Introduction to Research Data Management

Nicholas Syrotiuk
Research Data Manager
Durham University

22 October 2024

DOI: http://doi.org/nnz5


RDM concepts and terms
10:00 [1]
RDM core activities + Task 1
20:00 [2]
Session plan
CIS storage options
10:00 [3]
RDM tools + Task 2
20:00 [4]
RDM best practices
15:00 [5]
Summary and questions
15:00 [6]

Part 1A:

RDM CONCEPTS AND TERMS

What do we mean by

RDM?

RESEARCH

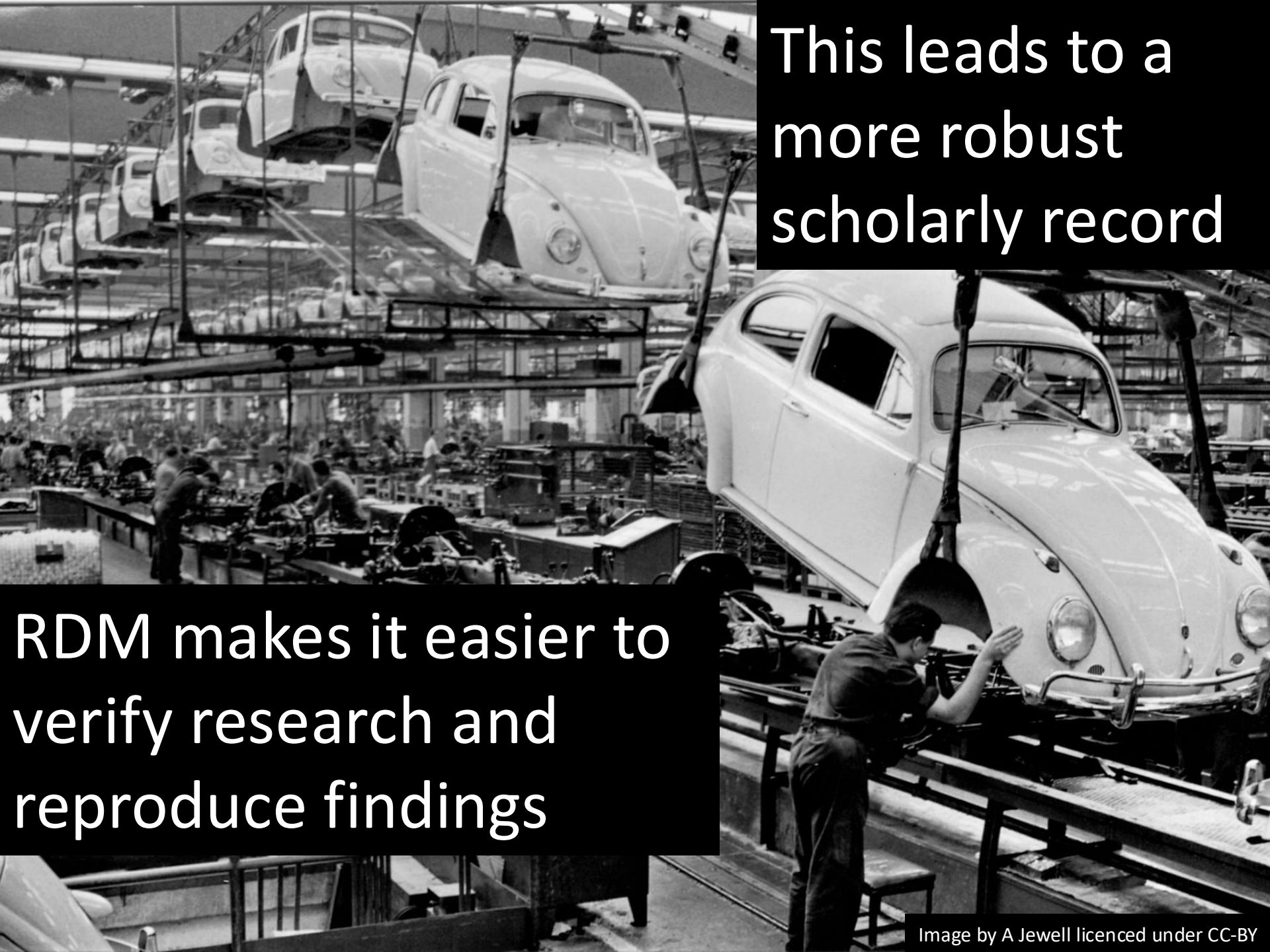
DATA

MANAGEMENT

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

RDM is a quality issue


RDM leads to increased transparency of the research process
Research data
Code
Methods
Documentation



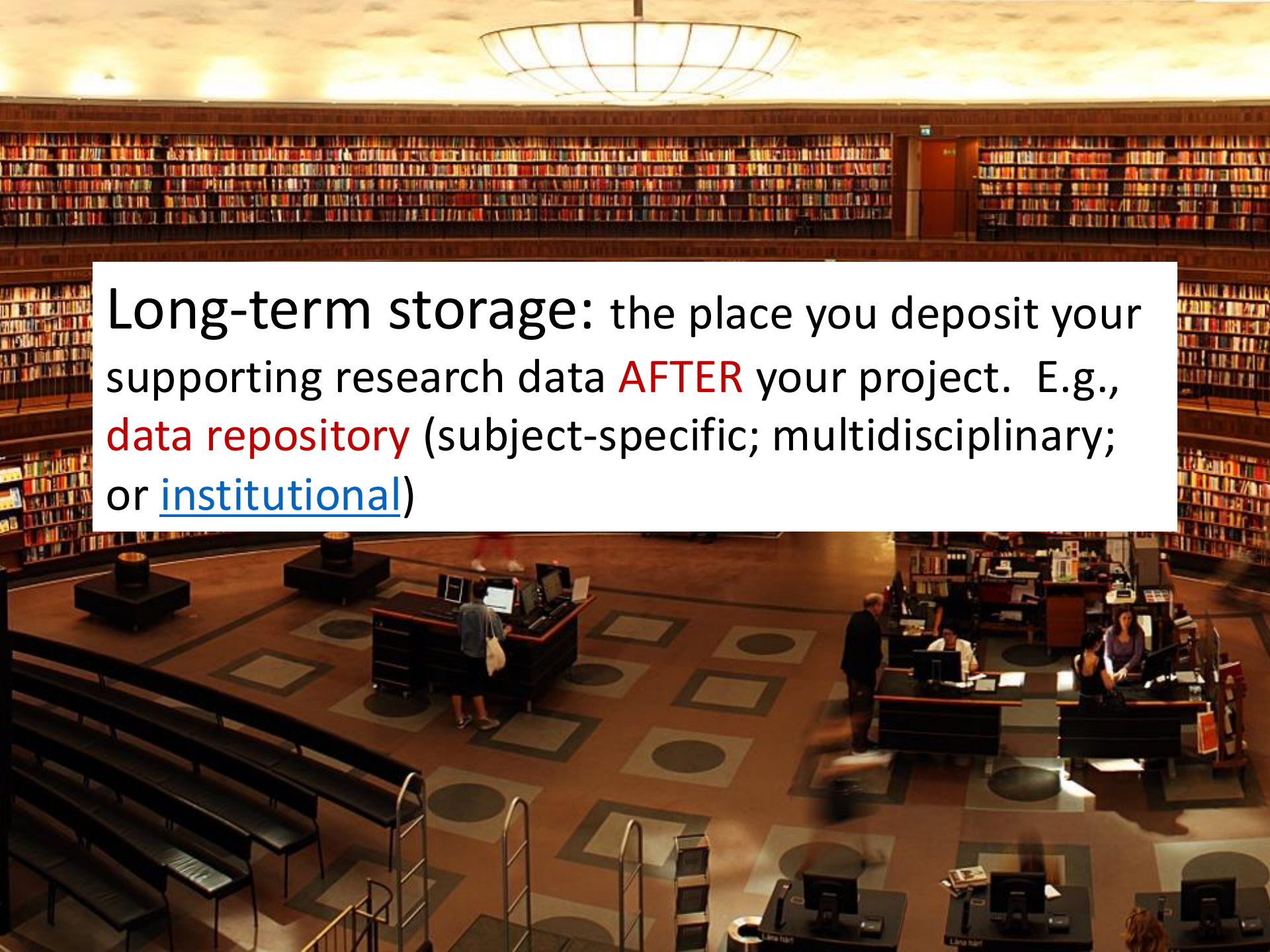
This leads to a more robust scholarly record

RDM makes 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


TERMINOLOGY


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



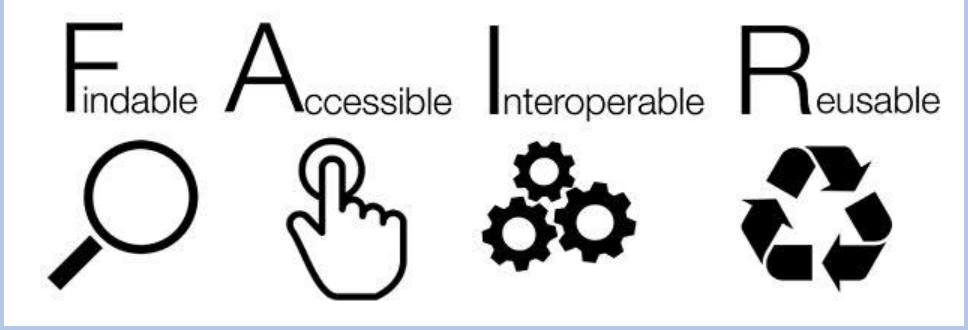
Long-term storage: the place you deposit your supporting research data AFTER your project. E.g., data repository (subject-specific; multidisciplinary; or institutional)

Data sharing has two meanings:

1. Post-research: sharing by publishing research data in a data repository
2. During research: sharing data with others

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

Part 1B:

WHY BOTHER WITH RDM?

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

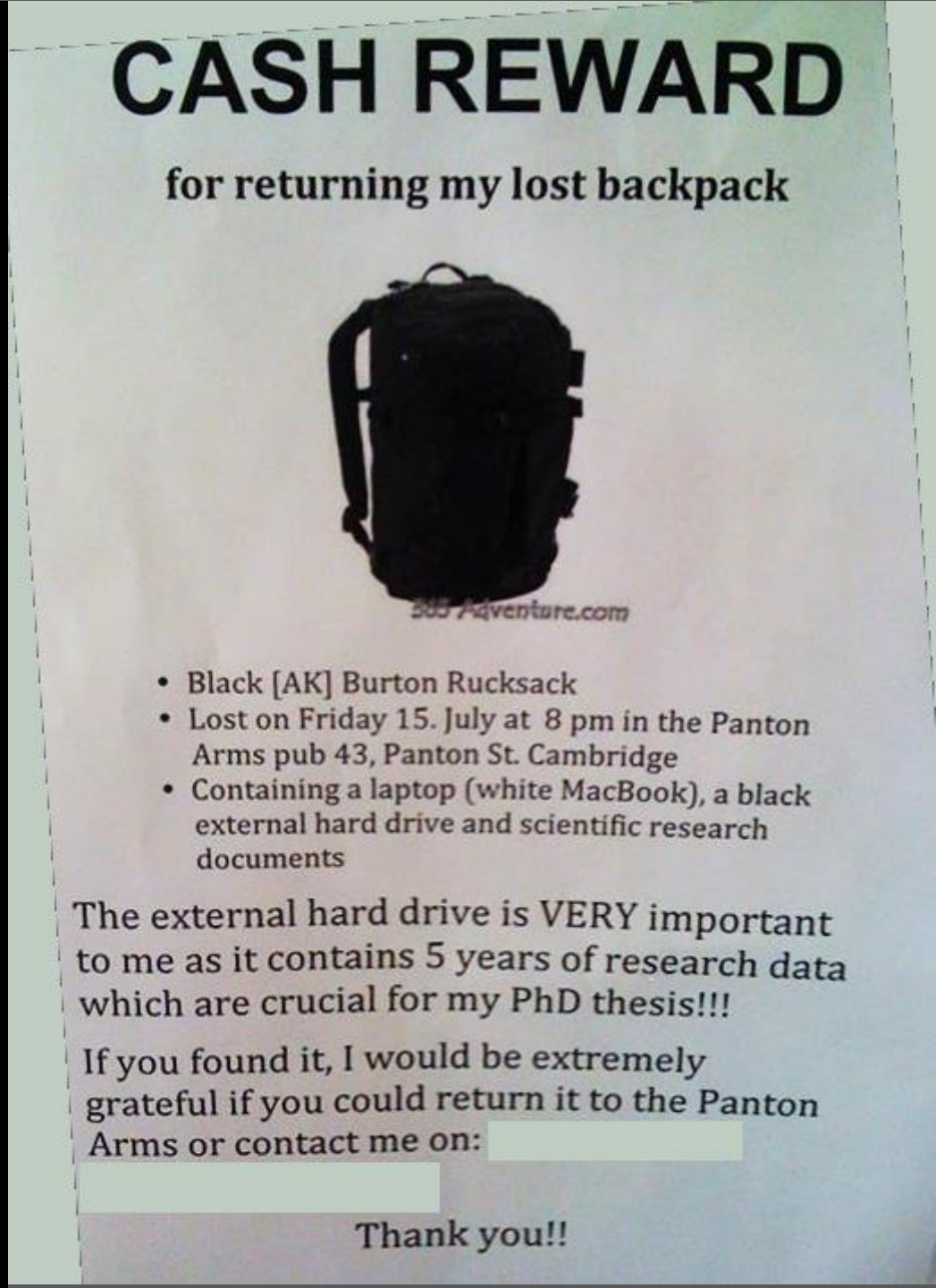
Friday January 15 2021, 12.01am, The Times

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?

"Availability of research data declines rapidly with article age"--Vines et al.

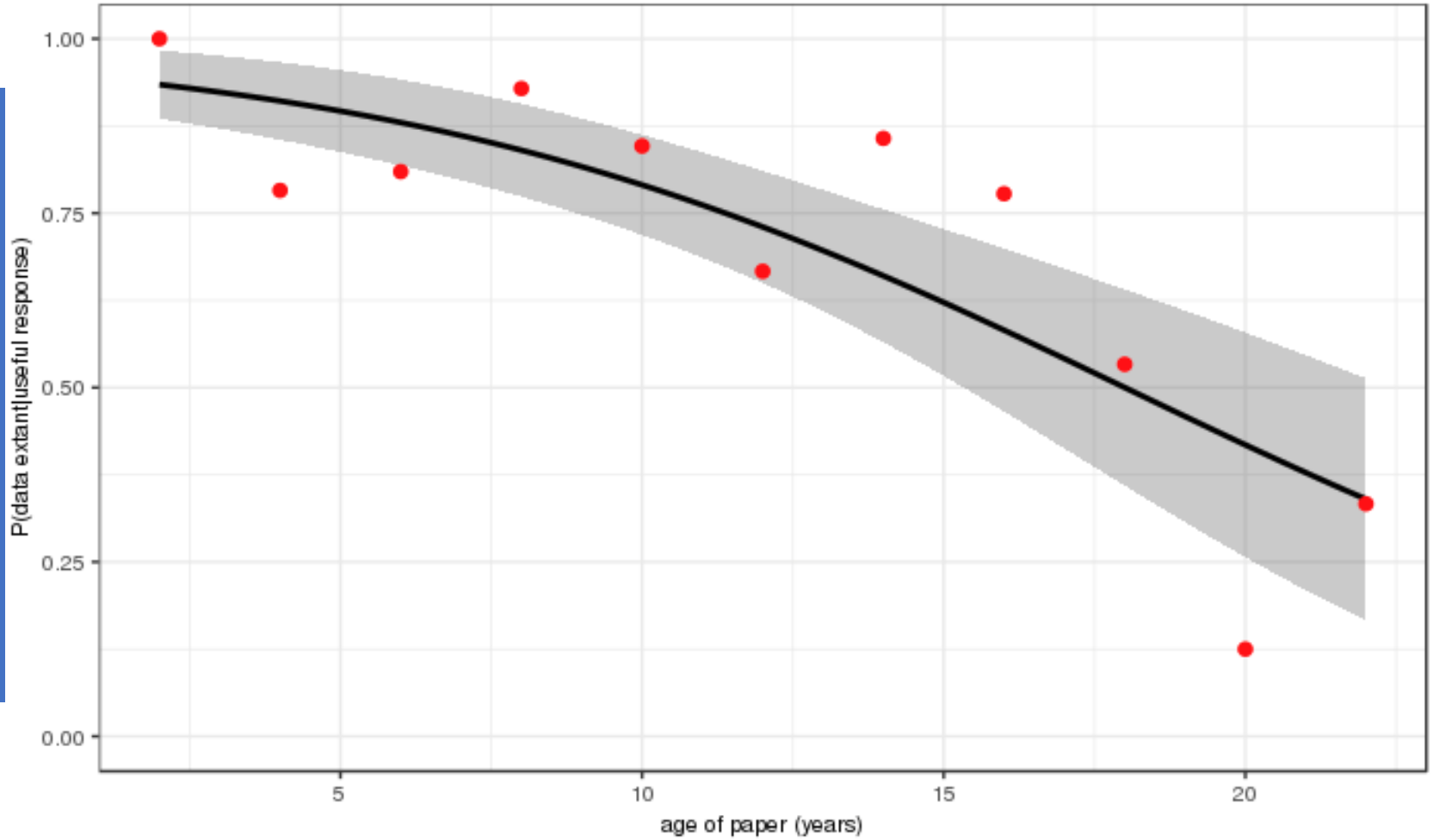
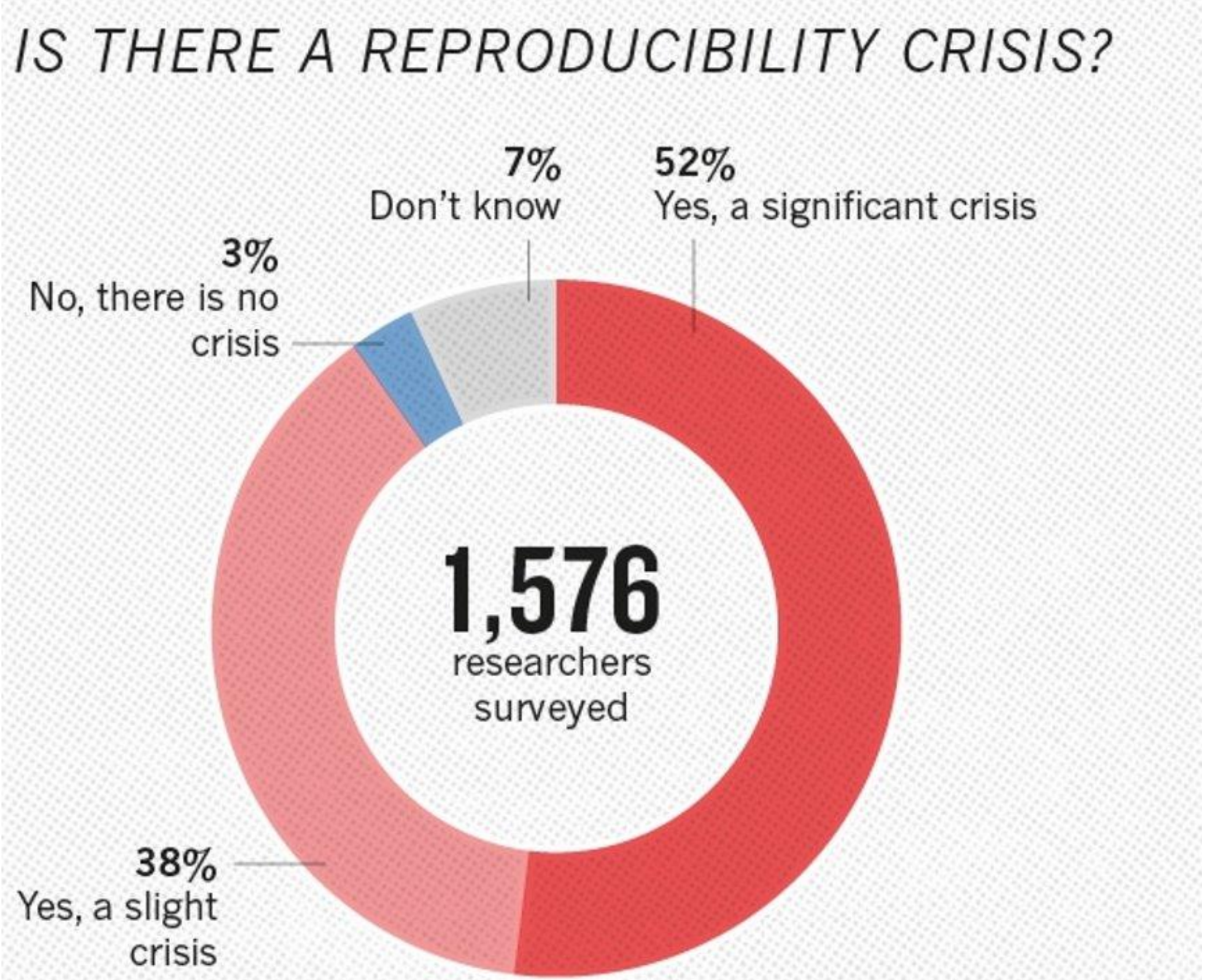


Image created by N Syrotiuk from the orignal research data and R script.



Source: Baker's paper published in 2016.

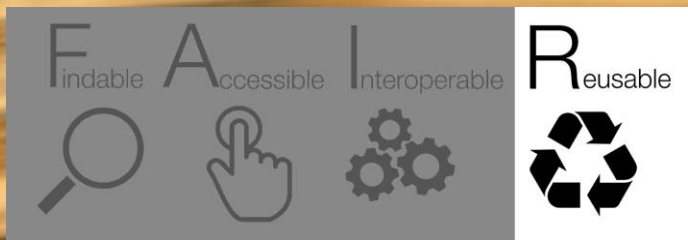
Part two:

RDM CORE ACTIVITIES

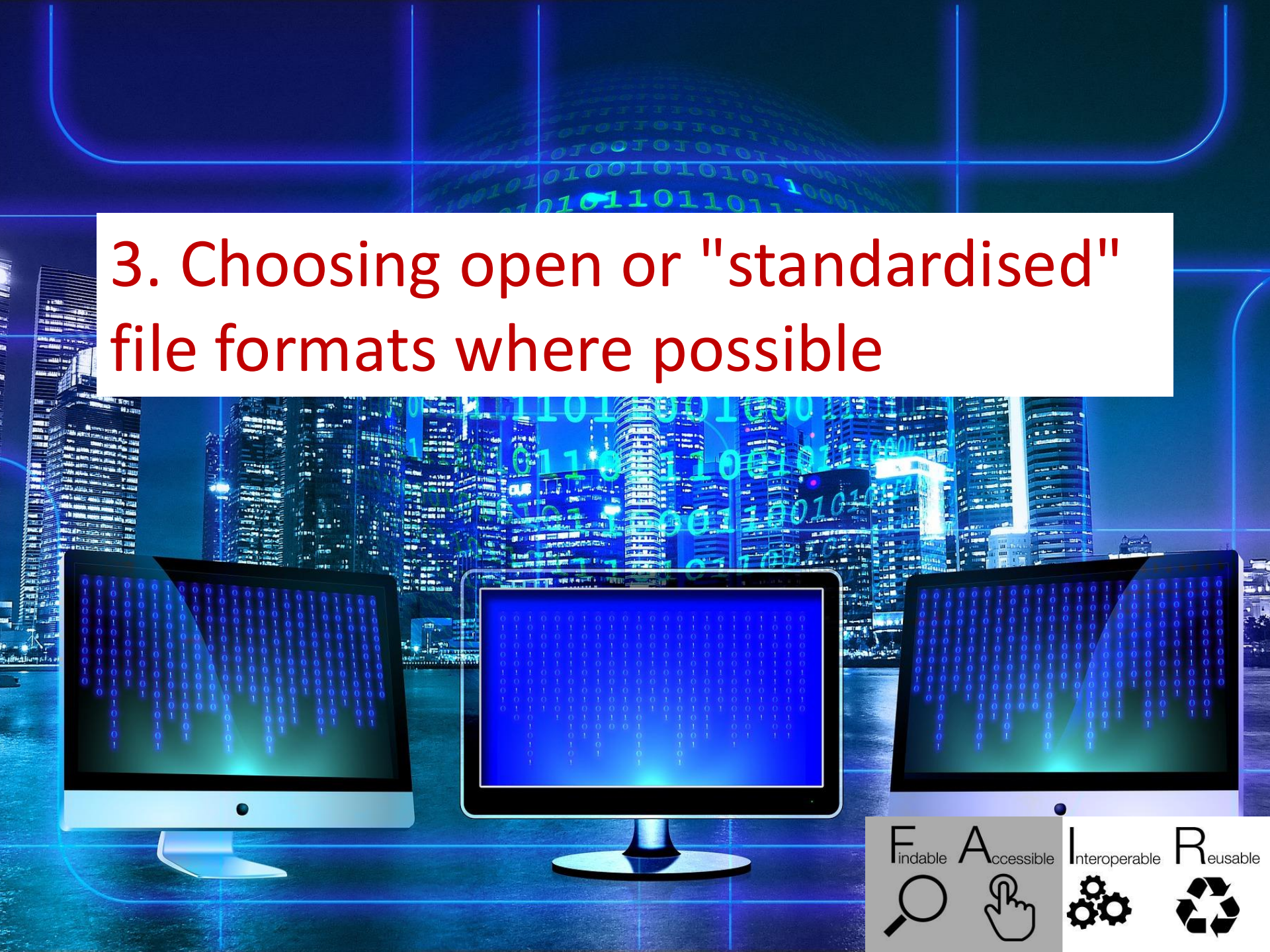
1. Planning and describing data-related work before it takes place

Findable Accessible Interoperable Reusable

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



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



4. Storing research data securely during a project


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

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

Common types of research data

- Survey data
- Recordings
- Text documents
- Tabular data
- Photographs
- Image data
- Big, new, novel or voluminous data
- Scientific measurements e.g, NMR data
- 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

Menti.com
Code: 2294 3666

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

Enter a word 25

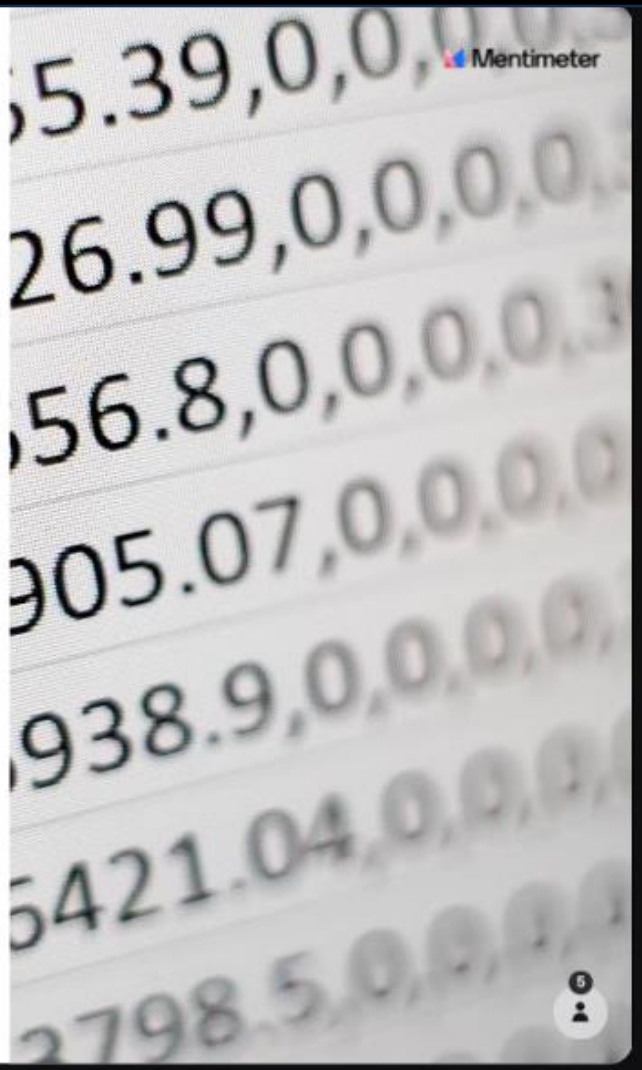
Enter another word 25

Submit

Results of Mentimeter survey

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

11 responses



Part three:

CIS STORAGE OPTIONS:

OneDrive, PRS and Hamilton

Microsoft OneDrive for Business


OneDrive for Business
1 TB of secure, encrypted personal storage

Easy to share with external collaborators

OneDrive for Business


Automatic replication
OneDrive for Business

Personal Research Storage (PRS)

Storage Manager

My Access | Manage Shares | Audit Log

Directories shared with you

Departmental Read Write ul/Staff

University Library Departmental Share

CIFS (Windows): `\\deptblue02.mds.ad.dur.ac.uk\ul\Staff`
CIFS (Mac): `smb://deptblue02.mds.ad.dur.ac.uk/ul/Staff`
CIFS (Linux): `//deptblue02.mds.ad.dur.ac.uk/ul/Staff`

Owners: MATTHEW,EDWIN PHILLIPS , NATALIE COSTELLO

Departmental Read Write ul/Systems

University Library Departmental Share

CIFS (Windows): `\\deptblue02.mds.ad.dur.ac.uk\ul\Systems`
CIFS (Mac): `smb://deptblue02.mds.ad.dur.ac.uk/ul/Systems`
CIFS (Linux): `//deptblue02.mds.ad.dur.ac.uk/ul/Systems`

Owners: MATTHEW,EDWIN PHILLIPS , NATALIE COSTELLO

Personal Read Write dch0ph/RDM

CIFS (Windows): `\\prsblue02.mds.ad.dur.ac.uk\dch0ph\RDM`
CIFS (Mac): `smb://prsblue02.mds.ad.dur.ac.uk/dch0ph/RDM`
CIFS (Linux): `//prsblue02.mds.ad.dur.ac.uk/dch0ph/RDM`
Linux Desktop Service: `/prs/prsblue02/silver3_05/dch0ph/RDM`

Owners: PAUL HODGKINSON

Personal Read Write pzvx49/chromatogram

CIFS (Windows): `\\prsblue02.mds.ad.dur.ac.uk\pzvx49\chromatogram`
CIFS (Mac): `smb://prsblue02.mds.ad.dur.ac.uk/pzvx49/chromatogram`
CIFS (Linux): `//prsblue02.mds.ad.dur.ac.uk/pzvx49/chromatogram`
Linux Desktop Service: `/prs/prsblue02/silver3_07/pzvx49/chromatogram`

Owners: Nicholas SYROTIUK

➔ Manage pzvx49/chromatogram

Shares you manage

Personal pzvx49

CIFS (Windows): `\\prsblue02.mds.ad.dur.ac.uk\pzvx49`
CIFS (Mac): `smb://prsblue02.mds.ad.dur.ac.uk/pzvx49`
CIFS (Linux): `//prsblue02.mds.ad.dur.ac.uk/pzvx49`
Linux Desktop Service: `/prs/prsblue02/silver3_07/pzvx49`

CIS tool for managing PRS

Storage Manager

My Access | Manage Shares | Audit Log

PRS: pzvx49 Manage pzvx49 >

CIFS (Windows): `\\prsblue02.mds.ad.dur.ac.uk\pzvx49`
CIFS (Mac): `smb://prsblue02.mds.ad.dur.ac.uk/pzvx49`
CIFS (Linux): `//prsblue02.mds.ad.dur.ac.uk/pzvx49`
Linux Desktop Service: `/prs/prsblue02/silver3_07/pzvx49`

Type
Personal
Tier
Silver 3
Capacity
1 TB
Current usage
81.2 GB
Usage checked
2024-04-28 05:48
Server
prsblue02
Owners
Nicholas SYROTIUK

Directory: chromatogram Manage chromatogram >

dna sequencing data

Read Write Nicholas SYROTIUK

Directory: prs Manage prs >

Personal Research Storage

Read Write Nicholas SYROTIUK
Read-only RICHARD,IAN HIGGINS

Directory: researchdata Manage researchdata >

Durham University research data

Read Write Nicholas SYROTIUK

Storage Manager. Please contact the Service Desk if you have difficulties using this system.

Personal Research Storage

Not encrypted

Complicated to share with external collaborators

Personal Research Storage

Hamilton supercomputer

Users get 600 GB storage by default

Hamilton supercomputer

Hamilton supercomputer

Can request additional storage

Must back up own research data

Hamilton supercomputer

Not encrypted

Hamilton supercomputer

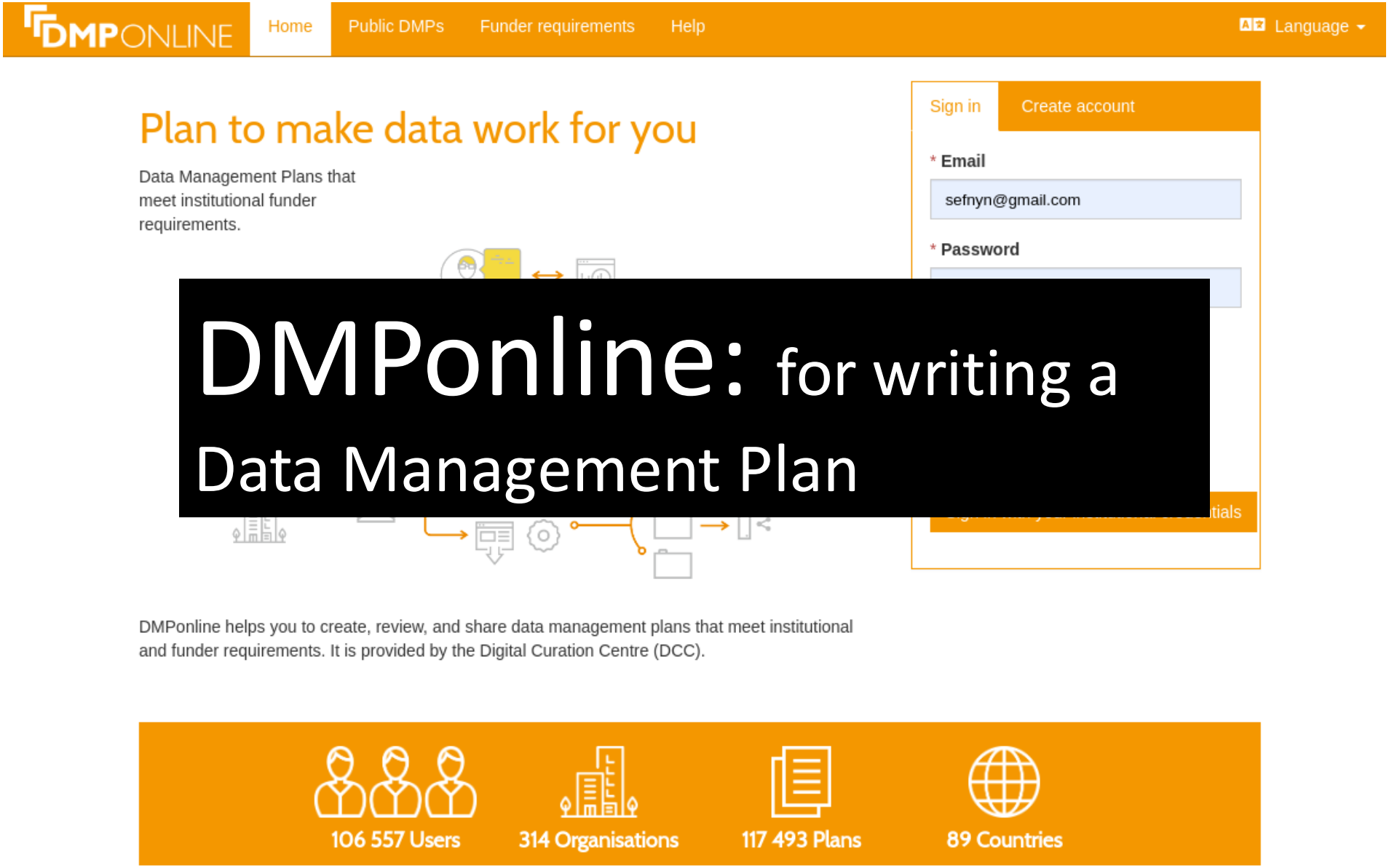

Hamilton supercomputer
Complicated to share with external collaborators

Comparison of storage options

OneDrive for Business	Personal Research Storage	Hamilton supercomputer
1 TB encrypted cloud storage	1 TB storage on site (not encrypted)	600 GB storage on site (not encrypted)
Easy to share with external collaborators	Complicated to share with external collaborators	Complicated to share with external collaborators
Automatic replication	Storage tier: Silver 3 Performance level: Silver Protection level: 3	No backup (full details)
Managed with file explorer	Managed with CIS Storage Manager	Data directory: /nobackup/...

Part four:

RDM TOOLS


DMPONLINE
Home
Public DMPs
Funder requirements
Help
Language
Plan to make data work for you
Data Management Plans that meet institutional funder requirements.
Sign in
Create account
* Email
sefnyn@gmail.com
* Password
DMPonline: for writing a Data Management Plan
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).
106 557 Users
314 Organisations
117 493 Plans
89 Countries

Research and GDPR Decision Tool

To help staff and students navigate through guidance on data protection.

For determining if General Data Protection Regulations apply


IT Service Catalogue
Durham University
Search for services
Home
Who Am I
Login
Storage Options Tool
CIS Storage Options:
for choosing the best short-term storage solution
The tool on
provide adv
appropriate
cases. This
and does n
such as DU
specific rea
solution ma
particular s
recommen
please thin
an option marked as less suitable, especially if sensitive information is involved. Do you currently use solutions flagged as avoid?
If you require more detailed advice please contact the IT Service Desk.
JPEG, etc.)
igh
or more)
Access
I need high-performance access to my files for real-time processing (i.e. no Internet latency)
I need to access my files when off campus
Guidance
Sharing

Collections

Durham University

Enter search terms | Go | All

About | Help | Contact | Login

Research data repository:

for long-term storage of research data

J.P.S. Badyal

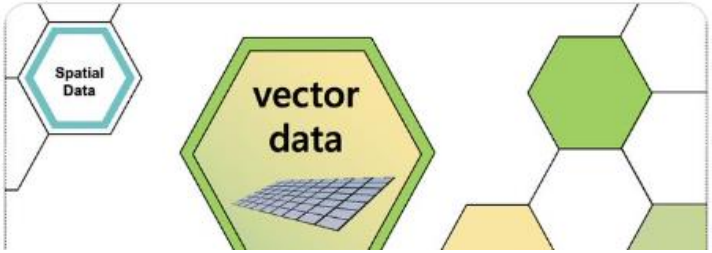
Tea–Essential Oil–Metal...

230108_DATA ARCHIVE Te...

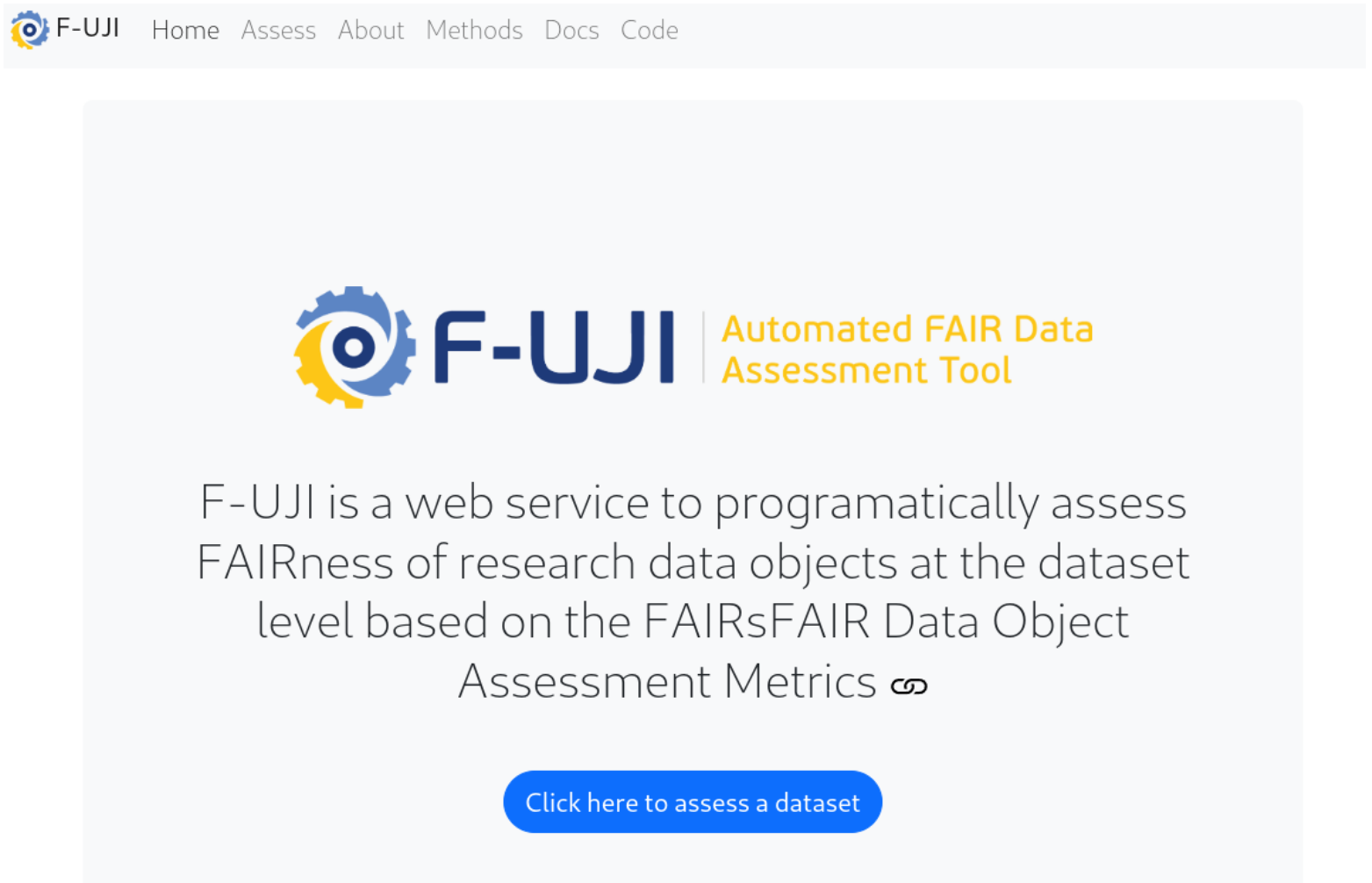
Functional surfaces, Hybrid nanocoating, Tea, Antibacterial materials, Virucidal agents, Coronavirus, COVID-19

A.P.

alike! Details below:



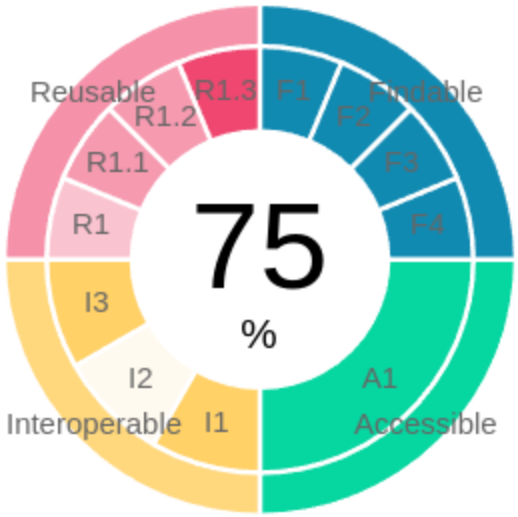
F-UJI: how F.A.I.R. is your dataset?



For DOI:

https://doi.org/10.5281/zenodo.7514328

Summary:



	Score earned:	Fair level:
Findable:	7 of 7	advanced
Accessible:	3 of 3	advanced
Interoperable:	3 of 4	moderate
Reusable:	5 of 10	moderate

Tool	Purpose	Core activity
DMPonline	For writing a Data Management Plan	Planning
Research and GDPR decision tool	For advice on working with personal research data	Planning and Secure storage
CIS Storage Options	For finding a short-term storage solution	Secure storage
Research Data Repository	For long-term storage of research data	Data deposit and Linking
F-UJI	For determining to what extent a dataset is F.A.I.R.	Data deposit

Task two:

SHARE WORK BY MAKING A DATA DEPOSIT


Live
Demonstration

Instructions

1. Browse to: collections.durham.ac.uk
2. Click "Share your work"
3. Login with CIS username
4. Click "Upload data and allocate DOI"
5. Tick "Agree to deposit agreement"
6. Select one file
7. Start upload
8. Apply metadata
9. Save record

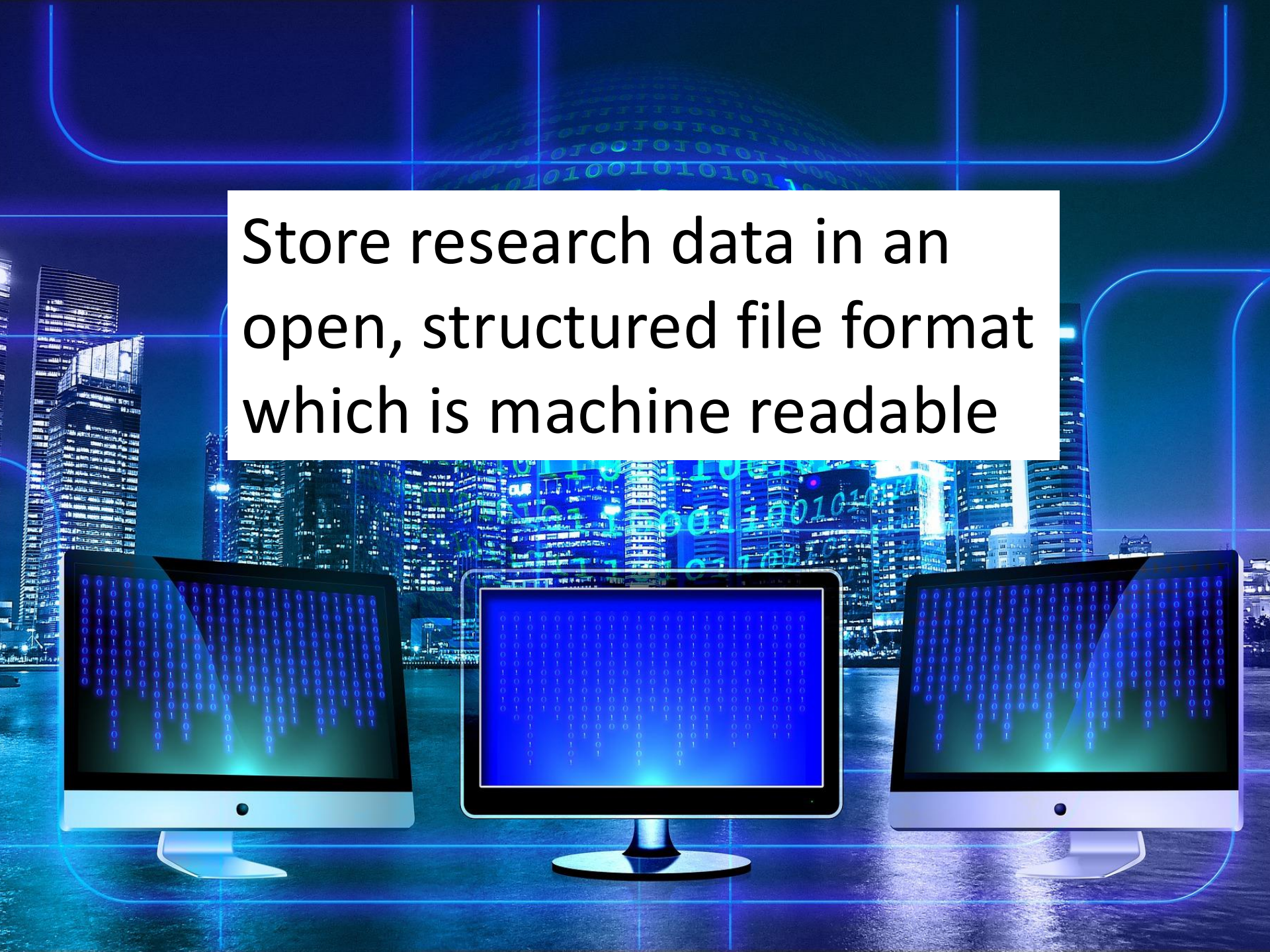
Part five:

RDM BEST PRACTICES

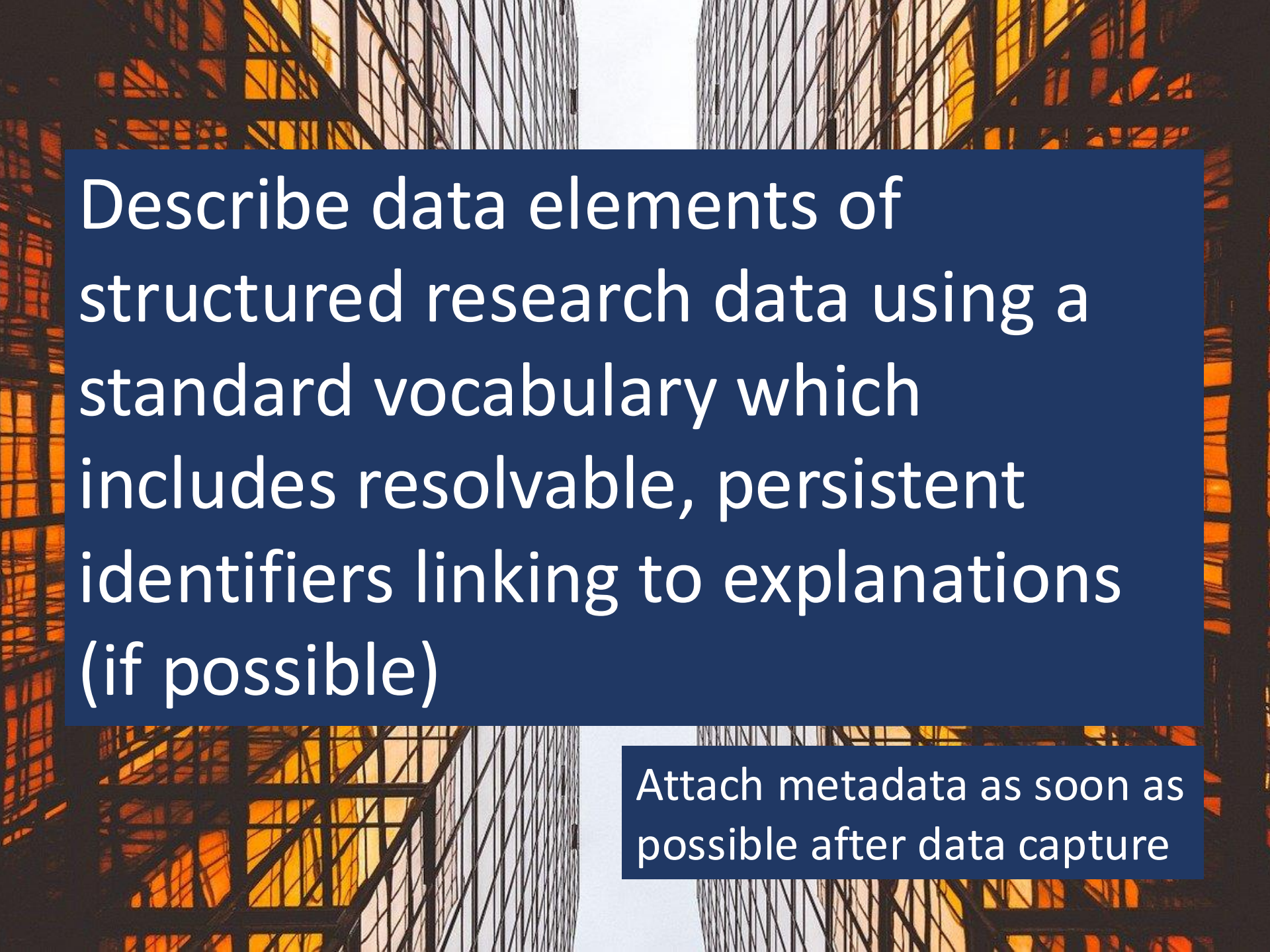

Write a Data Management Plan


Save and backup
raw data

Record all the steps used to process data


Store research data in an open, structured file format which is machine readable


Avoid
proprietary
file formats


Describe data elements of structured research data using a standard vocabulary which includes resolvable, persistent identifiers linking to explanations (if possible)
Attach metadata as soon as possible after data capture

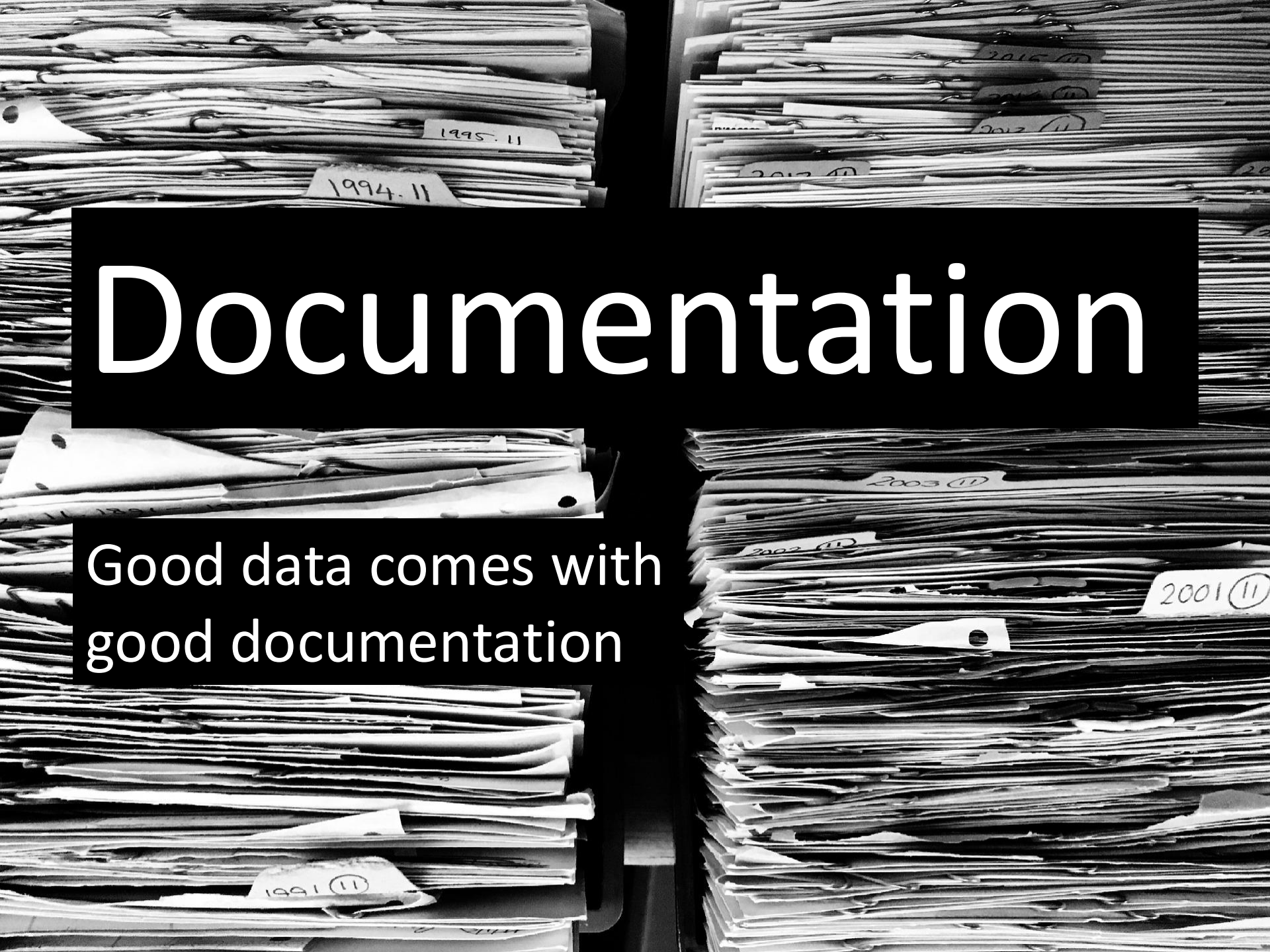
Persistent identifiers (PID)

Example 1: Human genes and genetic disorders
PID: http://omim.org/entry/173900

Landing page: Polycystic kidney disease 1

Example 2: Protein sequencing
PID: http://purl.uniprot.org/uniprot/P98161

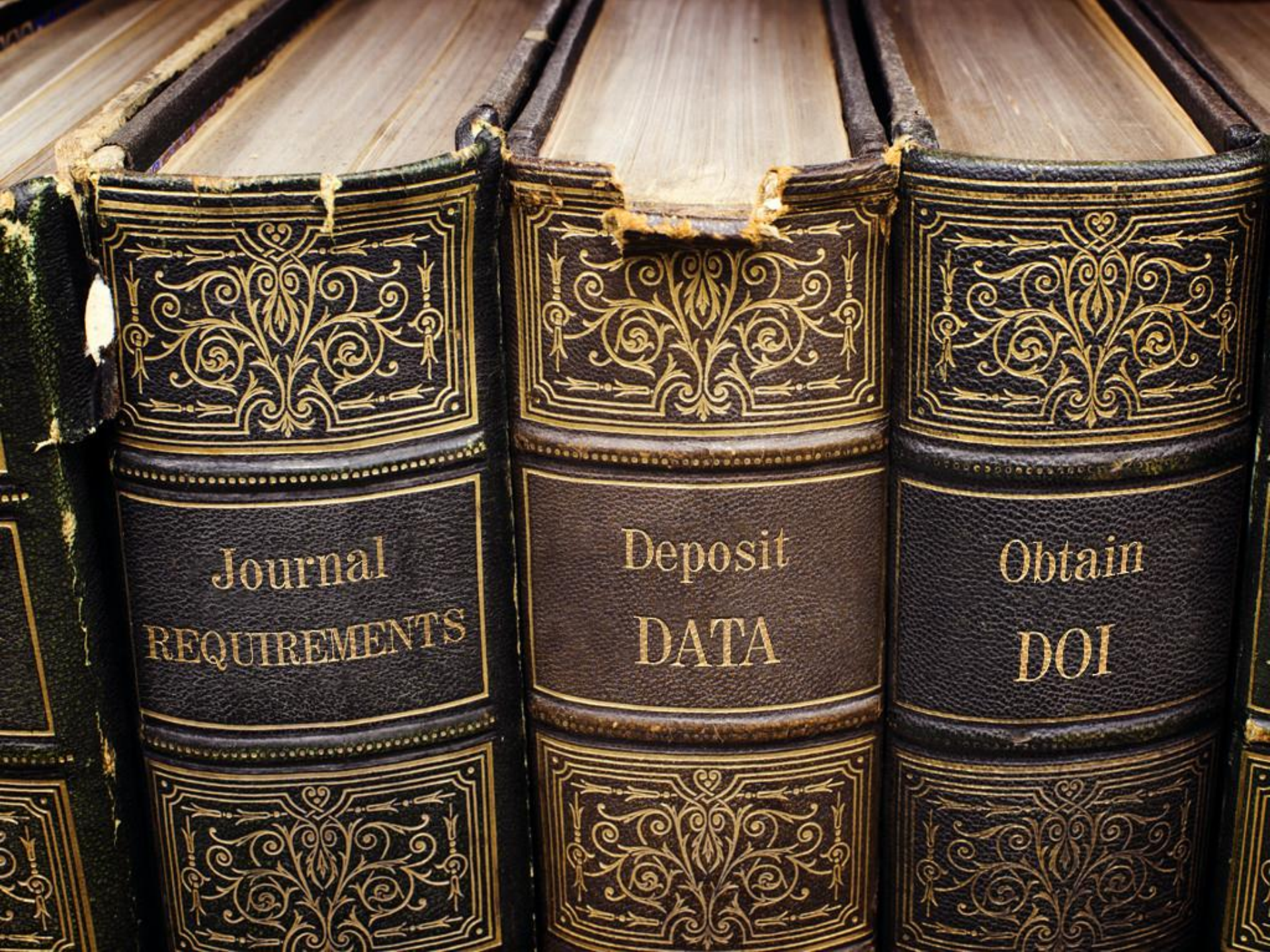
Landing page: The human polycystin-1 protein

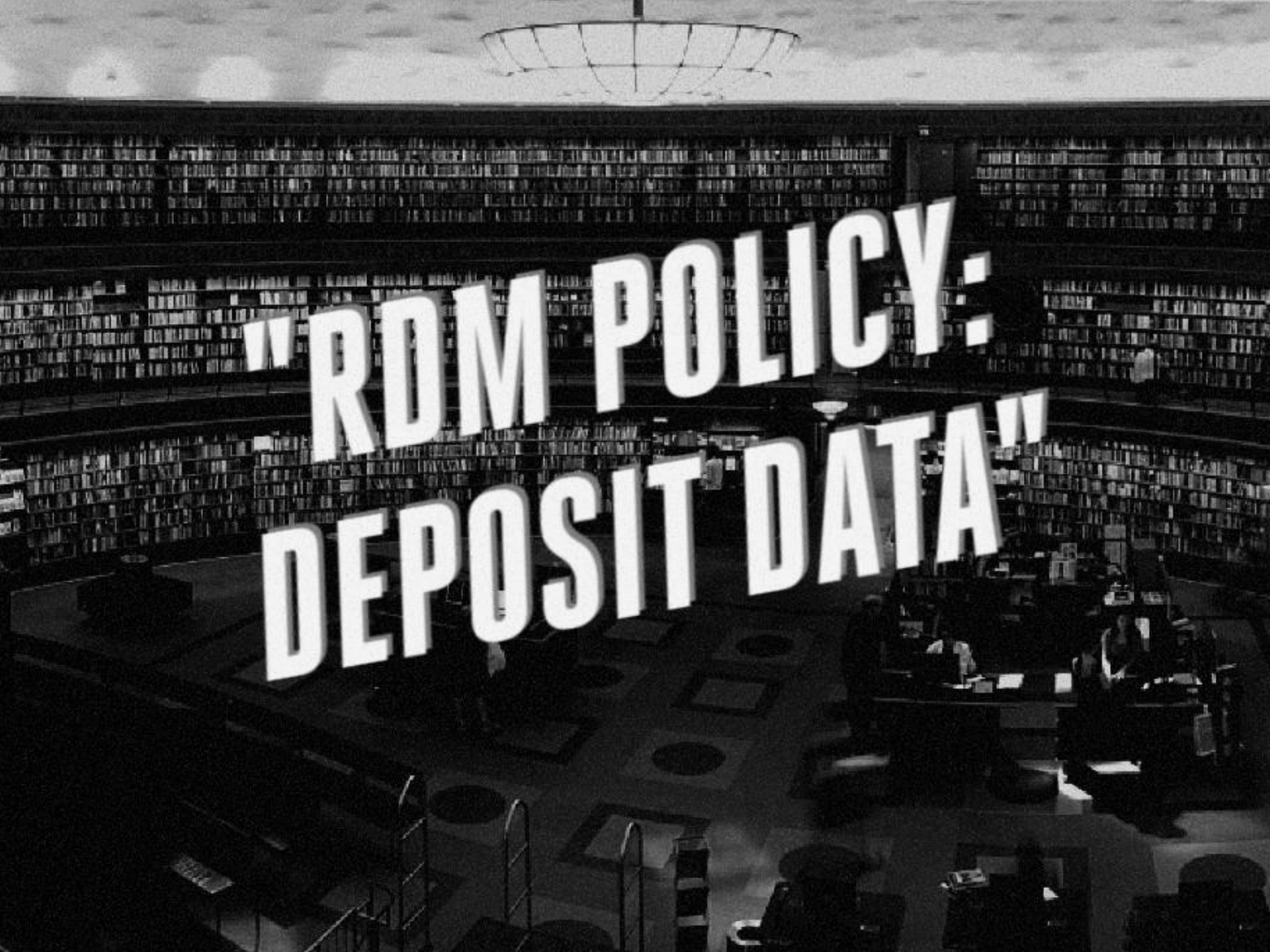


Documentation

Good data comes with good documentation


DEPOSIT DATA
FUNDER
REQUIREMENTS
1234567


Journal
REQUIREMENTS
Deposit
DATA
Obtain
DOI


"RDM POLICY:
DEPOSIT DATA"


Three types of repository
1
Subject-specific
2
Multi-disciplinary
3
Institutional

Benefits of sharing research data

Advances scholarship and improves replicability

Increases impact and citation rate of associated research

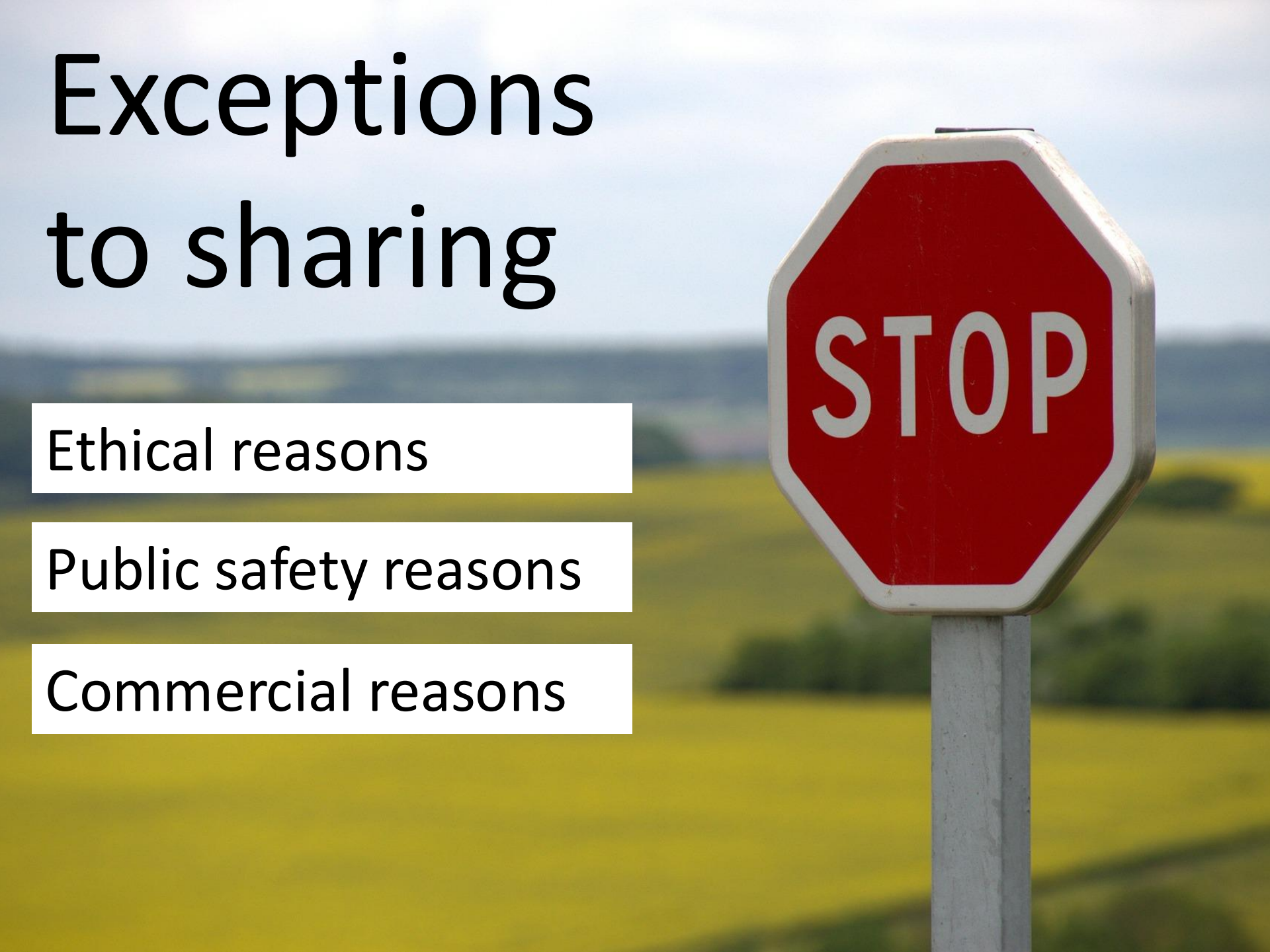
Helps other researchers more easily discover data relevant to their work and reduces duplication of effort

Promotes academic work in new ways

"Publicly available data was significantly associated with a 69% increase in citations ..."

DOI: https://doi.org/10.1371/journal.pone.0000308

Exceptions to sharing



- Ethical reasons
- Public safety reasons
- Commercial reasons

Data access statement: include this in your paper / thesis

Are your research data published openly?

1. Yes. Here's the DOI.

2. Yes, with restrictions as described in the Non-disclosure Agreement.

3. No, the data are too sensitive.

4. No new data were generated.

Part six:

SUMMARY

Basic do's and don'ts of research data management

DO	DON'T
Have a plan for managing research data	Make it up as you go along
Keep backups. Make this easy with automated syncing services like Dropbox, provided your data isn't too sensitive	Carry the only copy around on a memory card, your laptop, your phone, etc
Describe your data as you collect it. This makes it possible for others to interpret it, and for you to do the same a few years down the line	Leave this till the end. The quality of metadata decreases with time, and the best metadata is created at the moment of data capture
Save your work in open file formats, where possible, and use accepted metadata standards to enable like-with-like comparison	Invent new 'standards' where community norms already exist
Deposit your data in a data centre or repository, and link it to your publications	Be afraid to ask for help. This will exist both within institutions, and via national / European support organisations

Research Data Management: Best practice

Without intervention, data + time = no data

Following F.A.I.R. data principles and sharing research data can lead to making more progress as a research community collectively

Working reproducibly and writing good RDM documentation ultimately saves time

Publish (anonymised) research data in a repository in order to preserve it for the long term. Improves transparency and reproducibility. Note exceptions to sharing.

SUPPORT


Durham University
Library and Collections
Library Research Support
University Library and Collections / Guides / Library Research Support: Ope
Sear
Library Research Support: Open Research: Research Data Management
This guide is intended to provide advice and support on open access research, including guidance around Durham Research Online (DRO), open access publishing, research data management and related topics.
Home
What is Open Access?
University & Funder Policies
Research Publications Policy & Rights Retention
Paying for Open Access
Publisher Deals
Durham Research Online (DRO)
REF OA
PGRs & eTheses
Research Data Management
Open Access Week 2023
Monthly Activity & News
SafePod: Secure Data Access
What do we mean by Research Data Management?
Our researchers create, collect and measure large amounts of research data every day, and some of this data is used to support our
https://libguides.durham.ac.uk/open_research/rdm
spirit of open scholarship, researchers are expected to demonstrate they are working transparently and reproducibly. Researche

RDM training

Massive open online course on RDM

07 Nov: Writing a good Data Management Plan

Durham Research Methods Café

Every other Wednesday of term
All PGR/PGT students and DU staff Welcome

Conversation topic: When things go wrong!

28th February 2024, 1100-1230

Discussing what to do when research doesn't go as planned or appears to be failing.

Free Coffee/Tea and pastries!

Check out our Research Methods Café online for methods-related chat/announcements, and online access to the Café Conversations.

Location:
First Floor, Arthur Holmes Building: left of Calman Learning Centre, upstairs, turn left.
Door signed 'DRMC'.

Co-hosts:

- Advanced Research Computing (ARC),
- Durham Research Methods Centre (DRMC)
- Research Support, Durham Library


Thank you
Nicholas Syrotiuk
@DurhamRdm

QUESTIONS

Inkpath

Introduction to Research Data Management

From 22/10/2024 09:30 to 22/10/2024 11:00

--- PLEASE NOW MARK YOUR ATTENDANCE AT THIS ACTIVITY ---

Mobile App Users

1. Open the Inkpath App
2. Press this button
3. Scan this QR Code
4. Confirm attendance Complete

Web Users

1. Open **webapp.inkpath.co.uk**
2. Press this button Record Attendance
3. Enter this unique code

7976866137

4. Press SUBMIT
5. Confirm attendance Complete

References:

Baker, M. (2016): *1,500 scientists lift the lid on reproducibility.* Nature, 533:7604. DOI: http://doi.org/10.1038/533452a

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

Corti, L., Van den Eyden, V., Bishop, L., Woolard, M. (2020): *Managing and sharing research data : a guide to good practice.* 2nd ed. London: Sage.

Piwowar HA, Day RS, Fridsma DB (2007): *Sharing detailed research data is associated with increased citation rate.* PLoS ONE 2(3): e308. DOI: http://doi.org/10.1371/journal.pone.0000308

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

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

References continued:

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

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

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