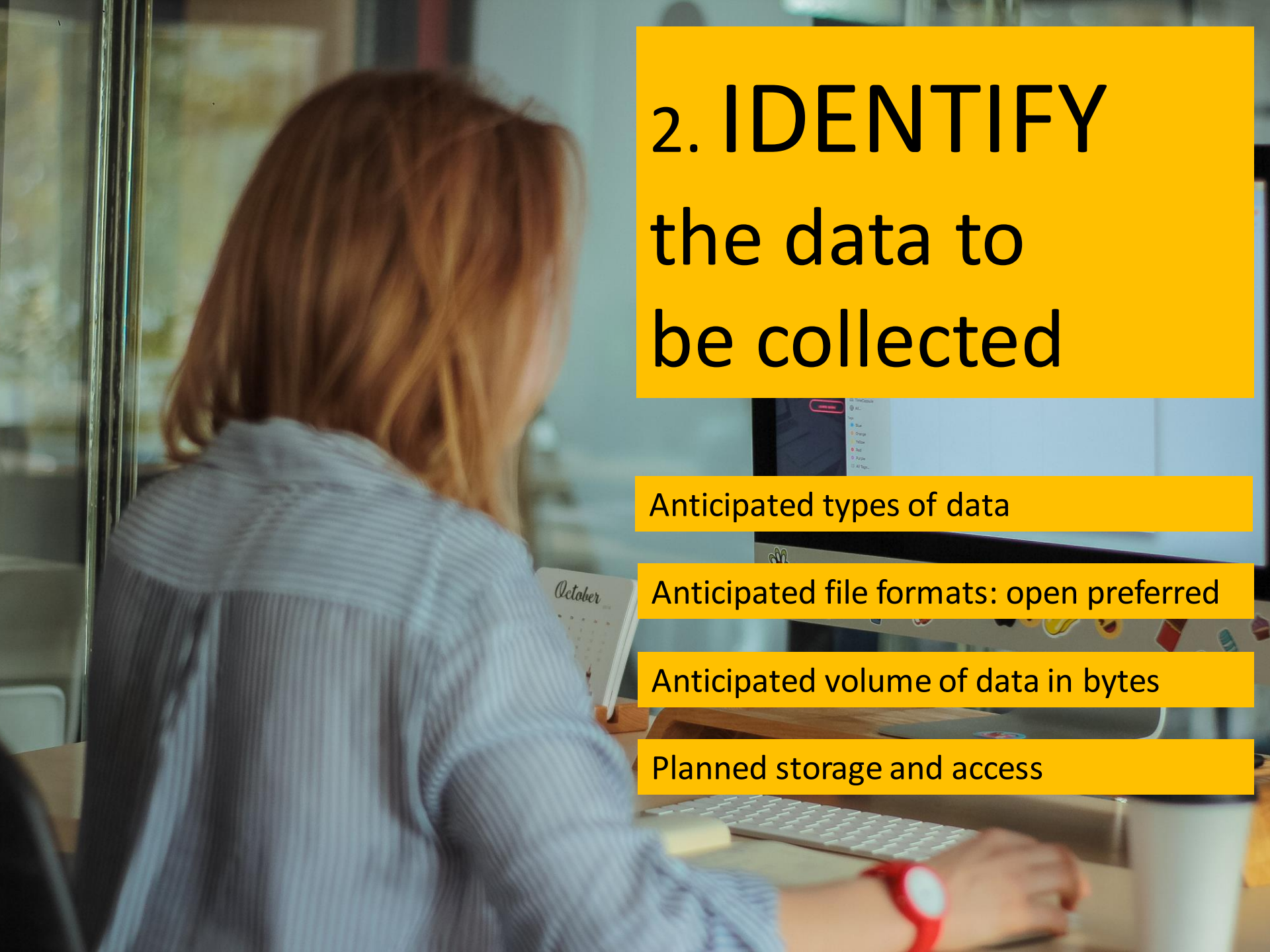
In order to recognise the intellectual contributions of researchers who generate, preserve and share key research datasets, all users of research data should **acknowledge the sources of their data** and abide by the terms and conditions under which they are accessed.

Principle 7

It is **appropriate to use public funds** to support the management and sharing of publicly funded research data. To maximise the research benefit which can be gained from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.


TEN RULES
FOR A GOOD
DMP


1. Determine
FUNDER
requirements

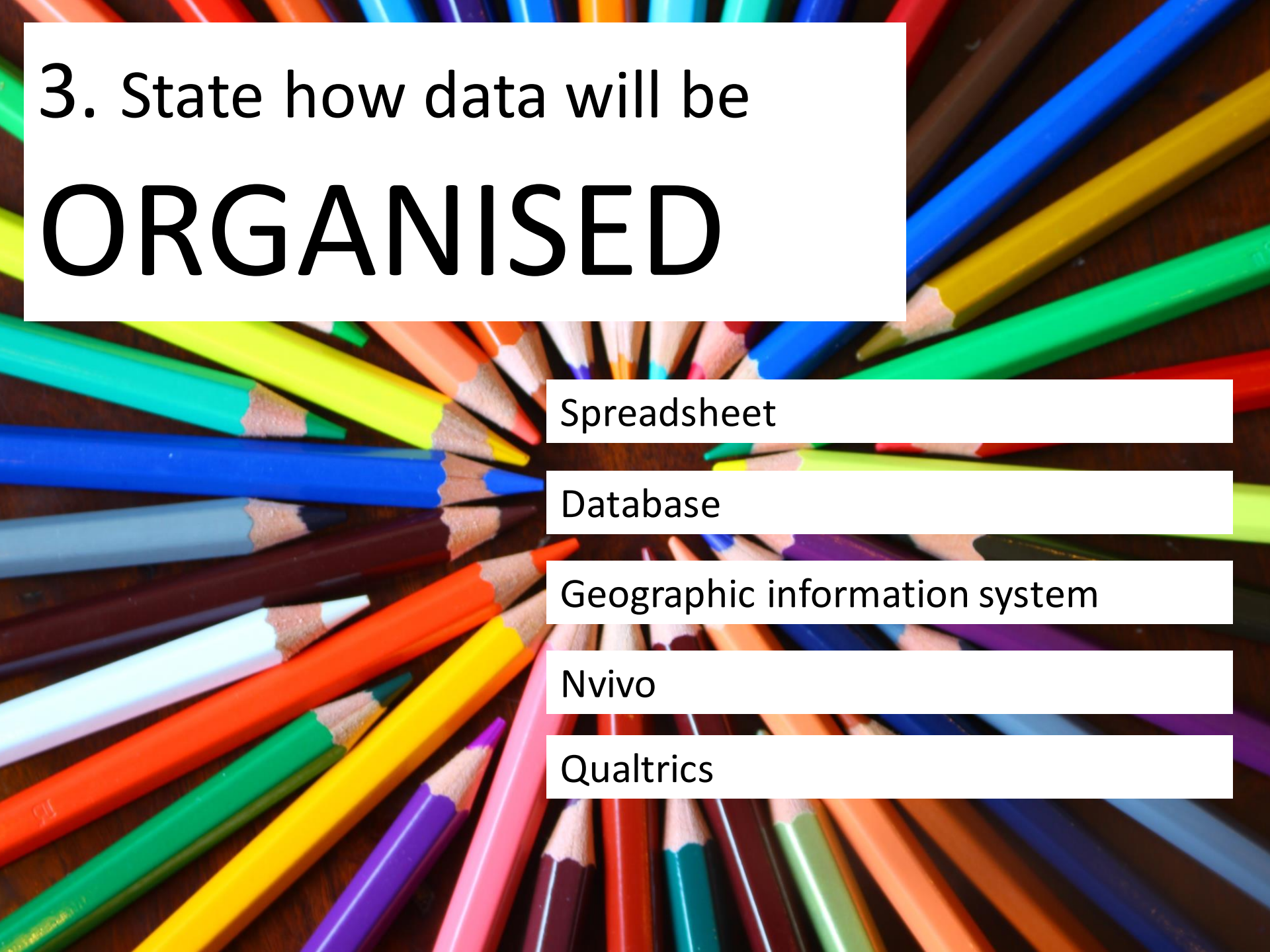


2. IDENTIFY the data to be collected

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

Consider describing research data in a table

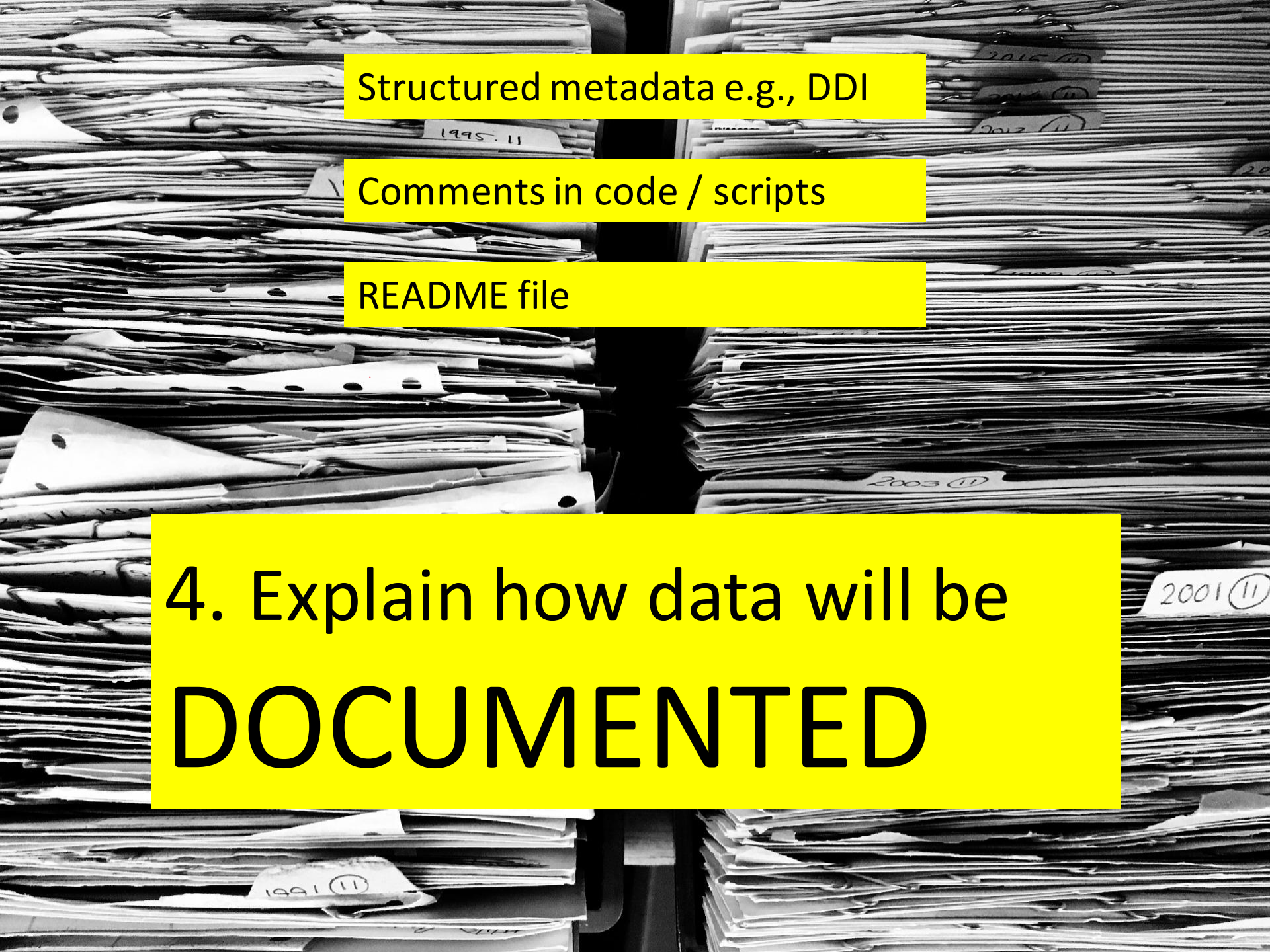
Data created or collected	Data type	Data format	Volume or duration	Planned storage or access
Raw ethnographic field notes	Notebooks	Paper	n/a	No shared access
Photographs	Digital images	JPEG	100 photographs * 4 Mb each = ca. 400 Mb	OneDrive for Business
Interviews	Sound recordings	MP3	20 interviews, 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



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 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 in an open data repository
Submit data to a journal
Publish a data paper
8. Describe how data will be
DISSEMINATED

9. Assign ROLES and RESPONSIBILITIES

- 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
- Long-term storage

Part two:

QUESTIONS

in chat channel or in person

Part three:

DEMONSTRATION OF DMPonline


Live
Demonstration

Part four: **Try writing a**

DATA MANAGEMENT PLAN

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

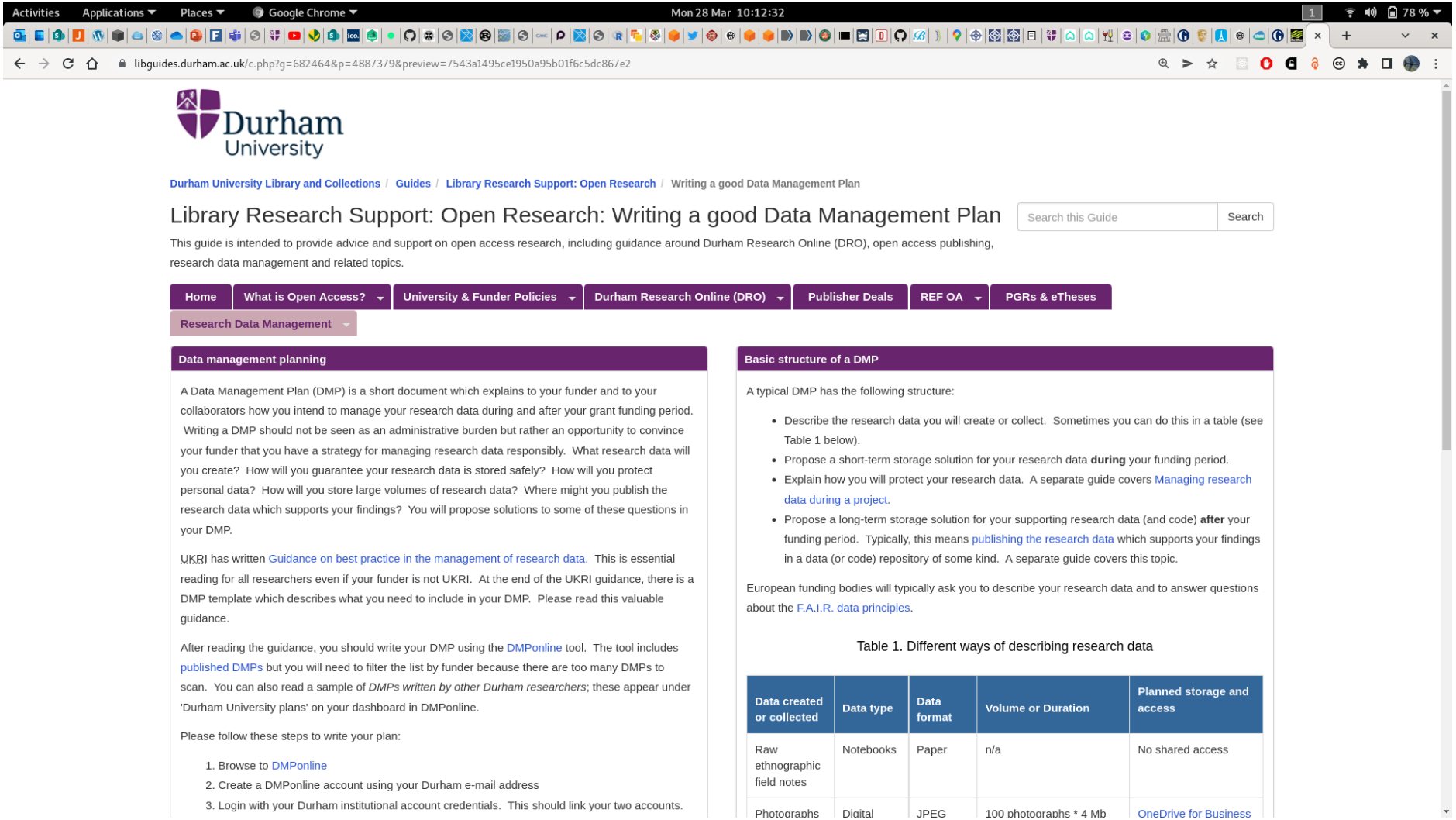
DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It is provided by the Digital Curation Centre (DCC).

59,972 Users

314 Organisations

65,212 Plans

89 Countries



Guidance on writing a good DMP


Thank you
Nicholas Syrotiuk
@DurhamRdm

References:

Vines, T. H., et al. (2014): "The availability of research data declines rapidly with article age," Current Biology 24(1): 94-97. DOI: https://doi.org/10.1016/j.cub.2013.11.014

Vines, T. H., et al. (2013): "The availability of research data declines rapidly with article age." Dryad Digital Repository. (dataset). DOI: https://doi.org/10.5061/dryad.q3g37

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. Retrieved from https://www.cessda.eu/DMGuide