Practical strategies for Research Data Management
6th November 2017
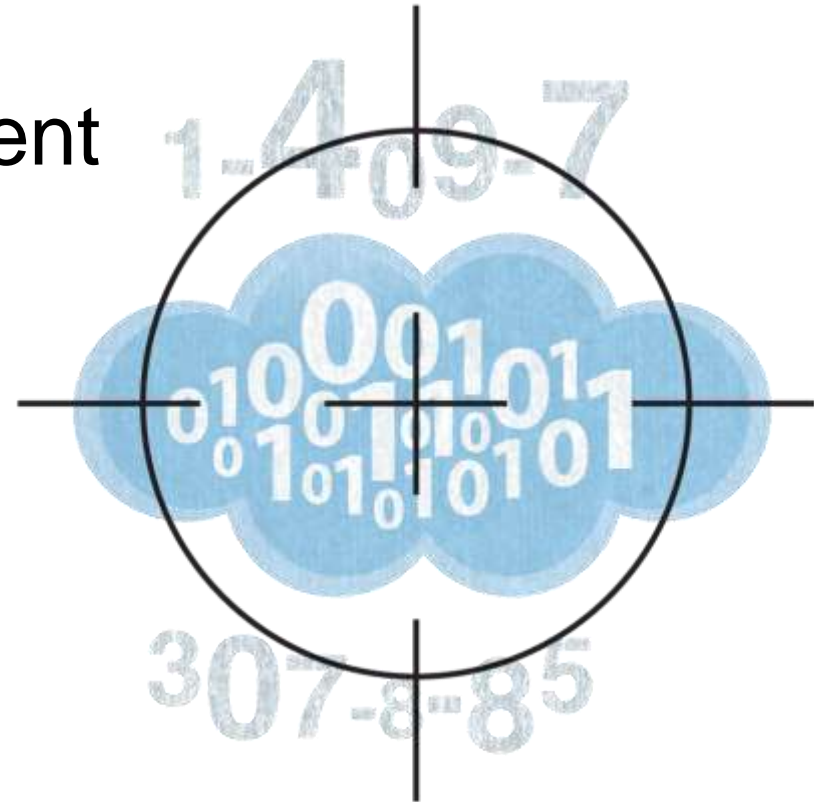
Dan Crane & Isabel Chadwick

Research Support Librarians

library-research-support@open.ac.uk

Overview of the workshop

- What is Research Data Management?
- Working with data
- Planning data management
- Sharing data
- Useful resources
- Questions?

What is Research Data Management?

"Research data management concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. It aims to ensure reliable verification of results, and permits new and innovative research built on existing information."
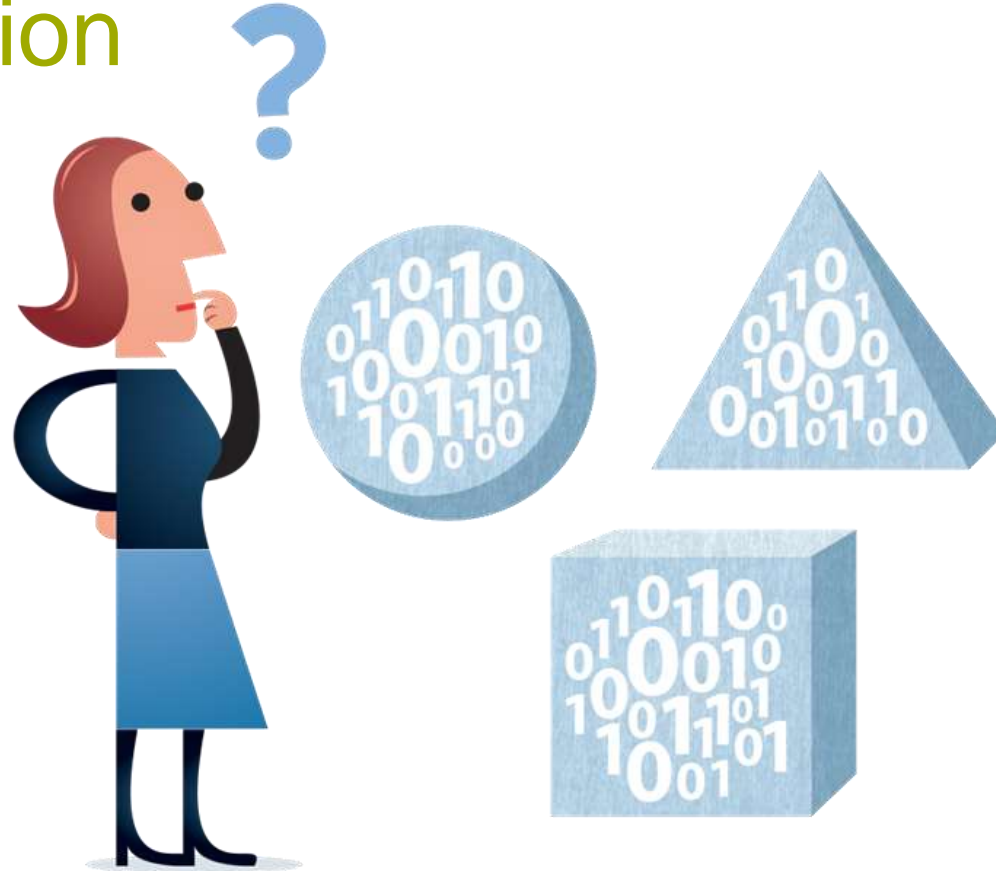
Digital Curation Centre (2011)
Making the Case for Research Data Management
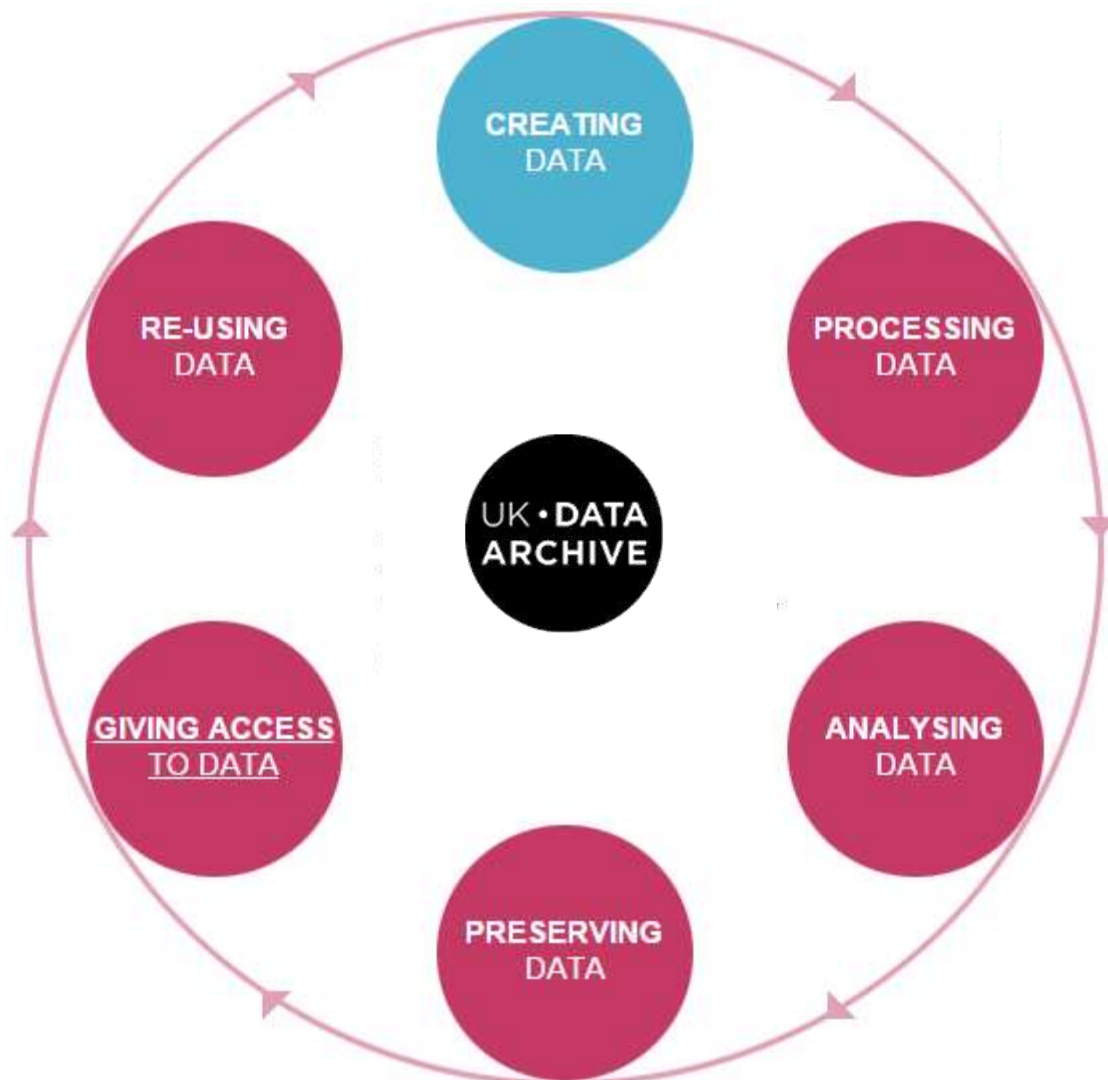http://www.dcc.ac.uk/sites/default/files/documents/publications/Making%20the%20case.pdf

What is Research Data Management?
Discussion

- Describe your research

- What type of data do you create/use?

- What data management challenges do you face?

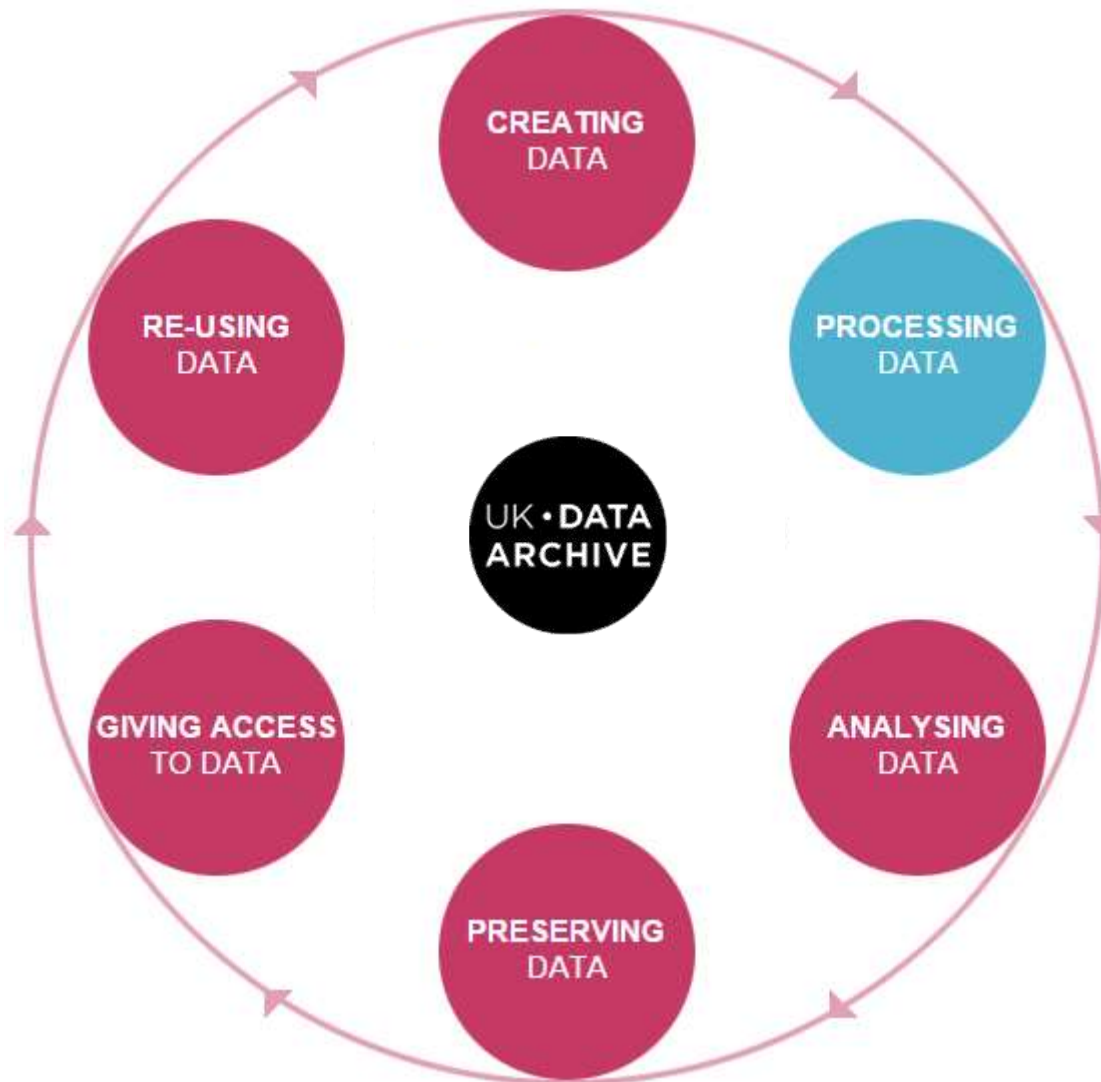# What is Research Data Management?
UK Data Archive Data Lifecycle model

Creating data

❑ Design research
❑ Plan data management
❑ Plan consent for sharing
❑ Locate existing data
❑ Collect data
❑ Capture and create metadata

What is Research Data Management?
UK Data Archive Data Lifecycle model

CREATING DATA

RE-USING DATA

PROCESSING DATA

UK · DATA ARCHIVE

GIVING ACCESS TO DATA

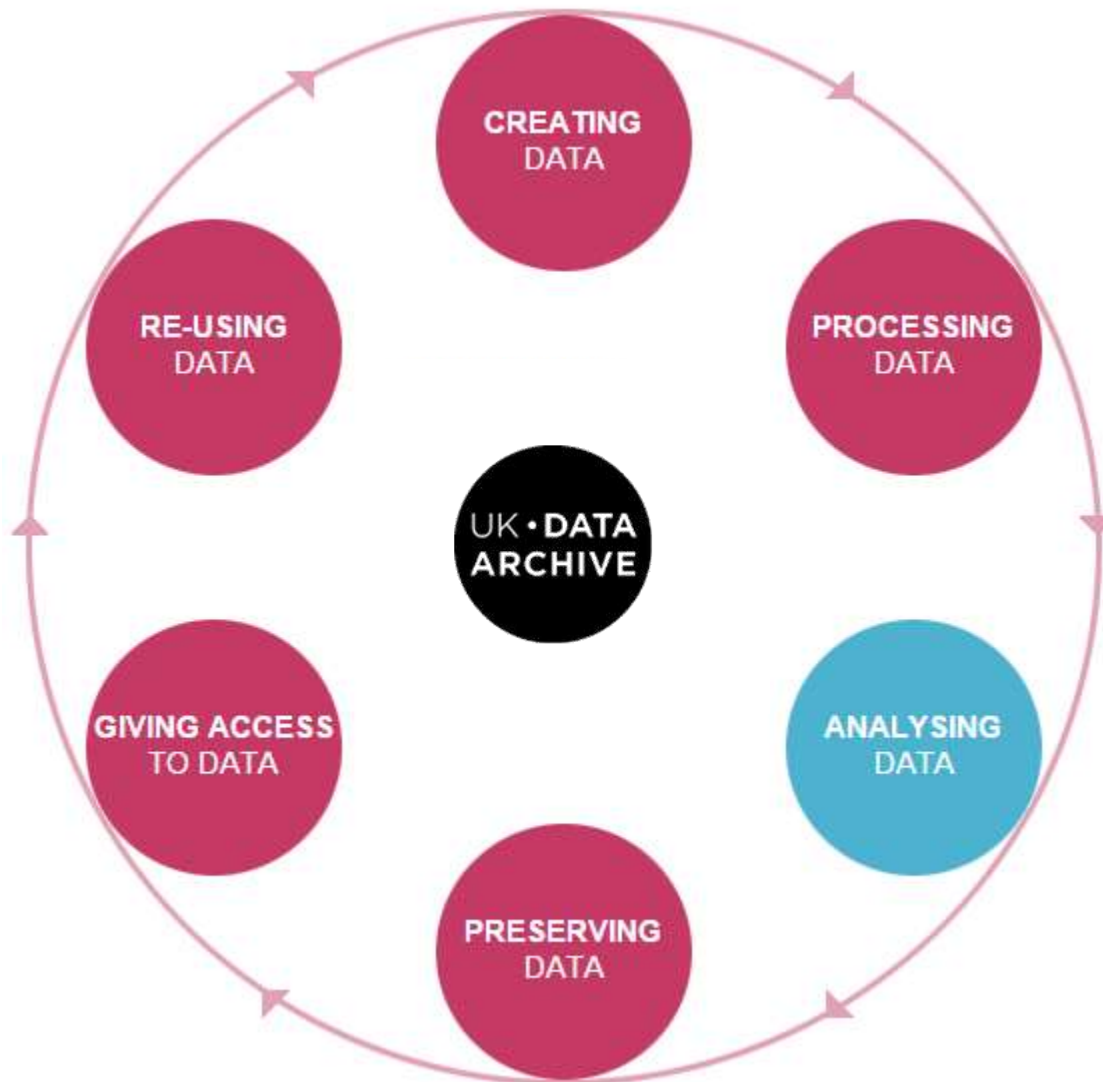ANALYSING DATA

PRESERVING DATA

Processing data

- ❑ Enter data, digitise, transcribe, translate
- ❑ Check, validate, clean data
- ❑ Anonymise data
- ❑ Describe data
- ❑ Manage and store data

http://www.data-archive.ac.uk/create-manage/life-cycle

What is Research Data Management?
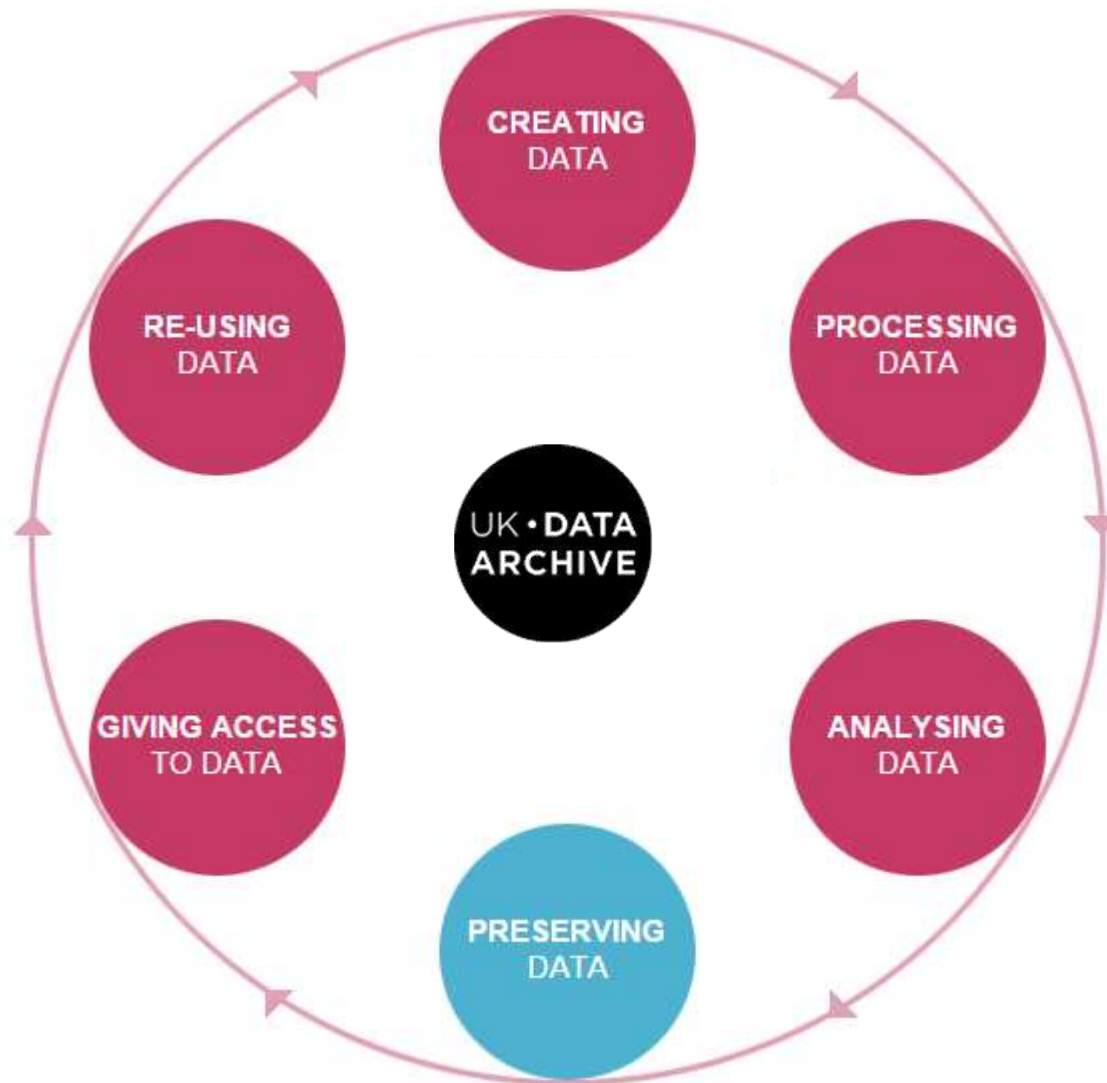## UK Data Archive Data Lifecycle model

Analysing data

- ❑ Interpret data
- ❑ Produce research outputs
- ❑ Author publications
- ❑ Prepare data for publications

What is Research Data Management?
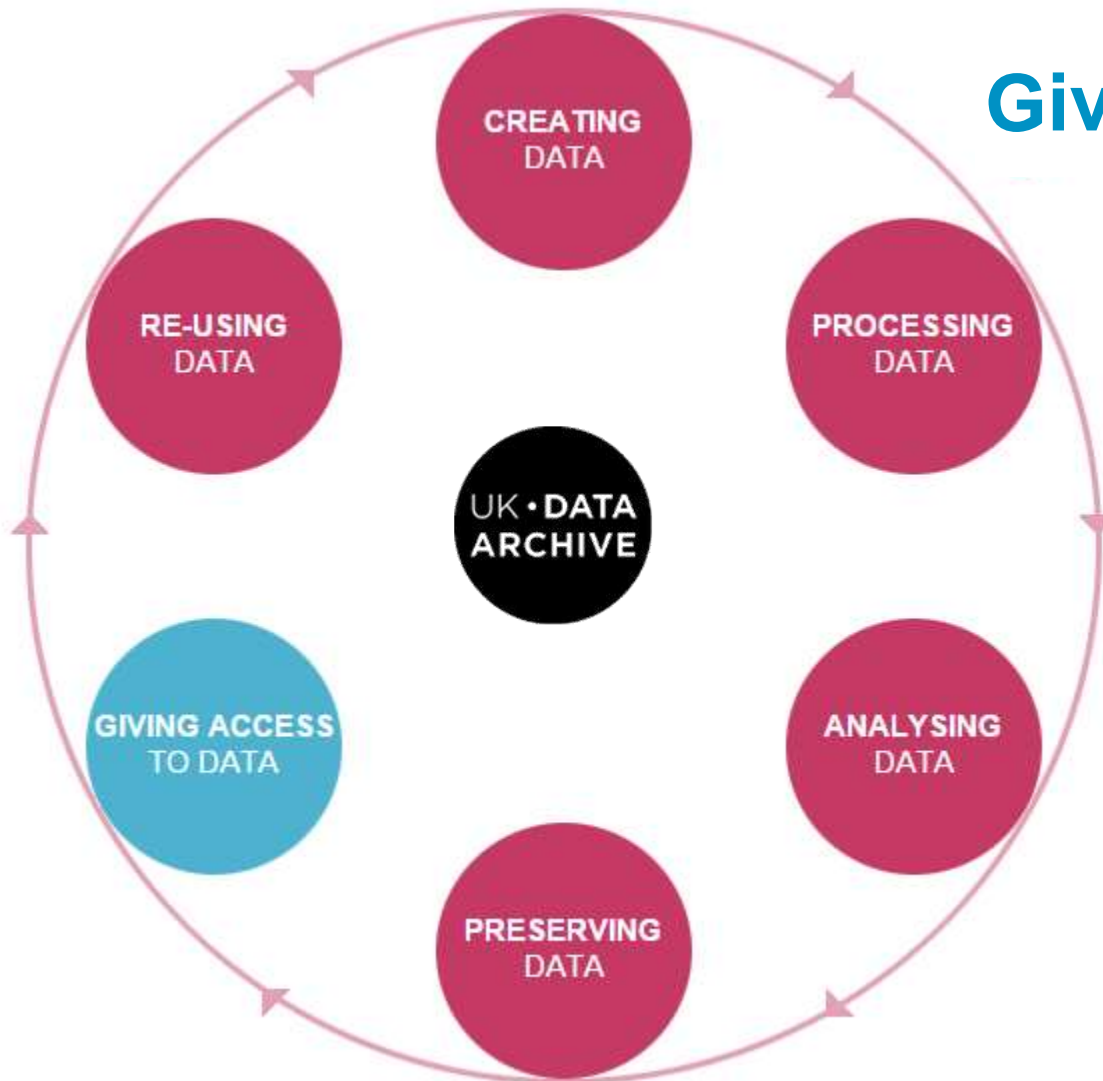## UK Data Archive Data Lifecycle model

Preserving data

❑ Migrate data to best format
❑ Migrate data to suitable medium
❑ Back-up and store data
❑ Create metadata and documentation
❑ Archive data

http://www.data-archive.ac.uk/create-manage/life-cycle

What is Research Data Management?
UK Data Archive Data Lifecycle model

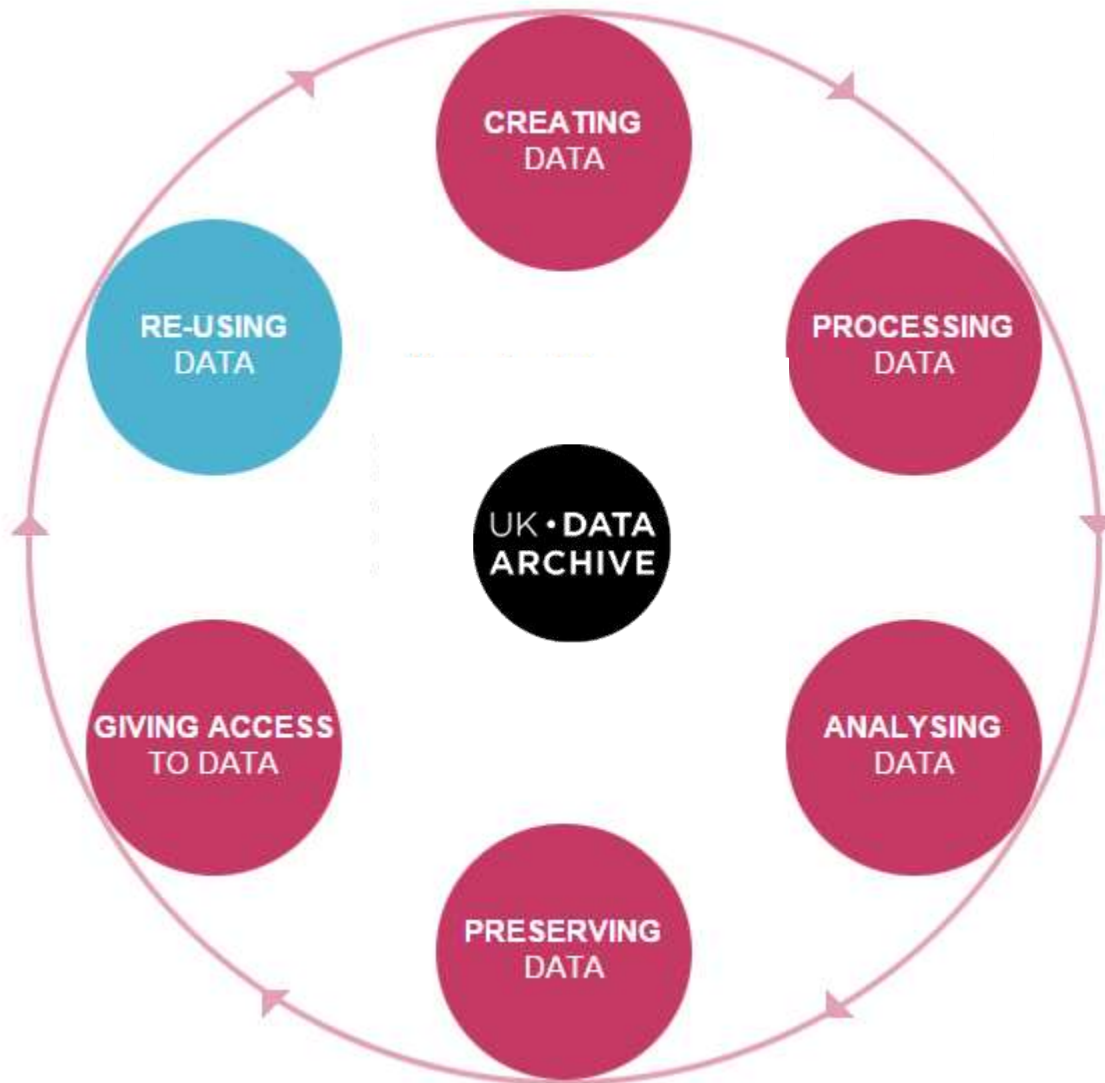## Giving access to data

- ❑ Distribute data
- ❑ Share data
- ❑ Control access
- ❑ Establish copyright
- ❑ Assign licences
- ❑ Promote data

http://www.data-archive.ac.uk/create-manage/life-cycle

What is Research Data Management?
UK Data Archive Data Lifecycle model

CREATING DATA

RE-USING DATA

PROCESSING DATA

UK · DATA ARCHIVE

GIVING ACCESS TO DATA

ANALYSING DATA

PRESERVING DATA

Re-using data

- ❑ Follow-up research
- ❑ New research
- ❑ Undertake research reviews
- ❑ Scrutinise findings
- ❑ Teach and learn

http://www.data-archive.ac.uk/create-manage/life-cycle

What is Research Data Management?
Why spend time and effort on this?

- So you can work efficiently and effectively
  - Save time and reduce frustration
  - Highlight patterns or connections that might otherwise be missed
- Because your data is precious
- To enable data re-use and sharing
- To meet funders' and institutional requirements

What is Research Data Management?
What does the OU expect?

"Research data must be managed to the highest standards throughout their lifecycle in order to support excellence in research practice."

"In keeping with OU principles of openness, it is expected that research data will be open and accessible to other researchers, as soon as appropriate and verifiable, subject to the application of appropriate safeguards relating to the sensitivity of the data and legal and commercial requirements."

OU Research Data Management Policy, November 2016

http://www.open.ac.uk/library-research-support/sites/www.open.ac.uk.library-research-support/files/files/Open-University-Research-Data-Management-Policy.pdf

What is Research Data Management?
What do funders expect?

"Good data management is fundamental to all stages of the research process and should be established at the outset."

"Open access to research data is an enabler of high quality research, a facilitator of innovation and safeguards good research practice."

Concordat on Open Research Data

What is Research Data Management?
What do funders expect?

Full Coverage · Partial Coverage · No Coverage

Research Funders	Policy Coverage		Policy Stipulations					Support Provided			
	Published outputs	Data	Time limits	Data plan	Access/ sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	◐	○	●	○	◐	◐
BBSRC	●	●	●	●	●	●	●	●	●	◐	●
CRUK	●	●	●	●	●	●	●	◐	●	○	○
EPSRC	●	●	●	◐	●	●	●	◐	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	◐
MRC	●	●	●	●	●	●	○	◐	●	○	◐
NERC	●	●	●	●	●	●	●	●	●	●	◐
STFC	●	●	●	●	●	●	●	◐	●	◐	◐
Wellcome Trust	●	●	●	●	●	●	●	●	●	◐	●

http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies

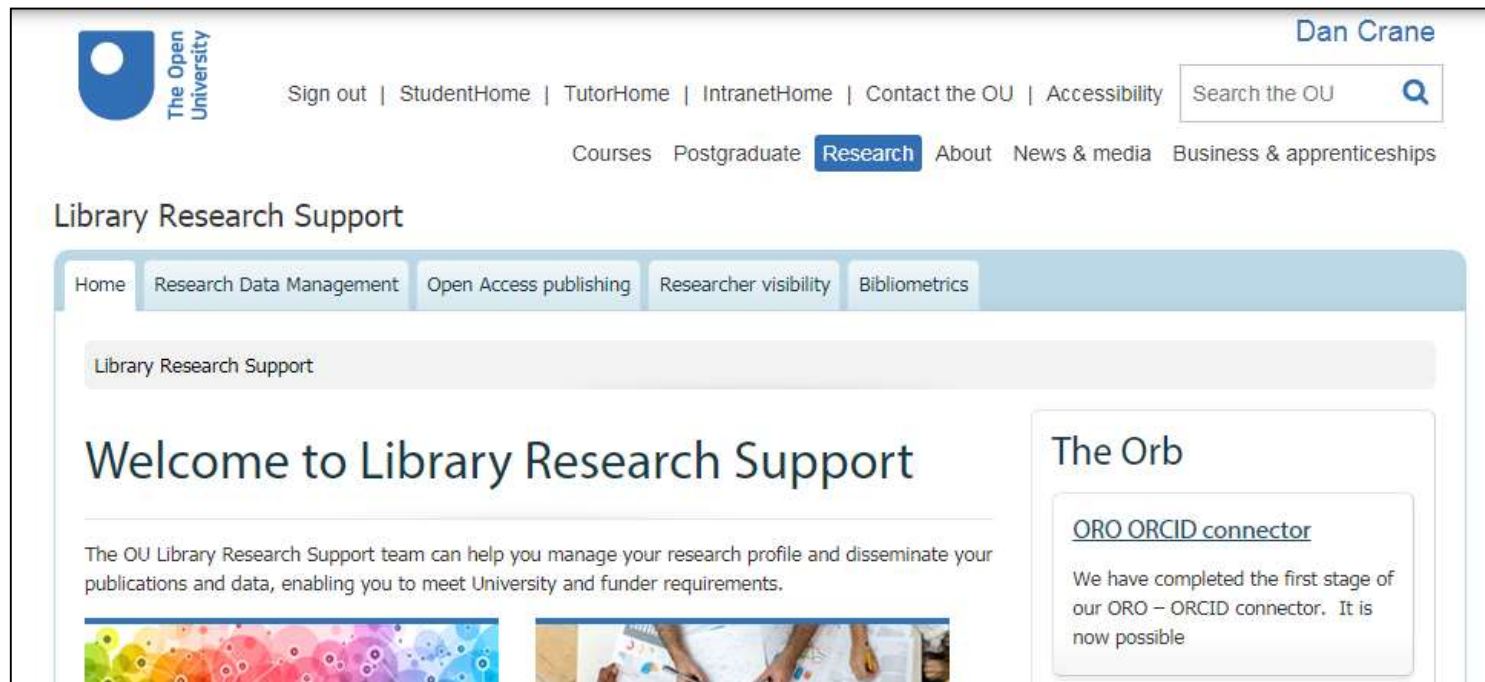# **What is Research Data Management?**
What does the OU **provide**?

- Support from the library research support team and website http://www.open.ac.uk/library-research-support/
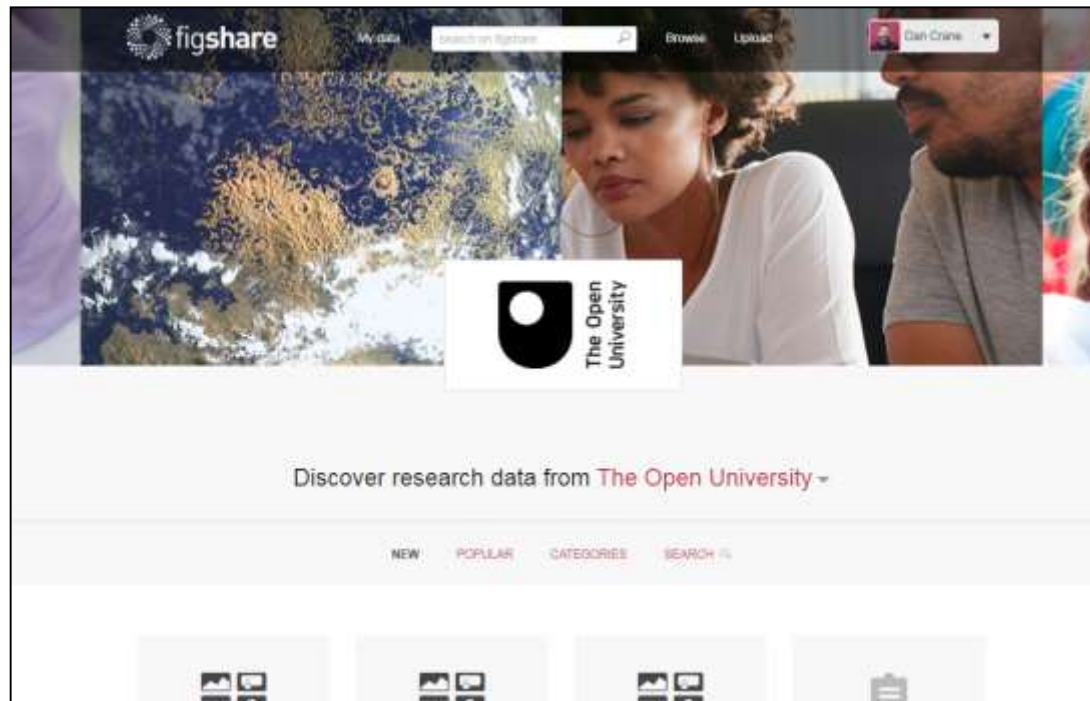
What is Research Data Management?
What does the OU **provide**?

- A repository, (ORDO) which meets funder requirements, facilitating secure, long-term storage of data https://ou.figshare.com/

Working with data

"Start as you mean to go on"

Thinking about the requirements at the beginning of the project will *limit the work needed during and at the end of the project.*

Working with data
Information security

Discuss the research data management issues raised by the scenarios.

What practical measures could have been taken to reduce risks to security?

Working with data
Storage options

There are several options available to you:

- OU networked file storage

- SharePoint

- OneDrive

- ORDO

- Cloud based services (DropBox, Google Drive etc.)

Tip: See the <u>comparison guide</u>

Working with data
Filing systems

Filing is more than saving files, it's making sure you can find them later in your project

- Naming
- Directory Structure
- File Types
- Versioning

All these help to keep your data safe and accessible.

Working with data
Naming conventions

Decide on a file naming convention **at the start of your project**. Useful file names are:

- consistent.
- meaningful to you and your colleagues.
- allow you to find the file easily.

Agree on the following elements of a file name:

- Vocabulary
- Punctuation
- Dates (YYYY-MM-DD)
- Order
- Numbers
- Version information

Ideally you should be able to tell what's in a file before opening it.
Tip: create a readme file detailing the naming scheme.

Working with data
Naming conventions
what you don't want…

Dan.doc

My paper.doc

Results.xls

Meeting.doc

20June.csv

IMPORTANT.pdf

Article_Manuscript October_FINAL.doc

Article_Manuscript October_FINAL FINAL.doc

Article_Manuscript October_FINAL FINALv1.doc

Article_Manuscript October_FINAL FINALv2.doc

Article_Manuscript October_FINAL FINALv2 last version.doc

Working with data
Naming conventions

Slides-RDM-PracticalStrategiesForRDM-2017-11.ppt

Slides-RDM-PracticalStrategiesForRDM-2017-11.ppt

type of document

general area of
work / topic

specific area of work / title

date
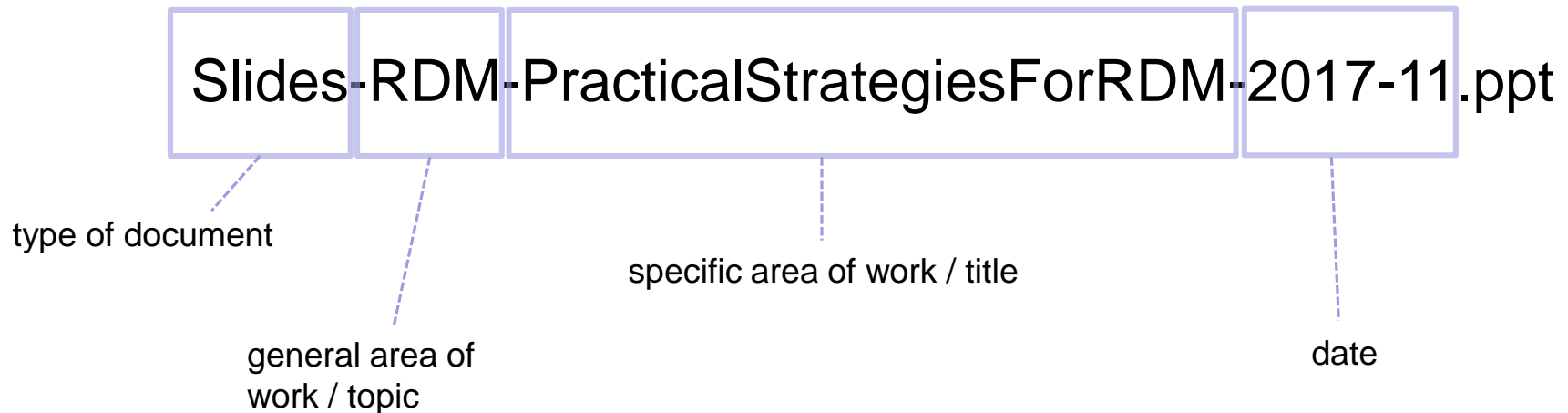
Working with data
File formats

- Unencrypted
- Uncompressed
- Non-proprietary/patent-encumbered
- Open, documented standard
- Standard representation (ASCII, Unicode)

Type	Recommended	Avoid for data sharing
Tabular data	CSV, TSV, SPSS portable	Excel
Text	Plain text, HTML, RTF PDF/A only if layout matters	Word
Media	Container: MP4, Ogg Codec: Theora, Dirac, FLAC	Quicktime H264
Images	TIFF, JPEG2000, PNG	GIF, JPG
Structured data	XML, RDF	RDBMS

Further examples: http://www.data-archive.ac.uk/create-manage/format/formats-table

Working with data
Metadata & documentation

- Metadata is additional information that is required to make sense of your files – it's data about data.

Search by Discipline

Biology

Earth Science

General Research Data

Physical Science

Social Science & Humanities

Guidance on disciplinary metadata standards: http://www.dcc.ac.uk/resources/metadata-standards

Think FAIR!

Findable

Accessible

Interoperable

Re-usable

Data FAIRport initiative: http://datafairport.org/

Working with data
Personal and Sensitive data

When working with research participants....

- Ensure you have obtained *valid consent*
- Inform your participants what will happen with the data during and after the project
- Consider who needs access to the data
- Can data be anonymised
- Consider controlling access if anonymisation or consent for sharing are impossible
- Pre-planning and agreeing with participants during the consent process, on what may and may not be recorded or transcribed, can be more effective than anonymisation

For more information, see the UK Data Archive guidance:
https://www.ukdataservice.ac.uk/manage-data/legal-ethical/consent-data-sharing/gaining-consent

Working with data
Personal and Sensitive data (2)

Managing sensitive data

- If possible, collect the necessary data without using personally identifying information
- There is a difference between *de-identification* and *anonymisation*
- De-identify or anonymise your data upon collection or as soon as possible thereafter
- Avoid transmitting unencrypted personal data electronically
- Consider whether you need to keep original collection instruments (recordings, surveys etc.) once they have been transcribed and quality assured

Working with data
External collaborators: IT Options

www.digitalbevaring.dk

- Shared areas or SharePoint
- Zendto
- Office 365 has OneDrive
- ORDO
- Be wary of Dropbox & similar

Remember the data storage for research projects comparison table:
http://www.open.ac.uk/library-research-support/sites/www.open.ac.uk.library-research-support/files/files/RDM-data-storage-options.pdf

Planning for data

So, there's a lot to think about…

www.digitalbevaring.dk
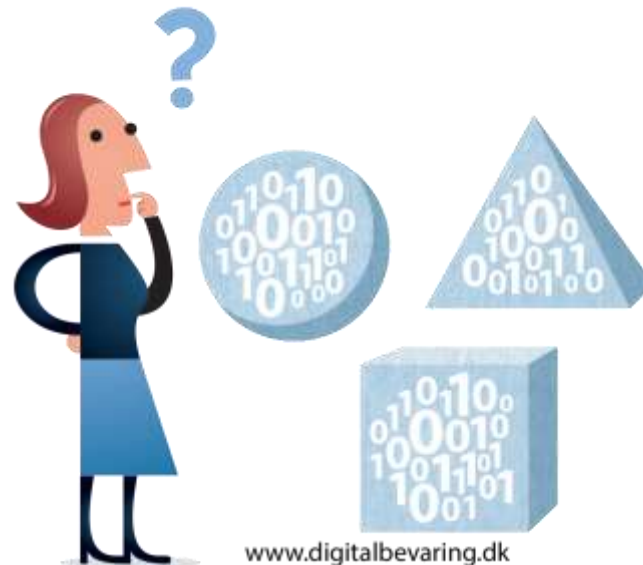
Planning for data

Data Management Plans are useful whenever you are creating data to:

- Make informed decisions to anticipate and avoid problems

- Avoid duplication, data loss and security breaches

- Develop procedures early on for consistency

- Ensure data are accurate, complete, reliable and secure

- Save time and effort – make your life easier!

Planning for data

Data Management Plan example

Data Collection

What data will you collect or create?

How will the data be collected or created?

Documentation and Metadata

What documentation and metadata will accompany the data?

Ethics and Legal Compliance

How will you manage any ethical issues?

How will you manage copyright and Intellectual Property Rights (IPR) issues?

Planning for data

Data Management Plan example

Storage and Backup

How will the data be stored and backed up during the research?

How will you manage access and security?

Selection and Preservation

Which data should be retained, shared, and/or preserved?

What is the long-term preservation plan for the dataset?

Data Sharing

How will you share the data?

Are any restrictions on data sharing required?

Responsibilities and Resources

Who will be responsible for data management?

What resources will you require to deliver your plan?

Planning for data
Which funders require a DMP?

| | Full Coverage | ◑ Partial Coverage | ○ No Coverage |

Research Funders	Policy Coverage		Policy Stipulations					Support Provided			
	Published outputs	Data	Time limits	Data plan	Access/ sharing	Long-term curation	Monitoring	Guidance	Repository	Data centre	Costs
AHRC	●	●	●	●	●	◑	○	●	○	◑	◑
BBSRC	●	●	●	●	●	●	●	●	●	◑	●
CRUK	●	●	●	●	●	●	●	◑	●	○	○
EPSRC	●	●	●	◑	●	●	●	◑	○	○	●
ESRC	●	●	●	●	●	●	●	●	●	●	◑
MRC	●	●	●	●	●	●	○	◑	●	○	◑
NERC	●	●	●	●	●	●	●	●	●	●	◑
STFC	●	●	●	●	●	●	●	◑	●	◑	◑
Wellcome Trust	●	●	●	●	●	●	●	●	●	◑	●

Note: Data Management Plans are now a requirement of all Horizon 2020 projects

www.dcc.ac.uk/resources/policy-and-legal/ overview-funders-data-policies

Planning for data
Tips

- Keep it simple, short and specific

- Seek advice - consult and collaborate

- Base plans on available skills and support

- Make sure implementation is feasible

- Justify any resources or restrictions needed
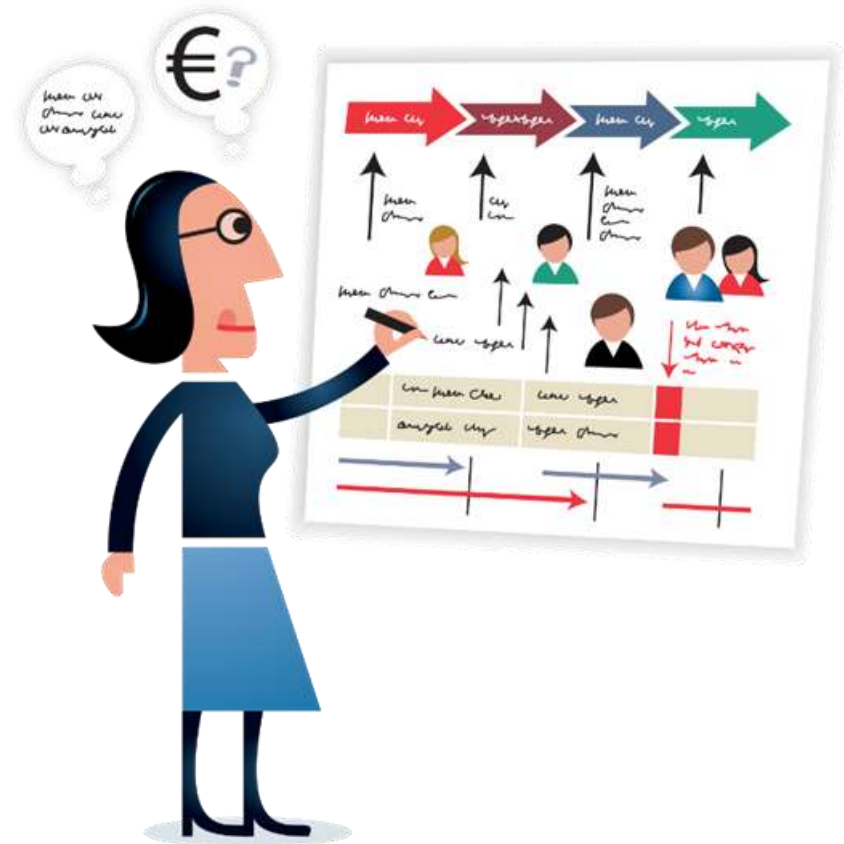
Planning for data
Activity

THE DATA LIFECYCLE

CREATING DATA

PROCESSING DATA

ANALYSING DATA

PRESERVING DATA

GIVING ACCESS TO DATA

RE-USING DATA
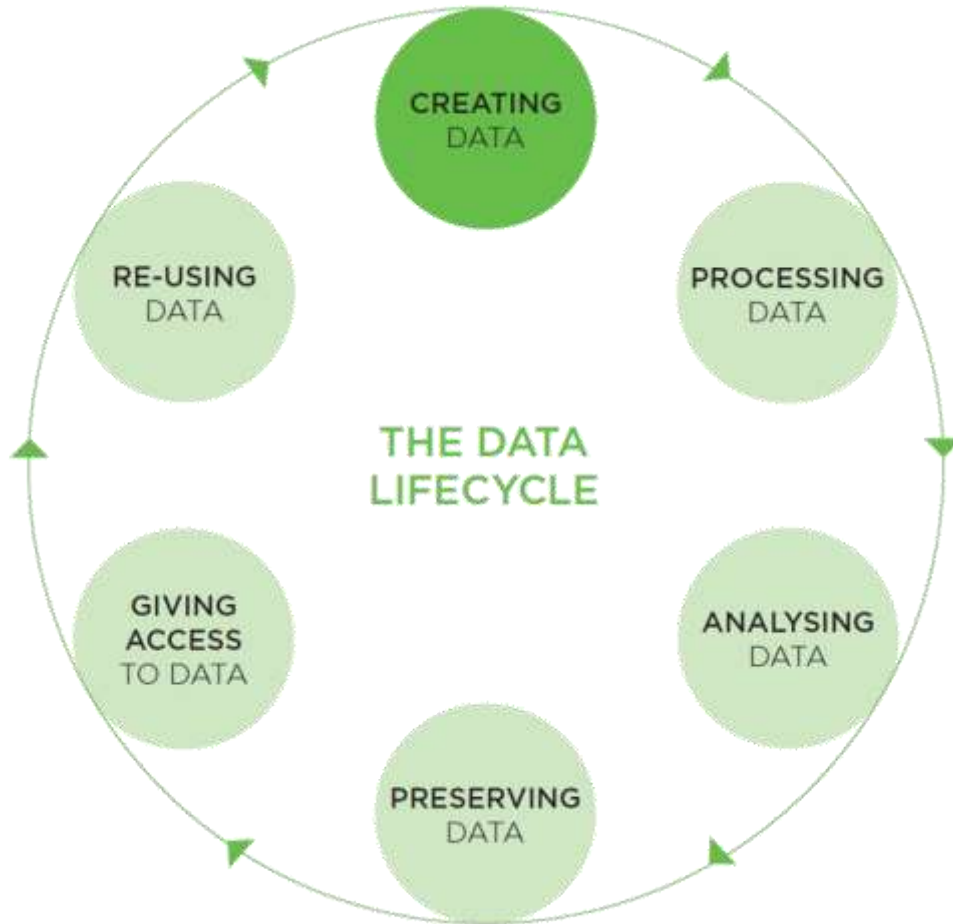
Think about your own research.

What actions would you need to perform on your data at each stage of the UKDA's Lifecycle model?

What would you put in your DMP?

Planning for data
DMPOnline

A web-based tool to help you write DMPs according to different requirements. DCC, funder and OU guidance.

https://dmponline.dcc.ac.uk

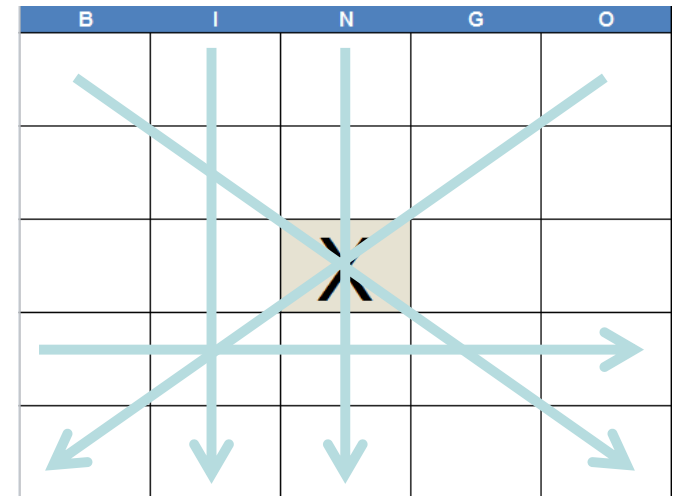# Now for a game…

Rules

- Take a bingo card and an example DMP.

- Each square contains a positive quality: good DMPs will do all/most of these.

- Read each square and if it is true for the example DMP, mark it with a cross.

- The first person to get five crosses in a row (vertical, horizontal, or diagonal) calls "Bingo!" and gets a prize.

Sharing data
A quick overview

- Benefits: transparency, re-use, impact

- Required by many funders and publishers

- Lots of tools and guidance to help researchers

- Come to training session 'Making your research data open' on 8th December

Library Services
How we can help

- Data Management Plan checking

- Support with setting up new projects

- Advice on preparation of data for sharing

- Data Repository (ORDO)

- Online guidance

- Enquiries

Email: library-research-support@open.ac.uk

Useful links

- The OU Library Research Support website: http://www.open.ac.uk/library-research-support/research-data-management
- Open Research Data Online (ORDO): https://ou.figshare.com
- Digital Curation Centre: http://www.dcc.ac.uk/
- DMP Online: https://dmponline.dcc.ac.uk/
- UK Data Archive: http://www.data-archive.ac.uk/
- MANTRA: http://datalib.edina.ac.uk/mantra/
- The Orb: http://open.ac.uk/blogs/the_orb
- OU Human Research Ethics Committee: http://www.open.ac.uk/research/ethics/
- OU Data Protection: http://intranet6.open.ac.uk/governance/data-protection/advice-and-resources *(if clicking on the link doesn't work, copy and paste the address)*
- OU Information Security: http://intranet6.open.ac.uk/it/main/information-security *(if clicking on the link doesn't work, copy and paste the address)*

Questions?
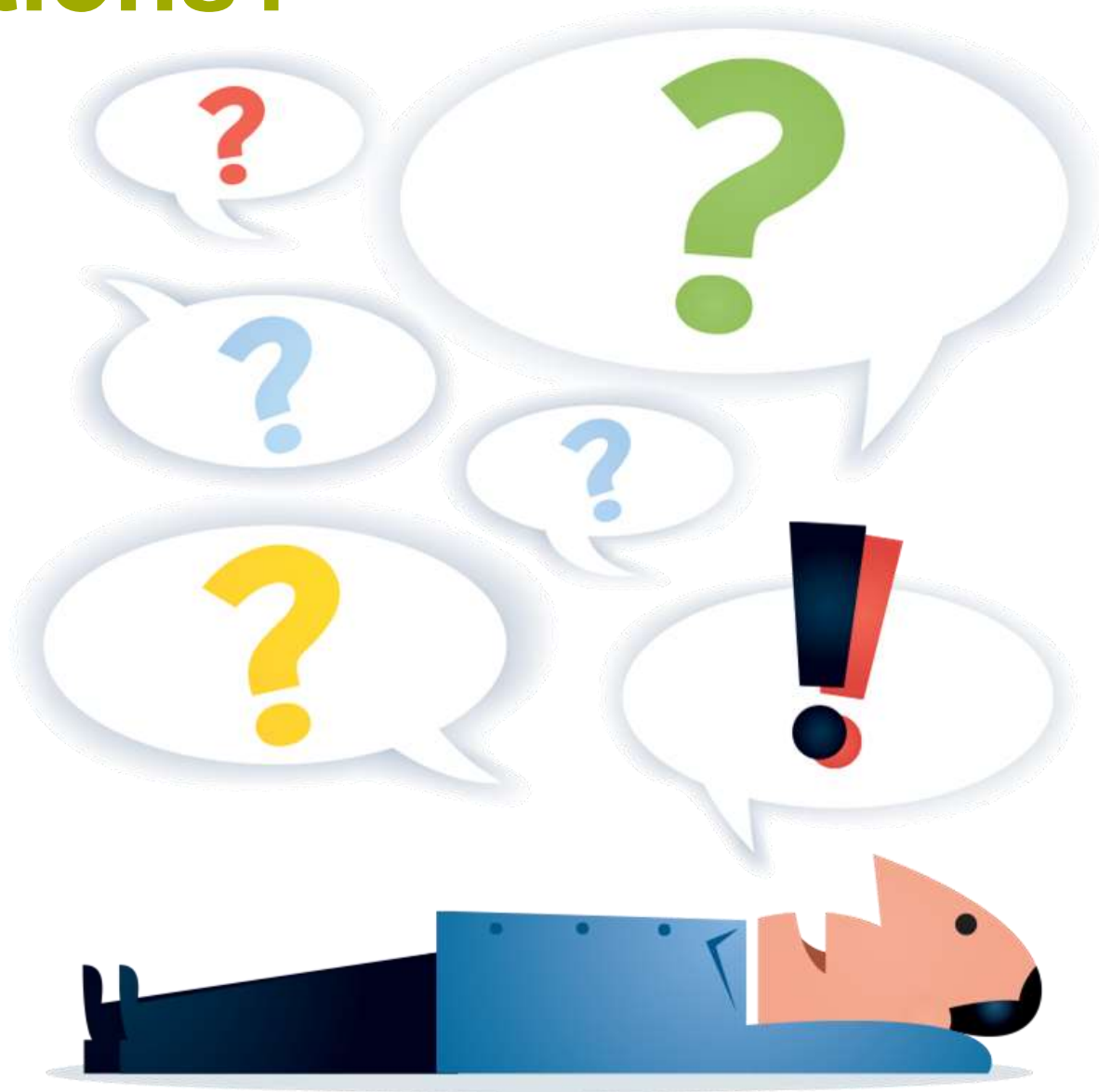
3 take home points

1. Start early to help you work better and protect your precious data

2. Write a Data Management Plan

3. Don't be shy. Ask for help!

Image credits

Unless otherwise stated, all images are by Jørgen Stamp at http://www.digitalbevaring.dk