

Kế hoạch/thực hành quản lý dữ liệu nghiên cứu

Ghi chép thông tin nghiên cứu (notebook)

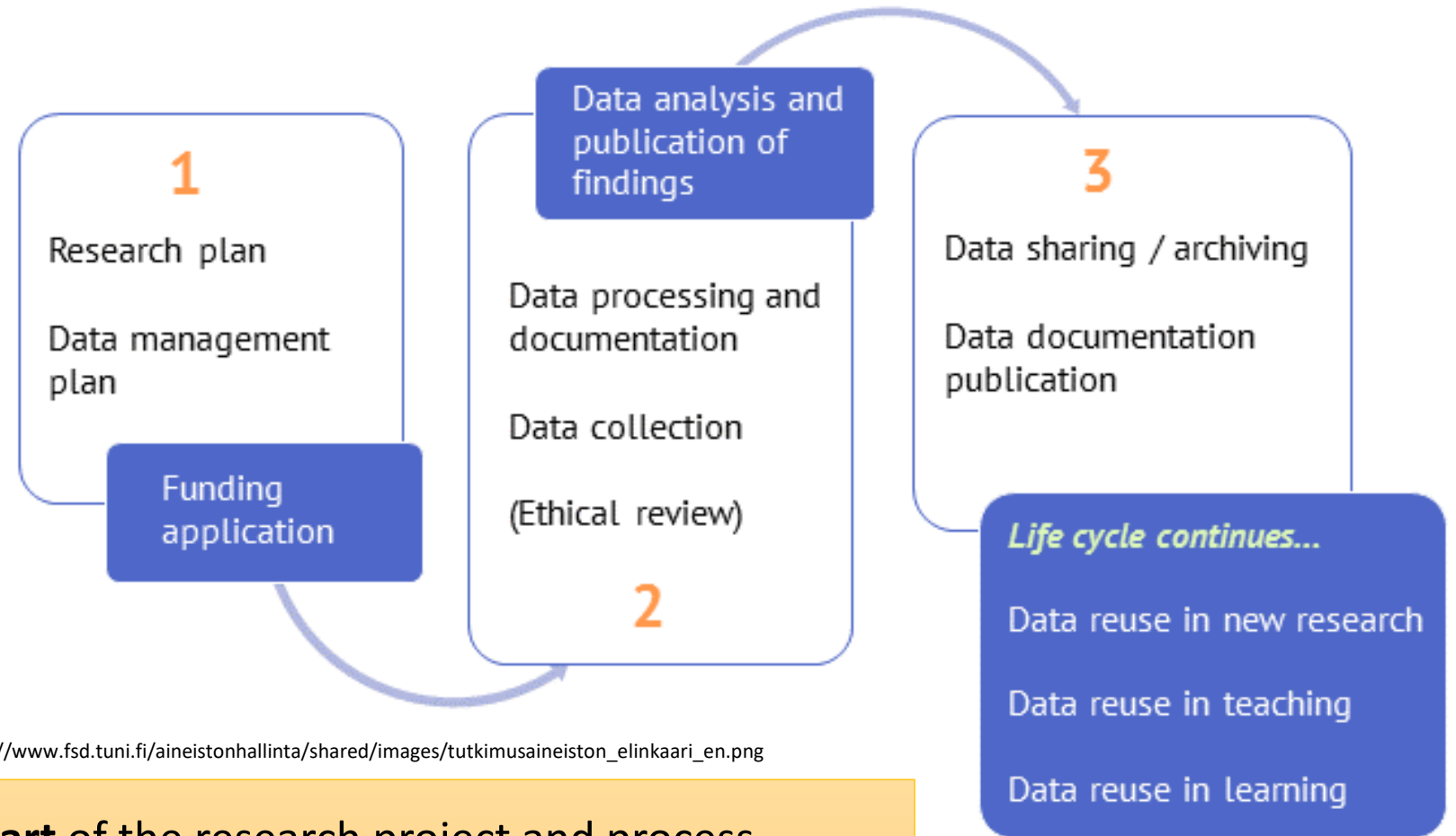
Nguyễn Thái Hoàng Tâm
email: nthtam@hcmus.edu.vn

Kế hoạch/thực hành quản lý dữ liệu nghiên cứu

Nguồn tài liệu chính sử dụng cho phần trình bày:

<https://www.fsd.tuni.fi/en/services/data-management-guidelines/>

“Vòng đời” của dữ liệu nghiên cứu - Research data life cycle



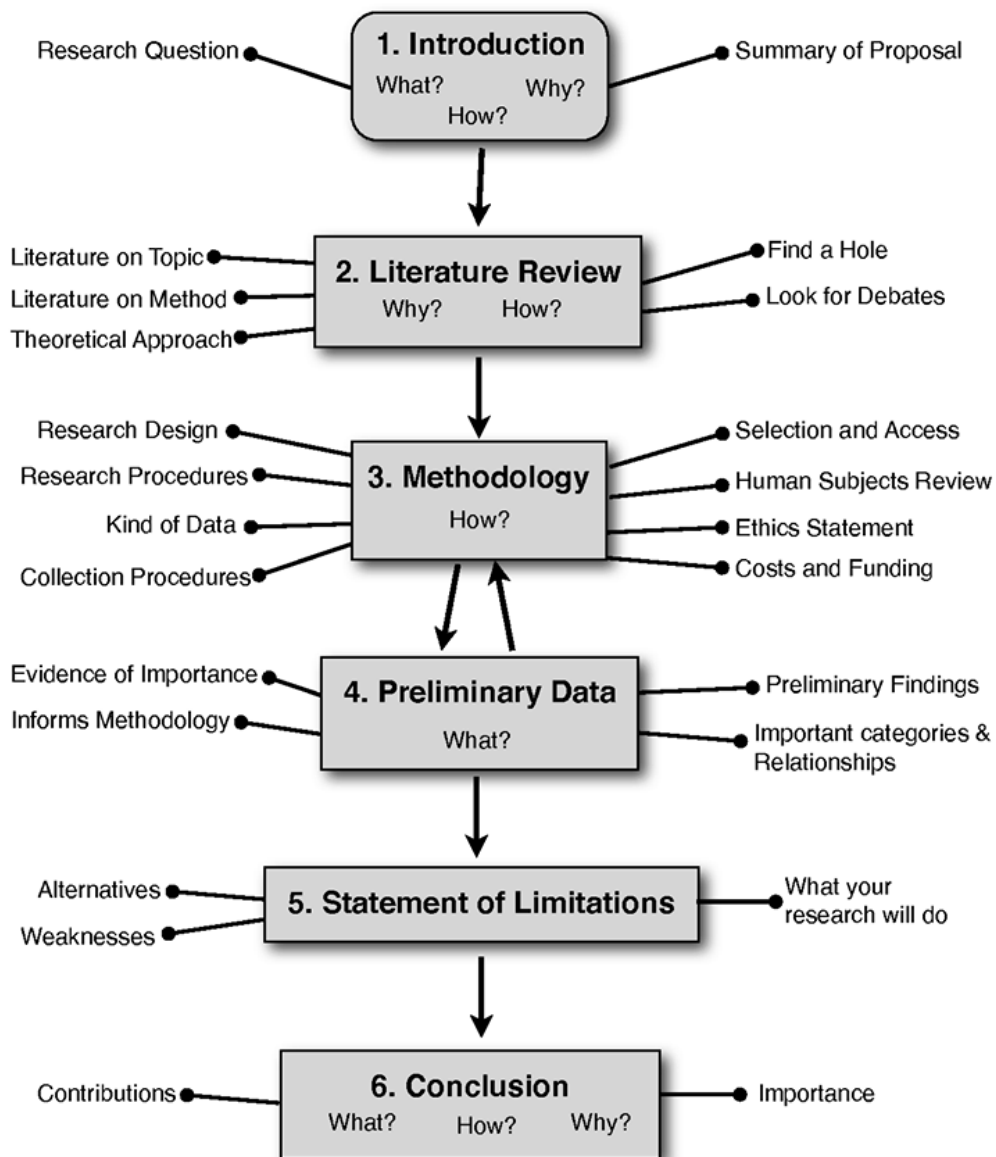
https://www.fsd.tuni.fi/aineistonhallinta/shared/images/tutkimusaineiston_elinkaari_en.png

1. Research data are **an essential part** of the research project and process
2. Archiving research data = academic **credit for researchers**
3. Reusing data is economic and saves resources.

Tại sao tôi phải quản lý dữ liệu và lập kế hoạch quản lý dữ liệu?

1. Rèn luyện các kỹ năng quản lý khoa học
2. Giảm nguy cơ thất lạc dữ liệu
3. Lường trước các vấn đề phức tạp về quyền sở hữu và quyền của người dùng
4. Hỗ trợ truy cập mở để tạo ra hiệu quả cộng tác làm việc.
5. Yêu cầu của nhà tài trợ/quỹ nghiên cứu (CBHD?/quỹ nc Việt Nam?)
6. Tiết kiệm thời gian và tiền bạc.
7. Kế hoạch và phương thức quản lý dữ liệu phản ánh các kỹ năng quản lý như một “leader”/lãnh đạo dự án

Research Proposal Flow Chart



Kế hoạch nghiên cứu (research plan) mô tả việc **xử lý dữ liệu** một cách khoa học, phân tích và phương pháp luận

→ **Kế hoạch quản lý dữ liệu (data management plan)** mô tả việc quản lý dữ liệu về mặt kỹ thuật và hành chính

1. Xác định dữ liệu → phân loại (dựa vào thông tin và mục tiêu sử dụng)
2. **Phương thức thu thập dữ liệu** (đảm bảo thông tin và tính hợp pháp của dữ liệu)
3. Hình thức-nơi lưu trữ - có hay không hình thành “metadata”?
4. Người thực chịu trách nhiệm/quyền sở hữu dữ liệu
5. **Tổ chức, sắp xếp dữ liệu, có chú thích cho từng cách tổ chức sắp xếp**
6. Lập bản kế hoạch để thực hiện → điều chỉnh phát sinh

Components of a Data Inventory

What is it?

- Title
- Author
- Description
- Date created

What does it consist of?

- How many files
- Where are they stored
- File size
- Are there multiple versions

Who Owns it?

- Department Owner
- Author
- Contact (Responsibility)

Who can access it?

- Restrictions in place
- Users who have access
- Protocol for handling requests

How Secure is it?

- Uniqueness
- Protections in place
- Back-up locations (2 or more)
- Procedure for format migration
- Retention period

Why is the data useful?

- Publications that use the data

Thông tin cơ bản cần biết về 1 dữ liệu

An important component of Research Data Management is performing a data inventory, a brainstorm of all the data that you will collect. Consider the following factors about your data:

- How will data be collected? In what formats?
- Will it be reproducible? What would happen if it got lost or became unusable later?
- Estimated size, growth and updating of data?
- What tools or software are needed?
- Will you use pre-existing data? From where?
- What is your storage and backup strategy?

Phương thức thu thập dữ liệu (đảm bảo thông tin và tính hợp pháp của dữ liệu)

Phụ thuộc nhiều vào định dạng của dữ liệu

| Loại dữ liệu | Định dạng lưu trữ | Lưu ý cho dữ liệu |
|---|---|---|
| Các dữ liệu chuyển đổi (Converting files) (StatTransfer software) | TIFF -> jpeg/Word → pdf | Dữ liệu dễ dàng bị mất do phần mềm chuyển đổi được lựa chọn sử dụng không phù hợp |
| Dữ liệu số hóa (bài báo) | → dạng lưu trữ pdf | |
| ma trận (dữ liệu dạng hàng, cột) và thống kê | các phần mềm thống kê đều lưu trữ siêu dữ liệu trong một tập dữ liệu | |
| Văn bản | TXT: Files saved as plain text (non-formatted, *.txt), RTF: Rich Text Format (*.rtf); DOC/DOCX: DOC files (*.doc, *.docx), ODT: OpenDocument Text (*.odt, *.fodt) | |
| Hình ảnh | TIFF duy trì được tất cả thông tin ảnh, không phụ thuộc vào hệ điều hành. BMP (sử dụng cho Window) (~ TIFF) JPEG (online, thông tin dễ dàng bị mất nếu chỉnh sửa hình ảnh). GIFT → PNG (web) | |
| Thu âm/hình | WAV (Waveform Audio File Format), MPEG-1/2 (MP3), MPEG-4 (H.264, mp4) | |

Vd: ma trận (dữ liệu dạng hàng, cột) và thống kê

GraphPad Prism - [THMG7 woVI_2017.pzf:IC50 of THMG7 wo vi_MCF7]

File Edit View Insert Change Arrange Window Help

The screenshot displays the top portion of the Prism 7 software window. The menu bar includes: Prism, File, Sheet, Undo, Clipboard, Analysis, Interpret, Change, Draw, Write, Text, Export, Print, Send, and Help. Below the menu bar is a toolbar with various icons for file operations (like saving, opening, and printing), editing (like undo, redo, and copy), and analysis (like graphing and fitting). The 'Text' menu is currently open, showing options for text formatting (bold, italic, underline, font color, background color) and text alignment (left, center, right, justified). The 'Help' menu is also visible, with an option to 'Open Prism Help' and a search bar.

| | | | | | | | | |
|--|--|--|--|--|--|--|--|--|
| Family | | | | | | | | |
| Data Tables | | | | | | | | |
| THMGT wo vi_MCF7 | | | | | | | | |
| THMGT wo vi_MCF7 - triplicate | | | | | | | | |
| THMGT wo vi_Fibro - triplicate | | | | | | | | |
| Summary IC50_SI | | | | | | | | |
| so sanh IC50_Fibro | | | | | | | | |
| So sanh SI | | | | | | | | |
| Copy of THMGT wo vi_Fibro - triplicate | | | | | | | | |
| Copy of THMGT_MCF7 | | | | | | | | |
| Info | | | | | | | | |
| Project info 1 | | | | | | | | |
| Results | | | | | | | | |
| IC50 of THMGT wo vi_MCF7 | | | | | | | | |
| IC50 of THMGT wo vi_MCF7 - triplicate | | | | | | | | |
| Nonlin fit of THMGT wo vi_Fibro - triplicate | | | | | | | | |
| 1way ANOVA of So sanh SI | | | | | | | | |
| 1way ANOVA of so sanh IC50_Fibro | | | | | | | | |
| Nonlin fit of Copy of THMGT_MCF7 | | | | | | | | |
| Graphs | | | | | | | | |
| THMGT wo vi_MCF7 | | | | | | | | |

| | | A | B | C | D | E | F | G |
|----|--------------------------|------------------|-----------------|------------------|----------------|------------------|-----------------|-------|
| | | THMGT 2017 | THMGT wo TH | THMGT wo MG | THMGT wo DB | THMGT wo TM | THMGT wo SD | Title |
| | | Y | Y | Y | Y | Y | Y | Y |
| 1 | IC50 determination | | | | | | | |
| 2 | Best-fit values | | | | | | | |
| 3 | IC50 | 0.7926 | 0.9334 | 0.8814 | 1.861 | 0.8143 | 0.9032 | |
| 4 | HILLSLOPE | 4.625 | 3.596 | 4.456 | 5.669 | 5.015 | 4.776 | |
| 5 | Std. Error | | | | | | | |
| 6 | IC50 | 0.05251 | 0.04947 | 0.05332 | 0.1699 | 0.05860 | 0.05294 | |
| 7 | HILLSLOPE | 0.9486 | 0.5963 | 1.021 | 1.632 | 1.266 | 1.238 | |
| 8 | 95% Confidence Intervals | | | | | | | |
| 9 | IC50 | 0.6792 to 0.9061 | 0.8266 to 1.040 | 0.7662 to 0.9966 | 1.494 to 2.228 | 0.6877 to 0.9408 | 0.7888 to 1.018 | |
| 10 | HILLSLOPE | 2.576 to 6.674 | 2.309 to 4.884 | 2.250 to 6.662 | 2.142 to 9.195 | 2.281 to 7.750 | 2.101 to 7.450 | |
| 11 | Goodness of Fit | | | | | | | |
| 12 | Degrees of Freedom | 13 | 13 | 13 | 13 | 13 | 13 | |
| 13 | R² | 0.9376 | 0.9474 | 0.9418 | 0.9783 | 0.9298 | 0.9437 | |
| 14 | Absolute Sum of Squares | 1489 | 1098 | 1456 | 660.7 | 1729 | 1488 | |
| 15 | Sy.x | 10.70 | 9.191 | 10.58 | 7.129 | 11.53 | 10.70 | |
| 16 | Number of points | | | | | | | |
| 17 | Analyzed | 15 | 15 | 15 | 15 | 15 | 15 | |

Cách thức tổ chức, sắp xếp dữ liệu, **có chú thích cho từng cách tổ chức** sắp xếp

Tổ chức, sắp xếp dữ liệu

Folder Structure

Folder Structure is an important part of Data Management. When created an organization structure, consider the following to improve your navigation and ease of access to data and files

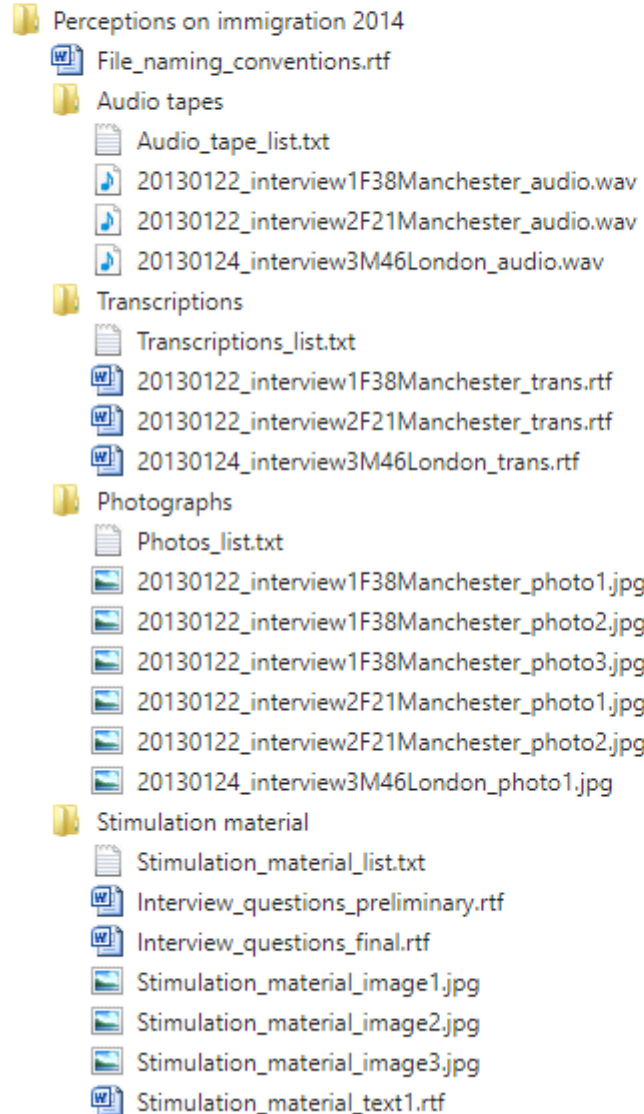
1. **Object type** - e.g. My Research Project, My Music, My Photos, etc.
2. **Organizational structure** - e.g. Department, sub-unit, or by individual.
3. **Combine**: structure directory by unit then type.

Best Practices for Naming Files

- Be Descriptive:
 - 75092238.txt is not very useful to colleagues and is easily forgettable.
 - 20120814_instrument8_rainyday_raw.txt (up to 255 characters) gives date, location condition and type of data collected.
- Don't rely on nesting in folders: 2012/august/instrument8/day14/raw.txt
 - Folder is useful for storage, but ultimately file naming conventions are more important to long term storage.
- Use consistent structure that falls into a useful order (for sorting) and decide on shared terminology
 - Speak with your colleagues/partners and agree on common terms and standards
- List versions alphanumerically, eg. v1, v2, v3 rather than *last*, *final*, *finalfinal*, *useTHISone*.
 - Number demonstrate a more logical order and cut down on your frustration.
- Use numerical dates, eg. YYYYMMDD rather than Dec09.
 - This will keep your files ordered by number rather than alphabetically. Otherwise, you'll have files organized by 2015APRIL, 2015DEC rather than 201501xx, 201502xx
- Some computers will not understand file names with UPPERCase letters, weird characters (/ , . # ?), or spaces between words.

Example 3: Structure of the data folder

In our example case, the data are varied and contain audio tapes of the interviews, interview transcripts, stimulation material shown to the research subjects, and photographs taken by the subjects. Please remember that background information must never be stored in the file name only (see the section Documenting background information).



Tổ chức, sắp xếp dữ liệu, **có chú thích cho từng cách tổ chức** sắp xếp

| Name | Date modified | Type | Size |
|--|-------------------|---------------|------|
| Experimental Results | 4/8/2020 11:22 AM | File folder | |
| Presentations | 4/8/2020 11:22 AM | File folder | |
| Proposals | 6/24/2020 7:10 AM | File folder | |
| References | 4/8/2020 11:22 AM | File folder | |
| Notes for main folder organization.txt | 2/12/2020 2:21 PM | Text Document | 1 KB |

The Notepad window displays the following text:

```
File Edit Format View Help
Folder "Experiment Results":
dùng cho việc upload các KQ thí nghiệm thô (raw) và sau xử lý phân tích

Folder "Presentation":
dùng cho việc upload các files (ppt, word) báo cáo định kỳ

Folder "Proposals":
dùng cho việc upload các file đề cương KLTN, KLTN, các quy trình thực nghiệm

Folder References:
dùng cho việc upload các file tài liệu được sử dụng cho các file trong phần Proposals và presentations

Đây là các folder yêu cầu chính. Trong các folder này, em tự tổ chức sắp xếp các folder con, các file.
Mỗi folder con cần có 1 file mô tả cách tổ chức folder và đặt tên file theo cấu trúc:time(ymd)_title_(...)
```

Readme.txt Files

Table 3. Relatively comprehensive, but unstructured metadata.

Written metadata

This experiment was designed to collect productivity, diversity, and soil data for Northern California grasslands. The results were published in a paper titled, "Soil nutrients and the relationship between diversity and productivity" (Doe and Smith 2003). Data were collected at two sites, the Coastal Hills Reserve and the Valley Oak Reserve, within the coastal mountains of Northern California. The area is primarily oak (*Quercus* spp.) savannah and grasslands on limestone soil. In spring of 2002, 10 1-m² plots were randomly distributed throughout a 100-km² area of each location. All plots were placed on grasslands. In each plot, plants were identified to species and then clipped at root level, dried, and weighed to obtain aboveground peak standing biomass. As most of the production is from annual plants, peak standing biomass can be used as an approximate measure of annual productivity. Approximately 0.5 g of soil was collected from the midpoint of each plot. This soil was taken back to the laboratory and analyzed for total nitrogen and phosphorus content.

All three species were observed in the plots. Nonnative plants observed included: *Avena fatua* and *Bromus hordeaceus*. Native plants included *Calochortus luteus*.

Data were collected by PI Jane Doe with assistance from graduate student John Smith in conjunction with the staff of the Coastal Hills Reserve and the Valley Oak Reserve. Collection of the data was funded by NSF grant No. 12345. Data may be used freely. Please acknowledge persons, grants, and reserves in any resulting publications.

Contact information:
Jane Doe
Department of Biology
Northern California University
University Town, CA 95666
(321) 654-0987
E-mail: doe@uncal.edu

When?
What?
Where?

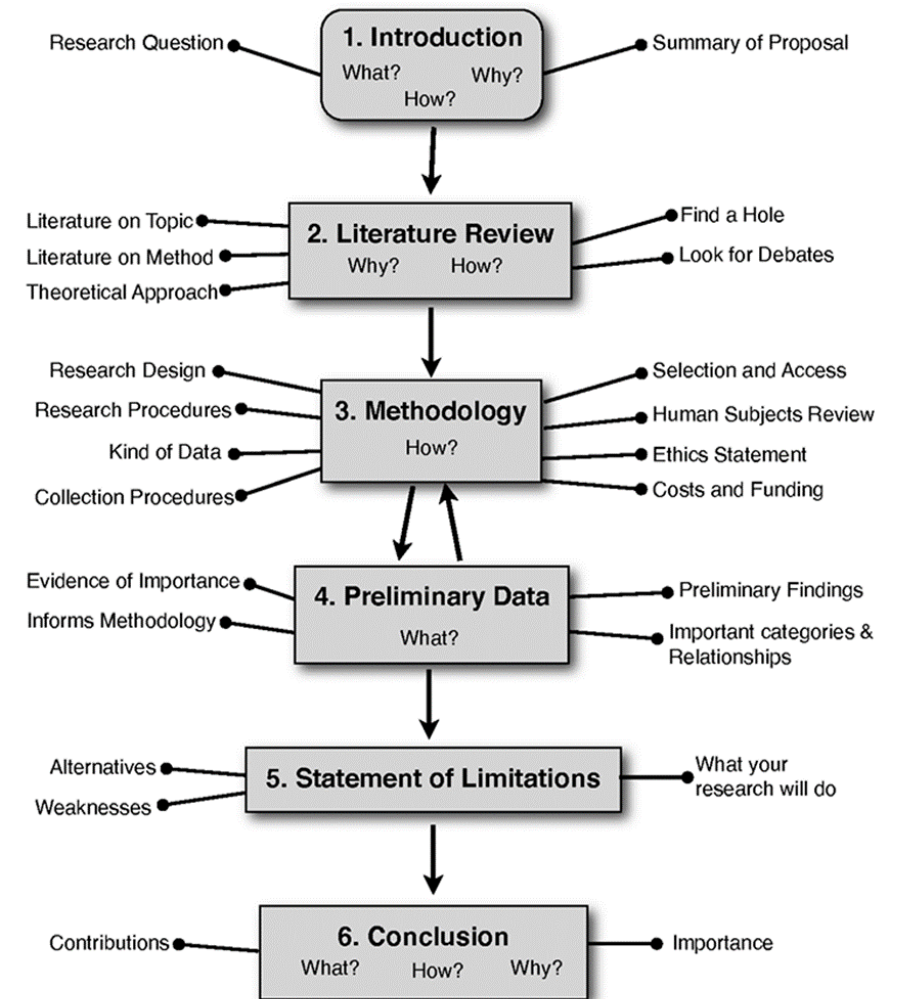
How?
Who?

Create a .txt file that offers a synopsis of the experiment that includes Time Frame, Subject, Location, Methods and Funding Institutions. This gives future researchers more context and documentation.

Các mục cơ bản của một bản kế hoạch quản lý dữ liệu (Data Management Plan)

1. General description of data/ Mô tả dữ liệu
2. Ethical and legal compliance/ Tính hợp pháp của dữ liệu
3. Documentation and metadata/ Tài liệu và dữ liệu tổng hợp
4. Storage and backup/ Lưu trữ và sao lưu dữ liệu
5. Opening, publishing and archiving the data after the research project/ Dữ liệu sau nghiên cứu
6. Data management responsibilities and resources/Trách nhiệm quản lý dữ liệu

Research Proposal Flow Chart



Bài tập “viết bản kế hoạch quản lý dữ liệu” cho đề tài nghiên cứu/KLTN

- Theo mẫu tham khảo, nội dung trình bày không quá 5 trang A4
- Nộp file kế hoạch cùng 1 bản tóm tắt đề cương nghiên cứu/KLTN sau khi SV nhận đề tài làm KLTN (sẽ cập nhật thời gian cụ thể cho SV)
- HV nộp về email: nthtam@hcmus.edu.vn;

| | |
|--|--------------------------------|
| 1. THÔNG TIN CHUNG | |
| Họ tên học viên | Tên đề tài/mã số chủ đề: |
| MSHV | |
| Thời gian thực hiện chủ đề | Cơ quản/Cá nhân quản lý đề tài |
| | Phương thức liên hệ: |
| 2. MÔ TẢ DỮ LIỆU | |
| ✓ Mô tả ngắn gọn các dữ liệu được thu thập hoặc được tạo ra hoặc tái sử dụng lại (có hợp pháp) | |
| ✓ Định dạng của những dữ liệu này là gì? Kích thước dữ liệu? | |
| ✓ Trình bày ngắn gọn cách thức các dữ liệu được thu thập, phân tích, xử lý có thể ảnh hưởng đến chất lượng/độ chính xác của dữ liệu và cách thức hạn chế những điều này? | |
| ✓ Quyền sở quản lý và sở hữu các dữ liệu? Dữ liệu được chia sẻ? Đối tượng được chia sẻ | |

| | |
|---|--|
| 3. PHƯƠNG THỨC LƯU TRỮ DỮ LIỆU | |
| 3.1. Trong thời gian thực hiện đề tài | |
| ✓ Mô tả ngắn gọn cách sắp xếp, nơi lưu trữ dữ liệu (trữ/sao lưu) cùng kích thước dữ liệu | |
| ✓ Có “metadata”? Nếu có, mô phương thức thực hiện/công cụ sử dụng Nếu không, mô tả ngắn gọn cách thức để tái sử dụng dữ liệu nghiên cứu | |
| ✓ Thời gian định kỳ sao lưu dữ liệu? | |
| ✓ Người thực hiện sao lưu? | |
| ✓ Kinh phí lưu trữ dữ liệu dự kiến? | |
| 3.2. Sau khi kết thúc đề tài | |
| ✓ Mô tả ngắn gọn những dữ liệu tiếp tục được lưu trữ? Và cách thức lưu trữ | |
| ✓ Mô tả ngắn gọn những dữ liệu tiếp tục được lưu trữ và được phép tái sử dụng? Và cách thức lưu trữ | |
| ✓ Quyền sở hữu/chia sẻ dữ liệu | |
| Thời gian lập bản kế hoạch | |
| Người lập bản | |

Ghi chép thông tin nghiên cứu (notebook)

Thông tin nghiên cứu được thu thập

Thông tin được tạo mới từ thực nghiệm (Laboratory notebook/Labbook)

Nguồn tài liệu chính sử dụng cho phần trình bày:

http://ocw.mit.edu/courses/biology/7-16-experimental-molecular-biology-biotechnology-ii-spring-2005/labs/lab_notebook.pdf

Ghi chép thông tin để làm gì?

... là một hình thức quản lý dữ liệu trong suốt quá trình nghiên cứu thực nghiệm

→ xem lại/phân tích/tổng hợp (thông tin/dữ liệu) → kết luận giả thuyết/xuất hiện vấn đề mới....

Content of the Laboratory Notebook

1. Table of Contents

- Title
- Date
- Page Number

Thực hiện sau

2. Date of Experiment

3. Title of Experiment (e.g., Serial knock-down of DIAP1 by RNAi)

4. Purpose (e.g. To determine the relationship between the amount of RNA- used against DIAP1 and the amount of cell death.)

5. Materials and Methods

- Protocol
- Written
- Pasted
- May refer to previous protocol in notebook (note any changes)
- Write before you begin procedure
- Amend as you go through the experiment
- List any calculations

- Có thể ghi 1 lần đầy đủ/chi tiết các mục 3, 4, 5 cho mỗi một mục tiêu/nội dung thực nghiệm, → lần sau có thể trích dẫn phần đã ghi, nếu có điểm khác thì cần ghi lại.
- Những lần lặp thực nghiệm cùng mục tiêu, có thể ghi thông tin mục 2, 6 và 7. Mục 5 chỉ ghi những thay đổi nhỏ (nếu có)

Content of the Laboratory Notebook

6. Observations and Results

- Everything that happens or doesn't happen is data.
- Any writing that will facilitate data entry should be planned out in advance
- Results may include:
 - Tables
 - Charts
 - Graphs
 - Printouts
 - Pictures
 - Gels
 - Films
 - Calculations

Thực hành hoàn chỉnh mục 6 và 7 sau mỗi lần thực nghiệm

7. Discussion and Conclusion

- Discuss results and implications of data.
- Prepare a conclusion. How did the experiment go?
- What to do next?

Electronic Lab Notebooks

Not keen on using a physical notebook? Need a digital Notebook for documenting your laboratory and experiments? Try these two Electronic Lab Notebooks.

Your Laboratory and Advisor may have alternatives, discounts and preferred tools in this process. Please consult with them for best options.

- [LabArchives.com](https://labarchives.com)

"LabArchives® is the leading secure and intuitive cloud-based Electronic Lab Notebook (ELN) application enabling researchers to easily create, store, share and manage their research data. Far more than an "ELN", LabArchives provides a flexible, extensible platform that can be easily customized to match your lab's workflow providing benefits to Principal Investigators, lab managers & staff, post doctoral fellows and grad students."

- [ResearchSpace.com \(Personal Edition\)](https://researchspace.com)

- Create and Import research data
- Organize Data
- Manage Samples
- Allows for importing of MsOffice Documents and Image files



Ưu/nhược điểm của ghi chép điện tử và bản giấy?

Ghi chép thông tin “đọc” tài liệu READING NOTEBOOK

LƯU Ý:

- Ghi chú mục tiêu/vấn đề tìm kiếm thông tin để tránh ghi chép thông tin xa mục tiêu ban đầu trong quá trình đọc
- Ghi chú tài liệu đọc cho mỗi phần ghi chép thông tin để tra cứu lại nguồn đọc

Literature Review Template

Definition: A literature review is an objective, critical summary of published research literature relevant to a topic under consideration for research. Its purpose is to create familiarity with current thinking and research on a particular topic, and may justify future research into a previously overlooked or understudied area.

A typical literature review consists of the following components:

1. Introduction:

- A concise **definition of a topic** under consideration (this may be a descriptive or argumentative thesis, or proposal), as well as the **scope** of the related literature being investigated. (Example: If the topic under consideration is ‘women’s wartime diaries’, the scope of the review may be limited to published or unpublished works, works in English, works from a particular location, time period, or conflict, etc.)
- The introduction should also note intentional **exclusions**. (Example: “*This review will not explore the diaries of adolescent girls.*”)
- Another purpose of the introduction is to state the **general findings** of the review (what do *most* of the sources conclude), and comment on the **availability** of sources in the subject area.

2. Main Body:

- There are a number of ways to organize the evaluation of the sources. **Chronological and thematic approaches** are each useful examples.
- Each work should be critically summarized and evaluated for its **premise, methodology, and conclusion**. It is as important to address inconsistencies, omissions, and errors, as it is to identify accuracy, depth, and relevance.
- Use logical connections and **transitions** to connect sources.

3. Conclusion

- The conclusion **summarizes the key findings** of the review in general terms. Notable commonalities between works, whether favourable or not, may be included here.
- This section is the reviewer’s opportunity to **justify a research proposal**. Therefore, the idea should be clearly re-stated and supported according to the findings of the review.