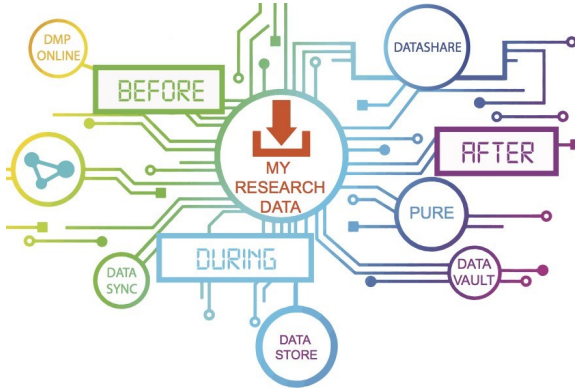


Data Management for your Research

6th October 2023

Simon Smith

Research Data Service

Course Content

1. Principles & Drivers
2. Benefits
3. Data Management Planning
4. Sensitive Data: Ethical Obligations
5. Legal Obligations
6. Active Data Storage & Back-Up
7. Organising and Documenting Your Data
8. Preserving Your Research Data

The Most Essential Essential

What are we talking about?

What does ‘data’ mean anyway?

More like a date?

Definitions

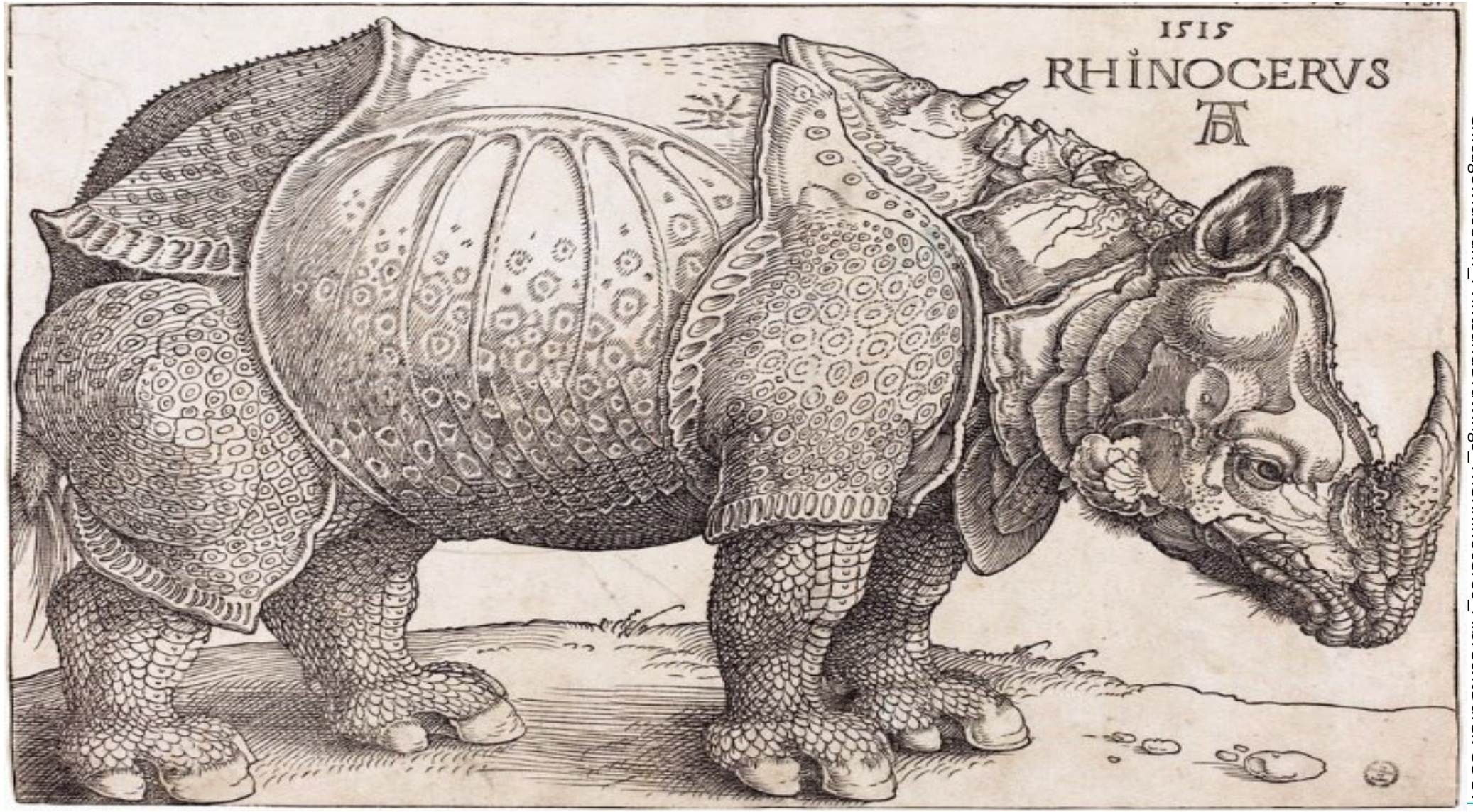
Research Data

Evidence that underpins the answer to the research question: the information necessary to support or validate a research project's observations, findings, or outputs.

Research Data Management

The active management and appraisal of research data over the lifecycle of a research project.

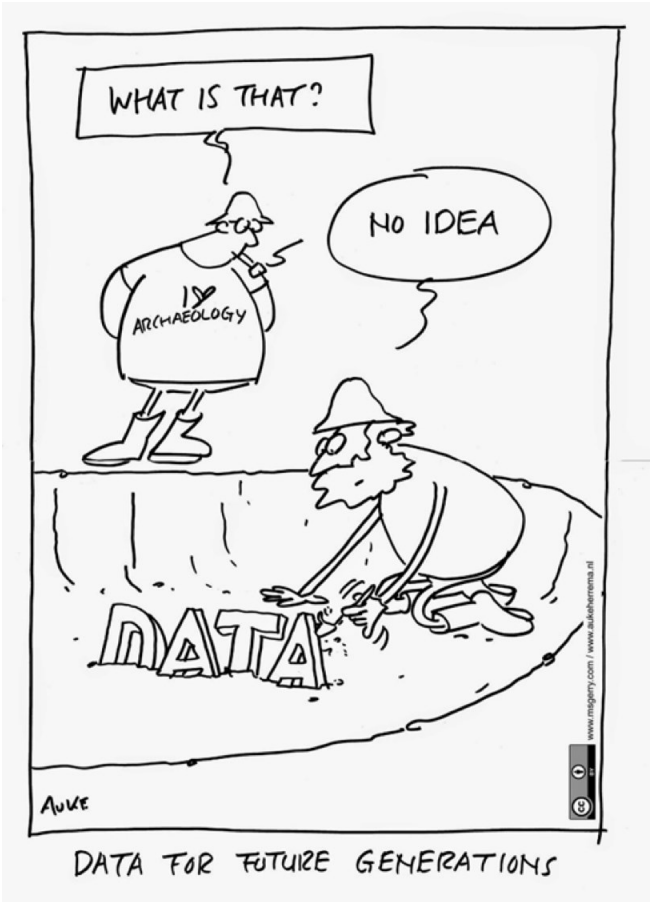
The Elephant in the Room



ChicagoMuseum_PrintsandDrawings_TheRhinoceros_AlbrechtDurer. CC BY

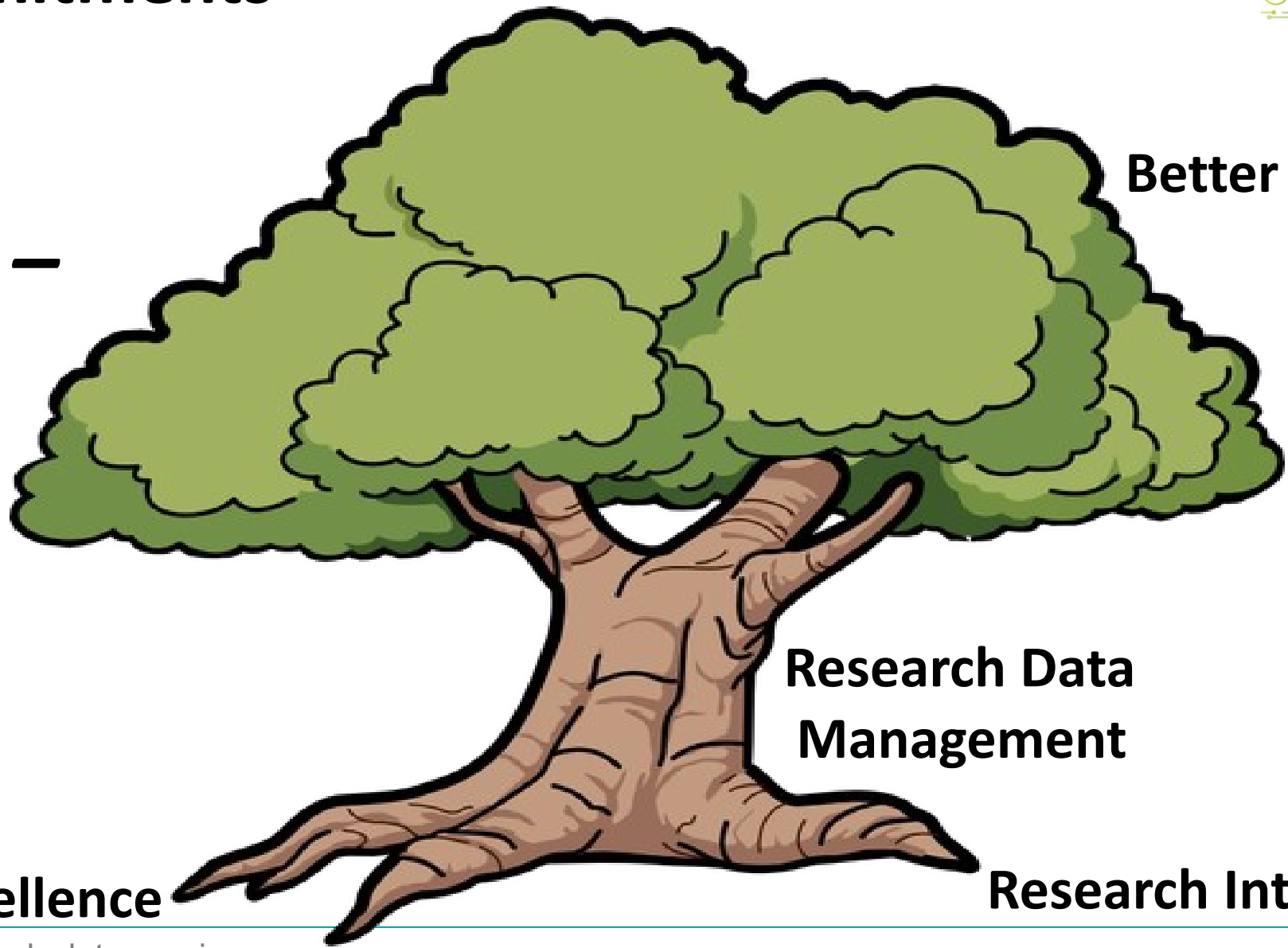
Part 1: Principles & Drivers

To begin at the beginning



Basic Commitments

Data Reposi – Tree



Better Data

Research Data Management

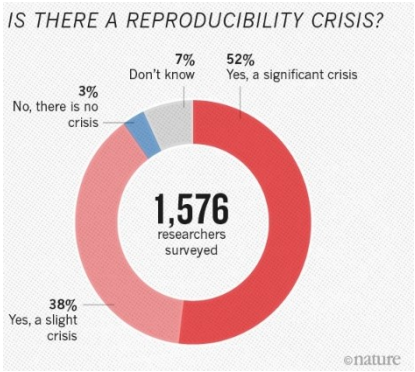
Research Excellence

Research Integrity

Guiding Principles

Integrity and Transparency

- Verifiability
- Reproducibility
- Re-use

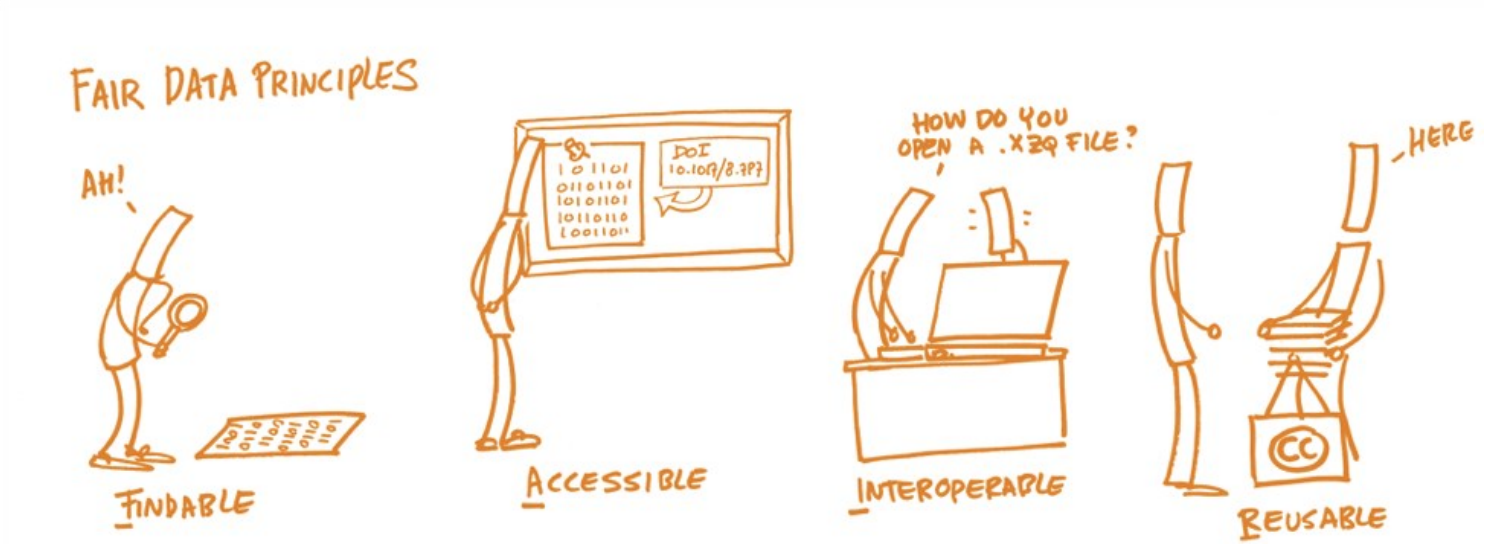


Research Data Are a Public Good

More Guiding Principles: FAIR Data

As open as possible, as closed as necessary

- Findable
- Accessible
- Interoperable
- Re-usable



Plan to make Data FAIR

Drivers


MY DATA
MY TAX MONEY
SOCIETY
MONEY FOR RESEARCH
DATA SHARING

Funders and Publishers

National & International Funders

UKRI

Horizon Europe

Wellcome

Big Publishers

- Elsevier
- Wiley
- Taylor & Francis
- Springer

PLOS Journals:

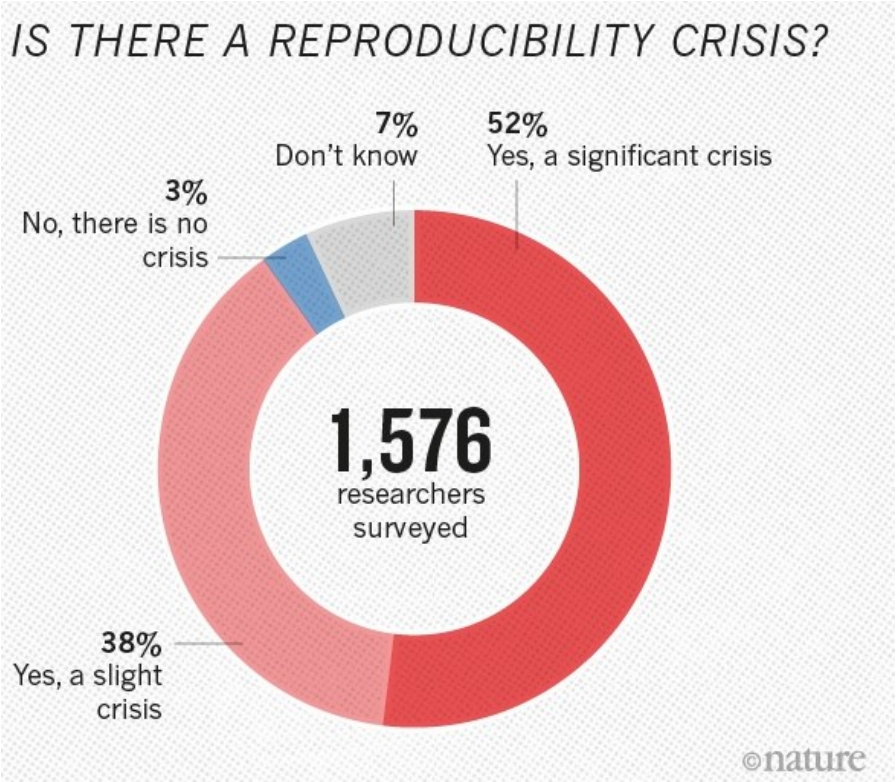
- Make all data necessary to replicate findings publicly available without restriction at the time of publication.
- Specific legal or ethical restrictions prohibit public sharing of a data set: authors must indicate how to obtain access to data.

University of Edinburgh

Research Data Management Policy

- Create a data management plan (DMP).
- Include research data management costs in grant proposals.
- Link datasets to other research outputs with persistent identifiers (DOI and ORCID).
- Don’t give exclusive rights to data to publishers.

Research Integrity



Baker, M. (2016) "1,500 scientists lift the lid on reproducibility", *Nature*, *533:7604*, http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970

Without access to research data, conclusions are at risk!

"… leaving data with authors means that almost all of it is lost over time, unavailable for validation of the original results or to use for entirely new purposes"

– Timothy Vines.

Part 2: Benefits

Why RDM is a Good Idea

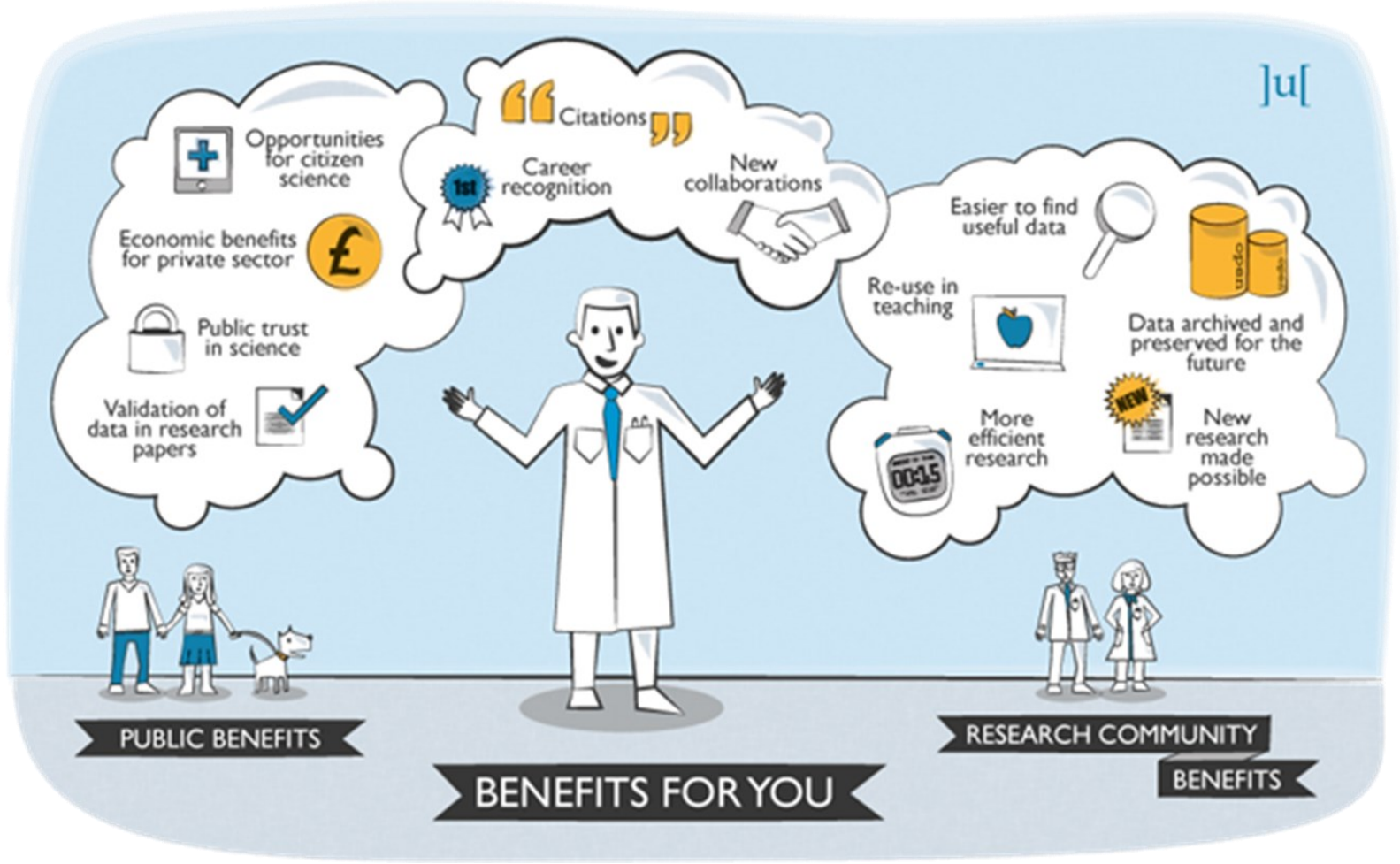


Image credit - Journal of Open Archaeology Data, CC-BY 3.0

Benefits Of Managing Your Research Data

Research Data are a Social Good

- Supports verification and replication
- Speeds up scientific progress
- Meet funder, journal, and institutional requirements
- New collaboration opportunities
- Promotes long-term sustainability of your data

Research Data Management is Good For You

Make it easy to:

- Find and re-use your own data
- Minimise risks
- Manage data quality
- Demonstrate integrity and transparency
- Share your data with other researchers

Research Data Management: the foundation of Good Research

RDM Good Practice

Ensure datasets are:

- Well-structured,
- Well-documented,
- Linked to publications via persistent identifiers (DOI)
- Licensed for re-use

And maximise opportunities for other researchers to

- Build on your research
- Find new uses for your data
- Have great ideas for future collaborations

RDM to Minimise Risks

Data Loss: nothing to see here

No famous case studies of lost research data at Edinburgh!

… because researchers are too embarrassed to admit to it.

But support staff, they *know.*

Data loss owing to hard drive failure and stolen laptops.

Always store data on networked drives: DataStore backed up automatically

RDM For Visibility and Impact

No-one wants to be the invisible researcher

Increase Your Academic Reach

The Open Data Citation Advantage (SPARC Europe, Feb. 2017)

- Sharing data increases citations in a range of disciplines

The Citation Advantage of Linking Publications to Research Data (Colavizza, G., et al., April 2020)

- Linking to data increases citation impact by up to 25%

Exercise: Feeling the Benefits

Scenario:

You are PI on a multi-million pound research project, the results of which will be of significant benefit to the public. Not only will your project make the world a better place, it will also seal your reputation as one of the great minds of your generation.

Nice.

A journalist contacts you to say they're going to publish a story on your research with a global media outlet. You and your research are going to get world-wide attention.

Really nice.

Exercise: Feeling the Benefits

Scenario:

It soon becomes clear, however, that the journalist thinks your research supports every crackpot conspiracy theory going:

The Illuminati, flat earthers, honest politicians.

Your reputation is going to be ruined. The university's reputation is going to be undermined. You will never get funding again.

What are you going to do?

Image by Hognatius: https://www.deviantart.com/hognatius/art/J-Jonah-Jameson-740221284

Exercise: Feeling the Benefits

Options and Questions:

- Publish your data

If not, why not? If yes, how and where?

- But you're 6 months away from publishing your findings.

Publishing your data makes it a citable object.

You can apply access conditions/restrictions until you are ready to publish.

Exercise: Feeling the Benefits

Options and Questions:

- How will you make sure the data is a) findable; and b) understandable?

Organise it and document it

Get a metadata record for it

Get a DOI to use in all your communications

Tweet or blog about it: tell the real story of your data

- But wait! The hard drive on your laptop has failed!

So what? That's not where you stored your data anyway: it's all backed up on the network.

Part 3: Data Management Planning

Start as you mean to go on

Data Management Planning: What? Why? How?

What: thinking about how you will

Organise, document, securely store, and back your data up

Why: to ensure your data is available for the purposes of:

Verification, Replication, Re-use

And to prevent:

Loss and unauthorised access

How: DMPOnline https://dmponline.ed.ac.uk

Writing a DMP for your Research

Hands-on workshop

- Understand the necessity/benefits of producing a DMP!
- Discover how to register for and use DMPonline!
- Draft a basic DMP!

Find course dates and register: https://www.ed.ac.uk/is/data-training

Part 4: Sensitive Data: Ethical Obligations

Getting it right for research participants

How Sensitive is Sensitive?

For our purposes:

- Identifiable individual people ('personal data')
- Rare or endangered species of plants or animals
- Posing a threat to others or to national security
- Commercially sensitive

UK GDPR: Special Category Personal Data e.g.

- Racial or ethnic origin
- Political opinions
- Genetic & Health Data

Obligations and Open Research

Basic Ethical Principles Prevent

- Direct or indirect harm to participants/subjects
- Breaching confidentiality agreements
- Breaching contractual arrangements
- Reputational damage: you and the University
- Breaking the Law

Working with Sensitive Data

Informed Consent

- Consent for sharing and re-use of data
- Avoid restrictive access conditions (e.g. promising to delete data)

Encryption

- Sensitive data on DataStore
- Hard disk on devices off campus (Required)

Training and Advice

Workshop: Working with Personal and Sensitive Data

- Course dates and registration: https://www.ed.ac.uk/is/data-training

Guidance on the UoE website: https://www.ed.ac.uk/infosec

- Encryption / VPNs
- Securing mobile devices
- Password managers, inc. LastPass
- Secure deletion

Ask your Supervisor or Ethics Committee

Part 5: Legal Obligations

I Fought the Law…

Duty of Confidentiality

Duty of confidentiality in UK common law:

Personal information shared in confidence **must not** be disclosed without legal authority or justification.

For example:

- Valid informed consent
- Overriding public interest
- Statutory basis or legal duty (e.g. by court order)

General Data Protection Regulation (GDPR)

GDPR covers all processing of personal data.

- Personal data: identifiable living persons
- Special category data: race, ethnic origin, politics, religion, genetics, sex life, health.

UK/EU GDPR applies to:

- Anyone in the UK/EU who processes personal data anywhere in the world
- Anyone outside the UK/EU who processes personal data on UK/EU citizens

Data Protection Act 2018: UK implementation of the GDPR.

General Data Protection Regulation (GDPR)

Data Protection Principles

- Process lawfully, fair and transparently
- Minimise amount of data held
- Keep to the original purpose
- Uphold accuracy
- Hold data no longer than necessary
- Ensure data integrity and confidentiality
- **Accountability!**

Training and Advice

On Learn:

- Data Protection Training
- Data Protection Training for Research
- Information Security Essentials

Data Protection Officer, Rena Gertz: dpo@ed.ac.uk

More information: https://www.ed.ac.uk/data-protection

Information Security: https://www.ed.ac.uk/infosec

Part 6: Active Data Storage & Back-Up

Keeping your data safe

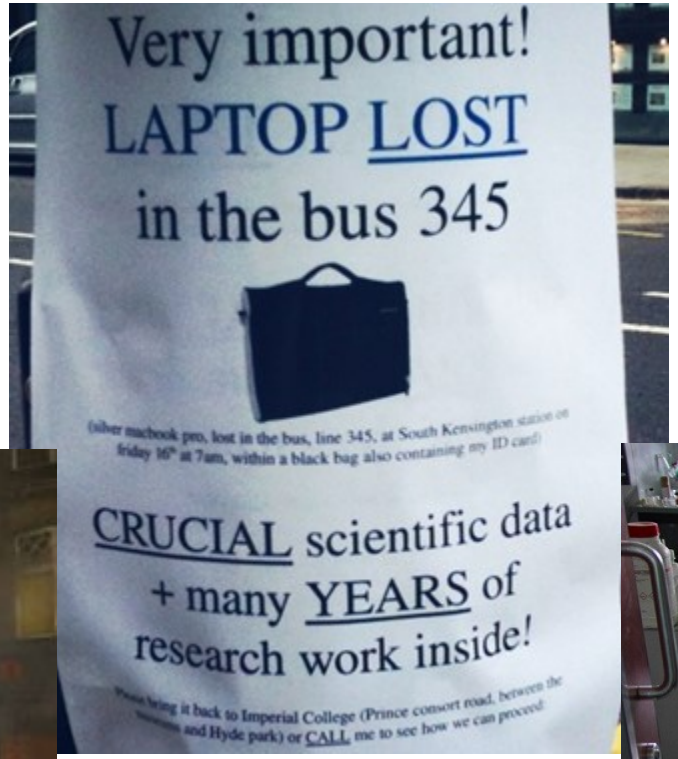
What Do We Mean By Active Data?

Active data is research data that is being

- Collected
- Created
- Processed
- Analysed

Active data is all the data you need to access quickly and easily.

Risks of Off-Network Storage



Basic Principles of Good Storage & Back-up

Do use managed networked services to ensure:

- Regular back-up
- Data security
- Accessibility

Don’t use portable storage media / devices to avoid:

- Losing data
- Unauthorised access
- Quality control problems

Central Data Storage Options

Option 1: DataStore

- Secure network Storage
- Multi-site back-up and disaster recovery
- Sensitive Data Requires Encryption

Option 2: OneDrive for Business

- Secure Cloud Storage
- Automatically Encrypted

Git at the University of Edinburgh

GitLab Community Edition (CE) environment

- Fully supported
- Version control tool
- Store and track changes to code and other documents.

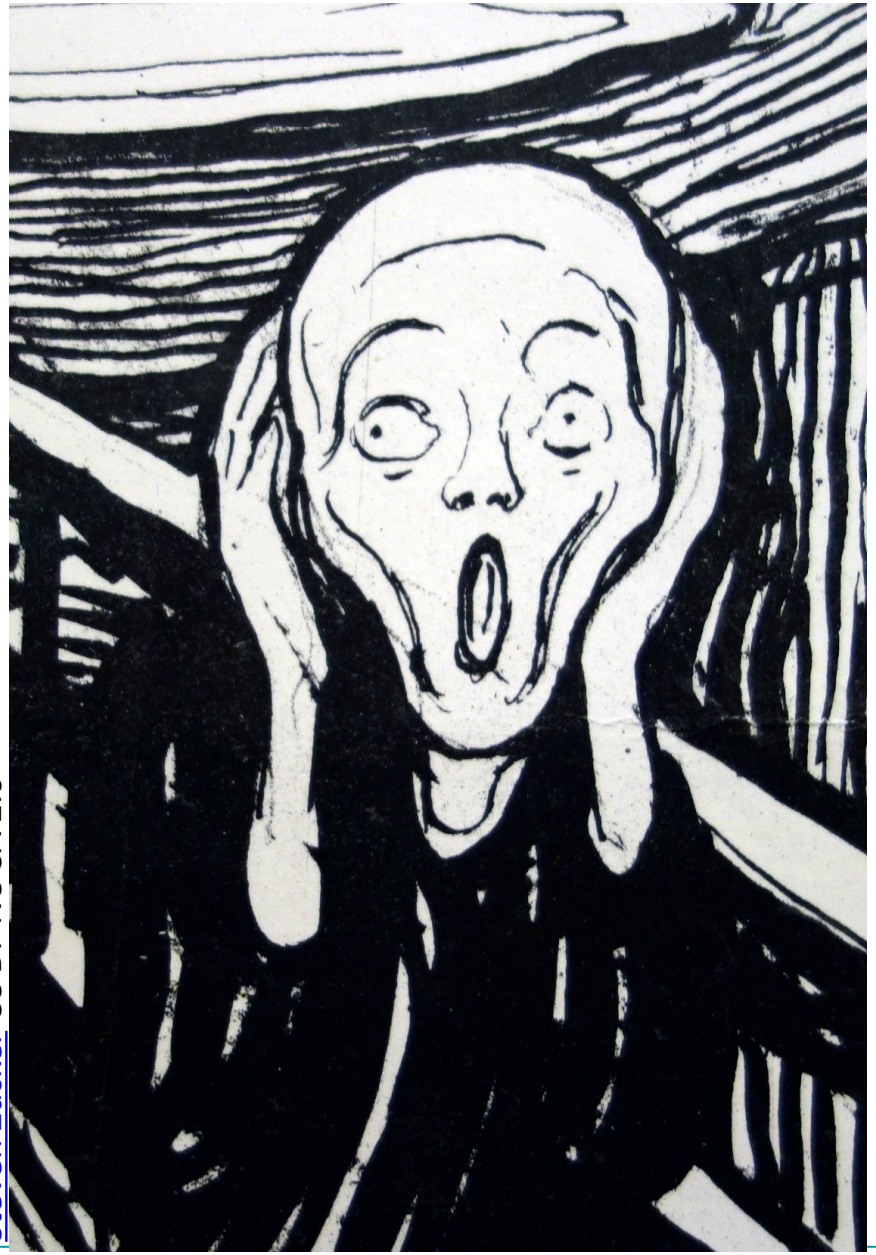
Access: https://git.ecdf.ed.ac.uk/

Documentation:
https://www.wiki.ed.ac.uk/display/ResearchServices/GitLab

Part 7: Organising & Documenting Your Data

In 2013, researchers in Minnesota found evidence that untidiness promotes creativity.

=



Steven Zucker CC BY-NC-SA 2.0

How to Lose Data or Render it Useless:

Human Error!

- Accidental deletion or misplaced files
- Overwriting and versioning
- Poorly described metadata
- Lack of documentation
- Defunct file formats

Organise and Document Your Data

To ensure that data can be

- Found
- Accessed
- Understood / Interpreted
- Analysed

Good For: Verification / Replication / Re-use

- Without the data, your findings cannot be verified
- If your findings cannot be verified, your work may be discredited

Organise and Document Your Data

To Help Out Future-You:

- Document data routinely and consistently
- Choose sustainable and interoperable file formats
- Use meaningful file-naming convention and versioning
- Implement a logical organisational structure for directories and files
- Clearly identify data at different stages: raw, working, and final

A File by Any Other Name…

Name that File:

- No Special Characters:
  “”£$%!¬&*^()+=[]{}~@:;#,.<>
- Underscore Spaces: _
  or %20 may replace the space
- Reverse Dates for sorting:
  YYYYMMDD
- Version Control:
  Date_FileName_v1, _v2, _v3

File Naming:

Honza Nedoma, CC BY-SA 4.0

Short but Meaningful

File Naming for Audit Trailing

File names can contain useful information:

Compare:

- File name 1. interview02.docx
- File name 2. I_P02_R01_20180731.docx

File Name 2 includes:

- I = interview (type of data)
- P[n] = participant ID - participant 02
- R[n] = researcher ID – researcher 01
- Date of interview (YYYYMMDD) – 20180731

Where's Data?

Discovery Metadata Makes Your Data Findable:

Metadata Record in Pure for Published Datasets

- Title
- Creator
- Description
- Collection Dates
- Subject
- Location
- Access Rights
- Keywords

ReadMe Files: How Easy?

ReadMe File: a plain text (.txt) file stored alongside datasets

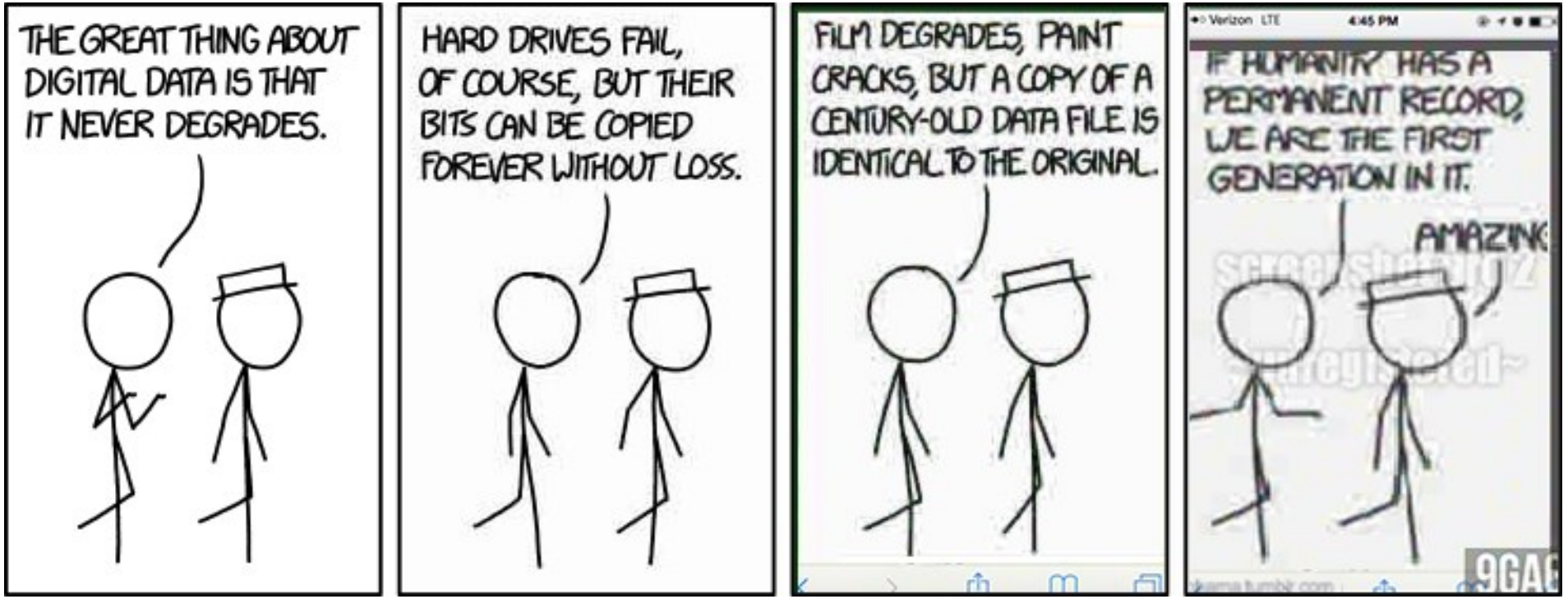
Describes the data: what, where, how, when

For Example:

- Collection/creation methodology
- Software and other tools
- Standards / calibration
- Variables
- Formatting
- Definitions, field codes, labels symbols, abbreviations
- Processing information: anonymisation procedures, quality assurance, etc.
- Analysis methods inc. links to code used

Part 8: Preserving & Sharing Your Research Data

Preserving is not just for pickles



https://xkcd.com/1683/

Digital Preservation & Data Sharing

Digital Preservation:

Ensures digital content remains:

- Alive
- Discoverable
- Accessible
- Usable

Data Sharing:

Ensures digital content remains available for the purposes of

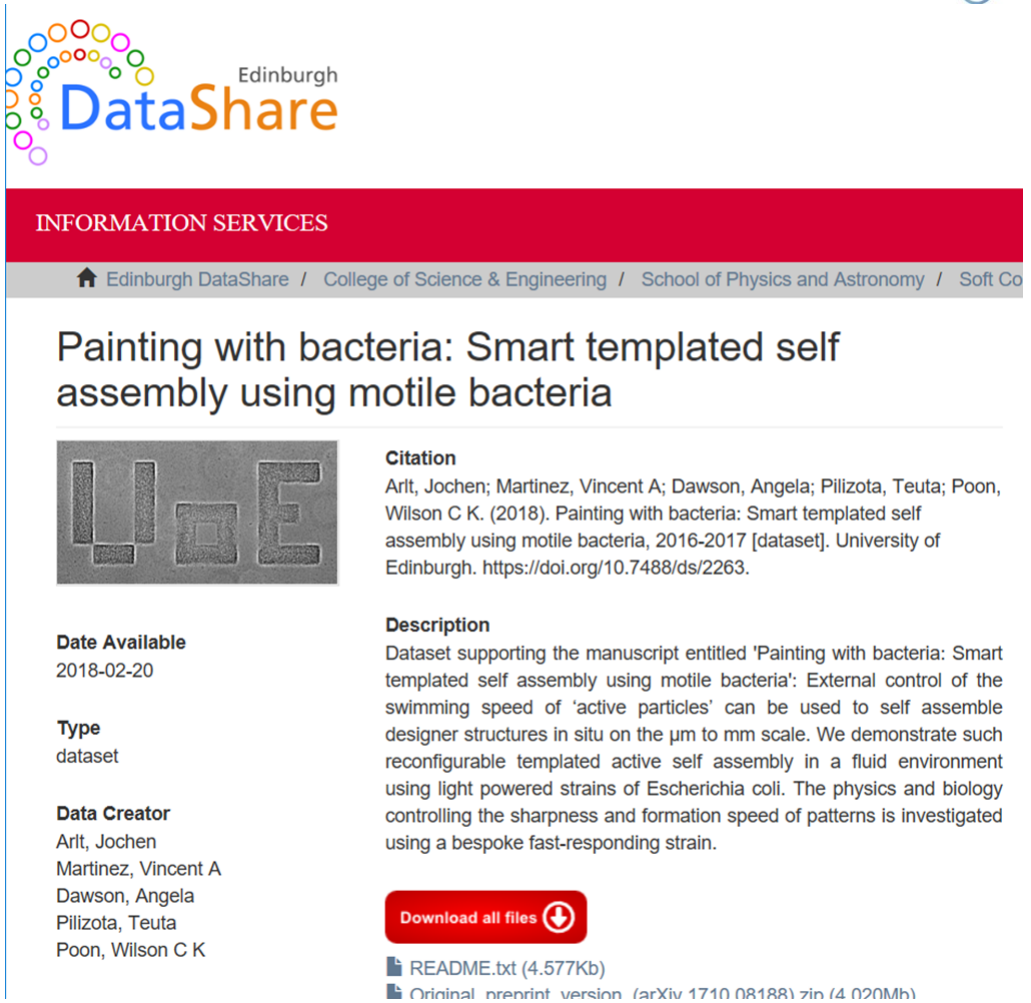
- Replication
- Verification
- Re-use

Support with Sharing Your Data

The University has a trusted data repository:

Edinburgh DataShare

https://datashare.ed.ac.uk



Edinburgh DataVault

A service to archive golden copy research data

- Large data (multiple terabytes)
- Sensitive data
- Secure system
- Restricted access to authorised UoE users
- Low cost
- Chain of custody and review process
- Discoverability via Pure

DATA VAULT

The deposit process has multiple steps, so allow plenty of time!

Other Repositories Are Available

zenodo

Search

Upload Communities

Log in Sign up

Featured communities

Need help uploading? Contact us

Transform to Open Science

Browse New upload

Transform to OPen Science (TOPS) is a $40 million, 5-year mission, led by NASA's Science Mission Directorate's Open-Source Science initiative. Within the TOPS mission, NASA is designating 2023 as the Year Of Open Science, a community initiative to spark change and inspire open science...

Curated by: nasatransformtoopen

Recent uploads

April 19, 2023 (v1.18.4) Software Open Access

Flowminder/FlowKit: 1.18.4

Jonathan Gray; maxalbert; James Harrison; Thingus; dependabot-support; Bhavin Panchal; Dan Williams; OwlHute;

View

Need help?

Contact us

Zenodo prioritizes all requested related to the COVID-19 outbreak.

Training

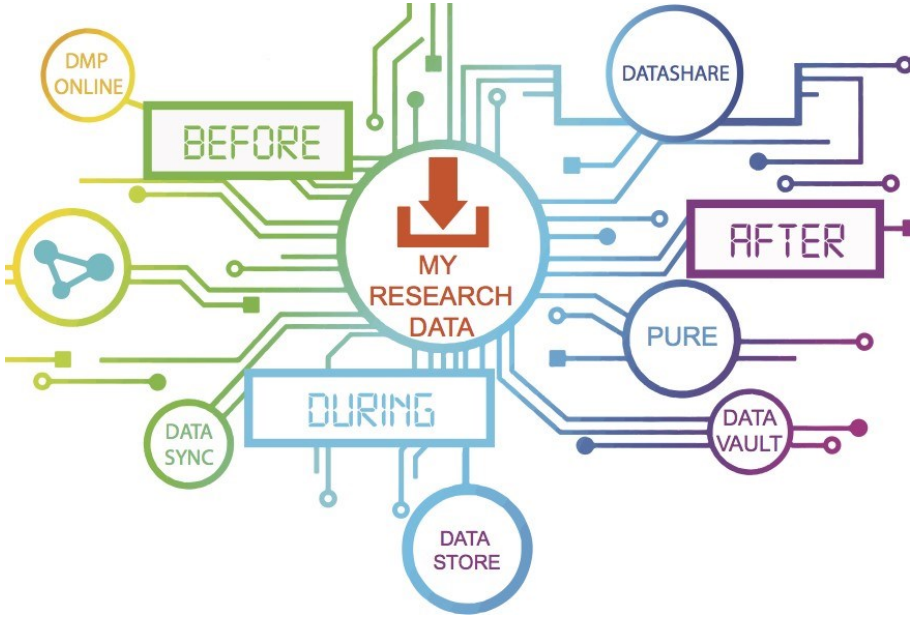
New course from the Research Data Service:

"Archiving your Research Data" featuring the University's Digital Archivist.

- Discover how to digitally preserve your data!
- Learn how to choose and use a research data repository!
- Explore how to make your data FAIR!

Find Course dates and register: https://www.ed.ac.uk/is/data-training

The End



General RDM queries & requests for help writing a DMP should be sent to data-support@ed.ac.uk

RDM website: http://www.ed.ac.uk/is/research-data-service

Training Courses: www.ed.ac.uk/is/data-training

RDM blog: https://libraryblogs.is.ed.ac.uk/datablog/