MANCHESTER 1824
The University of Manchester

Managing Your Research Data

CPD Masterclass

10 December 2014

Chris Gibson and Mary McDerby

MANCHESTER 1824
The University of Manchester

Objectives

- To appreciate why research data management is important
- To identify common data management issues
- To explore best practices and resources for managing these issues
- To understand how the RDM Service can help you identify data management resources, tools and best practices

---

MANCHESTER 1824
The University of Manchester

Format

- Introduction
- Collaborative exercises
  - Group discussion and feedback
  - Share experiences
- Summary and final questions

MANCHESTER 1824
The University of Manchester

What is research data (1)?

- No consensus on definition – varies according to discipline or research funder

"Research data is the evidence base on which academic researchers build their analytic or other work" (HEFCE)

"Recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings" (EPSRC)

MANCHESTER 1824
The University of Manchester

What is research data (2)?

"Research data is the material underpinning a research assertion" (Molloy)

Types of research data
- Observational
- Experimental
- Simulation
- Derived or compiled
- Reference

MANCHESTER 1824
The University of Manchester

Why manage research data?

- Increase your research impact
- Save time
- Preserve your data
- Maintain data integrity
- Meet funder and legal requirements
- Promote new discoveries

MANCHESTER 1824
The University of Manchester

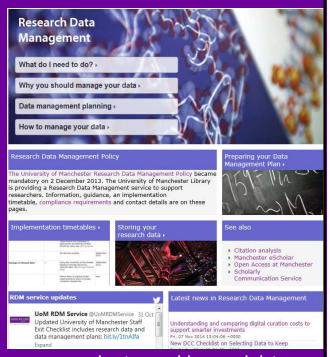
Drivers for RDM

- Increasing pressure
- Public funding
- RCUK Common Principles
- EPSRC Policy Framework
- Royal Society 'Science as an Open Enterprise'

MANCHESTER 1824
The University of Manchester

RDM at Manchester

- University RDM Policy
- Service Team
- Support network
- Research Data Storage Service
- Data Management Planning Tool
- Website and guidance



www.manchester.ac.uk/researchdata

Group exercise

- In groups, pick topics from this list:
  1. Collaborative projects
  2. Organising and versioning files
  3. Metadata and data documentation
  4. Storage, backup and security
  5. Sharing data
  6. Appraising data (what to keep / delete)
- Discuss what issues you have faced in your research in this area
- What solutions to the problems have you discovered?
- Feedback to group

Issue #1: Collaborative projects

- Challenges of team science
- Challenges with rotating personnel

Collaborative Projects: Best Practices

- Establish standard procedures, protocols and policies
  - Centralized approach to share best practices
  - Economies of scale
  - Lasting framework

See UK Data Service:
http://ukdataservice.ac.uk/manage-data/collaboration.aspx

Collaborative Projects: Best Practices

- Develop a data management framework:
  - assignment of data management responsibilities to named individuals
  - standard procedures e.g. file sharing and storage
  - guidelines e.g. filenaming, version control, data formats
  - standardised forms e.g. for consent procedures, ethical review and data management plans
  - standards and protocols e.g. data quality control standards, confidentiality agreements for data handlers
  - policies e.g. security policy for data storage and transmission
  - data copyright and ownership statements for the centres and for individual researchers
  - data sharing strategy e.g. via repository, data centre, website
  - information on funder requirements or policies on managing and sharing data that apply to projects or the centre

Collaborative Projects: Best Practices

- Develop a resources library
  - Relevant data policy and guidance documents, templates, tools, exemplars of good practice
  - Access via website, wiki, shared network drive
  - Local resources
    - Local regulations relating to data
    - IT procedures e.g. backup, data management
    - Guidelines e.g. data sharing, filenaming, version control, anonymisation
    - Standard forms e.g. consent forms and information sheets, ethical review
    - Data inventory e.g. to track data, related documentation and outputs
  - External resources
    - Funder research data policy
    - ICO Data Protection Act 1998 guidance

13

Issue #2: Records Management

- Inconsistently labeled files
- in multiple versions…
- inside poorly structured folders…
- stored on multiple media…
- in multiple locations…
- and in various formats…

14

Records Management: Best Practices

- Create file naming conventions
- Create directory structure naming conventions
- Version your files
- Choose appropriate file formats for preserving and sharing your data files

15

File Naming

- Brief and relevant
- No special characters, dots or spaces
- For separation use capitals or underscores_ instead of periods or spaces
- Use date format ISO 8601:YYYYMMDD
- Include a version number

16

File Naming

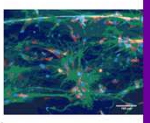
File name = **principal identifier** of file
Easy to: identify, locate, retrieve, access

Provides context e.g.:
✓version number e.g. *FoodInterview_1.1*
✓date e.g. *HealthTest_2011-04-06*
✓content description e.g. *BGHSurveyProcedures*
✓creator name e.g. *CommsPlanHLJ*

17

Data Naming Format

Histology



- Apex of the heart - A
- Basal side (base of the heart) - B

- Experiment #, A or B, and Section #
- Example: 0231A_216
- When sample is stained, an abbreviation for the type of stain used is added to the slide
- Example: 0231A_216_act (actinin)

18

Directory Structure

- Data and documentation files in separate folders
- Store like with like (file type):
  - Data files: organised by data type and then by research activity
  - Documentation files: by type of documentation file then by research activity
- Restrict the level of folders to 4 deep
- Max 10 sub-folders in each folder
- If you can't see all subfolders at a glance, create an alphabetic menu
- Use shortcuts instead of multiple copies

19

Directory Structure: Example

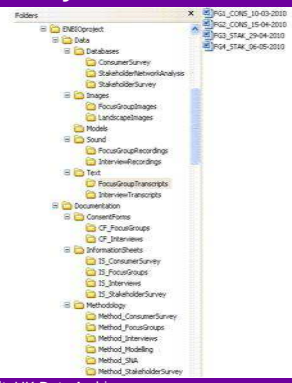


Image credit: UK Data Archive

20

Version Control Strategies

- Ensure you are working on the correct version of your data
  - Ideal: keep original (master) version of data file the same and save iterative versions
  - If you need to modify data files then save copy of every iteration
- Record file status/versions
- Record relationships between files
  e.g. data file and documentation; similar data files
- Keep track of file locations
  e.g. laptop vs. PC

21

Version Control: Single user

- File naming; unique file name with date or version number

File name	Changes to file
Interviewschedule_1.0	Original document
Interviewschedule_1.1	Minor revisions made
Interviewschedule_1.2	Further minor revisions
Interviewschedule_2.0	Substantive changes

- Version control table or file history alongside data file
- Version control facility within software e.g. Microsoft Word 2003

22

Version Control: Multiple Users

- Control rights to file editing: read/write permissions
  e.g. Microsoft Office
- Versioning/file sharing software:
  e.g. Google Drive, Amazon S3
- Manual merging of multiple entries/edits

23

Version Control: Multiple locations

- Synchronise files
  e.g. MS SyncToy software, DropBox
- Use remote desktop

24

Formats for Storing and Sharing Data

- Consider storing data in more than one format
  - E.g. The format the data was created in and a more open format (.xls vs .csv)
- Also important when archiving data at the end of a project
- Use a format that can be used by other researchers e.g. If you create a file using NVivo can the file be opened by researchers who don't have NVivo?
- Use a format that doesn't lose data – a lossless format
- Scan paper records (e.g. consent forms) to allow for more secure long term storage

25

What Format Should I Store my Data in? (Open Formats)

- What is an Open Format?
  - The specification has been published
  - The format can be used and implemented by anyone
- Advantages of Open Formats
  - Not limited to one piece of software
  - More chance of being able to use the format in the future
- UK Data Service recommended formats
  - http://ukdataservice.ac.uk/manage-data/format/recommended-formats.aspx
- Library of Congress sustainable formats
  - http://www.digitalpreservation.gov/formats/index.shtml

26

What Format Should I Store my Data in? (Lossless Formats)

- What is a Lossless Format?
  - Data is compressed without any being lost
  - Lossy compression formats discard data in the compression process
  - Used for preservation/archiving purposes
  - E.g., TIFF, PDF/A, MPEG, FLAC
- Advantage of Lossless Formats
  - The original data can be recreated as none has been lost

27

Issues with File Formats

- Size of files
  - Lossless formats tend to be larger (because they hold more information on the file)
- Version control
  - If the file is stored in more than one format do you update both formats when you make changes?
- Storage
  - At least one back up (and/or the master copy) should be kept separate to your working data

28

Possible issues when using different file formats

	A	B	C	D	E	F
1		Timber volumes in m3				
2	Year	1994	1995	1996	1997	1998
3	Date recorded	20/01/1995	23/01/1996	11/01/1997	16/01/1998	14/12/1998[1]
4	Logging private land	20346.345	47005.223	26001.754	11468.897	0.000
5	Logging forest reserves	4060.567	1777.783	804.997	0.000	3329.653
6	Logging state land	0.000	1200.000	559.162	2077.567	358.935
7	Total	61119.912	87065.006	64802.913	51354.464	5686.588
8						
9		Data missing				
10		Estimate				
11						
12	[1] temporary volumes					
13						

MS Excel format

	A	B	C	D	E	F
1		Timber volumes in m3				
2	Year	1994	1995	1996	1997	1998
3	Date recorded	20/01/1995	23/01/1996	11/01/1997	16/01/1998	14/12/1998[1]
4	Logging private land	20346.345	47005.223	26001.754	11468.897	0
5	Logging forest reserves	4060.567	1777.783	804.997	0	3329.653
6	Logging state land	0	1200	559.162	2077.567	358.935
7	Total	61119.912	87065.006	64802.913	51354.464	5686.588
8						
9		Data missing				
10		Estimate				
11						
12	[1] temporary volumes					
13						

Loss of annotation

Tab–delimited text format

29

Issue #3: Metadata

- How will someone make sense of your data e.g. the cells and values of your spreadsheet?
- How can you describe a data set to make it discoverable?
- What universal or disciplinary standards could be used to label your data?

30

What is Metadata?

"Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called data about data or information about information"
(NISO, Understanding Metadata 2004;1)

Metadata is a love note to the future (Martin Hamilton)

- Helps you and future users:
  - ✓ To understand your data in detail
  - ✓ Find, use, manage, and properly cite your data

31

Metadata makes data intelligible

- Describes the characteristics of a dataset at various levels:
  - Process used to collect data
    - e.g. recorded in published paper, codebook, lab book, log, diary
  - Data itself
    - e.g. data format, software used to read the data
  - Changes of the dataset in time
    - Data provenance – historical report of uses and edits of research data over time

32

Types of Metadata

Types of metadata	Goal	Example
Descriptive metadata	*The minimal metdata, required to find a digital object.* *If there are additional contextual metadata, a user will have a better idea on how to use the data*	*Author, title, abstract, date* *Contextuele metadata are for example location, time, data collection method (tools)*
Structural metadata	*These link the individual objects of a unity*	*Links to related digital objects, (e.g. the article written based on the linked research data)*
Technical metadata	*Information on the technical aspects of the data set*	*Data format, hardware/software used, calibration, version, authentication, encryption, metadata standard*
Administrative metadata	*Metadata focusing on user rights and management of digital objects*	*License, possible reasons for an embargo, waivers* *Search logs, user tracking*

Credit: Research Data Netherlands

33

How do I include metadata (1)?

- Embedded documentation
  - Information about a file or dataset can be included within the data or document itself. For digital datasets this means:
    - in separate files (e.g. text files)
    - integrated into data files (e.g. as a header or specified locations in the file)
  - Examples include:
    - Code, field and label descriptions
    - Descriptive headers or summaries
    - Transcripts
    - Recording information in the Document Properties function of a file (Microsoft)

34

Example of embedded documentation

- Directory: sam_monarch_wing_05242011
- Metadata for this directory
  - Creator: Katherine McNeill
  - Subject: monarch butterfly wing
  - Description: this directory contains Sashimi ESEM images of a monarch butterfly wing I took after finding a butterfly floating by the Charles River near MIT
  - Contributor: Anne Graham helped me with these images
  - Date: 05/24/2013
  - Type: image
  - Original Format: Sashimi Microscope format (.sam)
  - Relation: this is a directory that will contain multiple files
  - Coverage: By the Charles River in Cambridge, MA, MIT side
  - Rights: Monarch Butterfly Research Foundation (funder) owns the data (grant number: 00213)

35

How do I include metadata (2)?

- Supporting documentation
  - Information in separate files that accompanies data in order to provide context, explanation, or instructions on confidentiality and data use or reuse.
  - Examples include:
    - Working papers or laboratory notebooks
    - Questionnaires or interview guides
    - Final project reports and publications

36

How do I include metadata (3)?

- Catalogue metadata
  - Structured information used to identify and locate data that meets user's requirements via a catalogue.
  - Usually structured according to an international standard and associated with the data by repositories or data centres when materials are deposited with them
  - Examples include:
    - Title
    - Description
    - Abstract
    - Creator
    - Geographic location
    - Keywords

37

Metadata: Best Practices

- Document data at the start of a project and continue to add information as the project progresses:
  - File naming conventions and formats
  - How data is organised
  - How data was generated (including any equipment or software used)
  - How data has been altered or processed
- Include explanation of:
  - Variables and their units
  - Codes, abbreviations, missing values
  - Formats for dates, time, geographic coordinates, and other parameters
- Note where you got the data so that you and others can find it

38

Metadata: Best Practices

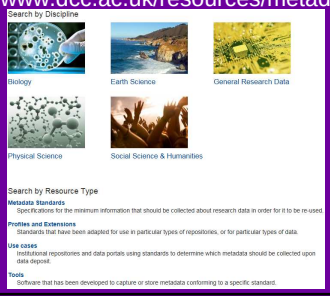
Common metadata fields associated with a data set:

- Title
- Creator
- Identifier
- Subject
- Funders
- Rights
- Access information
- Language
- Dates
- Location
- Methodology
- Data processing
- Sources
- List of file names
- File formats
- File structure
- Variable list
- Code lists
- Versions
- Checksums

39

Metadata: Best Practices

- Use an established metadata standard
  - DCC offers guidance: http://www.dcc.ac.uk/resources/metadata-standards



40

Issue #4: Storage, Backup and Security

- Where can you store your data?
- How much server space do you need now, and in the future?
- How many copies of data should you have?
- How often should data be backed up?
- Does the data need to be shared?
- Is the data sensitive or confidential?

41

Storage

- Research Data Storage Service
  - Upto 20TB replicated storage per project free at point of use
- Networked storage (P drive)
  - If store extra working copies then schedule synchronisation and/or backups
- Remote storage services (The Cloud)
  - Not for sensitive or confidential data; see UoM guidance: http://www.itservices.manchester.ac.uk/secure-it/cloud/
  - Read terms of service
- Physical storage (desktop and laptop)
  - Not for master copies of data
  - Security: physical; password-protect
  - Risks e.g. fire, theft, failure, replaced, upgraded, etc.
- External storage (removable media)
  - Not for master copies of data
  - Risks e.g. misplace, theft, damage, error prone

42

Backup

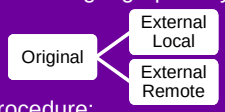
- Allows you to *restore* your data if it is lost:
  - Hardware or software malfunction
  - Environmental disaster (fire, flood)
  - Theft
  - Unauthorized access

Does your operating system and/or network administrator already have existing protocols and tools in place?

43

Backup: Best Practices

- 3 copies that are geographically distributed:



- Backup Procedure:
  - Full (replicates all files)
  - Differential incremental (copies only files changed since last incremental or full backup)
  - Cumulative incremental (copies only files changed since last full backup)
- Backup Frequency (as appropriate):
  - Daily
  - Weekly
  - Monthly
- Test your strategy!

44

Security: Best Practices

- Safeguard your data from malicious or accidental access to your system.
  - Access
  - Systems
  - Data Integrity

Does your network administrator have security services that meet your needs?

45

Security (access)

- Limit the availability of your data.
  - User ID/passwords
  - Role-based access rights
  - Limitations of wireless devices to protect access

46

Security (systems)

- Protect the hardware and software systems you use
  - Anti-virus software
  - Up-to-date versions of software and storage media
  - Firewalls
  - Intrusion detection software
  - Limit physical access

47

Security (data integrity)

- Protect the integrity of your data at the file level
  - Encryption
  - Electronic signatures
  - Watermarking

48

Issue #5: Sharing data

"Everyone in a research team should have a clear sense of their responsibilities in ensuring that … research data are of the highest quality; … are well documented so that other researchers can access, understand, use and add value to them … independently of the original investigators."

[MRC Guidance on Data Management Plans]

- Issues to consider:
  - Future 'shareability' of the data: format, software, anonymisation, documentation, ethics, consent & confidentiality
  - Description & documentation
  - Future rights management & licensing
  - Timescale for release
  - Infrastructure for sharing

49

Drivers for sharing

- Openness
  - Open science, open source, open standards, open knowledge, open government, open content, and… Open Data!
    - *"Open data is data that can be freely used, reused and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike." (Open Data Handbook)"*
- National research policies
  - **Research Councils UK** "*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 that does not harm intellectual property.*"
  - **Economic & Social Research Council** "*… publicly funded research data … valuable, long-term resources that, where practical, must be made available for secondary, scientific research.*"
  - **Medical Research Council** "*… publicly-funded research data … should be openly available to the maximum extent possible.*"
  - **Wellcome Trust** "*… aim(s) to ensure that the data generated by the research we support is managed and shared in a way that maximises the benefit to the public.*"

50

Drivers for sharing

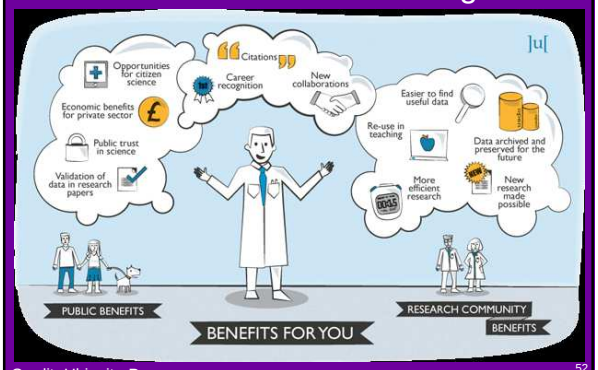
- UoM RDM Policy
  - "The University expects its researchers to make Relevant Data openly available to other researchers in a timely way, with as few restrictions as possible.

    It is recognised that some restrictions may be necessary, e.g. to protect intellectual property, respect confidentiality, or honour third party agreements, but these should be minimised as far as is practicable.

    It is also recognised that it may be appropriate to allow a limited period of privileged access to the data for the research team that acquired or created it."
- Why not!
  - "… we have the technologies to permit worldwide availability and distributed process of scientific data, broadening collaboration and accelerating the pace and depth of discovery …" [John Willbanks, VP Science, Creative Commons – in Wikipedia]

51

Benefits of Data Sharing



Credit: Ubiquity Press

52

Benefits of data sharing: example

"Some people just keep their completed research data to themselves. I put it on Open Access for a couple of reasons. First of all, philosophically I believe that science is something open and results should be reproducible. Open Access helps both aims. I also believe that in the end you'll get more citations for your work. In addition I have examples of people who could have simply lifted the data, gone away and done something with it and given me a citation for it; but actually they have come to me and said, "OK, I've got this data, which is yours, we're interested in it, but we need your expertise to interpret it" and then I get a co-authorship out of it as well."

Prof. Tim Naylor, Norman Lockyer Professor of Astrophysics and Head of Physics, University of Exeter

53

Barriers to sharing

- Barriers can relate to:
  - the Researcher - intellectual property issues
  - the Institution - unrealised commercial value
  - the Subject - confidentiality

54

Two Stages of Data Sharing

- Two stages of your project when you may share data
  - "Live" sharing during the project
  - Making your "completed" data available at the end of your project
- Different issues and ways of sharing data during these stages

55

Why Should You Share your Data?

- Benefits – "Live" data
  - Increased collaboration opportunities with colleagues
  - Increased exposure of your current work
  - Increased efficiency across research group
- Benefits – "Completed" data
  - Increased citation counts
  - Increased exposure for your work
  - Increased chance of collaboration in the future
  - Allows others to build on your research
- Policy
  - RCUK Common Principles on Data Policy
  - University Research Data Management Policy

56

Sharing Data Within a Research Project

- Considerations for organizing and sharing data:
  - Location of data
  - Folder structuring
  - File naming standards
  - Versioning
  - Formats
  - Responsibility
  - Communication
  - Documentation and Backup

57

Sharing Outside the Research Project

- Consult University policies
  - e.g. Research Data Management Policy, Intellectual Property Policy
- Consult funders' policies
- Consult your PI
- Consult your ethics office

58

How to Share Your Data: During Your Research

- With your supervisor; with project colleagues; with external interested parties
- Remote access or ZendTo at this University
- Cloud Storage – Dropbox, Googledrive, Skydrive etc. Not for sensitive or confidential data
- Email – issues with large data and/or sensitive data Potential version control problems.
- USB sticks – easily lost; can transfer viruses
- External hard drives – less suitable if collaborator is at a different institution
- Websites – lack of permanency. Need internet connection. May not have access rights to the site.
- FTP – Not secure. Data can be intercepted.
- Hard copy documents – one of a kind.

59

How to Share Your Data: At the End of Your Research

- Data repositories
  - Discipline specific e.g. Archaeology Data Service
  - (Inter)national e.g. UK Data Service
  - Discipline-agnostic e.g. Zenodo, Figshare
  - University repository (under development)
- Data journal e.g. see illustrative list at the 'Data@Mlibrary blog'
- Paper with supplementary data e.g. via Dryad
- Websites
  - Link from your University personal web pages to data in a repository
  - Link from academic network sites
    - Academia.edu, ResearchGate.net

60

Issues in Data Sharing

- Ethical and Data Protection Act
- Copyright and legal issues
- File size
- File format – open or proprietary
- Metadata
- Documentation of data
- What to share
- Quality control and versioning
- Discoverability of the data
- Re-use of data

61

Ethical and Data Protection Act Issues

- Not all data can be shared.
- You must ensure that you don't share data you are not allowed to:
  - Abide by your ethical approval
  - Abide by the Data Protection Act
    - Are you sharing this data securely?
    - Have you got consent to share the data?
    - Use Cloud Storage wisely – not for sensitive data
- Ethics advice from: Research Office
- DPA Advice: Records Management Office
- See also UK Data Service advice:
  - http://ukdataservice.ac.uk/manage-data/legal-ethical.aspx

62

Copyright and Legal Issues

- You must abide by any contract you or your project group have signed:
  - This may state that you are not allowed to share the data or it may include the conditions of sharing
- You must be aware of who owns the copyright for the data you are sharing:
  - You may not be allowed to share it
  - You must get permission from the copyright owner before sharing data
  - Also applies to data in your thesis
- Advice from JISC Digital Media on using images: http://www.jiscdigitalmedia.ac.uk/

63

File Size

- Large files cannot be emailed
- Some files may not fit onto USB sticks
- How do you know if a file has been received?
- Large files can take a long time to upload to Cloud Storage
- Use the University's ZendTo service (up to 20GB)
- Remote login to University systems

64

File Format

- Is the file format you are using widely used?
  - If not, can you migrate it to a more widely used format? E.g. .xlsx (Excel); .pdf
- Is the format you are using an "open" format or is it "proprietary"?
  - Open formats can be more easily accessed by other researchers e.g. SPSS files can be saved as .csv files; Word files can be saved as an Open Document format (.odt rather than .docx)
- Make sure you don't lose important information when migrating formats
- UKDS guidance: http://ukdataservice.ac.uk/manage-data/format/recommended-formats.aspx

65

Metadata

- Best practice: record metadata as you collect / create your data
- Have you provided information about the data with the data you share?
- It is needed for discoverability, reuse, reproducibility and verification, etc.
- For example:
  - Author
  - Title
  - Date of creation
  - Publisher
  - Abstract
  - Description of the data

66

Supporting Documentation

- Have you provided enough information for another researcher to be able to understand, retrieve, validate and re-use the data?
  - Where was the data created?
  - How was the data created?
  - What hardware and software were used?
  - What methodologies were used?
  - What assumptions did you make in your experiments?
  - Why are there anomalies in your data?
- Along with the metadata, the documentation should enable the data to be understood and reusable independently of any other publications, data etc.

67

What to Share?

- You don't need to share all your "live" data
  - Only data that is helpful and useful to the recipient
- What to archive?
  - Consider policy / legal requirements
  - In collaboration with your supervisor or PI develop a set of criteria:
    - Only the data supporting your publications?
    - Data that can reproduce your results?
    - Data that can validate your results?
    - How unique or significant is your data?

68

Quality Control and Versioning

- If working on a collaborative project ensure that you are all working on the correct version of your data
  - Use version control tables or name your document appropriately
- Will versioning affect how you will share your data?
  - For example, is it easier to control versions with cloud storage than email?
- How will you ensure that the data isn't corrupted or changed in the process of sharing?
  - Archives / repositories can provide a Persistent Identifier

69

Data Discoverability

- Data needs to be found if it is to be re-used
- Discoverability can be aided by placing your data in a repository
  - Indexed by Google
  - Researcher community knows they exist and can go to them for data
  - Need adequate (and accurate) documentation to fully aid discoverability
  - Persistent Identifier (e.g. DOI) – can be included in citations, emails, Tweets etc.
  - Academic networks and personal web pages can link to the data in a repository
- Reference your data in publications
  - See RCUK policy on Open Access (Expectations for Researchers, and section 3.3) http://www.rcuk.ac.uk/research/openaccess/policy/

70

Data Re-use

- Data citation is becoming more common
  - Get credit for all your research
- If others use your data it can increase your citation rates
  - Piwowar HA, Day RS, Fridsma DB (2007) Sharing detailed research data is associated with increased citation rate. PLoS ONE 2(3): e308. http://dx.doi.org/10.1371/journal.pone.0000308
- Sharing can mean that your data is re-used in areas you didn't think it could be
  - E.g. ships' logs are being used by climate scientists
- Prof Tim Naylor on data sharing :
  - 'I have examples of people who could have simply lifted the data, gone away and done something with it and given me a citation for it; but actually they have come to me and said, "OK, I've got this data, which is yours, we're interested in it, but we need your expertise to interpret it" and then I get a co-authorship out of it as well.'

71

Issue #6: What to Keep/Delete: How to Appraise Your Data

- What should I keep?
- How long should I keep it?
- How to appraise data?
- Who decides?

72

Why Not Keep Everything?

- Storage space – both digital and physical
- Cost
- Outdated copies/back ups
- Makes version control harder
- Harder to fulfil requests for information/data

73

What should you keep, and how long for?

- It depends...
  - Funder requirements
  - Legal requirements
  - Institutional requirements

74

RCUK

- RCUK Common Principles on Data Policy
  - "Data with acknowledged long-term value should be preserved and remain accessible and usable for future research."
- RCUK Policy on Open Access
  - "All papers must include details of the funding that supported the research and, if applicable, a statement on how the underlying research materials – such as data, samples or models – can be accessed."
- Individual funders specify different retention periods
  - e.g. "The AHRC requires a minimum of three years after the end of project funding for both preservation and sustainability, but in many, if not most, cases a longer period will be appropriate."
  - e.g. EPSRC: "Research organisations must ensure that data is securely preserved for a minimum of 10 years from the end of any researcher 'privileged access' or, if others have accessed the data, from last date on which access to the data was requested by a third party."

75

Data Protection Act

Personal data:
- Relate to living individual
- Individual can be identified from those data or from those data and other information

- Fair processing:
  - Not kept longer than necessary
  - In accordance with the rights of data subjects, e.g. Right to be informed about how data will be used, stored, processed, transferred, destroyed, right to access the information and data held

- Security
  - Protected against unauthorised access, data loss, damage to data
  - Not transferred abroad without adequate protection

- Only disclosed if consent has been given to do so (except legal duty)

76

Institutional requirements

- RDM Policy
  - "8. **Relevant Data** that are likely to have long-term value, including data that substantiate research findings or represent records of the University, must be preserved and curated, following **Guidelines**, for as long they remain of value, and for at least any period stipulated by the funder or other relevant authority."

77

How to Appraise your Data?

- DCC guides (http://www.dcc.ac.uk/resources/how-guides):
  - Five steps to decide what data to keep
  - How to appraise and select research data for curation
- Issues to consider include:
  - Is the dataset unique?
  - Can the data be reused?
    - e.g. IPR and human subjects issues
  - Can the data be easily replicated?
  - Has the data been effectively documented?
  - Does the data underpin your publications/thesis?

78

Who decides what to keep?

- Shared responsibility between researchers and their organisations. For example:
  - PI in consultation with research group
  - Single researcher in consultation with supervisor/s
  - If contract work then in line with contract
- Where appropriate, seek further guidance e.g. for potential legal and ethical restrictions:
  - UMIP for IPR or licensing issues
  - Records Management Office for Data Protection issues
  - Research Office or Research Governance, Ethics and Integrity Team for ethics issues

79

Disposing of Research Data

- Once you have selected your data for retention make sure everything else is disposed of securely
- For confidential data, consult University guidance on confidential waste
  - http://www.estates.manchester.ac.uk/services/operationalservices/envsvcs/waste/waste&recycleconfidential/
- Securely shred physical materials
  - e.g. paper, CD/DVD
- Physically destroy portable media
  - e.g. memory sticks
- Erasing hard drives
  - Deleting data does not erase a file. Files need to overwritten to ensure they are irretrievably deleted.
  - Seek guidance from IT Service Desk
- See UK Data Service and ICO guidance on data disposal:
  - http://ukdataservice.ac.uk/manage-data/store/disposal.aspx
  - https://ico.org.uk/for-the-public/online/deleting-your-data

80

Acknowledgements

- MIT
- NECDMC
- UK Data Service
- University of Cambridge
- University of Edinburgh
- University of Exeter

81

MANCHESTER 1824
The University of Manchester

THANK YOU

Contact details:

researchdata@manchester.ac.uk

http://www.manchester.ac.uk/researchdata

82