Research Data Management 101
Workshop #1: Introduction to RDM

Le Tuan Anh

LIFESPAN Research Centre, Singapore

11 July 2019

Data and Research Data

What is data?

Data and Research Data

What is data?

Data = Datums = Values of attributes

Data and Research Data

What is data?

Data = Datums = Values of attributes

- Singular/scalar
    - Symbols
    - Numbers
    - Characters
    - . . .

Data and Research Data

What is data?

Data = Datums = Values of attributes

- Singular/scalar
    - Symbols
    - Numbers
    - Characters
    - . . .
- Composite data
    - Data table
    - Word list
    - Images/videos
    - . . .

Research Data

- Experiment data (surveys, measurements, images, videos, etc.)

Research Data

- Experiment data (surveys, measurements, images, videos, etc.)
- Reviewed articles in literature review

Research Data

- Experiment data (surveys, measurements, images, videos, etc.)
- Reviewed articles in literature review
- Review notes

Research Data

- Experiment data (surveys, measurements, images, videos, etc.)
- Reviewed articles in literature review
- Review notes
- Citation information

Research Data

- Experiment data (surveys, measurements, images, videos, etc.)
- Reviewed articles in literature review
- Review notes
- Citation information
- Analysis scripts, used libraries

Research Data

- Experiment data (surveys, measurements, images, videos, etc.)
- Reviewed articles in literature review
- Review notes
- Citation information
- Analysis scripts, used libraries
- Research journal, documentation, produced articles, etc.

Research Data

- Experiment data (surveys, measurements, images, videos, etc.)
- Reviewed articles in literature review
- Review notes
- Citation information
- Analysis scripts, used libraries
- Research journal, documentation, produced articles, etc.
- **Metadata**
- . . .

RDM in workflow

Main aspects

- Organising: storing and retrieving
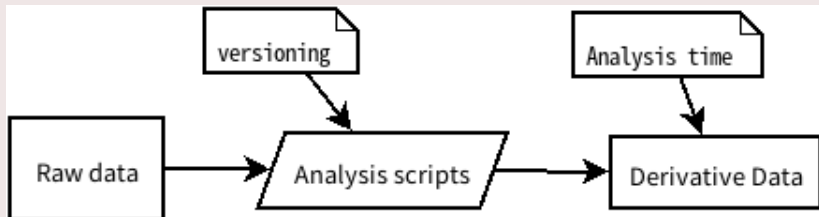- Versioning: tracking changes
- Archiving: storing data of finished tasks

RDM in workflow

Main aspects

- Organising: storing and retrieving
- Versioning: tracking changes
- Archiving: storing data of finished tasks

Storage

- Physical copies
  - Survey forms
  - Consent forms
  - Hard drives that store digital data
  - . . . .

RDM in workflow

Main aspects

- Organising: storing and retrieving
- Versioning: tracking changes
- Archiving: storing data of finished tasks

Storage

- Physical copies
  - Survey forms
  - Consent forms
  - Hard drives that store digital data
  - . . . .
- Digital copies
  - Files (also digitalised version of physical copies)
  - Folder (i.e. structure)

Analysis and Versioning

- Raw data (do not edit, archive immediately)
- Derivative data

Derivative data

Backup, or else . . .

FOR MY LOST LAPTOP

I am a Rutgers Chemistry 5th year PhD student. On April 19th afternoon, my LENOVO THINKPAD T420S laptop was stolen from room 203 of Wright-Rieman building. If you stole my laptop and now you are reading this letter, I would like to say that you can keep the computer and I would like to pay you money for my data under D drive. The data is my FIVE-YEAR work. I really need the data under the D drive, there is a folder named RESEARCH, under RESEARCH folder, there is a THESIS folder. I only need that folder for my thesis defense, which is coming very soon. I would like to pay you $1000 and use whatever way you offer to send you the money. The price is negotiable. My laptop password is 850713zd, my email address is [REDACTED] and phone number is [REDACTED] PLEASE contact me and I would appreciate it so so much!!!

Versioning, or else . . .

Versioning, or else . . .

File Types

Binary files A sequence of bytes, can be edited using a *Hex editor* (e.g. Hex Fiend for Mac OS, Notepad++/Hex plugin for Windows, GHex for Linux, etc.)

File Types

Binary files A sequence of bytes, can be edited using a *Hex editor* (e.g. Hex Fiend for Mac OS, Notepad++/Hex plugin for Windows, GHex for Linux, etc.)

Text files Stored using a particular encoding (UTF-8), can be edited using a *text editor* (VS Code, Notepad++, Sublime, Atom, Emacs, Geany, etc.)

File Types

Common research-related data

Source code text files

File Types

Common research-related data

Source code text files

Experiment data Mixed

- Data tables
- Images
- Audio
- Video

File Types

Common research-related data

Source code text files

Experiment data Mixed

- Data tables
- Images
- Audio
- Video

Literature Review Mixed

File Types

Common research-related data

Source code text files

Experiment data Mixed

- Data tables
- Images
- Audio
- Video

Literature Review Mixed

- Original articles (PDF/doc/web snapshot)

File Types

Common research-related data

Source code text files

Experiment data Mixed

- Data tables
- Images
- Audio
- Video

Literature Review Mixed

- Original articles (PDF/doc/web snapshot)
- Citation (jabref format)

File Types

Common research-related data

Source code text files

Experiment data Mixed

- Data tables
- Images
- Audio
- Video

Literature Review Mixed

- Original articles (PDF/doc/web snapshot)
- Citation (jabref format)
- Snippets

File Types

Common research-related data

Source code text files

Experiment data Mixed

- Data tables
- Images
- Audio
- Video

Literature Review Mixed

- Original articles (PDF/doc/web snapshot)
- Citation (jabref format)
- Snippets

Documentation Research journal, metadata, and articles, etc.:
mixed

Data handling 101

Security Password-protected archives on secure machines, disconnected from Internet

Data handling 101

Security Password-protected archives on secure machines, disconnected from Internet

Version control track your changes, *at integration level*

Data handling 101

Security Password-protected archives on secure machines, disconnected from Internet

Version control track your changes, *at integration level*

Backup Regularly, **HERE-NEAR-FAR**

Data handling 101

Security Password-protected archives on secure machines, disconnected from Internet

Version control track your changes, *at integration level*

Backup Regularly, **HERE**-**NEAR**-**FAR**

Naming convention

*topic_description_YYYYMMDD_LTA_v*0001*.tar.gz*

Topic : 30 characters

Description : 1-3 words

Timestamp : YYYYMMDD format

Initial : 2-4 characters

Versioning : v0001 format

Hand-on tasks

- Locate your HOME folder (Mac users: Command-Shift-H in Finder)

Hand-on tasks

- Locate your HOME folder (Mac users: Command-Shift-H in Finder)
- Creating a research workspace in your home folder

Hand-on tasks

- Locate your HOME folder (Mac users: Command-Shift-H in Finder)
- Creating a research workspace in your home folder
- Create a research project structure

Hand-on tasks

- Locate your HOME folder (Mac users: Command-Shift-H in Finder)
- Creating a research workspace in your home folder
- Create a research project structure
- Create raw data: 12 word files

Hand-on tasks

- Locate your HOME folder (Mac users: Command-Shift-H in Finder)
- Creating a research workspace in your home folder
- Create a research project structure
- Create raw data: 12 word files
- Copy some images

Hand-on tasks

- Locate your HOME folder (Mac users: Command-Shift-H in Finder)
- Creating a research workspace in your home folder
- Create a research project structure
- Create raw data: 12 word files
- Copy some images
- Create metadata file in markdown

To awesome people!

Thank you