



مجموعة جامعة  
الجميع الذكية  
للتعليم القابضة

MIDOCEAN  
UNIVERSITY

# Introduction to Research Methods

Prepared by Dr. Mona Moustafa Elashry

MGM 333

**Part 1: Starting the research process**

CH6: Research data management

**Lecture 7**

Chapter  
**06**

# Research data management





مجموعة جامعة  
الجميع الذكية  
للتعليم القابضة

MIDOCEAN  
UNIVERSITY

# Session Outline

Chapter  
**06**

## Session Outline

- ☐ Introduction- research data and its management challenges
- ☐ Why is research data management important?
- ☐ The research process, data lifecycles and research data management
- ☐ How do I manage my research data?
- ☐ Summary.



مجموعة جامعة  
الجميع الذكية  
للتعليم القابضة

MIDOCEAN  
UNIVERSITY

# Introduction-research data and its management challenges

- ❑ Research data management (RDM) comprises the activities that keep your data secure, accessible and usable.
- ❑ It is a fundamental part of good research practice.

- ❑ So, what is the research data that needs to be managed? Unfortunately, a single definition of research data does not exist.
- ❑ One generic definition states that research data is 'the evidence base on which . . . researchers build their analytic or other work' (Fry et al., 2008).

- ☐ Another defines it as being 'collected, observed, or created, for purposes of analysis to produce original research results'
- ☐ More detailed definitions depend on the research discipline and the nature of the research being undertaken..

**Table 6.1** Examples of research data types and formats

| Existing Data for Re-use                                                                                                                                         |                                                                                     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| Type                                                                                                                                                             | Format                                                                              |
| Data within articles, books, conference papers                                                                                                                   | Paper interlibrary loan<br>e-journal                                                |
| Books and documents: for historical study, for content analysis, etc.                                                                                            | Paper<br>Electronic                                                                 |
| Statistics: from published resources, from within the organization being researched                                                                              | Paper<br>Electronic<br>In a manipulable form in a database                          |
| Datasets from previous research                                                                                                                                  | Electronic<br>Published in archives/repositories/on websites                        |
| People's words already collected/recorded: e.g. diaries, audiovisual recordings, e-mails, podcasts, blog posts, tweets, discussion board entries, Facebook posts | Paper<br>Audiovisual<br>Electronic                                                  |
| Artefact, e.g. painting to be studied, designed object to be evaluated                                                                                           | Realia in many diverse physical formats such as specimens, samples, tools, textiles |

## New data collected/created

| Type                                                                                                                             | Format                                                                                                                                                                     |
|----------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Respondents' words: from interviews, focus groups, qualitative responses in questionnaires etc.                                  | Audiovisual<br>Paper<br>Electronic (e.g. Word file)<br>Original recording, transcript, anonymized transcript, analysed (e.g. in software such as NVivo), quoted in outputs |
| Numeric data (numbers, graphs, images): from questionnaires, measurements by instruments (e.g. gate counts, CCD readout, X-rays) | Paper<br>Audiovisual<br>Electronic<br>Analysed numeric data (e.g. in software such as Excel, SPSS)                                                                         |
| Meeting notes, researcher diary, researcher's thoughts                                                                           | Memory<br>Paper<br>Audiovisual<br>Electronic                                                                                                                               |

| Supporting documents and records                                                                                                                |                                   |
|-------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------|
| Type                                                                                                                                            | Format                            |
| Project documentation, e.g. proposal, research method protocols, research instrument templates, experimental set up, signed consent forms, etc. | Paper<br>Electronic               |
| Supporting documents and records                                                                                                                |                                   |
| Type                                                                                                                                            | Format                            |
| Communications, e.g. letters, e-mails, SMS, blog posts, tweets, etc.; both sent and received                                                    | Paper<br>Electronic<br>On the web |
| Outputs, e.g. project website, project blog, conference paper, journal article, etc.                                                            | Paper<br>Electronic<br>On the web |

- ❑ If we consider research data from different perspectives you can see the types of management challenges these raise:
- ❖ Re-using existing data or collecting/creating new data
  - Data that already exists will be managed by the data sources, therefore there is less need for you to manage it yourself.
  - However, data you create is your responsibility. It is important to be aware of the distinction between the two sources because this influences how you decide to manage your data.

## ❖ The form of the data – text, sound, image, artefact, performance

- Different forms require different management strategies

## ❖ The storage media – paper, analogue, digital, transient (i.e. captured only in human memory), in mobile/portable devices during field trips/data collection

- For any single project, data will be kept in different systems and in different physical locations.
- Each system requires different management techniques and a mapping of what is kept and where it is kept

## ❖ The supporting documents and records of the research process and activities

- Without these, research data can be meaningless or difficult to interpret, both by other people and by the researchers themselves after the passage of time.
- These documents/records need to be managed along with the data



مجموعة جامعة  
الجميع الذكية  
للتعليم القابضة

MIDOCEAN  
UNIVERSITY

# Why is research data management important?

- ❑ There are seven fundamental reasons why managing research data matters:
  - To comply with an agreed research requirement
  - To work effectively and efficiently
  - To protect the data
  - To enable data use and/or re-use
  - To share data with others
  - To preserve the data
  - As good research practice.

## ❑ A requirement

- Different stakeholders, such as research funders or universities, could place requirements on what you do with your research data.
- In such cases, you would have no choice but to manage the data in accordance with these requirements.

## ☐ To protect the data

- This means keeping your data safe, safeguarding it, protecting it.
- This covers many things from avoiding loss, theft, accidental deletion, corruption of files and virus attack to avoiding obsolescence, i.e. the inability to read the data storage device and/or format.
- 
- **Risks do materialize; real examples include the following:**

- **Risks do materialize; real examples include the following:**
  - An external hard drive containing a Masters dissertation dropped in transit and smashed; with no exact backup copy the dissertation had to be recovered by piecing together chapters in different files from various places.

- **Risks do materialize; real examples include the following:**
  - A PhD thesis chapter of approximately 12,000 words accidentally 'deleted'; a combination of key strokes brought up a blank document which the author, on exit at the save prompt, saved with the same file name.
  - It was only recovered through the careful work of computer forensics experts.

- **Risks do materialize; real examples include the following:**
  - An NVivo file, containing critical data analysis tracked, corrupting and no backup version stored.
  - MP3 interview files for a research project captured on a laptop placed in the boot of car which was broken into; the laptop was stolen and not recovered, the interview data was lost.

## ❑ To enable data use and/or re-use

- Data has value as an enabler of new opportunities and to analyze situations through time.
- In addition, re-using data makes good economic sense. There is a subtle but important distinction between use and re-use.

## ❑ To enable data use and/or re-use

- During a research project you use your data: you analyze it, synthesize it, produce outputs.
- After the research project you can continue using your data: it can be read, discussed, written about, used to generate new ideas and research questions.

## ❑ To enable data use and/or re-use

- Re-use means further analysis of the data to generate new data.
- You and possibly your colleagues can use it in new research projects. If data is shared then it can be used and/or re-used by others who were not involved in the original research; this falls into the realms of the next reason for RDM.

## ❑ To share data with others

- Often original research will benefit from appropriate sharing and in some instances this will be imperative to work as a research team.
- Data can only be effectively and efficiently shared if it is well managed to ensure that confidential data is protected but that data can be appropriately accessed as required.

## ❑ To share data with others

- In re-using data new data can be generated from the original data, for example by:
  - ✓ migrating data into a different format
  - ✓ creating new visualizations of the data
  - ✓ enhancing the data, e.g. by adding additional information
  - ✓ integrating the data with other data to create a combined dataset
  - ✓ applying new analyses and techniques to the data.

## ❑ To preserve the data

- If we are to be able to use/re-use data then preservation must be considered as early as possible, especially in the digital environment.
- Otherwise, as Rothenberg (1999, 42) stated over 15 years ago, ‘digital information lasts forever – or five years, whichever comes first.’ Improvements in the backward compatibility of software

## ❑ As good research practice

- A final reason why research data management is important is that it is simply part of good research practice, in the same way as is the initial literature review, project management, consent, methodological design, etc.



مجموعة جامعة  
الجميع الذكية  
للتعليم القابضة

MIDOCEAN  
UNIVERSITY

# The research process, data lifecycles and research data management

- ❑ The first thing to think about in managing research data is the 'big picture' i.e. the research process and the data lifecycle.
- ❑ The research process comprises a number of activities:

# The research process, data lifecycles and research data management

- planning the project
- collecting data (finding and accessing existing data to re-use it; obtaining new research data)
- processing and analyzing data
- producing outputs (sharing data with other researchers and the stages at which it may need to be shared and with whom; publishing results and scholarly articles, books, etc.)

- starting a new research project, informed by previous published research, and possibly re-using research data.
- There are many different interpretations and representations of the data lifecycle (Hodson, 2012).

- A simplified data lifecycle, applicable to any research project, comprises:
  - ✓ planning what to do, the data management plan
  - ✓ documenting the research process and activities, file structures and metadata for files, documents, etc.
  - ✓ storage (which aligns to the previous point) and access/confidentiality requirements

- A simplified data lifecycle, applicable to any research project, comprises:
  - ✓ appraisal to decide what to keep, and for how long to keep it, and what to securely delete/destroy
  - ✓ preservation
  - ✓ publication and sharing of data.



مجموعة جامعة  
الجميع الذكية  
للتعليم القابضة

MIDOCEAN  
UNIVERSITY

# How do I manage my research data?

- ❑ This section gives some guidance for managing data at the **different stages of the data lifecycle**, i.e. planning, documentation, appraisal, storage, preservation and publication

## ❑ The data management plan: planning what to do

- All research projects start with planning, determining the research question, the methodological approach, and how the data will be collected and analyzed.
- The planning stage is in some ways the most important stage. Decisions made here have an impact on every other stage of the lifecycle, so it is important to get it right from the start!

## ❑ The data management plan: planning what to do

- Whichever template is used, the data management plan is a living document – not something that is completed at the start of the research and then never revised.

## ❑ Documentation: naming and describing your data and its context

- Structuring and naming data consistently from the start is important. It helps with the identification, maintenance, management and sharing of data throughout a project.
- For example, if data is being collected or analyzed by more than one person it is very important to store data centrally, organized in a structured way that is readily understood with agreed naming conventions.

## ❑ Documentation: naming and describing your data and its context

- A file name can be an audit trail. For example, using the naming convention IPnRnYYYYMMDD, where I = interview, Pn = participant ID, Rn = Researcher ID and YYYYMMDD = date of the interview, a file named IP2R120121201.doc immediately identifies it as the interview of Participant 2 by Researcher 1 on 1 December 2012.

## ☐ Appraisal: selecting what to keep

- Although storage space and costs may not be an immediate issue, the 'keep everything' approach may not be viable in the longer term, or necessary.

## ❑ Appraisal: selecting what to keep

- As the volume of data kept increases it becomes harder to find what you want.
- So you need to decide what you must keep (legally or ethically), what you want to keep (for your own purposes or because the data will be of use to others) and what you can dispose of.

## ☐ Appraisal: selecting what to keep

- You also need to consider how you will keep data and for how long, and conversely how you will dispose of it.
- This is also the point at which you need to consider if the data could be re-used by you in another project, or if it could be shared for others to re-use.

## ❑ Appraisal: selecting what to keep

- These general criteria can help you to appraise your data:

- 1- Does the item fit with your organization's remit and priorities? Is there a legal requirement to keep, or delete, the item?
- 2- Is the item significant from a scientific, social or cultural viewpoint?
- 3 Is the item unique?
- 4 Is the item in a condition (covering technical, legal and ethical issues) that enables it to be kept, shared, and re-used?

## ❑ Appraisal: selecting what to keep

- These general criteria can help you to appraise your data:

5 Would it be feasible to replicate the item and if yes what would be the cost of doing so?

6 What is the cost of managing and preserving the item, and does its value merit such expenditure?

7 Is the metadata and contextual information needed to find, access and re-use the item comprehensive and correct?

## ☐ Storage: keeping data safe and secure

- You need to keep data safe, protecting it against corruption or loss. Mechanisms such as regular or automatic backups and anti-virus software help.
- You also need to keep data secure, controlling access to prevent unauthorized people accessing it and keeping personal details and data confidential

## ☐ Preservation: keeping data over time

- Depositing data in a specialist repository, will ensure its preservation but, unless preservation is considered before the decision to share it via a repository, there may be no data to share.
- You need to keep alert to any changes in the hardware and software that you use, and the effect this may have on existing files that you have.

## ☐ Preservation: keeping data over time

- The most practical way for you to deal with such changes is migration, i.e. copying your existing files onto new media or into new software versions.
- Keep a record that you have made such changes.

## ☐ Preservation: keeping data over time

- If data is to be kept long term, or placed in a repository, then it may need to be transferred into a format that is more suitable for preservation.
- 
- The Digital Preservation Coalition provides valuable advice on preservation strategies.

## ❑ Publication of outputs and sharing of data

- The outputs of a research project comprise the data itself as well as the scholarly works that present and discuss the findings, such as conference papers, journal articles and books.

## ☐ Publication of outputs and sharing of data

- If you decide to publish your data then there are a number of ways this can be done; for example, depositing it in a repository, publishing it as a data paper in a journal or as an enhanced publication, putting it up on a website.

## ☐ Publication of outputs and sharing of data

- When scholarly works are published their management becomes the responsibility of the publisher, and if versions are placed in another repository, the responsibility lies with the repository.



مجموعة جامعة  
الجميع الذكية  
للتعليم القابضة

MIDOCEAN  
UNIVERSITY

# Summary

- The nature of research data is extremely varied and managing it can be challenging.
- However, research data management is an essential part of the research process and integral to conducting good research.

- It is valuable to support researchers during the life of a research project and also enables data to be destroyed in a timely fashion, published and/or shared and preserved.
- It assures authenticity, reliability, integrity and usability and enables appropriate use or re-use.

- The data management plan is the starting point and key to adopting a fit-for-purpose, risk-managed approach to managing research data that is appropriate to its nature and potential value and meets any regulatory or other requirements.

- The processes of managing research data over its life can be applied to other data and information contexts and are transferable skills.



مجموعة جامعة  
الجميع الذكية  
للتعليم القابضة

MIDOCEAN  
UNIVERSITY



# Thank You

