

**BỘ GIÁO DỤC  
VÀ ĐÀO TẠO  
HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ**

**VIỆN HÀN LÂM  
KHOA HỌC VÀ CÔNG NGHỆ VN  
HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ**



**Nguyễn Hồng Việt**

**NGUYỄN HỒNG VIỆT**

**HỆ THỐNG THÔNG TIN**

**2023**

**NGHIÊN CỨU XÂY DỰNG KIẾN TRÚC DỮ LIỆU NGHIÊN  
CỨU KHOA HỌC VÀ CÔNG NGHỆ**

**LUẬN VĂN THẠC SĨ HỆ THỐNG THÔNG TIN**

***Hà Nội – 2023***

**BỘ GIÁO DỤC  
VÀ ĐÀO TẠO**

**VIỆN HÀN LÂM  
KHOA HỌC VÀ CÔNG NGHỆ VN**

**HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ**



**Nguyễn Hồng Việt**

**NGHIÊN CỨU XÂY DỰNG KIẾN TRÚC DỮ LIỆU NGHIÊN CỨU  
KHOA HỌC VÀ CÔNG NGHỆ**

Chuyên ngành: Hệ thống thông tin  
Mã số: 8480104

**LUẬN VĂN THẠC SĨ NGÀNH HỆ THỐNG THÔNG TIN**

NGƯỜI HƯỚNG DẪN KHOA HỌC :  
TS. NGUYỄN NHƯ SƠN

***Hà Nội – 2023***

**LỜI CAM ĐOAN**

*Tôi xin cam đoan đề tài nghiên cứu trong luận văn này là công trình nghiên cứu của tôi dựa trên những tài liệu, số liệu do chính tôi tự tìm hiểu và nghiên cứu. Chính vì vậy, các kết quả nghiên cứu đảm bảo trung thực và khách quan nhất. Đồng thời, kết quả này chưa từng xuất hiện trong bất cứ một nghiên cứu nào. Các số liệu, kết quả nêu trong luận văn là trung thực nếu sai tôi hoàn toàn chịu trách nhiệm.*

**Tác giả luận văn**

**Nguyễn Hồng Việt**

## LỜI CẢM ƠN

Lời đầu tiên, Tôi xin chân thành cảm ơn các thầy cô tại Học Viện Khoa học và Công nghệ đã giảng dạy, cung cấp kiến thức và tạo điều kiện hoàn thành chương trình học cao học.

Tôi xin bày tỏ lòng biết ơn sâu sắc tới Thầy TS. Nguyễn Như Sơn, Học Viện Khoa học và Công nghệ đã nhiệt tình hướng dẫn, tạo điều kiện thuận lợi cho tôi nghiên cứu khoa học, và giúp tôi có thể hoàn thành luận văn một cách tốt nhất.

Tôi xin cảm ơn sự chỉ bảo, giúp đỡ, đóng góp ý kiến của PGS.TS. Nguyễn Long Giang, PGS.TS. Nguyễn Việt Anh và PGS. TS. Nguyễn Đức Dũng đã dành cho tôi trong những năm qua.

Sau cùng, tôi muốn gửi tới tất cả những người thân trong gia đình và bạn bè lời cảm ơn chân thành nhất. Chính sự tin yêu mong đợi của gia đình và bạn bè đã tạo động lực cho tôi thực hiện thành công luận văn này.

Xin trân trọng cảm ơn!

**Tác giả luận văn**

**Nguyễn Hồng Việt**

## MỤC LỤC

LỜI CAM ĐOAN .....	3
LỜI CẢM ƠN .....	4
MỤC LỤC.....	5
DANH MỤC CÁC HÌNH.....	7
DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ CÁI VIẾT TẮT .....	9
MỞ ĐẦU .....	9
1.1 Lý do chọn đề tài.....	9
1.2 Mục tiêu nghiên cứu .....	11
1.3 Phạm vi nghiên cứu .....	12
1.4 Phương pháp nghiên cứu .....	12
1.5 Những đóng góp của luận văn .....	13
1.6 Bố cục của Luận văn .....	13
<b>CHƯƠNG 1. TỔNG QUAN NGHIÊN CỨU LIÊN QUAN ĐẾN KIẾN TRÚC DỮ LIỆU NGHIÊN CỨU KHOA HỌC VÀ CÔNG NGHỆ.....</b>	<b>14</b>
1.1. Tổng quan nghiên cứu.....	14
1.2. Tình hình nghiên cứu trong nước .....	17
1.3. Tình hình nghiên cứu ở nước ngoài.....	22
<b>KẾT LUẬN CHƯƠNG 1.....</b>	<b>28</b>
<b>CHƯƠNG 2. NGHIÊN CỨU XÂY DỰNG KIẾN TRÚC DỮ LIỆU NGHIÊN CỨU KHOA HỌC VÀ CÔNG NGHỆ.....</b>	<b>29</b>
2.1 Giới thiệu chung.....	29
2.2 Kiến trúc dữ liệu trong nghiên cứu khoa học và công nghệ.....	30
2.3 Nghiên cứu xây dựng kiến trúc dữ liệu nghiên cứu khoa học và công nghệ 35	
2.3.1 Khối tích hợp dữ liệu (Ingestion) .....	37
2.3.2 Khối lưu trữ (storage).....	41
2.3.3 Khối người dùng (user) .....	42
<b>KẾT LUẬN CHƯƠNG 2.....</b>	<b>43</b>
<b>CHƯƠNG 3: THỰC NGHIỆM MÔ HÌNH VÀ ĐÁNH GIÁ KẾT QUẢ.....</b>	<b>44</b>
3.1 Công cụ và môi trường thử nghiệm.....	44
3.2 Thực nghiệm thao tác trên hệ thống .....	45

<b>KẾT LUẬN CHƯƠNG 3</b> .....	54
<b>KẾT LUẬN</b> .....	55
<b>1. Kết quả đạt được</b> .....	55
<b>2. Hạn chế</b> .....	55
<b>3. Hướng nghiên cứu tiếp theo</b> .....	55
<b>TÀI LIỆU THAM KHẢO</b> .....	56
<b>1. Tài liệu tham khảo tiếng Việt</b> .....	56
<b>2. Tài liệu tham khảo tiếng Anh</b> .....	56

## DANH MỤC CÁC HÌNH

Hình 1.3.1: Hình mô phỏng vòng đời quá trình hình thành và lưu trữ dữ liệu DLNC .....	22
Hình 1.3.2: Mô hình tháp quản lý dữ liệu nghiên cứu cho các thư viện, được trình bày bởi Lewis [14].....	23
Hình 1.3.3: Hình ảnh thành tựu của MU (Monash University, 2019) về quản lý dữ liệu nghiên cứu khoa học và công nghệ từ năm 2006 đến năm 2017 ...	28
Hình 2.3.1: Mô hình kiến trúc của hệ thống .....	36
Hình 2.3.2: Khối tích hợp dữ liệu (Ingestion).....	37
Hình 2.3.3: Mô hình các nguồn dữ liệu khoa học và công nghệ của hệ thống .....	38
Hình 2.3.4: Chương trình mã nguồn mở Nifi (Logo của mã nguồn mở Nifi)	39
Hình 2.3.2.1: Máy chủ lưu trữ vật lý.....	42
Hình 2.3.3.1: Người dùng hệ thống .....	42
Hình 3.1.1: Công cụ và Môi trường thử nghiệm.....	45
Hình 3.2.1: Màn hình đăng nhập vào hệ thống System Introductions .....	46
Hình 3.2.2: Màn hình đăng nhập vào hệ thống Aspect.....	47
Hình 3.2.3: Màn hình đăng nhập vào hệ thống Documentations và Databases .....	47
Hình 3.2.4: Màn hình đăng nhập vào hệ thống Documentations và Databases .....	48
Hình 3.2.5: Màn hình đăng nhập vào hệ thống Contribute Documents .....	49
Hình 3.2.6: Màn hình giao diện kết quả sau khi được gửi lên hệ thống.....	49
Hình 3.2.7: Màn hình đăng nhập vào hệ thống Contribute Documents .....	50
Hình 3.2.8: Màn hình đăng nhập vào hệ thống để chọn và tải tập tin lên .....	50
Hình 3.2.9: Màn hình đăng nhập vào hệ thống để chọn up dữ liệu.....	51

Hình 3.2.10: Màn hình đăng nhập vào hệ thống để chọn tải dữ liệu để huấn luyện mô hình học máy .....	52
Hình 3.2.11: Màn hình hiển thị khi dữ liệu được huấn luyện thành công .....	53
Hình 3.2.12: Màn hình thông tin liên hệ tại Học Viện Khoa học và Công nghệ .....	53



**DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ CÁI VIẾT TẮT**

<b>STT</b>	<b>Từ viết tắt</b>	<b>Từ tiếng anh</b>	<b>Diễn giải/tạm dịch</b>
1	RDBMS	Relational Database Management System	Hệ quản trị cơ sở dữ liệu quan hệ
2	MRC	Mekong River Commission	Ủy hội sông Mê Công
2	PDIES	Procedures for Data and Information Exchange and Sharing	Quy trình trao đổi và chia sẻ dữ liệu và thông tin
3	ODV	Open Development Vietnam	Phát triển Mở Việt Nam
4	ESRC	Economic and Social Research Council	Hội đồng nghiên cứu Kinh tế và Xã hội
5	GIS	Geographic Information System	Hệ thống thông tin địa lý
6	IOT	Internet of Things	Internet vạn vật

## MỞ ĐẦU

### 1.1 Lý do chọn đề tài

*Thứ nhất*, trong giai đoạn hiện nay, dữ liệu nghiên cứu trong khoa học và Công nghệ là một nguồn tài nguyên quý báu để thúc đẩy sự phát triển và đổi mới trong các lĩnh vực, dữ liệu nghiên cứu trong khoa học và công nghệ có ảnh hưởng lớn đến sự phát triển của xã hội và nền kinh tế. Nghiên cứu trong lĩnh vực này có thể mang lại giá trị thực sự cho cộng đồng khoa học và công nghệ.

Dữ liệu trong nghiên cứu khoa học và Công nghệ luôn luôn thay đổi và phát triển. Việc quản lý, chia sẻ nguồn tài nguyên quý này là rất quan trọng và Dữ liệu trong nghiên cứu khoa học và Công nghệ có thể tạo ra cơ hội cho sự đổi mới và khám phá các khía cạnh mới của một lĩnh vực cụ thể hoặc giúp phát triển các phương pháp mới để nghiên cứu các hướng tiếp theo.

Dữ liệu trong nghiên cứu khoa học và Công nghệ cho phép các nhà nghiên cứu chia sẻ kiến thức của họ với cộng đồng khoa học và xã hội. Điều này làm cho tiến bộ khoa học trở nên mở cửa và có thể được đánh giá và kiểm tra bởi người khác.

*Thứ hai*, xuất phát từ thực tế tại Viện Hàn lâm Khoa học và Công nghệ Việt Nam (Sau đây viết tắt là Viện Hàn lâm) là một tổ chức quốc gia chuyên về lĩnh vực khoa học và công nghệ tại Việt Nam. Chức năng nhiệm vụ chính của Viện Hàn lâm bao gồm:

Nghiên cứu và phát triển khoa học công nghệ: Viện Hàn lâm thực hiện các dự án nghiên cứu khoa học và phát triển công nghệ, từ đó đóng góp vào sự phát triển của ngành khoa học và công nghệ ở Việt Nam. Đây có thể bao gồm nghiên cứu cơ bản, ứng dụng và phát triển sản phẩm công nghệ mới.

Phát triển nguồn nhân lực: Viện Hàn lâm đào tạo và phát triển nguồn nhân lực trong lĩnh vực khoa học và công nghệ. Điều này có thể bao gồm đào tạo các nhà khoa học, kỹ sư, và chuyên gia công nghệ thông qua các chương trình đào tạo, học bổng và khóa học đào tạo.

Viện Hàn lâm cung cấp tư vấn về các vấn đề liên quan đến khoa học và công nghệ cho chính phủ và các tổ chức có liên quan. Điều này giúp chính phủ và

các cơ quan quyết định có thông tin cơ bản để đưa ra các quyết định chính sách hợp lý trong lĩnh vực khoa học và công nghệ.

Tổ chức hội nghị, Hội thảo và sự kiện: Viện Hàn lâm thường xuyên tổ chức hội nghị, seminar, hội thảo và các sự kiện khác để tạo điều kiện cho việc trao đổi thông tin, kết nối các nhà nghiên cứu, và thúc đẩy sự phát triển trong lĩnh vực nghiên cứu khoa học và công nghệ.

Xuất bản và công bố: Viện Hàn lâm có thể xuất bản sách, bài báo, và tài liệu nghiên cứu để chia sẻ kiến thức và kết quả nghiên cứu với cộng đồng khoa học và công chúng.

Tổng quan, Viện Hàn lâm có vai trò quan trọng trong việc thúc đẩy phát triển của khoa học và công nghệ tại Việt Nam và đóng góp vào sự phát triển và cải thiện chất lượng cuộc sống của người dân như:

- Nghiên cứu tổng hợp các nguồn tài nguyên thiên nhiên, điều kiện tự nhiên và môi trường;
- Triển khai, ứng dụng và chuyển giao các kết quả nghiên cứu khoa học, công nghệ;
- Đề xuất và chủ trì thực hiện các nhiệm vụ khoa học và công nghệ trọng điểm cấp quốc gia theo phân công của cơ quan nhà nước có thẩm quyền;
- Tuyên truyền, phổ biến kiến thức, kết quả nghiên cứu khoa học và phát triển công nghệ đáp ứng yêu cầu phát triển thị trường khoa học, công nghệ.

Hiện tại Viện Hàn lâm gồm 28 Viện nghiên cứu, 06 Đơn vị giúp việc cho Chủ tịch, 08 đơn vị sự nghiệp và 01 Công ty TNHH.

Do vậy có rất nhiều các dự án, đề tài các sản phẩm trong nghiên cứu khoa học và Công nghệ, có rất nhiều dữ liệu trong nghiên cứu được hình thành từ các nhà Khoa học, các công trình, đề tài (Đề tài cấp nhà nước, cấp viện và các đề tài khoa học trẻ ...) Hiện tại dữ liệu nghiên cứu khoa học đó chưa được lưu trữ hay quản lý một cách có hệ thống.

Bản thân tôi nhận thấy là một cán bộ tham gia trực tiếp hoạt động ứng dụng Công nghệ thông tin của Trung tâm Tin học và Tính toán, Viện Hàn lâm Khoa học và Công nghệ Việt Nam nên đòi hỏi phải luôn nghiên cứu tìm tòi, học hỏi

để nghiên cứu xây dựng kiến trúc dữ liệu trong nghiên cứu khoa học và công nghệ vào hoạt động đào tạo và nghiên cứu khoa học tại Viện Hàn lâm.

Xuất phát từ những vấn đề nêu trên tôi đã chọn đề tài **“Nghiên cứu xây dựng kiến trúc dữ liệu nghiên cứu khoa học và Công nghệ”**.

## **1.2 Mục tiêu nghiên cứu**

- Phân tích thực trạng, Nghiên cứu xây dựng kiến trúc dữ liệu nghiên cứu khoa học và công nghệ.

- Trên cơ sở kiến trúc dữ liệu nghiên cứu khoa học và công nghệ, tác giả đưa ra xây dựng mô hình thử nghiệm lưu trữ quản lý dữ liệu nghiên cứu Khoa học và Công nghệ tại Học viện khoa học và công nghệ.

Cụ thể:

Mục tiêu nghiên cứu về kiến trúc dữ liệu trong lĩnh vực khoa học và công nghệ là tạo ra các hệ thống và cơ sở dữ liệu hiệu quả, linh hoạt và bảo mật để quản lý và phân tích dữ liệu liên quan đến nghiên cứu và phát triển công nghệ. Dưới đây là một số mục tiêu cụ thể trong nghiên cứu này:

Tối ưu hóa cơ sở dữ liệu khoa học và công nghệ: Phát triển các kiến trúc dữ liệu và hệ thống quản lý dữ liệu hiệu quả để lưu trữ và truy xuất thông tin từ các dự án nghiên cứu và công nghệ. Mục tiêu là tối ưu hóa hiệu suất, giảm thiểu thời gian tìm kiếm và truy xuất dữ liệu, và tăng cường khả năng tương tác với dữ liệu.

Tích hợp dữ liệu đa nguồn: Phát triển các phương pháp và công nghệ để tích hợp dữ liệu từ nhiều nguồn khác nhau, bao gồm dữ liệu từ các dự án nghiên cứu, cơ sở dữ liệu công nghệ, tài liệu nghiên cứu, và dữ liệu mô phỏng. Mục tiêu là tạo ra một hệ thống toàn diện và thống nhất của thông tin.

Bảo mật và quyền riêng tư: Đảm bảo an toàn và bảo mật dữ liệu là một mục tiêu quan trọng. Nghiên cứu này nên tập trung vào phát triển các biện pháp bảo mật để bảo vệ dữ liệu quan trọng và đảm bảo tuân thủ các quy định về quyền riêng tư.

Khả năng truy cập và chia sẻ dữ liệu: Xây dựng kiến trúc dữ liệu có khả năng chia sẻ dữ liệu một cách dễ dàng giữa các nhóm nghiên cứu và các tổ

chức. có thể bao gồm việc phát triển tiêu chuẩn và giao thức để chia sẻ dữ liệu khoa học và công nghệ.

**Phân tích dữ liệu và học máy:** Sử dụng các công cụ và kỹ thuật phân tích dữ liệu và học máy để tạo ra thông tin cấp cao từ dữ liệu khoa học và công nghệ. Mục tiêu là phát triển các mô hình dự đoán, phân loại, và trích xuất tri thức từ dữ liệu.

**Hỗ trợ quyết định:** Tạo ra các công cụ và ứng dụng dựa trên dữ liệu để hỗ trợ quyết định trong lĩnh vực khoa học và công nghệ, bao gồm việc dự đoán xu hướng nghiên cứu, định hình chính sách công nghệ và hướng dẫn phát triển sản phẩm.

Những mục tiêu này giúp tạo ra một hệ thống dữ liệu khoa học và công nghệ mạnh mẽ và thông minh, góp phần thúc đẩy sự phát triển và tiến bộ trong các lĩnh vực này, và hỗ trợ quyết định và định hướng trong nghiên cứu và phát triển công nghệ.

### **1.3 Phạm vi nghiên cứu**

- Giới hạn phạm vi về nội dung: Nghiên cứu xây dựng kiến trúc dữ liệu nghiên cứu khoa học và Công nghệ (Nghiên cứu dữ liệu nghiên cứu khoa học và Công nghệ tại Học viện khoa học và công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

- Giới hạn phạm vi quãng thời gian diễn biến của đối tượng nghiên cứu: 02 năm (từ tháng 5 năm 2021 đến tháng 5 năm 2023).

- Giới hạn phạm vi không gian khảo sát, nghiên cứu: Nghiên cứu trường hợp Học viện Khoa học và Công nghệ – Viện Hàn lâm.

### **1.4 Phương pháp nghiên cứu**

- Phương pháp nghiên cứu tài liệu: Nghiên cứu hệ thống văn bản pháp luật và các văn bản liên quan đến dữ liệu Nghiên cứu khoa học và công nghệ, đồng thời nghiên cứu các báo cáo, giáo trình để kế thừa kết quả đã được nghiên cứu.

- Phương pháp phân tích và tổng hợp: Trên cơ sở các nguồn tài liệu, dữ liệu, báo cáo về dự án, hội thảo, hội nghị ... tổng hợp và phân tích dữ liệu đã thu thập được.

- Phương pháp thảo luận nhóm: Thảo luận với các học viên cùng khoá, các nghiên cứu sinh của Khoa Công nghệ thông tin (đặc biệt là những nghiên cứu sinh đang công tác tại Viện Hàn lâm) và các cán bộ đang công tác tại Viện Hàn lâm.

### **1.5 Những đóng góp của luận văn**

- Chỉ ra được thực trạng và các nghiên cứu liên quan đến dữ liệu nghiên cứu khoa học và công nghệ tại Việt Nam và một số nước trên thế giới.

- Xây dựng ứng dụng phần mềm lưu trữ, xử lý dữ liệu nghiên cứu khoa học và Công nghệ tại Học viện KHCN, Viện Hàn lâm.

### **1.6 Bố cục của Luận văn**

Ngoài các phần mở đầu, kết luận, danh mục tài liệu tham khảo và các phụ lục, luận văn gồm 3 chương:

**Chương 1:** Tổng quan các nghiên cứu về dữ liệu nghiên cứu khoa học và công nghệ

**Chương 2:** Nghiên cứu xây dựng kiến trúc dữ liệu nghiên cứu khoa học và công nghệ

**Chương 3:** Thực nghiệm mô hình và đánh giá kết quả

# CHƯƠNG 1. TỔNG QUAN NGHIÊN CỨU LIÊN QUAN ĐẾN KIẾN TRÚC DỮ LIỆU NGHIÊN CỨU KHOA HỌC VÀ CÔNG NGHỆ

## 1.1. Tổng quan nghiên cứu

Nghiên cứu xây dựng kiến trúc dữ liệu nghiên cứu khoa học và công nghệ (Nghiên cứu dữ liệu nghiên cứu khoa học và Công nghệ tại Học viện KHCN, Viện Hàn lâm).

Hiện nay, lượng dữ liệu nghiên cứu trong khoa học và công nghệ được tạo ra từ các dự án khoa học và công nghệ trong nước là rất lớn. Tuy nhiên, dữ liệu nghiên cứu thu thập được từ các dự án này chỉ được thể hiện dưới dạng siêu dữ liệu (thuyết minh đề tài, bài báo, bài trình bày tại hội nghị, báo cáo kết quả nghiên cứu). Hầu hết các dữ liệu thô như dữ liệu khảo sát được thu thập trong quá trình nghiên cứu, dữ liệu ảnh/Video hoặc dữ liệu giám sát IoT đều chưa được thu thập. Do đó, nhu cầu cấp thiết là xây dựng một nền tảng quản lý dữ liệu nghiên cứu khoa học và công nghệ là rất cấp thiết và quan trọng cho các nhóm nghiên cứu, tổ chức và công ty tại Việt Nam.

Khái niệm về DLNC hiện nay còn nhiều tranh luận để đưa ra một khái niệm thống nhất, vì hình thức và nội dung của dữ liệu ở các lĩnh vực khác nhau thì khác nhau.

Khái niệm về DLNC cũng có thể được hiểu và diễn giải ở nhiều cách khác nhau, năm 2005, Quỹ Khoa học Quốc gia [5] đã xuất bản tập báo cáo có tựa đề “Phát triển bộ sưu tập số mang tính chất dài hạn: khuyến khích nghiên cứu và phát triển giáo dục trong thế kỷ XXI” (Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century) đã đề cập tới việc có thể phân chia DLNC bởi nguồn gốc của chúng được tạo ra, bởi vì các dữ liệu này có thể là các dữ liệu về thí nghiệm, máy tính hay quan sát. Trong đó dữ liệu quan sát có thể là các quan trắc về nhiễu độ hay quan sát về thái độ của người bỏ phiếu trước cuộc tổng tuyển cử. Dữ liệu về máy tính có thể là kết quả từ thực hiện mô hình mô phỏng hoặc có thể là các thông tin về cấu hình như phần cứng, phần mềm. Dữ liệu về nghiên cứu, thí nghiệm ví dụ các mẫu biểu hiện gen hay tốc độ phản ứng hoá học.

DLNC có thể ở các hình thức sau:

- Văn bản, tài liệu, chữ, hay các bảng tính.
- Ghi chú trong phòng thí nghiệm, đi thực nghiệm, nhật ký.
- Bảng câu hỏi, bản ghi chép tay.
- Băng ghi âm, ghi hình.
- Hình ảnh, phim ảnh.
- Phản ứng thí nghiệm.
- Slide, hiện vật, mẫu vật, mẫu.
- Bộ sưu tập các đối tượng số được tạo ra và thu thập trong quá trình nghiên cứu.
- Hồ sơ dữ liệu.
- Nội dung cơ sở dữ liệu bao gồm cả hình ảnh, âm thanh, văn bản.
- Thuật toán, các kịch bản.
- Các phương pháp và quy trình công việc.
- Các quy trình tiêu chuẩn và giao thức.

Tuỳ thuộc vào mỗi cơ quan mục đích sử dụng khác nhau sẽ có sự phân biệt và sử dụng DLNC khác nhau. Chính vì điều này đã phân biệt sự khác nhau giữa dữ liệu xuất bản và DLNC ở bảng dưới đây:

Sự khác nhau giữa thông tin xuất bản và DLNC

Vai trò của việc quản lý DLNC đối với các cơ quan, tổ chức và cơ quan thư viện - thông tin

Phải nói rằng, đứng sau thành công của các nhà khoa học là thành công của các cơ quan nghiên cứu, tổ chức đã đồng hành cùng các nhà nghiên cứu đó. Đối với các trường đại học và các viện nghiên cứu, việc đưa ra một chiến lược cụ thể trong việc quản lý nguồn tài nguyên có giá trị này đem lại rất nhiều lợi ích về mặt kinh tế cũng như tiếng tăm. Theo tác giả Elsevier [12] điều này có thể được nhìn thấy ở một vài khía cạnh sau:



- Quản lý DLNC đem tới sự cải thiện về chất lượng và số lượng của các kết quả đầu ra vì tốc độ nghiên cứu sẽ tăng lên cùng với độ tin cậy của các bài báo nghiên cứu sẽ cải thiện khi DLNC được truy cập mở, sẵn có cho việc đánh giá.

- Tăng việc xuất bản và tăng chỉ số ảnh hưởng: Sự sẵn có của DLNC sẽ làm tăng việc xuất bản, dẫn tới sự bùng nổ tranh luận của nhiều quan điểm, chia sẻ dẫn tới làm tăng sự ảnh hưởng của bài viết, tạo ra nhiều trích dẫn.

- Tăng cường việc hợp tác: Dữ liệu được chia sẻ là một khởi đầu cho việc hợp tác, như các nhà nghiên cứu tìm kiếm để hiểu làm thế nào kết quả được thu thập và trao đổi các quan điểm trong việc phân tích và giải thích một vấn đề.

Để có thể khai thác toàn bộ tiềm năng của DLNC, đòi hỏi mỗi cơ quan, trung tâm thông tin - thư viện cần phải đưa ra những chiến lược phát triển và hợp tác hợp lý, trong đó đề cao việc thực hiện chiến lược và chính sách truy cập mở, đảm bảo việc hanh thông về tài chính và hỗ trợ về mặt công nghệ để tài liệu được truy cập mở và dễ dàng tìm thấy. Tuy nhiên, cùng một lúc cũng đảm bảo có các chính sách về bảo hộ quyền sở hữu và thông tin cá nhân, quyền riêng tư cùng với các tiêu chuẩn về an ninh, an toàn dữ liệu để xây dựng hệ thống phục vụ việc sử dụng và tra cứu dữ liệu. Ngoài ra, cần phải có chiến lược tái sử dụng thông tin, phục vụ hoạt động DLNC sẽ được phát hiện và tái sử dụng thông qua các xuất bản phẩm, trong đó hoạt động cốt lõi là thực hiện một giải pháp xuất bản phẩm tiên tiến xem xét việc công bố các dữ liệu, phần mềm và phương pháp trích dẫn. Quá trình kiểm duyệt, cài đặt và tạo ra các định danh đối tượng số duy nhất cho các dữ liệu và dữ liệu có liên quan. Hơn thế nữa, nhiệm vụ của các trung tâm, cơ quan thư viện - thông tin trong hoạt động này là phải đảm bảo dữ liệu được kết nối với nhau, nên chúng có thể dễ dàng được tìm kiếm bởi người dùng hoặc các đối tác trong cơ quan đó hay cộng đồng người dùng ngoài cơ quan. Để thúc đẩy dữ liệu có thể được tìm thấy, các chương trình liên kết dữ liệu cần phải được cài đặt và thiết lập. Ví dụ, trường Đại học Havard liên kết những bài báo của các nhà nghiên cứu trong trường với các cơ sở dữ liệu ở cả viện nghiên cứu và các cơ quan lưu trữ mà có liên kết với nhau về nhiều ngành nghề. Việc hỗ trợ tìm kiếm một lúc trên nhiều cơ sở dữ liệu hiệu quả cũng là một vấn đề cần phải được xem xét.

Vậy lợi ích của hoạt động quản lý DLNC đối với các nhà nghiên cứu là gì? Đơn giản với những nhà nghiên cứu, có một lý lẽ được đặt ra là có lẽ không cần phải thuyết phục với họ tầm quan trọng của DLNC là gì bởi vì hơn ai hết họ hiểu được việc này, tuy nhiên có nhiều nhà nghiên cứu vẫn không hiểu được tầm quan trọng và tiềm năng của việc chia sẻ nguồn tài liệu này với cộng đồng nghiên cứu mà có thể được liệt kê ở những điểm chính sau:

- Tăng cường sự hợp tác: Chia sẻ là hợp tác và đặc biệt điều này giúp cho việc kiểm soát, thẩm định lại kết quả của các nghiên cứu.

- Tăng các chỉ số ảnh hưởng và hoạt động xuất bản: Việc chia sẻ các DLNC và quản lý chúng một cách hiệu quả giúp quá trình nghiên cứu được rút ngắn, nhiều công trình được xuất bản hơn, nâng tầm nghiên cứu, tăng các chỉ số trích dẫn tài liệu và uy tín của nhà nghiên cứu.

- Tăng sự phân phối của các nghiên cứu: Truy cập tới các nguồn DLNC giúp tăng việc phân phối các nghiên cứu này tới nhiều cơ sở dữ liệu khác nhau, thúc đẩy sự trao đổi và tái sử dụng dữ liệu.

Tình hình lưu trữ dữ liệu của nghiên cứu sinh, và học viên cao học tại Học Viện KHCN hiện nay

Theo điều tra qua câu hỏi của các Học viên cao học và Nghiên cứu sinh tại Học viện KHCN học viên nhận thấy dữ liệu thu về từ việc điều tra qua bảng hỏi của đa số nghiên cứu sinh và Học viên được chọn lọc ở các ngành học khác nhau trong, quan điểm của nghiên cứu sinh là các hình thức lưu trữ dữ liệu chủ yếu là tự lưu trữ dữ liệu trên các thẻ nhớ, ổ cứng, hay sử dụng các dịch vụ miễn phí, nhưng giới hạn về dung lượng, việc lưu trữ trên các thiết bị tự do chiếm đa số tới 80%, lưu trữ không giới hạn về dung lượng nhưng việc phải lưu nhiều lần trên các hình thức lưu trữ khác nhau, không đảm bảo về mặt an toàn và bảo mật dữ liệu, khi tìm lại dữ liệu có thể gây khó khăn, thông tin dễ bị lấy cắp, là một trong những hạn chế đáng kể của hình thức này

## **1.2. Tình hình nghiên cứu trong nước**

Hiện nay nguồn dữ liệu nghiên cứu khoa học và công nghệ là rất quan trọng và Nhằm thúc đẩy việc tạo lập, sử dụng và khai thác dữ liệu nghiên cứu

khoa học và công nghệ mở cũng như khuyến khích sự quan tâm và tham gia của các nhà nghiên cứu trong việc chia sẻ dữ liệu khoa học và công nghệ dùng chung.

Hiện nay tình hình nghiên cứu về cấu trúc dữ liệu trong nghiên cứu khoa học và công nghệ tại Việt Nam đã có những phát triển đáng kể trong những năm gần đây, nhưng vẫn còn nhiều thách thức cần đổi mới. Dưới đây là một số điểm nổi bật về tình hình nghiên cứu này tại Việt Nam:

**Nghiên cứu ứng dụng cấu trúc dữ liệu:** Các nghiên cứu về cấu trúc dữ liệu tại Việt Nam thường tập trung vào các ứng dụng thực tế, chẳng hạn như trong lĩnh vực công nghiệp, y tế, và quản lý tài nguyên. Các ứng dụng này có thể bao gồm việc tối ưu hóa quy trình sản xuất, phân tích dữ liệu y tế, và quản lý tài sản.

**Giáo dục và đào tạo:** Cấu trúc dữ liệu là một phần quan trọng của các khoá học máy tính và công nghệ thông tin tại các trường đại học và viện nghiên cứu ở Việt Nam. Các khoá học và chương trình đào tạo về cấu trúc dữ liệu và thuật toán đã được phát triển để đào tạo các nhà nghiên cứu và chuyên gia trong lĩnh vực này.

**Phát triển ứng dụng và công nghệ:** Các công ty công nghệ tại Việt Nam cũng đang đầu tư vào nghiên cứu và phát triển về cấu trúc dữ liệu để phục vụ các ứng dụng công nghiệp và thương mại phục vụ cho đơn vị của mình. Điều này bao gồm việc phát triển phần mềm và ứng dụng mới liên quan đến xử lý và quản lý dữ liệu.

**Phòng thí nghiệm và dự án nghiên cứu:** Nhiều phòng thí nghiệm và dự án nghiên cứu tại Việt Nam đã tập trung vào cấu trúc dữ liệu và thuật toán, đặc biệt là trong các lĩnh vực như trí tuệ nhân tạo, xử lý ngôn ngữ tự nhiên, và phân tích dữ liệu lớn.

**Hợp tác quốc tế:** Việt Nam cũng đã hợp tác với các tổ chức và trường đại học quốc tế để thúc đẩy nghiên cứu về dữ liệu nghiên cứu khoa học. Nhiều dự án nghiên cứu chung và chương trình đào tạo đã được thiết lập.

Tuy nhiên, còn nhiều thách thức đang đối diện trong việc phát triển nghiên cứu về cấu trúc dữ liệu nghiên cứu khoa học và công nghệ tại Việt Nam. Một số trong số bao gồm:

Tài nguyên và nguồn lực: Thiếu tài nguyên và nguồn lực về công nghệ thông tin đầu tư vào nghiên cứu và phát triển trong lĩnh vực này.

Chất lượng giáo dục: Cần nâng cao chất lượng giáo dục và đào tạo trong lĩnh vực nghiên cứu, xây dựng quản lý cũng như chia sẻ dữ liệu nghiên cứu để đáp ứng nhu cầu của thị trường công nghệ hiện nay và nghiên cứu.

Quản lý dữ liệu: Cần phát triển các chính sách và quy định về quản lý dữ liệu, đặc biệt là khi xử lý dữ liệu cá nhân hoặc nhạy cảm.

Phát triển cộng đồng nghiên cứu: Việc xây dựng và phát triển cộng đồng nghiên cứu về cấu trúc dữ liệu cũng là một thách thức quan trọng.

Tóm lại, tình hình nghiên cứu về dữ liệu nghiên cứu trong khoa học và công nghệ tại Việt Nam đã và đang phát triển, nhưng cần sự đầu tư và nỗ lực từ cơ quan nhà nước, Bộ ban ngành, các Học viện, Viện nghiên cứu, trường đại học và cùng với các nhà khoa học, các nhà nghiên cứu để vượt qua các thách thức và tận dụng tiềm năng của lĩnh vực này.

Một số Hội thảo, Hội nghị trao đổi về cách thức quản lý Dữ liệu nghiên cứu khoa học và công nghệ kể đến như:

Ở Việt Nam cũng có một số hội thảo, hội nghị cách quản lý, quản trị và lưu trữ dữ liệu nghiên cứu trong khoa học và công nghệ như:

- Vào ngày 02 tháng 10 năm 2021 Công ty Cổ phần Thông tin và Công nghệ Số (IDT Vietnam) cùng với Trường đại học Nguyễn Tất Thành đã phối hợp cùng tổ chức hội nghị, hội thảo trực tuyến bàn về việc Quản lý dữ liệu trong nghiên cứu Khoa học và Công nghệ tại các trường Đại học hiện nay và là sự kết hợp giữa Công ty Cổ phần Thông tin và Công nghệ Số (IDT Vietnam) và một số trường Đại học như , Đại học Nguyễn Tất Thành, Đại học RMIT, Đại học Việt Đức, Đại học Đà Lạt ... Cùng trao đổi ở buổi hội thảo, hội nghị này có rất nhiều các nhà nghiên cứu và khoa học cùng chia sẻ rất nhiều nội dung

và bàn luận về việc quản lý, lưu trữ dữ liệu nghiên cứu trong Khoa học và Công nghệ mà rất nhiều các độc giả quan tâm.

- Vào ngày 21 tháng 6 năm 2022 Hội nghị hội thảo giữ Cục Thông tin Khoa học và Công nghệ quốc gia đã phối hợp với Viện công nghệ thông tin-Viện Hàn lâm được tổ chức tại thành phố Hồ Chí Minh, cùng tham dự có Quỹ Đổi mới sáng tạo VinGroup Viện Nghiên cứu dữ liệu lớn hội nghị, Hội thảo Dữ liệu nghiên cứu nghiên cứu trong lĩnh vực khoa học và công nghệ để chia sẻ, dùng chung trong thời kỳ nguyên số. Tại buổi Hội thảo có ông Trần Đức Hiến, Cục trưởng Cục Thông tin Khoa học và Công nghệ quốc gia, ông Nguyễn Long Giang, Phó viện trưởng Viện CNTT, Viện Hàn lâm và có rất nhiều các các Giáo sư, Tiến sĩ và các nhà nghiên cứu đến từ các lĩnh vực khác nhau từ các Học viện, viện nghiên cứu chuyên ngành, trường đại học trên cả nước

Hay một số Sáng kiến quản lý và chia sẻ dữ liệu nghiên cứu ở Việt Nam cụ thể:

Ở cấp quốc gia, Thủ tướng Chính phủ đã ký Quyết định số 677/QĐ-TTg ngày 18/5/2017 về phê duyệt Đề án “*Phát triển Hệ tri thức Việt số hóa*”. Đề án đặt ra các mục tiêu như “*1. Xây dựng Hệ tri thức Việt số hóa thông qua việc tổng hợp, hệ thống hóa, việt hóa, số hóa, lưu trữ và phổ biến tri thức trong mọi lĩnh vực, trước hết là hỗ trợ cho giáo dục đào tạo, đổi mới sáng tạo và các lĩnh vực liên quan trực tiếp đến đời sống của người dân như pháp luật, y tế, kỹ thuật sản xuất...; 2. Tạo môi trường thuận lợi thu hút mọi người dân và doanh nghiệp tham gia, với vai trò vừa khai thác vừa đóng góp để làm giàu các tài nguyên tri thức số hóa của Việt Nam; 3. Khơi dậy, lan tỏa niềm đam mê khoa học và công nghệ, khát vọng sáng tạo, cống hiến của mọi người, mọi doanh nghiệp, đặc biệt là thế hệ trẻ, đội ngũ trí thức và các doanh nghiệp công nghệ thông tin trong việc tạo lập, làm giàu và phổ biến tri thức; 4. Từng bước góp phần phát triển công nghiệp nội dung số của Việt Nam, định hướng việc sử dụng tri thức của người dùng trên môi trường mạng*” (Bộ Khoa học và Công nghệ, 2017).

Ở cấp bộ, thực hiện Nghị định số 11/2014/NĐ-CP ngày 18/02/2014 của Chính phủ về Hoạt động thông tin KH&CN và Thông tư số 10/2017/TT-BKHCN ngày 28/6/2017, Bộ Khoa học và Công nghệ đã xây dựng hệ thống 10 cơ sở dữ liệu về khoa học và công nghệ quốc gia. Cụ thể là thông tin về “*Các*

*tổ chức khoa học và công nghệ; cán bộ nghiên cứu khoa học và phát triển công nghệ; nhiệm vụ khoa học và công nghệ; công bố khoa học và chỉ số trích dẫn khoa học; thống kê khoa học và công nghệ; công nghệ, công nghệ cao, chuyển giao công nghệ; thông tin về khoa học và công nghệ trong khu vực và trên thế giới; doanh nghiệp KH&CN; thông tin sở hữu trí tuệ và tiêu chuẩn đo lường chất lượng” (Bộ Khoa học và Công nghệ, 2018).*

Tại Việt Nam cũng có rất nhiều Viện nghiên cứu, Học viện và các Trường đại học đã và đang xây dựng nền tảng cho việc quản lý và chia sẻ dữ liệu nghiên cứu khoa học để thúc đẩy sự phát triển trong lĩnh vực này. Dưới đây là một số ví dụ:

Viện Hàn lâm thực hiện chức năng nghiên cứu cơ bản về khoa học tự nhiên và phát triển công nghệ; cung cấp luận cứ khoa học cho công tác quản lý khoa học, công nghệ và xây dựng chính sách, chiến lược, quy hoạch phát triển kinh tế, xã hội; đào tạo nhân lực khoa học, công nghệ có trình độ cao theo quy định của pháp luật. Trong đó phải kể đến Viện Công nghệ thông tin là đơn vị trong Viện Hàn lâm luôn đi đầu trong việc xây dựng các dự án về chia sẻ dữ liệu nghiên cứu khoa học và là đầu mối để cùng với các Đơn vị khác như các trường đại học các viện nghiên cứu khác để nghiên cứu xây dựng kiến trúc dữ liệu và các giải pháp quản lý và chia sẻ dữ liệu trên phạm vi toàn cầu.

Viện Công nghệ Thông tin và Truyền thông (ICT): ICT thuộc Quỹ Phát triển Khoa học và Công nghệ Quốc gia (NAFOSTED) của Việt Nam. Viện cũng đang tập trung vào nghiên cứu và phát triển công nghệ thông tin, bao gồm quản lý dữ liệu và ứng dụng trong khoa học và công nghệ.

Trung tâm Thông tin Khoa học và Công nghệ Quốc gia (VISTEC): VISTEC là một tổ chức chuyên về quản lý và cung cấp thông tin trong lĩnh vực khoa học và công nghệ tại Việt Nam. Trung tâm này cung cấp các dịch vụ như cơ sở dữ liệu, thư viện số và các giải pháp quản lý dữ liệu.

Trung tâm Nghiên cứu Khoa học Dữ liệu và Ứng dụng (DARCI): DARCI là một trung tâm nghiên cứu tại Trường Đại học Công nghệ Thông tin, Đại học Quốc gia Hà Nội. Trung tâm này tập trung vào nghiên cứu và phát triển các giải pháp về dữ liệu khoa học và công nghệ.

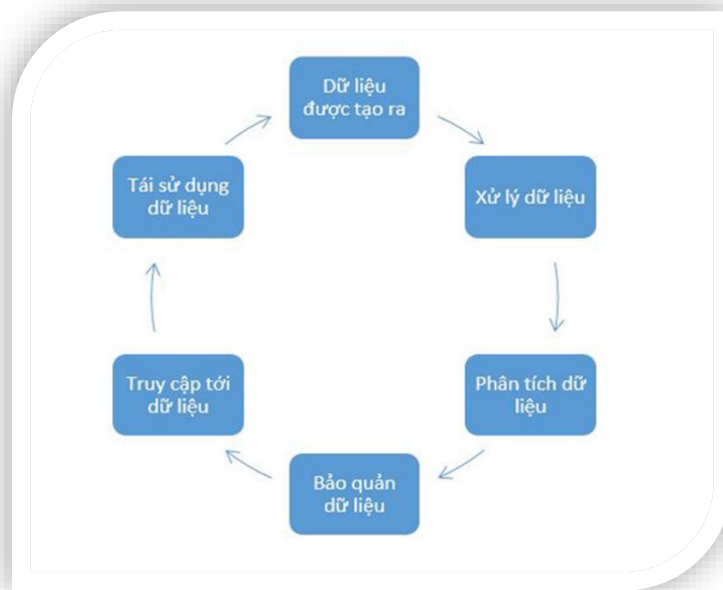
Viện Công nghệ Thông tin và Truyền thông (ICT): ICT thuộc Đại học Quốc gia Hà Nội cũng đang tập trung vào nghiên cứu và phát triển trong lĩnh vực công nghệ thông tin và truyền thông, bao gồm quản lý dữ liệu và các ứng dụng liên quan.

Viện Công nghệ Thông tin (ITI): ITI là một viện nghiên cứu và phát triển công nghệ thông tin có trụ sở tại TP.HCM. Là đơn vị thực hiện nghiên cứu trong nhiều lĩnh vực, bao gồm xử lý và quản lý dữ liệu khoa học và công nghệ.

Các tổ chức này cùng với nhiều Học viện, viện nghiên cứu, trường đại học và các công ty để thúc đẩy nghiên cứu và quản lý dữ liệu nghiên cứu trong lĩnh vực khoa học và công nghệ, từ việc phát triển công nghệ mới đến cung cấp cơ sở dữ liệu và giải pháp quản lý dữ liệu.

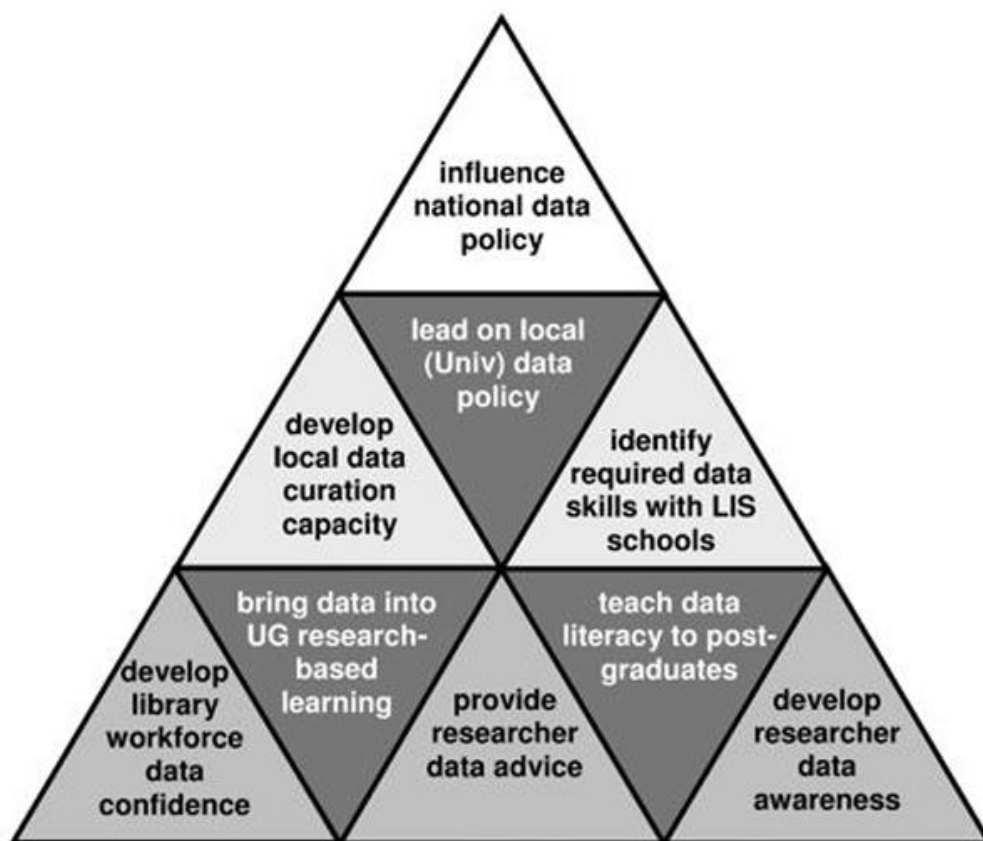
### 1.3. Tình hình nghiên cứu ở nước ngoài

Theo Hiệp hội lưu trữ dữ liệu của Vương quốc Anh [13] Dữ liệu nghiên cứu trong khoa học và công nghệ được hình thành tạo ra từ những quá trình đầu tiên của việc tiến hành xem xét, xử lý, phân tích, bảo quản, truy cập và tái sử dụng dữ liệu để tạo ra thông tin mới.



Hình 1.3.1: Hình mô phỏng vòng đời quá trình hình thành và lưu trữ dữ liệu DLNC

Theo nhà nghiên cứu Martin Lewis [15] ông đưa tháp như hình dưới để quản lý dữ liệu nghiên cứu cho các thư viện



*Hình 1.3.2: Mô hình tháp quản lý dữ liệu nghiên cứu cho các thư viện, được trình bày bởi Lewis [14]*

"Dữ liệu nghiên cứu, không giống như các loại thông tin khác, được thu thập, quan sát hoặc tạo ra, với mục đích phân tích để sản xuất các kết quả nghiên cứu ban đầu" (Đại học Edinburgh, Vương quốc Anh). Theo Jodi Reeves Flores và cộng sự trong nghiên cứu "Libraries and the Research Data Management Landscape" thì trên thế giới, các tổ chức, thể chế và chính phủ đã và đang nhận ra tầm quan trọng của việc quản lý dữ liệu nghiên cứu (DLNC). Điều này được thể hiện thông qua các mối quan tâm ngày càng tăng về lưu trữ tài liệu và bảo quản DLNC. Phong trào quản lý DLNC đã được hình thành dựa trên sự gia tăng của các yêu cầu, nhiệm vụ, kỹ thuật và một lượng lớn các công cụ hỗ trợ. Các hoạt động quản lý DLNC đã góp phần đảm bảo giá trị lâu dài và hữu ích cho các phân tích và nghiên cứu tiếp theo. Tuy nhiên, do sự thiếu sót của các bên có liên quan, nhiều vấn đề trong quá trình quản lý DLNC dần lộ rõ và đòi hỏi nhiều nỗ lực khắc phục.



Dữ liệu nghiên cứu trong khoa học và công nghệ trên toàn cầu đang phát triển rất nhanh. Dữ liệu đã trở thành một phần quan trọng của hầu hết các lĩnh vực nghiên cứu, và các nhà nghiên cứu trên khắp thế giới đang tập trung vào nghiên cứu và phát triển các phương pháp, công nghệ, và ứng dụng mới liên quan đến dữ liệu. Dưới đây là một số xu hướng và chủ đề nghiên cứu quan trọng về dữ liệu nghiên cứu trong lĩnh vực khoa học và công nghệ:

Học máy và Trí tuệ nhân tạo (AI): Sự phát triển của học máy và trí tuệ nhân tạo đã mở ra nhiều cơ hội mới trong việc phân tích và sử dụng dữ liệu trong các lĩnh vực như nhận dạng hình ảnh, xử lý ngôn ngữ tự nhiên, và dự đoán.

Big Data Analytics: Xử lý và phân tích dữ liệu lớn (big data) là một chủ đề nghiên cứu quan trọng. Các phương pháp và công cụ mới được phát triển để hiểu và tận dụng thông tin từ dữ liệu có khối lượng lớn.

Dữ liệu y tế và khoa học sức khỏe: Dữ liệu y tế từ các bệnh viện, thiết bị y tế thông minh, và các nguồn khác đã trở thành một trường nghiên cứu quan trọng để cải thiện chẩn đoán, điều trị, và quản lý sức khỏe.

Dữ liệu thời tiết và khí hậu: Hiểu rõ và dự đoán biến đổi khí hậu và các biến đổi thời tiết đang là một phần quan trọng của nghiên cứu về môi trường và khí hậu.

Dữ liệu vũ trụ: Nghiên cứu về dữ liệu từ các vệ tinh và thiết bị vũ trụ đang giúp mở rộng hiểu biết về vũ trụ và hành tinh khác.

Dữ liệu trong ngành công nghiệp: Các ngành công nghiệp, chẳng hạn như sản xuất và loại hình năng lượng, đang tận dụng dữ liệu để tối ưu hóa quy trình sản xuất và tiết kiệm tài nguyên.

An ninh và bảo mật dữ liệu: Bảo vệ dữ liệu trước các mối đe dọa về an ninh và quyền riêng tư là một phần quan trọng của nghiên cứu về dữ liệu.

Dữ liệu xã hội và hành vi con người: Nghiên cứu về dữ liệu xã hội và hành vi con người trên mạng xã hội và trực tuyến đang giúp hiểu rõ hơn về xã hội và văn hóa.

Dữ liệu và sự phát triển bền vững: Dữ liệu được sử dụng để theo dõi và đánh giá các mục tiêu phát triển bền vững của Liên Hợp Quốc, như giảm nghèo đói và bảo vệ môi trường.

Quản lý dữ liệu và chính trị dữ liệu: Nghiên cứu về các khía cạnh quản lý dữ liệu, bao gồm chính trị dữ liệu, luân phiên dữ liệu, và quyền riêng tư đang trở nên ngày càng quan trọng.

Các loại dữ liệu nghiên cứu trong khoa học và công nghệ thường thay đổi nhanh chóng do sự phát triển liên tục của công nghệ và yêu cầu của xã hội. Các nhà nghiên cứu trên khắp thế giới đang cùng nhau đối mặt với những thách thức và cơ hội đầy tiềm năng trong việc sử dụng, quản lý dữ liệu để tái tạo ra dữ liệu nghiên cứu mới để giải quyết các vấn đề quan trọng của thế giới hiện đại.

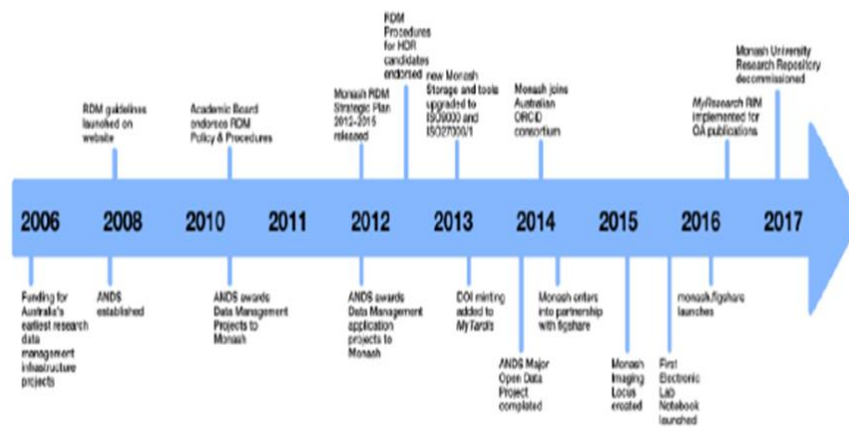
Theo một số nghiên cứu trên thế giới khác như:

Cox and Pinfield (2014) cho rằng quản lý dữ liệu nghiên cứu bao gồm nhiều quá trình và hoạt động khác nhau gắn chặt với vòng đời của dữ liệu, từ thiết kế và tạo lập dữ liệu, lưu trữ, bảo mật, bảo quản, tìm kiếm, chia sẻ và tái sử dụng lại. Các công đoạn này chịu sự chi phối bởi năng lực công nghệ, quy tắc đạo đức, các quy định của pháp luật và chính sách của chính phủ sở tại, do vậy hoạt động quản lý dữ liệu nghiên cứu sẽ được điều chỉnh cho phù hợp với từng bối cảnh khác nhau. Dữ liệu phần lớn được thu thập hoặc tạo ra từ quá trình thực hiện các dự án nghiên cứu trong trường đại học hoặc viện nghiên cứu. Vì vậy, thư viện ở các cơ quan này có vai trò rất quan trọng trong việc hỗ trợ quản lý dữ liệu nghiên cứu bởi vì đội ngũ cán bộ thư viện là những người có kiến thức, kỹ năng và kinh nghiệm trong việc tổ chức, lưu trữ và cung cấp khả năng tìm kiếm thông tin theo nhu cầu của người sử dụng. Mặt khác, cán bộ thư viện thường có mối liên hệ chặt chẽ với giảng viên tại các khoa trong trường đại học để hỗ trợ hoạt động giảng dạy, nghiên cứu của họ nên sẽ có điều kiện thuận lợi để tham gia tư vấn, hướng dẫn quản lý dữ liệu cho các dự án nghiên cứu. Quản lý và chia sẻ dữ liệu nghiên cứu được ưu tiên phát triển mạnh ở các trường đại học trên thế giới trong những năm gần đây (Cox & Pinfield, 2014). Hoạt động này hỗ trợ các nhà nghiên cứu quản lý và chia sẻ rộng rãi dữ liệu thu thập được đến cộng đồng nghiên cứu nhằm tăng cường tính minh bạch và khả

năng có thể kiểm chứng được của kết quả nghiên cứu, nâng cao mức độ ảnh hưởng của công trình nghiên cứu và tìm kiếm cơ hội hợp tác với các dự án nghiên cứu trong cùng lĩnh vực. Đặc biệt, nó cung cấp khả năng tái sử dụng dữ liệu và thông tin một cách hiệu quả, tránh nghiên cứu trùng lặp giúp tiết kiệm rất nhiều thời gian thu thập thông tin và chi phí cho cộng đồng các nhà khoa học chuyên ngành hoặc liên ngành. Nhận thức tầm quan trọng của hoạt động này, Ủy ban Châu Âu (EU) đã đưa sáng kiến về dữ liệu và quản lý dữ liệu nghiên cứu mở vào Chương trình Nghiên cứu và Đổi mới mang tên Horizon 2020, thực hiện trong 7 năm (2014 - 2020) có ngân sách gần 80 tỉ euro (Thestrup & Kruse, 2017). Theo đó, các dự án nghiên cứu nhận ngân sách từ Horizon 2020 bắt buộc phải có kế hoạch quản lý và chia sẻ dữ liệu nghiên cứu, tuân thủ nguyên tắc FAIR (Findable-tìm kiếm được, Accessible-truy cập được, Interoperable-chia sẻ được và Re-usable-tái sử dụng được). Buchholtz et al. (được trích dẫn trong bài viết của Thestrup & Kruse, 2017) ước tính rằng dữ liệu lớn và truy cập mở có thể đóng góp tới 1,9 % GDP của Châu Âu vào năm 2020. Các dự án liên quan đến thu thập, tổ chức, quản lý và chia sẻ dữ liệu nghiên cứu cũng đã được đẩy mạnh ở nhiều nước. Trong nghiên cứu của mình, Pryor et al. (2014) đã khảo sát một số dự án quản lý dữ liệu nghiên cứu nổi bật tại Hoa Kỳ, Anh và Úc nhằm so sánh và làm rõ các chiến lược thực hiện dự án khác nhau cũng như đúc kết được nhiều bài học quý giá thu được từ quá trình này. Tại Hoa Kỳ, dịch vụ quản lý dữ liệu nghiên cứu được triển khai chính thức bởi Thư viện Sheridan thuộc Trường Đại học Johns Hopkins (JHU) vào tháng 7 năm 2011 nhằm hỗ trợ các nhà nghiên cứu xây dựng kế hoạch quản lý dữ liệu cho các dự án nghiên cứu theo yêu cầu của cơ quan tài trợ là Quỹ Nghiên cứu Khoa học Quốc gia Hoa Kỳ (Pryor et al., 2014). Đến nay, Dịch vụ quản lý dữ liệu của JHU (JHUDS) cung cấp cho các nhà nghiên cứu, giảng viên và sinh viên sự hỗ trợ liên quan đến quản lý và chia sẻ dữ liệu, sử dụng GIS và dữ liệu bản đồ, sử dụng các công cụ và phần mềm để xử lý dữ liệu, tìm kiếm và sử dụng dữ liệu trong kho lưu trữ của JHU. Ba yếu tố chính đóng góp vào việc lập kế hoạch và triển khai thành công JHUDS là xây dựng ngân sách linh hoạt, nhân sự có kỹ năng và kinh nghiệm và mô hình quản lý dữ liệu nhiều phân tầng đáp ứng được nhu cầu hạ tầng kỹ thuật hiện tại và tương lai (Pryor et al., 2014). Tại Anh, Dịch vụ dữ liệu UK (UKDS) được khởi xướng vào năm 2012 do Hội đồng

ngiên cứu Kinh tế và Xã hội (ESRC) tài trợ nhằm xây dựng dịch vụ cung cấp dữ liệu thống nhất và đặc thù của ESRC tại Anh (Pryor et al., 2014). Mục tiêu chính của dự án là hỗ trợ người dùng có thể truy cập dễ dàng đến dữ liệu phù hợp với nhu cầu, có thể tìm kiếm được để đẩy mạnh các nghiên cứu về kinh tế và xã hội. Đồng thời, hoạt động này cũng nhằm tăng cường nhận thức về thực hành quản lý dữ liệu nghiên cứu cho các nhà nghiên cứu và những người tạo ra dữ liệu. UKDS được đặt tại Trung tâm Lưu trữ Dữ liệu quốc gia, Đại học Essex cùng với sự hợp tác chặt chẽ với các đối tác như Đại học Manchester, Đại học Southampton. Cấu trúc tổng thể của dịch vụ dựa trên mô hình chức năng của Hệ thống Thông tin Lưu trữ Mở (Open Archival Information System -OAIS Reference Model). Sự ra đời của UKDS được đánh giá là ảnh hưởng lớn đến nền tảng quản lý dữ liệu nghiên cứu trong các cơ sở giáo dục bậc cao ở Anh. Cụ thể là thay đổi nhận thức và kỹ năng quản lý dữ liệu của nhà nghiên cứu; cơ quan tài trợ dự án nghiên cứu yêu cầu xây dựng kế hoạch quản lý và truy cập mở đến dữ liệu; chính phủ thúc đẩy công khai kết quả nghiên cứu; các nhà xuất bản quy định cung cấp dữ liệu đi kèm với bài báo khoa học; và cộng đồng đòi hỏi tăng cường tái sử dụng dữ liệu (Pryor et al., 2014). Hiện nay, UKDS đang tiếp tục hỗ trợ các nhà nghiên cứu, giảng viên và sinh viên truy cập, sử dụng, chia sẻ dữ liệu về kinh tế, xã hội và dân số cũng như phát triển các tiêu chuẩn, thực hành tốt nhất về quản lý dữ liệu. Tại Úc, Đại học Monash (MU) là nơi chủ trì các dự án quốc gia liên quan đến thử nghiệm và phát triển nền tảng quản lý dữ liệu nghiên cứu (Pryor et al., 2014). Cụ thể là MU chịu trách nhiệm xây dựng Dịch vụ Dữ liệu Quốc gia Úc (ANDS), tạo lập cấu trúc cho mô hình quản lý dữ liệu nghiên cứu, lập kế hoạch chiến lược 2012-2015 và chính sách cho quản lý dữ liệu nghiên cứu cùng với các hướng dẫn và quy trình thực hiện, triển khai các chương trình đào tạo kỹ năng quản lý dữ liệu, và triển khai các giải pháp quản lý dữ liệu và siêu dữ liệu kèm theo. MU đã sớm thông qua Chiến lược quản lý thông tin của mình từ năm 2006 và thành lập Ủy ban Quản lý dữ liệu nghiên cứu cũng như chỉ định người điều phối sáng kiến này. Quyết định của MU tài trợ kinh phí để cung cấp nơi lưu trữ miễn phí dữ liệu nghiên cứu đã thúc đẩy các nhà nghiên cứu thu thập và quản lý dữ liệu các dự án của mình. Năm 2012, MU đã phê duyệt Chiến lược và Kế hoạch quản lý dữ liệu nghiên cứu. Những quyết định quan trọng từ lãnh đạo cấp cao MU đã thúc đẩy những

sáng kiến về quản lý và chia sẻ dữ liệu nghiên cứu liên tục được thực hiện như đề cập trong hình 3 (Monash University, 2019) [8].



Hình 1.3.3: Hình ảnh thành tựu của MU (Monash University, 2019) về quản lý dữ liệu nghiên cứu khoa học và công nghệ từ năm 2006 đến năm 2017

## KẾT LUẬN CHƯƠNG 1

Trong Chương 1, Học viên đã nêu nên tổng quan về dữ liệu nghiên cứu khoa học và công nghệ; Các nghiên cứu trong và ngoài nước về dữ liệu nghiên cứu khoa học và Công nghệ và cách thức quản lý.

Học viên cũng đã tìm hiểu mô hình kiến trúc dữ liệu nghiên cứu khoa học và công nghệ trong và ngoài nước để so sánh, các phương thức, cách quản lý, lưu trữ để chia sẻ dữ liệu nghiên cứu

Chương 1 đã cung cấp cơ sở lý luận để tác giả phân tích, nhận diện và xây dựng cấu trúc dữ liệu nghiên cứu khoa học và Công nghệ trong Chương 2

## CHƯƠNG 2. NGHIÊN CỨU XÂY DỰNG KIẾN TRÚC DỮ LIỆU NGHIÊN CỨU KHOA HỌC VÀ CÔNG NGHỆ

### 2.1 Giới thiệu chung

Nghiên cứu xây dựng kiến trúc dữ liệu nghiên cứu Khoa học và Công nghệ nói chung là một công việc rất quan trọng, Hiện nay dữ liệu nghiên cứu khoa học và công nghệ nhằm thúc đẩy việc tạo lập, sử dụng và khai thác dữ liệu khoa học và công nghệ mở cũng như khuyến khích sự quan tâm và tham gia của các nhà nghiên cứu trong việc chia sẻ dữ liệu khoa học và công nghệ dùng chung.

Trong khuôn khổ luận văn thạc sĩ và thời gian hạn chế, Học viên đề xuất nghiên cứu quản lý, lưu trữ dữ liệu nghiên cứu khoa học và công nghệ như: Luận văn, luận án và các sản phẩm nghiên cứu tại Học viện KHCN

#### *\* Giới thiệu chung về Học viện KHCN:*

Học viện KHCN thành lập theo Quyết định số 1691/QĐ-TTg ngày 22/9/2014 của Thủ tướng Chính phủ, theo đó Học viện KHCN trực thuộc Viện Hàn lâm, có chức năng đào tạo và cấp bằng thạc sĩ, tiến sĩ về các chuyên ngành khoa học tự nhiên và công nghệ; nghiên cứu khoa học.

Nhiệm vụ nhằm tạo điều kiện phát huy năng lực của đội ngũ cán bộ khoa học, cơ sở vật chất của Viện Hàn lâm, nhằm gắn kết tốt hơn nữa giữa nghiên cứu, đào tạo với tư vấn chính sách, góp phần phát triển nguồn nhân lực khoa học, công nghệ chất lượng cao phục vụ sự nghiệp công nghiệp hóa, hiện đại hóa đất nước.

Dưới đây là một số điểm chung về Học viện KHCN:

Mục tiêu đào tạo: Mục tiêu chính của Học viện KHCN là đào tạo sau đại học và phát triển nguồn nhân lực chất lượng cao cho các lĩnh vực khoa học và công nghệ. Những ngành học đa dạng như Khoa học máy tính, Kỹ thuật Điện tử, Sinh học, Công nghệ Thông tin, Kỹ thuật Hóa học, và nhiều lĩnh vực khác.

Nghiên cứu và phát triển: Học viện KHCN có sứ mệnh quan trọng trong việc thực hiện các dự án nghiên cứu và phát triển trong các lĩnh vực khoa học và công nghệ.

Các cấp học tại Học viện KHCHN bao gồm đào tạo sau đại học (Thạc sĩ và Tiến sĩ) và sau đó là các chương trình sau Tiến sĩ Post-doc. Trong đó có các khoa và các Viện chuyên ngành thuộc Viện Hàn lâm cùng tham gia vào công tác đào tạo, quản lý và tổ chức các ngành học khác nhau. Hiện nay Học viện KHCHN đang tổ chức đào tạo gồm 12 chuyên ngành bao gồm như Công nghệ thông tin và Viễn thông, Khoa học trái đất, Hóa học, Sinh học, Khoa học vật liệu, Khoa học môi trường ...

Hợp tác quốc tế: Học viện KHCHN thực hiện chương trình hợp tác với các trường đại học và tổ chức nghiên cứu quốc tế để cung cấp cơ hội học tập và nghiên cứu cho Học viên và giảng viên (Bao gồm các học viên đến từ các nước khác)

Xuất bản nghiên cứu: Học viện KHCHN có các ấn phẩm và sản phẩm nghiên cứu khoa học để công bố nghiên cứu của các giảng viên và học viên. Điều này giúp góp phần vào sự phát triển và chia sẻ kiến thức trong cộng đồng khoa học và công nghệ.

## **2.2 Kiến trúc dữ liệu trong nghiên cứu khoa học và công nghệ**

Kiến trúc dữ liệu nghiên cứu khoa học và công nghệ là một khía cạnh quan trọng để tổ chức và quản lý thông tin, dữ liệu và tri thức liên quan đến các hoạt động nghiên cứu và phát triển công nghệ. Điều này giúp đảm bảo rằng thông tin được lưu trữ, truy cập và chia sẻ một cách hiệu quả, từ đó thúc đẩy sự tiến bộ trong lĩnh vực này. Dưới đây là một số khía cạnh quan trọng của kiến trúc dữ liệu nghiên cứu khoa học và công nghệ:

- Phân loại và Tổ chức Dữ liệu: Đầu tiên, kiến trúc dữ liệu giúp phân loại và tổ chức dữ liệu một cách hợp lý. Điều này có thể bao gồm việc xác định các danh mục chính, chẳng hạn như dự án nghiên cứu, tài liệu khoa học, dữ liệu thử nghiệm, mã nguồn mở, và nhiều khía cạnh khác. Dữ liệu nên được phân loại sao cho dễ dàng tìm kiếm và truy cập.

- Tiêu chuẩn Hóa Dữ liệu: Một kiến trúc dữ liệu tốt cần định rõ các tiêu chuẩn và quy tắc cho việc lưu trữ và định dạng dữ liệu. Điều này đảm bảo tính nhất quán và khả năng tương thích giữa các tài liệu và dữ liệu khác nhau. Tiêu

chuẩn hóa cũng bao gồm việc xác định các nguyên tắc về đặt tên, định dạng, mã hóa và metadata.

- Quản lý Dữ liệu và Phiên bản: Một phần quan trọng của kiến trúc dữ liệu là quản lý dữ liệu và phiên bản. Nghiên cứu và phát triển thường liên quan đến việc thực hiện nhiều phiên bản và thay đổi dữ liệu. Cần có cơ chế để theo dõi và quản lý các phiên bản này một cách hiệu quả.

- Tìm kiếm và Truy cập Dữ liệu: Kiến trúc dữ liệu nên cung cấp cơ chế tìm kiếm và truy cập dữ liệu dễ dàng. Các công cụ tìm kiếm, cơ sở dữ liệu, và giao diện người dùng thân thiện giúp người sử dụng dễ dàng tìm thấy thông tin cần thiết và truy cập vào dữ liệu một cách thuận tiện.

- Chia sẻ và Hợp tác: Một khía cạnh quan trọng khác của kiến trúc dữ liệu là khả năng chia sẻ và hợp tác. Dữ liệu nghiên cứu và công nghệ thường cần được chia sẻ với cộng đồng nghiên cứu và các bên liên quan khác. Kiến trúc dữ liệu nên cung cấp các cơ chế để chia sẻ một cách an toàn và kiểm soát quyền truy cập.

- Bảo mật và Quản lý Quyền truy cập: Bảo mật dữ liệu là một vấn đề quan trọng. Kiến trúc dữ liệu nên có các biện pháp bảo mật để đảm bảo rằng dữ liệu không bị truy cập trái phép. Đồng thời, cần có cơ chế quản lý quyền truy cập để kiểm soát ai có thể truy cập và sửa đổi dữ liệu.

- Dữ liệu Liên kết và Phân tích: Kiến trúc dữ liệu cũng nên hỗ trợ khả năng liên kết dữ liệu từ các nguồn khác nhau và thực hiện phân tích dữ liệu một cách hiệu quả. Điều này có thể giúp tạo ra các thông tin mới và cung cấp cái nhìn sâu hơn về các vấn đề nghiên cứu và công nghệ.

Tóm lại, kiến trúc dữ liệu trong nghiên cứu khoa học và công nghệ đóng vai trò quan trọng trong việc tổ chức, quản lý và tận dụng dữ liệu nghiên cứu khoa học một cách hiệu quả để thúc đẩy sự phát triển trong lĩnh vực này.

Tất nhiên, dưới đây là một phân tích chi tiết hơn về các khía cạnh quan trọng của kiến trúc dữ liệu trong nghiên cứu khoa học và công nghệ:

Phân loại và Tổ chức Dữ liệu:



- Dự án Nghiên cứu: Tổ chức theo dự án để lưu trữ thông tin về mỗi dự án nghiên cứu. Bao gồm mô tả dự án, thành viên tham gia, kế hoạch và kết quả.

- Tài liệu Khoa học: Lưu trữ tài liệu như bài báo, báo cáo, sách liên quan đến nghiên cứu. Mỗi tài liệu có thể được gán metadata (Siêu dữ liệu) như tác giả, năm xuất bản, tạp chí, từ khoá.

- Dữ liệu Thử nghiệm: Lưu trữ dữ liệu thử nghiệm, kết quả thí nghiệm, số liệu và ghi chú kỹ thuật. Có thể sử dụng các hệ thống cơ sở dữ liệu hoặc kho lưu trữ để quản lý.

- Mã Nguồn Mở: Đối với các dự án công nghệ, lưu trữ mã nguồn mở và tài liệu liên quan để cho phép người khác hiểu và đóng góp vào dự án.

#### Tiêu chuẩn Hóa Dữ liệu:

- Định dạng Dữ liệu: Xác định định dạng chuẩn cho các loại dữ liệu khác nhau, chẳng hạn như hình ảnh, văn bản, âm thanh, video ...

- Metadata: Gắn metadata cho mỗi tài liệu và dữ liệu, bao gồm thông tin về nguồn gốc, tác giả, ngày tạo, quyền sở hữu và mô tả.

#### + Quản lý Dữ liệu và Phiên bản:

- Hệ thống Quản lý Phiên bản (VCS): Sử dụng hệ thống như Git để quản lý phiên bản mã nguồn và tài liệu. Điều này giúp theo dõi sự thay đổi và phục hồi phiên bản trước.

- Sao lưu Định kỳ: Thực hiện sao lưu dữ liệu và tài liệu định kỳ để đảm bảo an toàn dữ liệu trong trường hợp sự cố.

#### + Tìm kiếm và Truy cập Dữ liệu:

- Hệ thống Tìm kiếm: Xây dựng hệ thống tìm kiếm mạnh mẽ cho phép người dùng dễ dàng tìm thấy thông tin và dữ liệu cần thiết.

- Giao diện Người dùng Thân thiện: Tạo giao diện người dùng dễ sử dụng, giúp người dùng truy cập và tìm kiếm dữ liệu một cách nhanh chóng.

#### Chia sẻ và Hợp tác:

- Phân quyền Truy cập: Xác định các cấp độ quyền truy cập dữ liệu để kiểm soát người dùng có thể xem, chỉnh sửa hoặc chia sẻ dữ liệu.

- Nền tảng Chia sẻ: Tạo nền tảng cho phép người dùng chia sẻ dữ liệu một cách an toàn với người khác, bao gồm cả đối tác ngoài tổ chức.

+ Bảo mật và Quản lý Quyền truy cập:

- Mã hóa: Áp dụng mã hóa để bảo vệ dữ liệu quan trọng, đặc biệt là khi chia sẻ dữ liệu qua mạng.

- Xác thực và Ủy quyền: Sử dụng phương pháp xác thực mạnh mẽ và quản lý quyền truy cập để đảm bảo chỉ người có quyền mới có thể truy cập dữ liệu.

Dữ liệu Liên kết và Phân tích:

- Liên kết Dữ liệu: Xác định cách liên kết các dữ liệu khác nhau để tạo ra thông tin mới và tạo ra cái nhìn sâu hơn về vấn đề nghiên cứu.

- Công cụ Phân tích: Cung cấp các công cụ phân tích dữ liệu để hiểu rõ hơn về mô hình, xu hướng và mối quan hệ trong dữ liệu.

Nhớ rằng kiến trúc dữ liệu không chỉ là một cấu trúc tĩnh, mà là một quá trình liên tục, cần điều chỉnh và cải tiến để phản ánh sự phát triển trong lĩnh vực nghiên cứu Khoa học và công nghệ.

Hiện nay thực trạng dữ liệu nghiên cứu khoa học và Công nghệ hiện nay đã và đang trở thành hạ tầng quan trọng của nghiên cứu phát triển, đặc biệt trong bối cảnh công cuộc chuyển đổi số đang diễn ra mạnh mẽ. Đối với xã hội, dữ liệu đang từng bước trở thành tài sản cho các tổ chức và cá nhân, trở thành yếu tố đem lại lợi thế cạnh tranh cho các tổ chức và doanh nghiệp. Đối với công tác nghiên cứu khoa học, nhu cầu về dữ liệu quá khứ là hết sức cần thiết để phục vụ các nội dung phân tích, xây dựng mô hình và đưa ra các dự báo.

Để Quản lý và Nguyên tắc xây dựng, cập nhật, quản lý, khai thác và sử dụng dữ liệu nghiên cứu khoa học và công nghệ tôi dựa trên những nguyên tắc sau:

- Cơ sở dữ liệu nghiên cứu khoa học và công nghệ được xây dựng, quản lý theo nguyên tắc tập trung, thống nhất từ trung ương đến địa phương trên cơ

sở phân định quyền hạn, trách nhiệm cụ thể của các tổ chức, cá nhân có liên quan.

- Việc cập nhật, khai thác và sử dụng dữ liệu nghiên cứu khoa học và công nghệ phải bảo đảm đầy đủ, chính xác và kịp thời, thống nhất từ trung ương đến địa phương.

- Bảo đảm an toàn, an ninh thông tin và lưu trữ lâu dài, đúng mục đích, tạo thuận lợi cho cơ quan, tổ chức, cá nhân có yêu cầu cung cấp thông tin từ dữ liệu nghiên cứu khoa học và công nghệ.

- Bảo đảm quyền của các tổ chức, cá nhân được tiếp cận, khai thác, sử dụng thông tin trong dữ liệu nghiên cứu khoa học và công nghệ phù hợp với quy định của Đơn vị.

Xây dựng và quản lý dữ liệu nghiên cứu khoa học và công nghệ thực hiện theo các nội dung:

- Xây dựng và phát triển cơ sở hạ tầng kỹ thuật, phần mềm Cơ sở dữ liệu dùng chung và cơ sở dữ liệu thành phần, bảo đảm kết nối và chia sẻ dữ liệu thống nhất trên toàn bộ hệ thống;

- Thu thập, xử lý và cập nhật thông tin vào các cơ sở dữ liệu thành phần của dữ liệu nghiên cứu khoa học và công nghệ;

- Tổng hợp và tích hợp thông tin từ các cơ sở dữ liệu thành phần của dữ liệu nghiên cứu Khoa học và Công nghệ;

- Quản lý quyền truy cập và quyền cập nhật thông tin trong dữ liệu nghiên cứu Khoa học và Công nghệ;

- Theo dõi, giám sát tình hình sử dụng dữ liệu nghiên cứu Khoa học và Công nghệ;

- Bảo đảm an toàn, an ninh;

- Đào tạo nhân lực và hỗ trợ vận hành khai thác dữ liệu nghiên cứu Khoa học và Công nghệ.

Tại Học viện Khoa học và Công nghệ có rất nhiều các dữ liệu nghiên cứu khoa học và Công nghệ có thể ở các hình thức sau:

- Các bài luận án của các nghiên cứu sinh; và các công trình nghiên cứu sau tiến sĩ (Port doc);

- Các bài luận văn của các Học viên cao học;

- Các dữ liệu như: Bài giảng, kết quả đề tài của các Giảng viên, các nhà Khoa học tại Học viện KHHCN, các viện chuyên ngành thuộc Viện Hàn lâm.

### **2.3 Nghiên cứu xây dựng kiến trúc dữ liệu nghiên cứu choa học và công nghệ**

Việc quản lý dữ liệu nghiên cứu khoa học hiện nay có nhiều hạn chế và khó khăn, đặc biệt khi so sánh với các phương tiện và hệ thống quản lý dữ liệu tự động hiện đại. Dưới đây là một số mặt hạn chế của quản lý dữ liệu nghiên cứu bằng cách cũ:

Tính chính xác và bảo mật thấp: Quản lý dữ liệu thủ công dễ dàng dẫn đến sai sót và thiếu tính chính xác. Điều này có thể gây ra sự mất mát hoặc biến đổi dữ liệu không mong muốn. Ngoài ra, việc bảo mật dữ liệu cũng khó hơn khi không có các cơ chế tự động để bảo vệ dữ liệu.

Thời gian và công sức: Quản lý dữ liệu cũ đòi hỏi nhiều thời gian và công sức. Việc tổ chức, lưu trữ, và bảo quản dữ liệu một cách thủ công có thể làm mất thời gian và gây ra sự phiền toái.

Khả năng tìm kiếm và truy xuất hạn chế: Quản lý dữ liệu có thể làm cho việc tìm kiếm và truy xuất thông tin trong dữ liệu trở nên khó khăn hơn, đặc biệt khi dự án nghiên cứu có quy mô lớn và dữ liệu phức tạp.

Khó khăn trong chia sẻ dữ liệu: Chia sẻ dữ liệu trong dự án nghiên cứu hoặc với cộng đồng nghiên cứu khác có thể trở nên phức tạp hơn và không hiệu quả khi bạn quản lý dữ liệu.

Rủi ro mất dữ liệu: Nếu không có việc sao lưu dữ liệu định kỳ hoặc nếu không có phương tiện tự động để bảo vệ dữ liệu, rủi ro mất dữ liệu quan trọng sẽ cao hơn.

Khó khăn trong quản lý phiên bản: Quản lý phiên bản (version control) của dữ liệu và tài liệu có thể trở nên phức tạp và dễ gây nhầm lẫn khi thực hiện quản lý cũ.

Sự cản trở cho việc hợp tác: Quản lý dữ liệu bằng cách thủ công có thể gây khó khăn cho việc hợp tác với các thành viên khác trong nhóm nghiên cứu, đặc biệt khi họ sử dụng các phương tiện và hệ thống quản lý dữ liệu tự động.

Thách thức về quyền riêng tư và tuân thủ: Quản lý quyền riêng tư và tuân thủ các quy định pháp lý liên quan đến dữ liệu trở nên khó khăn hơn khi không có các công cụ tự động để hỗ trợ.

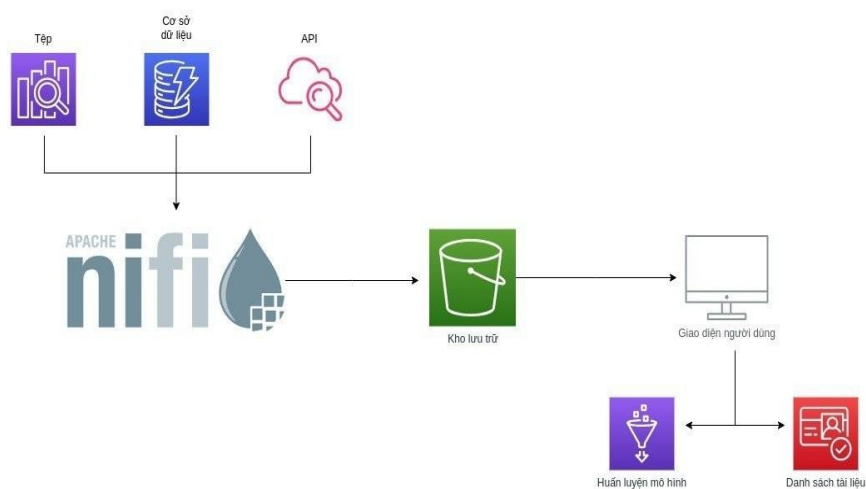
Khó khăn trong quản lý dữ liệu lớn: Với các dự án nghiên cứu lớn và phức tạp, việc quản lý dữ liệu thủ công trở nên không thể thực hiện được hoặc đòi hỏi nhiều nguồn lực lớn.

Trong môi trường nghiên cứu hiện đại, nhiều nhà nghiên cứu đã chuyển sang sử dụng các hệ thống quản lý dữ liệu tự động và công cụ để giảm bớt các hạn chế này và đảm bảo tính chính xác, bảo mật, và khả năng tìm kiếm dữ liệu tốt hơn.

Xuất phát từ hạn chế đó tác giả đề xuất mô hình quản lý dữ liệu tại Học viện KHCN

Mô tả kiến trúc dữ liệu nghiên cứu khoa học và Công nghệ tại Học viện Khoa học và Công nghệ

Hình dưới đây sẽ chỉ rõ ra cách thức cũng như vai trò của các quá trình tạo lập xử lý và quản lý dữ liệu nghiên cứu khoa học và Công nghệ tại Học viện Khoa học và Công nghệ.

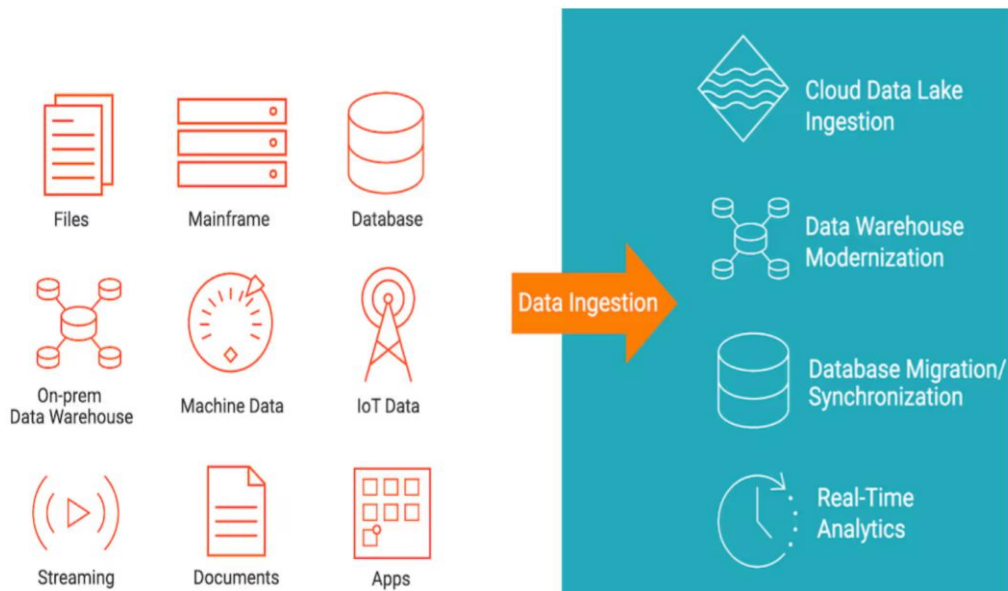


Hình 2.3.1: Mô hình kiến trúc của hệ thống

Các phần tiếp theo tác giả mô tả chi tiết các khối chức năng có trong hệ thống bao gồm: Khối tích hợp dữ liệu, khối lưu trữ, khối người dùng.

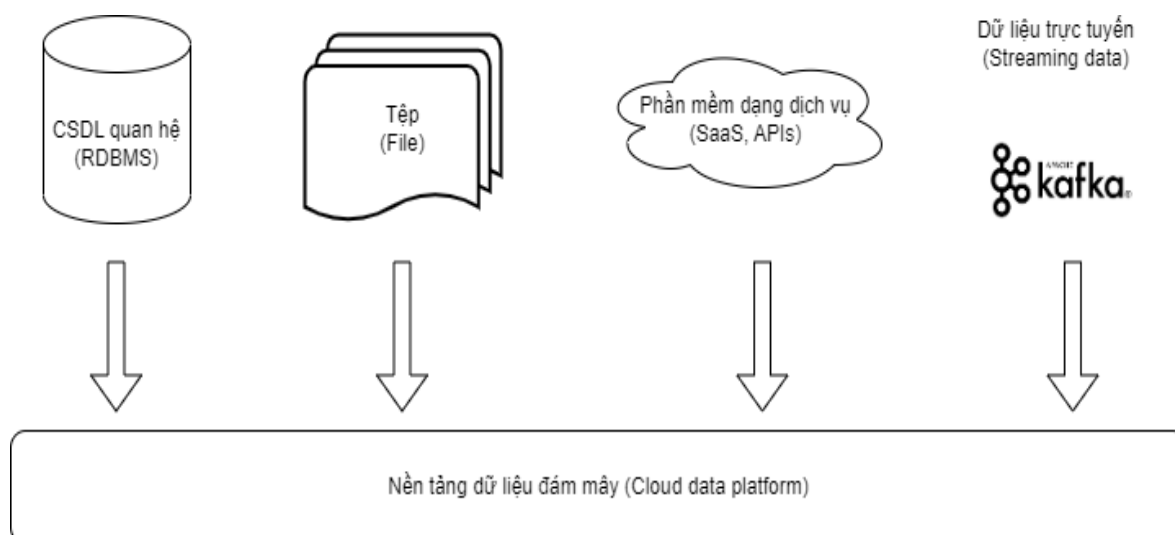
### 2.3.1 Khối tích hợp dữ liệu (Ingestion)

Khối này có nhiệm vụ kết nối với các nguồn dữ liệu khác nhau và lưu dữ liệu vào khối lưu trữ dữ liệu. Quá trình tích hợp dữ liệu này bao gồm một số tác vụ chính như: chuyển đổi dữ liệu từ các nguồn dữ liệu sang nền tảng lưu trữ dữ liệu mà vẫn giữ nguyên nội dung và định dạng của dữ liệu (việc bảo toàn dữ liệu này rất quan trọng đối với các dữ liệu được xử lý lại sau này); ghi lại các số liệu thống kê và trạng thái của dữ liệu sau đó lưu thông tin vào khối lưu trữ dữ liệu.



Hình 2.3.2: Khối tích hợp dữ liệu (Ingestion)

Trong các ứng dụng thực tế có rất nhiều nguồn dữ liệu khác nhau, tuy nhiên hệ thống chỉ tập chung sử dụng chủ yếu vào bốn nguồn dữ liệu chính là: cơ sở dữ liệu quan hệ (RDBMS), dữ liệu tệp (file), dữ liệu của các phần mềm dưới dạng dịch vụ (SaaS, API), dữ liệu trực tuyến (Streaming).



Hình 2.3.3: Mô hình các nguồn dữ liệu khoa học và công nghệ của hệ thống

- *Tích hợp dữ liệu từ tệp (file)*: Dữ liệu tệp (File) là loại dữ liệu phổ biến xuất hiện trong các nguồn dữ liệu. Hệ thống cung cấp hai phương pháp chuyển dữ liệu file vào nền tảng lưu trữ. Phương pháp thứ nhất là sử dụng giao thức truyền tệp (FTP) hoặc FTP tiêu chuẩn được hỗ trợ bởi nhiều công cụ ETL. Phương pháp thứ hai là sử dụng lưu trữ đám mây thay vì máy chủ FTP. Các tệp nguồn được lưu trữ tại một đám mây cục bộ và hệ thống sẽ thực hiện sao chép từ đám mây nguồn sang đám mây đích của nền tảng lưu trữ.

- *Tích hợp dữ liệu từ CSDL quan hệ (RDBMS)*: Hệ thống cung cấp các phương pháp: tích hợp từ RDBMS sử dụng SQL; tích hợp từ cơ sở dữ liệu NoSQL (BigData); tích hợp siêu dữ liệu (metadata) cho RDBMS và NoSQL.

- *Tích hợp từ RDBMS sử dụng SQL*: Sử dụng các câu lệnh truy vấn dữ liệu để thực hiện truy vấn dữ liệu từ CSDL nguồn và lưu trữ vào nền tảng.

- *Tích hợp từ CSDL NoSQL (BigData)*: Sử dụng các công cụ (Tool) của các CSDL NoSQL để định nghĩa một quy trình (pipeline) tích hợp dữ liệu. Các CSDL NoSQL phổ biến mà hệ thống hỗ trợ bao gồm: MONGODB, CASSANDRA.

- *Tích hợp các siêu dữ liệu (metadata) từ hệ thống nguồn (RDBMS, NoSQL)* bằng các công cụ được xây dựng, bảo đảm tính toàn vẹn dữ liệu từ hệ thống nguồn vào nền tảng lưu trữ dữ liệu.

- *Tích hợp dữ liệu từ dữ liệu dòng (thời gian thực)*: Hệ thống sử dụng giải pháp *Apache Kafka* thực hiện tích hợp dữ liệu dòng (streams) có yếu tố thời gian thực từ các ứng dụng, điển hình là các ứng dụng IoT.

- *Tích hợp dữ liệu từ các ứng dụng SaaS*: Ứng dụng SaaS ngày càng trở nên phổ biến trong giai đoạn hiện nay. Trong hệ thống, việc tích hợp với các ứng dụng SaaS được thực hiện bằng cách sử dụng API qua giao thức HTTP(s).

*Về công nghệ sử dụng*, hệ thống sử dụng nền tảng mã nguồn mở *Nifi* để thực hiện tích hợp dữ liệu. *Apache NiFi* là một trong những giải pháp mã nguồn mở phổ biến cho phép kết nối với nhiều nguồn dữ liệu khác nhau và đưa dữ liệu vào nền tảng dữ liệu. *NiFi* sử dụng kiến trúc có thể cho phép tạo các trình kết nối mới bằng Java.

*Apache NiFi* là một phần mềm mã nguồn mở viết bằng ngôn ngữ Java, được tạo ra để tự động hóa luồng dữ liệu giữa các hệ thống phần mềm với nhau. Phần mềm được xây dựng từ năm 2006 dựa trên phần mềm *NiagaraFiles* phát triển bởi lập trình viên NSA, sau đó được chuyển sang mã nguồn mở vào năm 2014.



*Hình 2.3.4: Chương trình mã nguồn mở Nifi (Logo của mã nguồn mở Nifi)*

Chức năng của phần mềm mã nguồn mở *Nifi* được biết đến như:

*Apache NiFi* là một hệ thống mã nguồn mở được phát triển bởi Apache Software Foundation, được thiết kế để quản lý, tự động hóa và chuyển đổi dữ liệu giữa các hệ thống khác nhau. Chức năng chính của *Apache NiFi* bao gồm:



Thu thập và Ingest Dữ liệu: NiFi cho phép bạn thu thập dữ liệu từ nhiều nguồn khác nhau như cơ sở dữ liệu, máy chủ web, thiết bị IoT, logs hệ thống, và nhiều nguồn dữ liệu khác. Nó hỗ trợ nhiều giao thức và định dạng dữ liệu khác nhau.

Xử lý và Chuyển đổi Dữ liệu: NiFi cho phép bạn thực hiện xử lý dữ liệu trong chuyển đổi dữ liệu từ định dạng này sang định dạng khác, lọc dữ liệu, làm sạch dữ liệu, và thậm chí thực hiện tính toán phức tạp trên dữ liệu.

Điều khiển và Quản lý Dữ liệu: NiFi cung cấp một giao diện người dùng trực quan để quản lý dữ liệu và luồng dữ liệu. Bạn có thể theo dõi, đánh giá và kiểm soát các luồng dữ liệu trong thời gian thực.

Các loại RDBMS: Oracle, MySql, Postgre, ...

Các loại DB NoSQL: Mongo, HBase, Cassandra, ...

Từ các nguồn web như: HTTP, web-socket

Lấy hoặc đẩy dữ liệu streaming vào Kafka

Ngoài việc nhập và xuất dữ liệu thì NiFi còn các chức năng như routing dữ liệu theo thuộc tính và nội dung, xử lý dữ liệu như: lọc, chỉnh sửa, thêm bớt nội dung của dữ liệu trước khi đưa đến nơi lưu trữ.

Ba nhóm tính năng nổi bật của Nifi bao gồm khả năng quản lý luồng dữ liệu; việc sử dụng, vận hành một cách dễ dàng; và khả năng mở rộng.

*Khả năng quản lý luồng dữ liệu:*

Đảm bảo an toàn: Mỗi đơn vị dữ liệu trong luồng sẽ được biểu diễn bởi một Object có tên là FlowFile. Nó sẽ ghi lại tất cả các thông tin về dữ liệu trong luồng như đang được xử lý bởi khối nào, đang được chuyển đi đâu, ... Lịch sử xử lý của một FlowFile lại được lưu trữ trong Provenance Repo để có thể truy vết. Kết hợp với cơ chế Copy- on -Write, NiFi lưu trữ lại dữ liệu tại từng bước trong luồng trước khi xử lý, giúp dễ dàng chạy lại dữ liệu.

Data Buffering: tính năng này giúp giải quyết vấn đề tốc độ không đồng bộ giữa hai hệ thống khác nhau. Nó hoạt động dựa theo cơ chế Queue giữa hai

khối xử lý trong luồng. Dữ liệu này sẽ được giữ trên RAM, nhưng nếu nó vượt qua ngưỡng mình cài thì dữ liệu sẽ được đưa xuống ổ cứng.

Thiết lập độ ưu tiên: trong một số trường hợp cần xử lý dữ liệu này trước khi xử lý những dữ liệu khác.

Hỗ trợ đánh đổi giữa tốc độ và khả năng chịu lỗi: Có những luồng dữ liệu cần đảm bảo tuyệt đối về tính toàn vẹn và an toàn của dữ liệu chấp nhận độ trễ cao. Và có những luồng ta lại cần chuyển được dữ liệu tới đích trong thời gian ngắn nhất có thể. NiFi sẽ hỗ trợ bạn cài đặt để cân bằng giữa hai yếu tố này.

*Độ phức tạp trong quá trình sử dụng:*

Việc tạo ra một luồng dữ liệu sẽ được thực hiện hoàn toàn trên giao diện WEB, và bằng một số thao tác kéo thả của người dùng sẽ nhanh chóng tạo được một luồng hoạt động đơn giản.

Tính tái sử dụng cũng được hỗ trợ, người dùng có thể tạo ra một template chứa một luồng cơ bản để sử dụng lại khi cần.

Theo dõi trực quan lịch sử xử lý của dữ liệu khi cần kiểm tra lỗi.

Chạy lại được cả dữ liệu tại từng bước xử lý

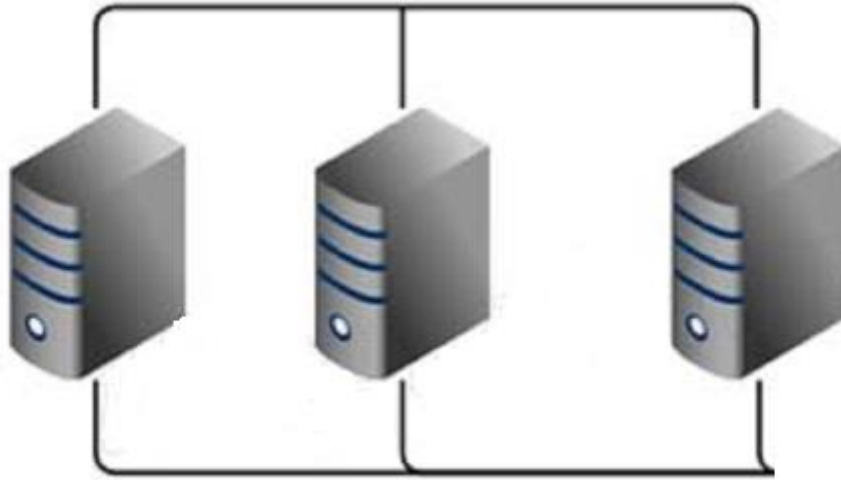
Người dùng có thể lập trình được một thành phần xử lý, điều khiển, ... trong NiFi khi cần. Ví dụ như một khối mã hóa hoặc giải mã dữ liệu.

*Khả năng mở rộng:*

Đây là một tính năng quan trọng của các ứng dụng trong các hệ thống phân tán là khả năng mở rộng. Nếu một luồng dữ liệu trên một server NiFi có thể xử lý được 100MB/s, nhưng yêu cầu thực tế lại lên đến 500MB/s thì các bạn có thể cài đặt một cụm gồm nhiều server để xử lý dữ liệu một cách song song mà không cần nâng cấu hình của server.

### **2.3.2 Khối lưu trữ (storage)**

Khối lưu trữ dữ liệu chịu trách nhiệm lưu trữ dữ liệu cho sử dụng dài hạn, ngắn hạn. Về công nghệ sử dụng, hệ thống sử dụng ở cứng để lưu trữ dữ liệu

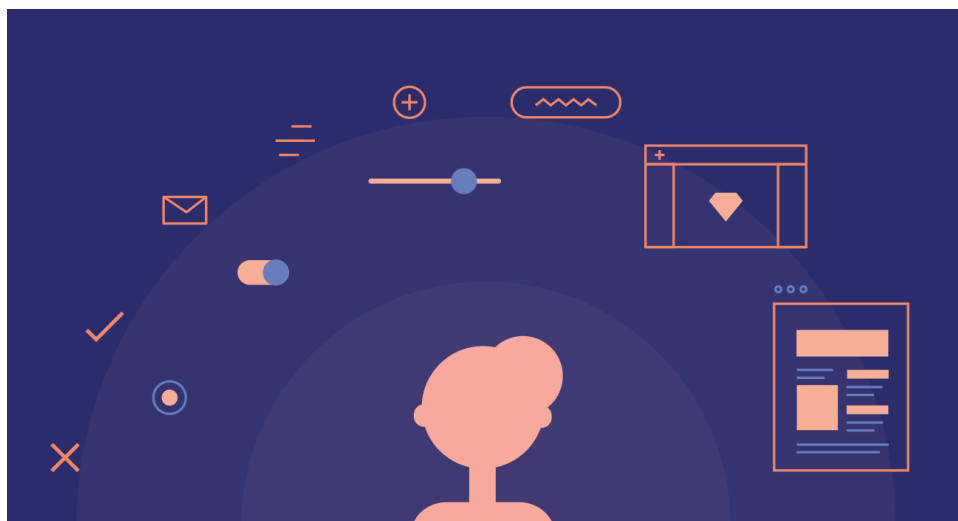


*Hình 2.3.2.1: Máy chủ lưu trữ vật lý*

### **2.3.3 Khối người dùng (user)**

Khối này có nhiệm vụ cung cấp kết quả của quá trình xử lý, phân tích dữ liệu cho các đối tượng người dùng khác nhau. Đối tượng sử dụng là những người muốn truy cập dữ liệu từ bộ lưu trữ (storage) mà không cần thông qua kho dữ liệu. Để thực hiện được điều này, hệ thống triển khai một API riêng biệt cho phép người dùng truy cập dữ liệu theo thời gian thực. Cách tiếp cận này giúp hệ thống đáp ứng được nhu cầu sử dụng dữ liệu ngày càng tăng của người dùng

- Người dùng sẽ được thao tác với hệ thống trên nền tảng Web, một số hoạt động người dùng có thể thực hiện trên nền tảng này



*Hình 2.3.3.1: Người dùng hệ thống*

## **KẾT LUẬN CHƯƠNG 2**

Chương 2 đã nêu lên các đặc điểm của hệ thống quản lý dữ liệu trong phạm vi nghiên cứu của luận văn qua đó học viên đề xuất xây dựng mô hình quản lý dữ liệu cho đối tượng nghiên cứu ở đây học viên chọn mô hình nghiên cứu dữ liệu nghiên cứu khoa học và công nghệ tại Học viện KHCN, Viện Hàn lâm và ngoài ra Chương 2 này cũng giới thiệu chi tiết các khối chức năng chính có trong hệ thống bao gồm Khối tích hợp dữ liệu (Ingestion) , Khối lưu trữ (Storage), Khối người dùng (User)

Các kết quả thực nghiệm và đánh giá sẽ được trình bày trong Chương 3.

## CHƯƠNG 3: THỰC NGHIỆM MÔ HÌNH VÀ ĐÁNH GIÁ KẾT QUẢ

### 3.1 Công cụ và môi trường thử nghiệm

- Framework Nextjs để tạo Web

Next.js là một framework được phát triển bởi Zeit, được sử dụng để xây dựng các ứng dụng web động trên Node.js. Next.js được thiết kế để giúp các nhà phát triển xây dựng các trang web tốc độ cao, đẹp mắt và dễ sử dụng.

Next.js cung cấp một số tính năng quan trọng như tự động code splitting, tự động cấu hình SEO, các tính năng server-side rendering và client-side hydration. Tự động code splitting cho phép Next.js chia các tập tin JavaScript thành các nhóm nhỏ hơn, giúp giảm kích thước tải xuống cho trang web và tăng tốc độ tải trang.

Next.js cũng cung cấp tính năng server-side rendering, giúp các trang web tải nhanh hơn và tốt hơn cho SEO. Client-side hydration cho phép Next.js tải nhanh hơn và trải nghiệm người dùng tốt hơn trên các trình duyệt.

Next.js cung cấp một API để tạo các trang với URL tĩnh, giúp các trang web của bạn có thể được share được và tốt hơn cho SEO. Next.js cũng hỗ trợ các tính năng xử lý form, cho phép nhà phát triển xử lý form dễ dàng và nhanh chóng.

- Ngôn ngữ lập trình: TypeScript là một phiên bản cao hơn của JavaScript, được thiết kế để xây dựng các ứng dụng lớn và phức tạp. Nó kế thừa nhiều khái niệm từ Java và C#, TypeScript là ngôn ngữ tĩnh (Static typed) có nghĩa là nó nghiêm ngặt và có trật tự trái ngược với free-type. Nó còn được bổ sung thêm lớp hướng đối tượng mà điều này không có ở Javascript.

- Thư viện axios để call API là quá trình gửi yêu cầu từ một ứng dụng hoặc hệ thống đến một API (Application Programming Interface) để trao đổi dữ liệu và thực hiện các chức năng cụ thể. Trong ngữ cảnh này, API calls đề cập đến việc gọi hoặc truy cập các phương thức, hàm hoặc endpoint của một API để lấy dữ liệu, cập nhật thông tin, hoặc thực hiện một hành động nào đó.

- Khi gọi API, ứng dụng gửi yêu cầu thông qua một giao thức (như HTTP) và nhận lại phản hồi từ API chứa dữ liệu hoặc kết quả được yêu cầu. Các API calls có thể được thực hiện bằng cách sử dụng các phương thức như GET,

POST, PUT, DELETE để truy xuất, tạo, cập nhật hoặc xóa dữ liệu từ một nguồn dữ liệu được quản lý bởi API.



Hình 3.1.1: Công cụ và Môi trường thử nghiệm

### 3.2 Thực nghiệm thao tác trên hệ thống

Giao diện của hệ thống này được xây dựng trên nền tảng *Nextjs*, ngôn ngữ lập trình *typescript*. Hệ thống được xây dựng với 5 chức năng chính.

- *System Introductions*: Phần này giới thiệu những điều cơ bản về hệ thống như mục đích, chức năng, v.v

- *Aspects*: Khối chức năng này hiển thị thông tin dữ liệu đã được lưu trữ. Hệ thống lưu trữ thông tin 12 lĩnh vực nghiên cứu và giảng dạy của Học viện Khoa học và Công nghệ. Những dữ liệu này được lưu dưới 3 định dạng chính đó là hình ảnh hoặc video (Image/Video), văn bản (Documentations) và cơ sở dữ liệu (Databases).

- *Contribute Documents*: Tác vụ này cho phép các giảng viên và học viên đóng góp những dữ liệu có ích, phục vụ cho quá trình nghiên cứu. Ngoài ra, học viên cũng có thể lưu trữ bản luận văn, luận án của học viên đó giúp cho việc quản lý của Học viện Khoa học và Công nghệ trở lên thuận tiện hơn.

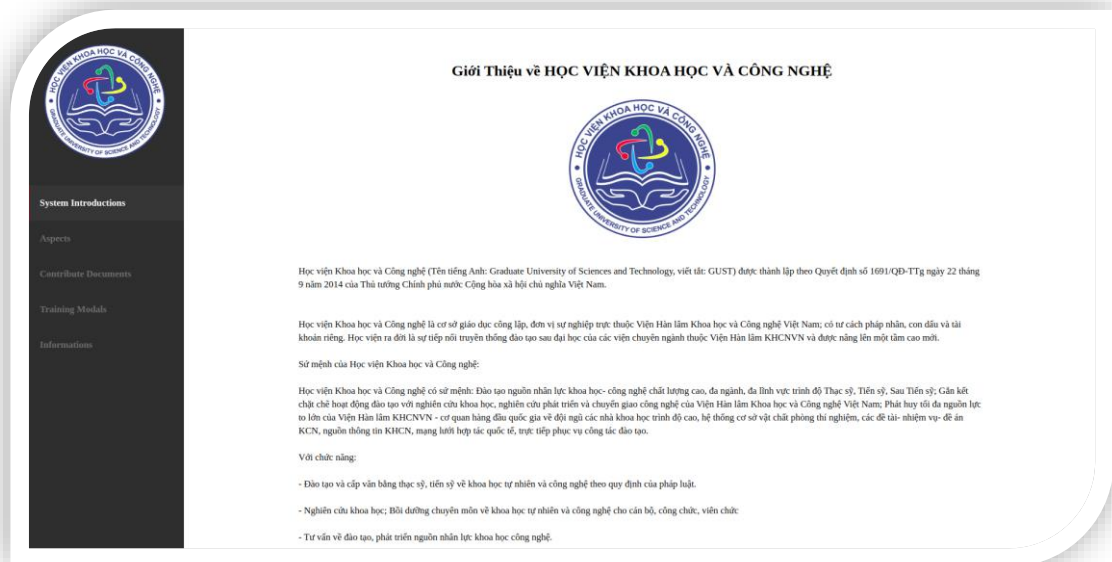
- *Training Models*: Khối chức năng này cho phép các nhà khoa học có thể huấn luyện ra các mô hình học máy từ dữ liệu sẵn có của người dùng, hoặc dữ liệu đóng góp của các nhà khoa học khác.

- *Informations*: Các nhà khoa học nếu có thắc mắc gì về hệ thống này có thể liên hệ theo thông tin trong tác vụ này.

Phần tiếp theo giới thiệu chi tiết cách thức hoạt động của 5 chức năng chính trong hệ thống.

### a) System Introductions (Giới thiệu hệ thống)

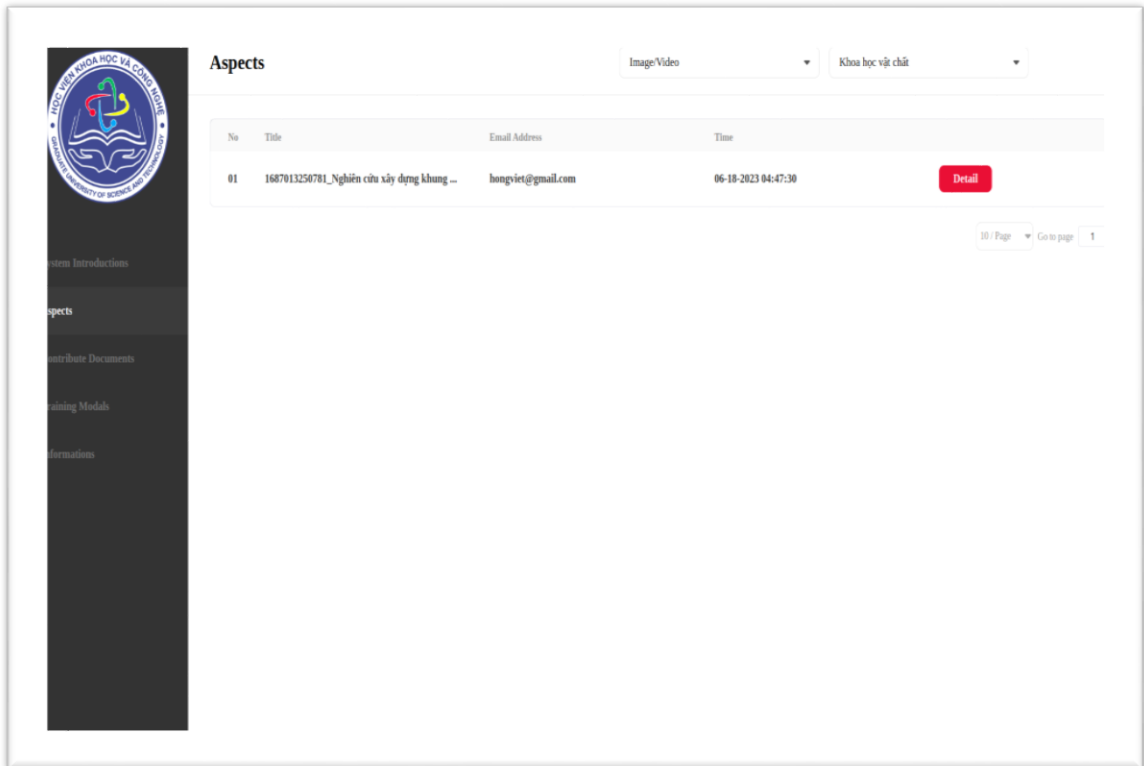
Màn hình *System Introductions* là giao diện mặc định sẽ được hiển thị mỗi khi có người dùng truy cập vào hệ thống. Tại đây sẽ có các thông tin giới thiệu chi tiết về Học viện Khoa học và Công nghệ cũng như các thông tin chính của hệ thống.



Hình 3.2.1: Màn hình đăng nhập vào hệ thống *System Introductions*

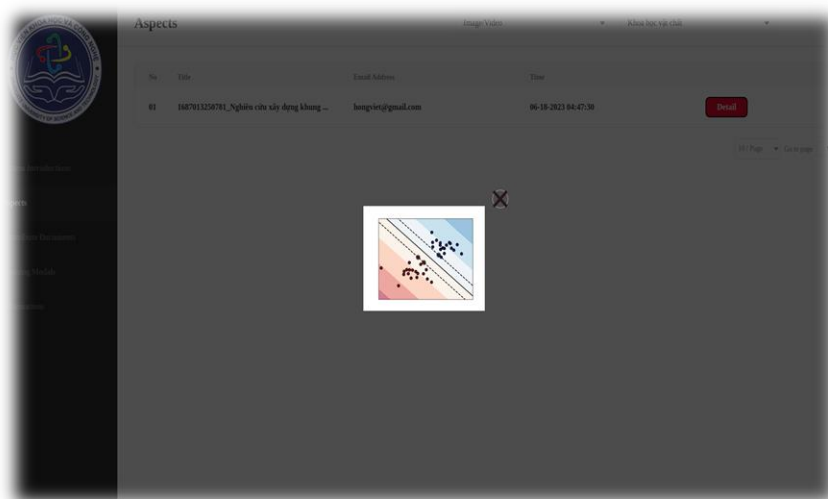
### b) Aspects (các lĩnh vực đang nghiên cứu và giảng dạy)

Các nhà khoa học có thể xem các tài liệu tương ứng với 12 lĩnh vực đang được nghiên cứu và giảng dạy tại Học viện Khoa học và Công nghệ bằng cách chọn mục **Aspects** trên thanh công cụ phía trái màn hình. Người dùng có thể chọn lĩnh vực quan tâm và định dạng của dữ liệu. Lúc này danh sách các tài liệu sẽ được hiển thị dựa trên bộ lọc mà người dùng đã chọn.



Hình 3.2.2: Màn hình đăng nhập vào hệ thống *Aspect*

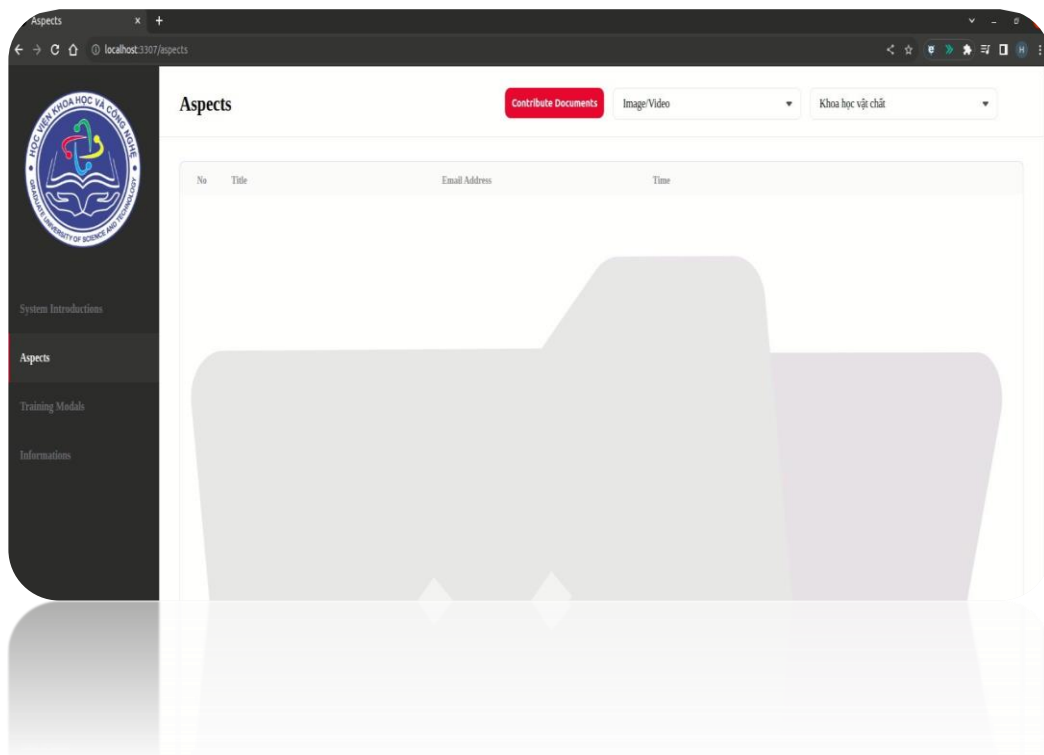
Mặc định khi lần đầu vào màn hình **Aspect** sẽ là hiển thị tài liệu định dạng *Image/Video* cho lĩnh vực *Công nghệ thông tin*. Đối với kiểu dữ liệu là *Image/Video*, người dùng có thể xem trực tiếp nội dung hình ảnh học video bằng cách chọn nút *Detail*.



Hình 3.2.3: Màn hình đăng nhập vào hệ thống *Documentations* và *Databases*



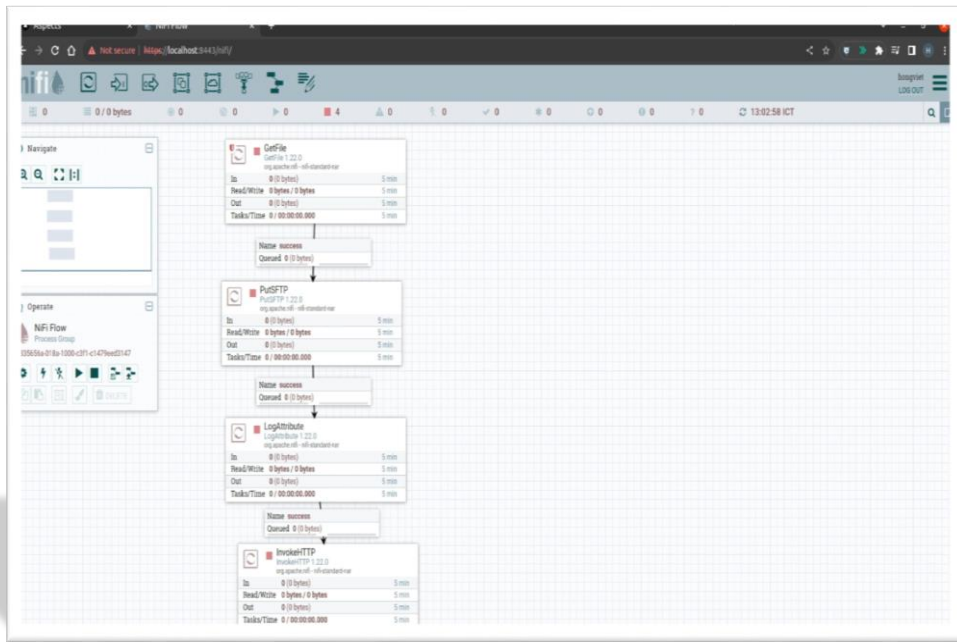
Đối với kiểu dữ liệu là *Documentations* và *Databases*, người dùng có thể tải tài liệu bằng cách chọn nút *Download*, một màn hình sẽ hiện ra để người dùng có thể chọn nơi lưu trữ dữ liệu.



Hình 3.2.4: Màn hình đăng nhập vào hệ thống *Documentations* và *Databases*

### c) **Contribute Documents**

Tại màn hình **Contribute Documents (Đóng góp dữ liệu)**, người dùng sẽ cần phải đăng nhập tài khoản cá nhân để thực hiện thao tác trên màn hình này. Tài khoản này sẽ được Học viện Khoa học và Công nghệ cấp khi học tập và làm việc tại các tổ chức liên quan. Giao diện ban đầu của màn hình này như sau:



Hình 3.2.5: Màn hình đăng nhập vào hệ thống Contribute Documents  
Sau khi hoàn thành quá trình gửi, kết quả sẽ hiển thị như hình dưới:



Hình 3.2.6: Màn hình giao diện kết quả sau khi được gửi lên hệ thống  
Sau khi nhấn chọn nút Login, người dùng sẽ được chuyển tới màn hình đăng nhập

Username \*

luuhieu

Password \*

Log in

Hình 3.2.7: Màn hình đăng nhập vào hệ thống Contribute Documents

Khi đăng nhập theo tài khoản cá nhân đã được cấp, người dùng có thể trở lại màn hình **Contribute Documents** để tải lên những tài liệu mà người dùng muốn chia sẻ.

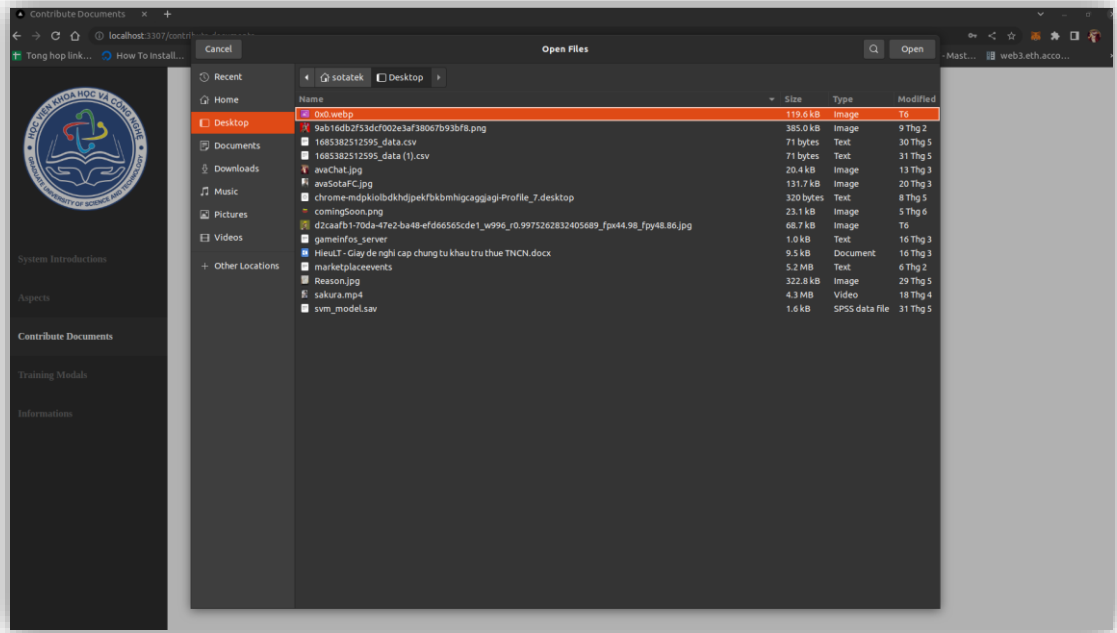
Khoa học vật chất

Select File

Upload

Hình 3.2.8: Màn hình đăng nhập vào hệ thống để chọn và tải tập tin lên

Để có thể tải dữ liệu lên, người dùng cần chọn lĩnh vực muốn đóng góp bằng cách nhấn chọn *Select File* để lựa chọn tệp tài liệu.



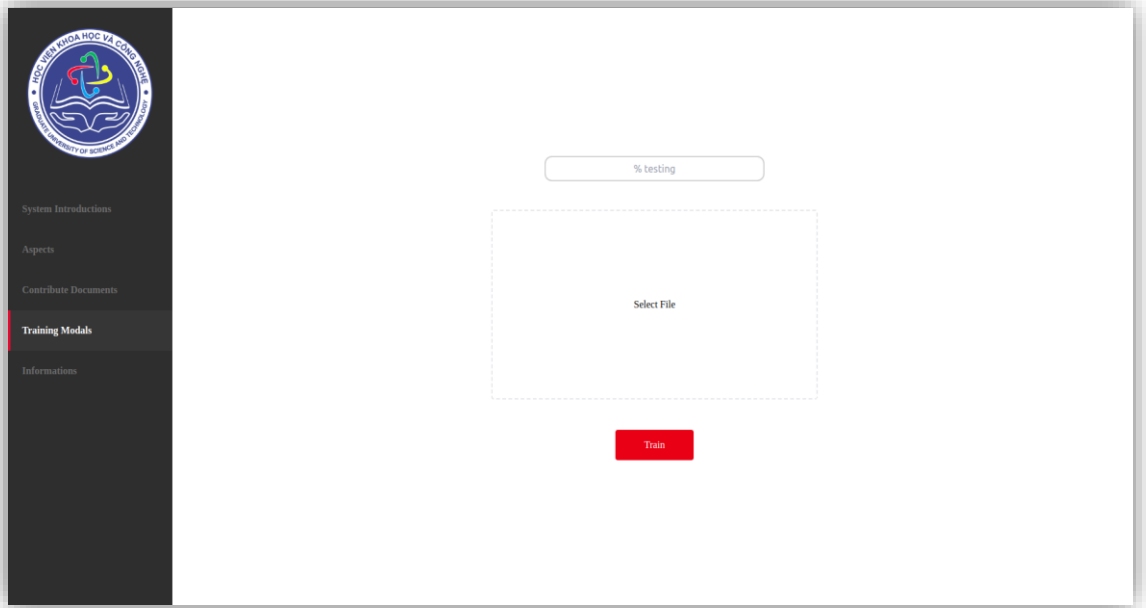
Hình 3.2.9: Màn hình đăng nhập vào hệ thống để chọn up dữ liệu

Lưu ý những tệp dữ liệu được chấp nhận được giới hạn trong các định dạng như: ảnh/video (png, jpg, mp4), văn bản (txt, docs), cơ sở dữ liệu (csv).

Sau khi bấm *Upload*, tài liệu sẽ được tải lên và lưu vào kho dữ liệu, người dùng có thể xem tài liệu vừa tải lên tại màn hình *Aspects*.

#### d) Training Models (Phần huấn luyện mô hình học máy)

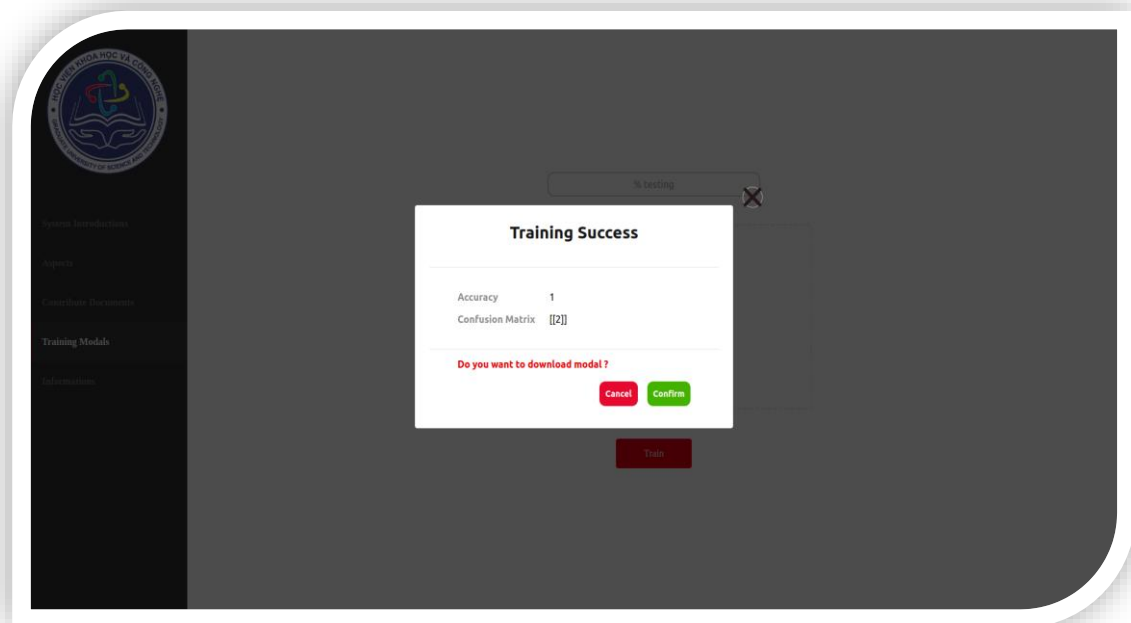
Tại màn hình này, người dùng có thể huấn luyện các mô hình học máy, cụ thể trong hệ thống này cung cấp việc huấn luyện mô hình phân lớp SVM (Support Vector Machine là bài toán đi tìm mặt phân cách sao cho margin tìm được là lớn nhất, đồng nghĩa với việc các điểm dữ liệu an toàn nhất). Tương tự như khối chức năng đóng góp dữ liệu, người dùng phải đăng nhập mới có thể tiếp tục thao tác trên màn hình này. Sau khi người dùng đăng nhập thành công, màn hình chính sẽ hiện ra như sau:



Hình 3.2.10: Màn hình đăng nhập vào hệ thống để chọn tải dữ liệu để huấn luyện mô hình học máy

Đầu tiên, người dùng cần lựa chọn tệp dữ liệu dùng để huấn luyện mô hình (lưu ý sử dụng định dạng csv) bằng cách nhấn chọn *Select File*. Sau khi chọn tệp dữ liệu xong, người dùng cần cài đặt số % dữ liệu cho tập kiểm thử. Sau khi đã hoàn thành các thao tác cần thiết, nhấn chọn *Upload* và hệ thống sẽ tự động huấn luyện mô hình cho người dùng.

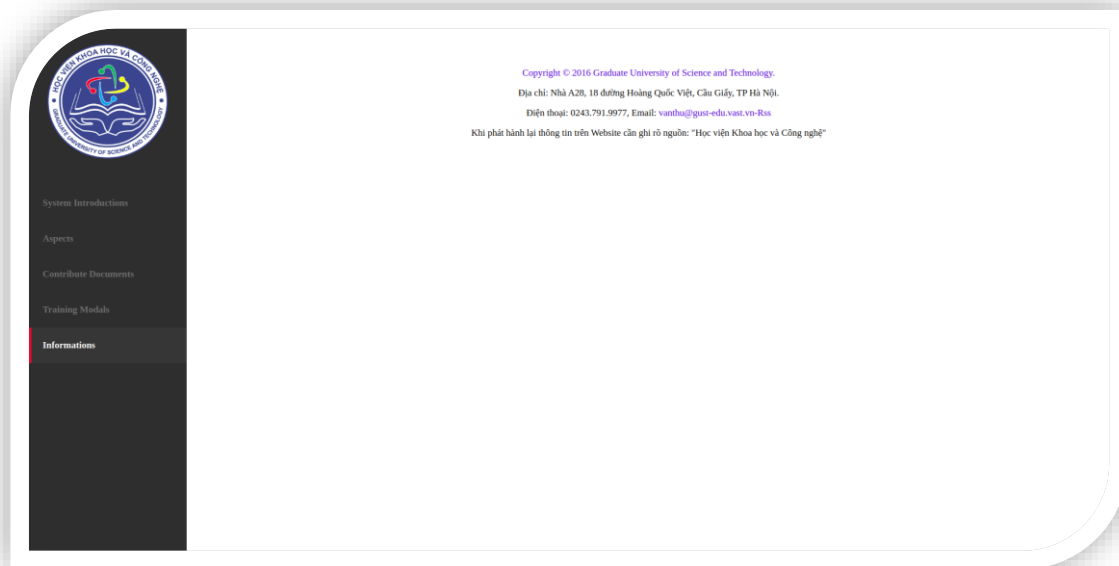
Sau khi quá trình huấn luyện mô hình hoàn tất, màn hình sẽ hiển thị thông báo về thông tin mô hình và đề xuất người dùng tải mô hình về máy để sử dụng. Ngoài ra, màn hình cũng sẽ hiển thị thêm một số thông tin trong quá trình huấn luyện mô hình như độ chính xác (Accuracy) và Confusion Matrix.



Hình 3.2.11: Màn hình hiển thị khi dữ liệu được huấn luyện thành công

### e) Informations

Các thông tin liên hệ sẽ được hiển thị khi người dùng nhấn chọn mục **Informations** trên thanh công cụ phía trái màn hình chính.



Hình 3.2.12: Màn hình thông tin liên hệ tại Học Viện Khoa học và Công nghệ

### KẾT LUẬN CHƯƠNG 3

Trong Chương 3, Học viên đã xây dựng ứng dụng thực nghiệm mô hình và đánh giá kết quả dữ liệu nghiên cứu khoa học và Công nghệ tại Học viện KH&CN.

- Công cụ và môi trường thử nghiệm xây dựng ứng dụng:
  - Framework Nextjs để tạo Web
  - Ngôn ngữ lập trình: TypeScript là một phiên bản cao hơn của JavaScript, được thiết kế để xây dựng các ứng dụng lớn và phức tạp
  - Thư viện axios để call API là quá trình gửi yêu cầu từ một ứng dụng hoặc hệ thống đến một API (Application Programming Interface) để trao đổi dữ liệu và thực hiện các chức năng
    - Hệ quản trị CSDL SQL – Web Apache
    - Tích hợp các nguồn dữ liệu sử dụng Apache Nifi
- Thực nghiệm thao tác trên hệ thống:
  - Giao diện của hệ thống này được xây dựng trên nền tảng Nextjs, ngôn ngữ lập trình typescript. Hệ thống được xây dựng với 5 chức năng chính (*System Introductions, Aspects, Contribute Documents, Training Models, Informantions*)
  - Hệ thống ứng dụng học viên xây dựng với mục đích là quản lý, lưu trữ dữ liệu nghiên cứu khoa học và công nghệ tạo Học viện KH&CN.

Đây cũng là tiền đề để học viên đưa ra một số hướng nghiên cứu tiếp theo trong việc xây dựng cấu trúc quản lý, sử dụng dữ liệu nghiên cứu khoa học và Công nghệ tại Học viện KH&CN.

## KẾT LUẬN

### 1. Kết quả đạt được

- Luận văn đã trình bày tóm tắt các nghiên cứu liên quan đến dữ liệu nghiên cứu Khoa học ở phạm vi trong và ngoài nước.

- Xây dựng được hệ thống có khả năng thu thập dữ liệu từ nhiều nguồn, lưu trữ và hiển thị dữ liệu.

- Xây dựng ứng dụng và huấn luyện mô hình từ những dữ liệu đã thu thập được.

### 2. Hạn chế

- Hệ thống hiện tại đang tập trung vào 4 nguồn dữ liệu phổ biến, chưa bao quát được nhiều nguồn dữ liệu khác.

- Không gian lưu trữ dữ liệu còn hạn chế.

- Phản ứng dựng xây dựng các mô hình từ dữ liệu thu thập được chưa đa dạng.

- Với kiến thức trong một số lĩnh vực khác còn hạn chế và thời gian tìm hiểu và nghiên cứu không dài nên đề tài không tránh khỏi những khuyết điểm và thiếu sót. Nếu có cơ hội cũng như có thời gian dài hơn tác giả sẽ tiếp tục hoàn thiện và nghiên cứu sâu hơn về vấn đề này

### 3. Hướng nghiên cứu tiếp theo

Trong tương lai, nghiên cứu này có thể thực hiện theo các hướng mới để cải thiện những hạn chế của hệ thống:

- Mở rộng khả năng tiếp nhận dữ liệu từ các nguồn khác ngoài 04 (bốn) nguồn dữ liệu phổ biến đã được nêu trong luận văn.

- Cải thiện không gian lưu trữ dữ liệu thông qua các dịch vụ lưu trữ dữ liệu.

- Mở rộng thêm các mô hình ứng dụng cho các nhà khoa học để tận dụng nguồn dữ liệu sẵn có.

- Nghiên cứu xây dựng hệ thống có thể tích hợp chung với hệ thống tại cổng thông tin điện tử tại Học viện KHCN, Viện Hàn lâm, Để quản lý, lưu trữ dữ liệu nghiên cứu khoa học có khả năng truy cập và chia sẻ



## TÀI LIỆU THAM KHẢO

### 1. Tài liệu tham khảo tiếng Việt

[1]. Bộ Khoa học và Công nghệ (2017). Đề án “Phát triển Hệ tri thức Việt số hóa”. Truy cập từ: <https://www.most.gov.vn/vn/tin-tuc/12064/de-an-phat-trien-he-tri-thuc-viet-so-hoa.aspx> [truy cập ngày 05/04/2023];

[2]. Bộ Khoa học và Công nghệ (2018). Cơ sở dữ liệu Quốc gia về H&CN: Minh bạch hóa hoạt động KH&CN. Truy cập từ: <https://www.most.gov.vn/vn/tin-tuc/14324/co-so-du-lieu-quoc-gia-ve-khcn--minh-bach-hoa-hoat-dong-khcn.aspx> [truy cập ngày 05/04/2023];

[3]. Nền tảng quản trị dữ liệu nghiên cứu khoa học và công nghệ mở phục vụ chia sẻ, dùng chung ở quy mô quốc gia. Truy cập từ: <https://iti.vnu.edu.vn/nen-tang-quan-tri-du-lieu-nghien-cuu-khoa-hoc-va-cong-nghe-mo-phuc-vu-chia-se-dung-chung-o-quy-mo-quoc-gia/> [truy cập thứ 2 ngày 02/06/2023];

[4]. Quản lý dữ liệu nghiên cứu: Một cách tiếp cận quản trị dữ liệu cho các thư viện Việt Nam <https://hvtc.edu.vn/tabid/558/catid/143/id/29661/Quan-ly-du-lieu-nghien-cuu-Mot-cach-tiep-can-quan-tri-du-lieu-cho-cac-thu-vien-Viet-Nam/Default.aspx> [truy cập ngày 08/06/2023] ;

[5]. Một số vấn đề trong quản lý dữ liệu nghiên cứu của các thư viện. truy cập từ: <https://hvtc.edu.vn/tabid/558/catid/143/id/32539/Mot-so-van-de-trong-quan-ly-du-lieu-nghien-cuu-cua-cac-thu-vien/Default.aspx> [truy cập thứ 2 ngày 16/06/2023].

### 2. Tài liệu tham khảo tiếng Anh

[6] National Science Foundation. Long-Lived Digital data collections: Enabling research and education in the 21st century. <http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>;

[7]. Cox, A. M., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science*, 46(4), 299-316. doi:10.1177/0961000613492542;

[8]. Mekong River Commission (2001). Procedures for Data and Information Exchange and Sharing. Retrieved from:

<http://vnmc.gov.vn/Upload/Documents/Procedures-Data-Info-Exchange-n-Sharing.pdf> [Accessed 5 Jun. 2019];

[9]. <https://www.monash.edu/library/researchers/researchdata/about/achievements> [Accessed 5 Jun. 2019];

[10]. Open Development Vietnam (2019). About us. Retrieved from: <https://vietnam.opendevlopmentmekong.net/about-us/> [Accessed 5 Jun. 2019];

[11]. Pryor, G., Jones, S., & Whyte, A. (2014). *Delivering Research Data Management Services: Fundamentals of Good Practice*. London: Facet Publishing;

[12]. RMIT University (2016). Research data management policy process. Retrieved from: <https://www.rmit.edu.au/about/governance-and-management/policies/research-policy/research-data-management-process> [Accessed 5 Jun. 2023];

[13]. UK Data Archive. Research data lifecycle: [http:// www.data-archive.ac.uk/create-manage/life-cycle](http://www.data-archive.ac.uk/create-manage/life-cycle) [Accessed 5 Jun. 2023];

[14]. Elsevier. Research data management: <http://www.elsevier.com/about/open-science/research-data/research-data-management>. [Accessed 5 Jun. 2023];

[15]. Martin Lewis. Libraries and the management of research data: [http://eprints.whiterose.ac.uk/11171/1/LEWIS\\_Chapter\\_v10.pdf](http://eprints.whiterose.ac.uk/11171/1/LEWIS_Chapter_v10.pdf). [Accessed 07 Jun. 2023].

Số: 948/QĐ-HVKHCN

Hà Nội, ngày 18 tháng 8 năm 2023

**QUYẾT ĐỊNH**  
**Về việc thành lập Hội đồng đánh giá luận văn thạc sĩ**

**GIÁM ĐỐC**  
**HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ**

Căn cứ Quyết định số 303/QĐ-VHL ngày 01/3/2023 của Chủ tịch Viện Hàn lâm Khoa học và Công nghệ Việt Nam về việc ban hành Quy chế Tổ chức và hoạt động của Học viện Khoa học và Công nghệ;

Căn cứ Thông tư số 15/2014/TT-BGDĐT ngày 15/5/2014 của Bộ trưởng Bộ Giáo dục và Đào tạo ban hành Quy chế đào tạo trình độ thạc sĩ;

Căn cứ Quyết định số 775/QĐ-HVKHCN ngày 21/11/2016 của Giám đốc Học viện Khoa học và Công nghệ ban hành Quy chế đào tạo trình độ thạc sĩ;

Căn cứ Quyết định số 850/QĐ-HVKHCN ngày 31/05/2021 của Giám đốc Học viện Khoa học và Công nghệ về việc công nhận học viên cao học khóa 2021A - Đợt 1 năm 2021;

Căn cứ Quyết định số 1458/QĐ-HVKHCN ngày 21/09/2022 của Giám đốc Học viện Khoa học và Công nghệ về việc công nhận đề tài và người hướng dẫn luận văn thạc sĩ;

Xét đề nghị của Trưởng khoa Khoa Công nghệ thông tin và Viễn thông, Trưởng phòng Đào tạo.

**QUYẾT ĐỊNH:**

**Điều 1.** Thành lập Hội đồng đánh giá luận văn thạc sĩ cho học viên Nguyễn Hồng Việt với đề tài:

**Nghiên cứu xây dựng kiến trúc dữ liệu nghiên cứu khoa học và công nghệ**

Ngành: Hệ thống thông tin

Mã số: 8 48 01 04

Danh sách thành viên Hội đồng đánh giá luận văn kèm theo Quyết định này.

**Điều 2.** Hội đồng có trách nhiệm đánh giá luận văn thạc sĩ theo đúng quy chế hiện hành của Bộ Giáo dục và Đào tạo, Học viện Khoa học và Công nghệ. Quyết định có hiệu lực tối đa 60 ngày kể từ ngày ký và phải đảm bảo thời hạn đào tạo theo quy định của Học viện. Hội đồng tự giải thể sau khi hoàn thành nhiệm vụ.

**Điều 3.** Trưởng phòng Tổ chức - Hành chính và Truyền thông, Trưởng phòng Đào tạo, Trưởng phòng Kế toán, Trưởng Khoa Công nghệ thông tin và Viễn thông, các thành viên có tên trong danh sách Hội đồng và học viên cao học có tên tại Điều 1 chịu trách nhiệm thi hành Quyết định này. /s/

Nơi nhận:

- Như Điều 3;
- Lưu hồ sơ học viên;
- Lưu: VT, ĐT, PQ.12.

  
GIÁM ĐỐC  
  
GS.TS. Vũ Đình Lâm



**DANH SÁCH HỘI ĐỒNG ĐÁNH GIÁ LUẬN VĂN THẠC SĨ**

(Kèm theo Quyết định số 948/QĐ-HVKHCN ngày 18/08/2023 của Giám đốc Học viện Khoa học và Công nghệ)

Chợ luận văn của học viên: Nguyễn Hồng Việt

Tên đề tài: **Nghiên cứu xây dựng kiến trúc dữ liệu nghiên cứu khoa học và công nghệ**

Ngành: Hệ thống thông tin Mã số: 8 48 01 04

Người hướng dẫn: TS. Nguyễn Như Sơn, Viện Công nghệ thông tin,  
Viện Hàn lâm Khoa học và Công nghệ Việt Nam

TT	Họ và tên, học hàm, học vị	Ngành	Cơ quan công tác	Trách nhiệm trong Hội đồng
1.	PGS.TS. Nguyễn Long Giang	Hệ thống thông tin	Viện Công nghệ thông tin, Viện Hàn lâm KHCN VN	Chủ tịch
2.	PGS.TS. Trần Thị Ngân	Hệ thống thông tin	Trường Quốc tế, Đại học Quốc gia Hà Nội	Phản biện 1
3.	PGS.TS. Ngô Quốc Tạo	Hệ thống thông tin	Viện Công nghệ thông tin, Viện Hàn lâm KHCN VN	Phản biện 2
4.	TS. Trần Đức Nghĩa	Hệ thống thông tin	Viện Công nghệ thông tin, Viện Hàn lâm KHCN VN	Ủy viên - Thư ký
5.	TS. Nguyễn Mạnh Hùng	Hệ thống thông tin	Học viện Kỹ thuật Quân sự, Bộ Quốc phòng	Ủy viên

Hội đồng gồm 05 thành viên./.

## **BẢN NHẬN XÉT LUẬN VĂN THẠC SĨ**

Họ và tên người nhận xét: **Trần Thị Ngân** Học hàm, học vị: PGS. TS.

Chức danh trong Hội đồng: Phản biện 1

Cơ quan công tác: Trường Quốc tế, Đại học Quốc gia Hà Nội

Họ và tên học viên: **Nguyễn Hồng Việt**

Tên đề tài: **Nghiên cứu xây dựng kiến trúc dữ liệu nghiên cứu khoa học và Công nghệ**

Ngành: Hệ thống thông tin

Mã số: 8 48 01 04

### **NỘI DUNG NHẬN XÉT**

#### ***1. Tính cấp thiết, tính thời sự, ý nghĩa khoa học và thực tiễn của đề tài luận văn:***

Vấn đề quản lý cơ sở dữ liệu đã và đang nhận được nhiều quan tâm từ những nhà quản lý, nhà khoa học và có nhiều nghiên cứu được thực hiện. Đề tài luận văn nghiên cứu xây dựng kiến trúc dữ liệu NCKH có tính thời sự và có ý nghĩa thực tiễn tại đơn vị là Học viện Khoa học và công nghệ, Viện hàn lâm KH&CN Việt Nam.

#### ***2. Sự không trùng lặp của đề tài nghiên cứu; tính trung thực, rõ ràng và đầy đủ trong trích dẫn TLTK:***

Theo người nhận xét, đề tài nghiên cứu là không trùng lặp với các đề tài, bài báo đã công bố trước đây.

Tài liệu tham khảo được trích dẫn trong luận văn. Tuy nhiên, học viên cần rà soát các tài liệu cho đúng định dạng, cách trích dẫn và cách ghi thông tin tài liệu.

#### ***3. Sự phù hợp giữa tên đề tài với nội dung nghiên cứu cũng như với chuyên ngành và mã số đào tạo:***

Tên đề tài phù hợp với nội dung nghiên cứu được trình bày trong luận văn. Tên đề tài phù hợp với chuyên ngành và mã số đào tạo.

#### ***4. Độ tin cậy và tính hiện đại của phương pháp nghiên cứu đã sử dụng để hoàn thành luận văn:***

Phương pháp nghiên cứu sử dụng để hoàn thành luận văn là tin cậy và hiện đại.

#### ***5. Kết quả nghiên cứu của luận văn:***

Luận văn trình bày kết quả nghiên cứu xây dựng kiến trúc dữ liệu KHCN tại Học viện KH&CN, Viện Hàn lâm KH&CN Việt Nam. Đồng thời, mô hình được cài

đặt thử nghiệm để đánh giá tính khả thi, sự phù hợp của kiến trúc đề xuất với quy trình thực tế tại đơn vị.

**6. Những hạn chế, thiếu sót của luận văn về nội dung, hình thức và các câu hỏi**

Học viên cần rà soát lại luận văn và chỉnh sửa theo một số góp ý dưới đây:

- Chỉnh sửa lại cách diễn đạt trong luận văn để đảm bảo đây là một sản phẩm khoa học. Ngôn ngữ sử dụng phù hợp, cách diễn đạt cần rõ ràng, dễ hiểu hơn.
- Viết lại về Mục tiêu, Phạm vi nghiên cứu để thể hiện đúng các mục tiêu cần đạt được và phạm vi thực hiện của đề tài một cách ngắn gọn, rõ ràng.
- Học viên xem xét bố cục lại các mục trong luận văn để có được các nội dung logic và khoa học hơn.
- Luận văn cần mô tả rõ hơn về quá trình xây dựng kiến trúc dữ liệu. Giải thích cơ sở cần thiết lập các khối như trong hình 3.
- Sử dụng chữ in hoa cho đúng quy định.

**7. Nếu tác giả chưa viết bài báo khoa học thì nội dung của luận văn có thể được viết thành các bài báo để gửi đăng trên tạp chí khoa học, sách chuyên ngành hoặc tuyển tập công trình hội nghị khoa học cấp quốc gia, quốc tế hay không?**

Nội dung của luận văn có thể được viết thành một công trình khoa học nhưng cần có sự đầu tư nhiều hơn của học viên cho kết quả nghiên cứu.

**8. Kết luận chung**

- Luận văn về cơ bản đáp ứng các yêu cầu đối với một luận văn Thạc sĩ;
- Luận văn có thể đưa ra bảo vệ để Học viên nhận học vị Thạc sĩ;

Hà Nội, ngày ... tháng 08 năm 2023

**Người nhận xét**  
(Ký, ghi rõ họ tên)



**PGS. TS. Trần Thị Ngân**

**BẢN NHẬN XÉT LUẬN VĂN THẠC SĨ**

Họ và tên người nhận xét: **Ngô Quốc Tạo** Học hàm, học vị: PGS. TS.

Chức danh trong Hội đồng: **Phản biện 2**

Cơ quan công tác: Viện Công nghệ thông tin, Viện Hàn lâm KHCN VN

Họ và tên học viên: **Nguyễn Hồng Việt**

Tên đề tài: **Nghiên cứu xây dựng kiến trúc dữ liệu nghiên cứu khoa học và Công nghệ**

Ngành: **Hệ thống thông tin**

Mã số: **8 48 01 04**

**NỘI DUNG NHẬN XÉT**

**1. Tính cấp thiết, tính thời sự, ý nghĩa khoa học và thực tiễn của đề tài luận văn:**

**Thứ nhất:** Trong giai đoạn hiện nay, dữ liệu nghiên cứu trong khoa học và Công nghệ là một nguồn tài nguyên quý báu để thúc đẩy sự phát triển và đổi mới trong các lĩnh vực, dữ liệu nghiên cứu trong khoa học và công nghệ có ảnh hưởng lớn đến sự phát triển của xã hội và nền kinh tế. Nghiên cứu trong lĩnh vực này có thể mang lại giá trị thực sự cho cộng đồng khoa học và công nghệ.

**Thứ Hai:** Xuất phát từ thực tế tại Viện Hàn lâm Khoa học và Công nghệ Việt Nam (KHCNVN) là một tổ chức quốc gia chuyên về lĩnh vực khoa học và công nghệ tại Việt Nam. Chức năng nhiệm vụ chính của Viện Hàn lâm KHCNVN bao gồm:

- Nghiên cứu và phát triển khoa học công nghệ: Viện Hàn lâm KHCNVN thực hiện các dự án nghiên cứu khoa học và phát triển công nghệ, từ đó đóng góp vào sự phát triển của ngành khoa học và công nghệ ở Việt Nam.
- Phát triển nguồn nhân lực: Viện Hàn lâm KHCNVN đào tạo và phát triển nguồn nhân lực trong lĩnh vực khoa học và công nghệ.
- Tổ chức hội nghị, Hội thảo và sự kiện: Viện Hàn lâm KHCNVN thường xuyên tổ chức hội nghị, seminar, hội thảo và các sự kiện khác để tạo điều kiện cho việc trao đổi thông tin, kết nối các nhà nghiên cứu, và thúc đẩy sự phát triển trong lĩnh vực nghiên cứu khoa học và công nghệ.

Xuất bản và công bố: Viện Hàn lâm KHCNVN có thể xuất bản sách, bài báo, và tài liệu nghiên cứu để chia sẻ kiến thức và kết quả nghiên cứu với cộng đồng khoa học và công chúng.

**2. Sự không trùng lặp của đề tài nghiên cứu so với các công trình khoa học, luận văn đã công bố ở trong và ngoài nước; tính trung thực, rõ ràng và đầy đủ trong trích dẫn tài liệu tham khảo:**

**Mục tiêu nghiên cứu**

- Phân tích thực trạng, nghiên cứu xây dựng kiến trúc dữ liệu nghiên cứu khoa học và công nghệ.

- Phân tích thực trạng, Nghiên cứu xây dựng kiến trúc dữ liệu trong nghiên cứu KHCN.

- NCS đưa ra xây dựng mô hình thử nghiệm lưu trữ quản lý dữ liệu nghiên cứu Khoa học và Công nghệ tại Học viện KHCN.

3. Sự phù hợp giữa tên đề tài với nội dung nghiên cứu cũng như với chuyên ngành và mã số đào tạo:

- Chỉ ra được thực trạng và các nghiên cứu liên quan đến dữ liệu nghiên cứu KH&CN tại Việt Nam và một số nước trên thế giới

- Xây dựng ứng dụng phần mềm lưu trữ, xử lý dữ liệu nghiên cứu khoa học và Công nghệ tại Học viện KHCN, VHLKHCN VN.

**4. Độ tin cậy và tính hiện đại của phương pháp nghiên cứu đã sử dụng để hoàn thành luận văn:**

- Tài liệu: Nghiên cứu hệ thống văn bản pháp luật và các văn bản liên quan đến dữ liệu Nghiên cứu khoa học và Công nghệ, đồng thời nghiên cứu các báo cáo, giáo trình, v.v...

- Phân tích và tổng hợp: Trên cơ sở các nguồn tài liệu, dữ liệu, báo cáo về dự án, hội thảo, hội nghị.

- Thảo luận nhóm: Thảo luận với các học viên cùng khoá, các NCS của Khoa Công nghệ thông tin và các cán bộ đang công tác tại VHLKHCNVN.

**5. Kết quả nghiên cứu của luận văn:**

**a) Kết quả của luận văn**

- Luận văn đã trình bày tóm tắt các nghiên cứu liên quan đến dữ liệu nghiên cứu Khoa học ở phạm vi trong và ngoài nước.

- Xây dựng được hệ thống có khả năng thu thập dữ liệu từ nhiều nguồn, lưu trữ và hiển thị dữ liệu.

- Xây dựng ứng dụng và huấn luyện mô hình từ những dữ liệu đã thu thập được.

**b) Hướng nghiên cứu tiếp theo**

Trong tương lai, nghiên cứu này có thể thực hiện theo các hướng mới để cải thiện những hạn chế của hệ thống:

- Mở rộng khả năng tiếp nhận dữ liệu từ các nguồn khác ngoài 04 (bốn) nguồn dữ liệu phổ biến đã được nêu trong luận văn.

- Cải thiện không gian lưu trữ dữ liệu thông qua các dịch vụ lưu trữ dữ liệu.

- Mở rộng thêm các mô hình ứng dụng cho các nhà khoa học để tận dụng nguồn dữ liệu sẵn có.

6. Những hạn chế, thiếu sót của luận văn về nội dung, hình thức và các câu hỏi



**a) Hạn chế**

- a) Hệ thống hiện tại đang tập trung vào 4 nguồn dữ liệu phổ biến, chưa bao quát được nhiều nguồn dữ liệu khác.
- b) Không gian lưu trữ dữ liệu còn hạn chế.
- c) Phần ứng dụng xây dựng các mô hình từ dữ liệu thu thập được chưa đa dạng.
- d) Sai lỗi biên soạn “Nghiên cứu xây dựng kiến trúc dữ liệu nghiên cứu khoa học và công nghệ” và một số từ khác.
- e) Các Hình vẽ nên đánh số theo chương.
- f) Viết lại các kết luận của các chương và lại kết luận luận văn thể hiện sự đóng góp của học viên.
- g) Tại sao nói về Viện Hàn Lâm nhiều nhưng lại chỉ nghiên cứu dữ liệu của Học Viện KHCN?
- h) So sánh mô hình của Học Viện Hành Chính về quá trình hình thành và lưu trữ DLNC với của học viên?

7. Nội dung của luận văn chưa thể viết thành các bài báo để gửi đăng trên TCKH, sách chuyên ngành hoặc tuyển tập công trình hội nghị khoa học cấp quốc gia, quốc tế.

**8. Kết luận chung**

Luận văn của học viên đáp ứng các yêu cầu của luận văn thạc sĩ ngành: Hệ thống thông tin, mã số: 8 48 01 04. Luận văn có thể đưa ra bảo vệ để nhận học vị Thạc sĩ.

Cần chỉnh sửa luận văn theo ý kiến của người đọc.

Hà Nội, ngày 23 tháng 08 năm 2023

**Người nhận xét**



**PGS. TS. Ngô Quốc Tạo**

Hà Nội, ngày 25 tháng 8 năm 2023

### BIÊN BẢN HỌP HỘI ĐỒNG ĐÁNH GIÁ LUẬN VĂN THẠC SĨ

Thực hiện Quyết định số 948/QĐ-HVKHCN ngày 18/08/2023 của Giám đốc Học viện Khoa học và Công nghệ về việc thành lập Hội đồng đánh giá luận văn thạc sĩ của học viên Nguyễn Hồng Việt.

Tên đề tài: Nghiên cứu xây dựng kiến trúc dữ liệu nghiên cứu khoa học và công nghệ

Ngành: Hệ thống thông tin

Mã số: 8 48 01 04

Hôm nay, ngày 25/8/2023 Hội đồng đã họp tại Phòng 1705, Học viện Khoa học và Công nghệ vào lúc 14h00, Hội đồng gồm 05 thành viên:

- |                              |                   |
|------------------------------|-------------------|
| 1. PGS.TS. Nguyễn Long Giang | Chủ tịch hội đồng |
| 2. TS. Trần Đức Nghĩa        | Thư ký - Ủy viên  |
| 3. PGS.TS. Trần Thị Ngân     | Phản biện 1       |
| 4. PGS.TS. Ngô Quốc Tạo      | Phản biện 2       |
| 5. TS. Nguyễn Mạnh Hùng      | Ủy viên           |

Thành viên vắng mặt: ... không có ... (Phản biện hoặc ủy viên, đã có bản nhận xét đồng ý cho phép học viên được bảo vệ trước Hội đồng đánh giá luận văn thạc sĩ).

### NỘI DUNG LÀM VIỆC

- Đại diện cơ sở đào tạo đọc quyết định thành lập Hội đồng đánh giá luận văn
- Chủ tịch Hội đồng, điều khiển phiên họp
- Thư ký HĐ, đọc lí lịch khoa học và bảng điểm của học viên
- Học viên trình bày luận văn trước Hội đồng
- Phản biện 1:

..... luận văn của học viên đã được đánh giá đạt yêu cầu để bảo vệ luận văn.  
..... Cần chú ý bổ sung phần quá trình xây dựng luận văn trong quá trình bảo vệ.  
..... luận văn của học viên đã được đánh giá đạt yêu cầu để bảo vệ luận văn.  
..... Cần chú ý bổ sung phần quá trình xây dựng luận văn trong quá trình bảo vệ.



Tổng số điểm: 40

Điểm trung bình: 8

Điểm thưởng công trình công bố: 0

Tổng điểm đánh giá luận văn và thưởng công trình công bố: 8

- Kết luận của Hội đồng:

+ Luận văn ... đạt yêu cầu ... (đạt/không đạt yêu cầu)

+ Tính không trùng lặp nội dung và tên đề tài với các công trình công bố:

... luận văn không trùng lặp với công trình công bố đã đăng tải.

11. Chủ tịch Hội đồng, công bố kết quả, yêu cầu học viên chỉnh sửa luận văn với các nội dung sau:

... chỉnh sửa nội dung luận văn theo các ý kiến của hội đồng.  
... Mã tài liệu luận văn chỉnh sửa đưa vào hệ thống.  
... Luận văn đạt yêu cầu.

SAO Y BẢN CHÍNH

Hà Nội, ngày tháng năm

TL. GIÁM ĐỐC

TRƯỞNG PHÒNG TỔ CHỨC HÀNH CHÍNH VÀ TRUYỀN THÔNG

Đỗ Thị Liên Thanh



Buổi họp đã kết thúc vào 16 giờ 30 phút ngày 25/8/2023

Hà Nội, ngày 25 tháng 8 năm 2023

THƯ KÝ HỘI ĐỒNG

*Handwritten signature of Trần Đức Nghĩa*

TS. Trần Đức Nghĩa

CHỦ TỊCH HỘI ĐỒNG

*Handwritten signature of Nguyễn Long Giang*

PGS.TS. Nguyễn Long Giang

XÁC NHẬN CỦA HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ  
KT. GIÁM ĐỐC  
PHÓ GIÁM ĐỐC



*Handwritten signature of Nguyễn Thị Trung*  
Nguyễn Thị Trung

**BẢN GIẢI TRÌNH CHỈNH SỬA LUẬN VĂN  
THEO KẾT LUẬN CỦA HỘI ĐỒNG ĐÁNH GIÁ LUẬN VĂN THẠC SĨ**

Họ tên học viên: Nguyễn Hồng Việt

Lớp: ITT21A01

Tên đề tài luận văn: "Nghiên cứu xây dựng kiến trúc dữ liệu nghiên cứu khoa học và công nghệ".

Chuyên ngành: Hệ thống thông tin

Mã số: 8480104

Người hướng dẫn khoa học: TS. Nguyễn Như Sơn

Ngày bảo vệ luận văn: 26/08/2023

Căn cứ biên Bản họp Hội đồng đánh giá luận văn thạc sĩ, học viên đã chỉnh sửa luận văn như sau:

STT	Nội dung đề nghị bổ sung, chỉnh sửa	Nội dung đã bổ sung, chỉnh sửa
1.	Luận văn cần được chỉnh sửa có tính khoa học hơn Cần mô tả rõ hơn quá trình xây dựng chức năng, ứng dụng	Tiếp thu chỉnh sửa : + Trình bày lại toàn bộ luận văn 1 cách khoa học, chặt chẽ, logic, ngôn ngữ sử dụng phù hợp và dễ hiểu hơn +Đã bổ sung và thiết lập lại các khối và mô tả rõ hơn về quá trình xây dựng kiến trúc dữ liệu tại hình 3
2.	Viết lại về Mục tiêu, Phạm vi nghiên cứu để thể hiện đúng các mục tiêu cần đạt được và phạm vi thực hiện của đề tài một cách ngắn gọn, rõ ràng.	- Luận văn đã viết lại về Mục tiêu, Phạm vi nghiên cứu theo đúng các mục tiêu cần đạt được tại Mục 1.2 Phần Mở đầu (trang 11). - Đã bổ sung nội dung Phạm vi nghiên cứu 3 tại Mục 1.3 Phần Mở đầu (trang 12).
3.	Sử dụng chữ in hoa cho đúng quy định	Tiếp thu chỉnh sửa. Đã rà soát và chỉnh sửa chữ in hoa đúng quy định



4.	<ul style="list-style-type: none"> <li>- Tại trang 8 sai về số lượng đơn vị tại Viện Hàn lâm KHCNVN</li> <li>- Tại Trang 27 Sai về số lượng chuyên ngành tại Học viện KHCN</li> <li>- Các Hình vẽ nên đánh theo số chương</li> </ul>	<ul style="list-style-type: none"> <li>- Tiếp thu chỉnh sửa cụ thể:</li> <li>+ Đơn vị trực thuộc Viện Hàn lâm KHCNVN gồm 28 Viện , 06 Đơn vị giúp việc cho Chủ tịch, 08 đơn vị sự nghiệp công lập và 01 Công ty TNHH</li> <li>+ Học viện KHCN hiện nay có 12 chuyên ngành đào tạo.</li> <li>+ Các Hình vẽ đã được chỉnh sửa đánh số theo Chương</li> </ul>
5.	Viết lại các kết luận của các chương và kết luận luận văn thể hiện sự đóng góp của học viên	Đã chỉnh sửa, bổ sung các kết luận của các chương để thể hiện sự đóng góp của học viên
6.	Ứng dụng bị đan xen Tiếng Anh và Tiếng Việt	Đã chỉnh sửa ứng dụng hệ thống là các mục là Tiếng Anh

\*Lưu ý: Trong trường hợp Hội đồng yêu cầu xin ý kiến của 02 phân biện sau báo vệ, học viên cần xin chữ ký của 02 phân biện xác nhận.

Hà Nội, ngày 31 tháng 08 năm 2023

CHỦ TỊCH HỘI ĐỒNG

TẬP THỂ HƯỚNG DẪN

HỌC VIÊN





PGS.TS. Nguyễn Long Giang

TS. Nguyễn Như Sơn

Nguyễn Hồng Việt

XÁC NHẬN CỦA HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

KT. GIÁM ĐỐC  
PHÓ GIÁM ĐỐC



Nguyễn Thị Trung