

Dữ liệu nghiên cứu mở và các nguyên tắc dữ liệu FAIR

Open Research Data and FAIR data principles

Tóm tắt: Dữ liệu nghiên cứu mở và các nguyên tắc dữ liệu FAIR đã được nhắc tới trong Khuyến nghị Khoa học Mở của UNESCO như là tiêu chuẩn trong quản lý tốt dữ liệu nghiên cứu. Bài viết nêu lên các định nghĩa cơ bản các khái niệm và gợi ý cách hiểu đúng các khái niệm đó và từng bước triển khai chúng vào thực tế cuộc sống.

Từ khóa: Dữ liệu nghiên cứu mở; các nguyên tắc dữ liệu FAIR; tìm thấy được, truy cập được, tương hợp được, sử dụng lại được (FAIR).

Summary: Open research data and the FAIR data principles have been mentioned in the UNESCO Open Science Recommendation as standards for good management of research data. The article gives basic definitions of concepts and suggests how to properly understand those concepts and step by step implement them into real life.

Keywords: Open Research data; FAIR data principles; findable, accessible, interoperable, reusable (FAIR)

1. Đặt vấn đề

Ngày 23/11/2021, tại phiên thứ 41 Hội nghị Toàn thể UNESCO, 193 quốc gia thành viên đã thông qua Khuyến nghị Khoa học Mở[1], biến Khoa học Mở thành một xu thế không thể đảo ngược của thế giới ngày nay. Như một công cụ thiết lập tiêu chuẩn quốc tế đầu tiên về khoa học mở, “Khuyến nghị này đưa ra định nghĩa chung, các giá trị, các nguyên tắc và các tiêu chuẩn được chia sẻ cho khoa học mở ở mức quốc tế và đề xuất một tập hợp các hành động có lợi cho việc vận hành khoa học mở một cách công bằng và bình đẳng cho tất cả mọi người ở các mức cá nhân, cơ sở, quốc gia, khu vực và quốc tế” (Xem mục 2 của Phần I. Mục tiêu và các mục đích của Khuyến nghị).

Trong số các nguyên tắc và tiêu chuẩn được chia sẻ cho khoa học mở ở mức quốc tế được nhắc tới ở trên, có các nguyên tắc dữ liệu FAIR dành cho việc quản lý tốt các dữ liệu nghiên cứu mở.

2. Dữ liệu nghiên cứu mở là gì?

Dữ liệu nghiên cứu mở là một trong năm thành phần của **kiến thức khoa học mở**. Vì vậy, định nghĩa kiến thức khoa học mở sẽ được nêu trước định nghĩa dữ liệu nghiên cứu mở, như bên dưới đây (xem mục 7 của Phần II. Định nghĩa Khoa học Mở):

“**Kiến thức khoa học mở** tham chiếu tới việc **truy cập mở** tới các xuất bản phẩm khoa học, **dữ liệu nghiên cứu, siêu dữ liệu**, tài nguyên giáo dục mở, phần mềm và mã nguồn và phần cứng mà chúng sẵn sàng **trong phạm vi công**

cộng hoặc có bản quyền và được cấp phép theo một giấy phép mở cho phép truy cập, sử dụng lại, tái mục đích, tùy chỉnh và phân phối theo các điều kiện nhất định, được cung cấp cho tất cả các tác nhân ngay lập tức hoặc nhanh nhất có thể,... và miễn phí.”

Định nghĩa **dữ liệu nghiên cứu mở** được nêu như sau (Xem mục 7b của Phần II. Định nghĩa Khoa học Mở):

“**Dữ liệu nghiên cứu mở** bao gồm, trong số những điều khác, dữ liệu dạng số và tuân tự, cả thô và đã được xử lý, và siêu dữ liệu kèm theo, cũng như các điểm số, các hồ sơ văn bản, các hình ảnh và âm thanh, các giao thức, mã phân tích và tiến trình có thể được mở cho bất kỳ ai để sử dụng, sử dụng lại, giữ lại và phân phối lại, **tuân thủ thừa nhận ghi công**. Dữ liệu nghiên cứu mở là sẵn sàng ở định dạng kịp thời và thân thiện với người sử dụng, người và máy đọc được và hành động được, phù hợp với các nguyên tắc của điều hành và quản trị dữ liệu tốt, ấy là các nguyên tắc Tìm thấy được, Truy cập được, Tương hợp được, Sử dụng lại được – FAIR (Findable, Accessible, Interoperable, Reusable), được sự giám tuyển và duy trì thường xuyên hỗ trợ.”

Từ định nghĩa các khái niệm ở trên, có thể đưa ra vài nhận xét ban đầu như sau:

1. Dữ liệu nghiên cứu mở là một thành phần của kiến thức khoa học mở. Vì vậy, nó mang theo đầy đủ các đặc tính được nêu trong định nghĩa của kiến thức khoa học mở. Cụ thể, dữ liệu nghiên cứu mở:



Hình 1: Dữ liệu nghiên cứu mở - 1 trong 5 thành phần của kiến thức khoa học mở

a) sẵn sàng trong phạm vi công cộng

- phạm vi công cộng là nơi các tài nguyên/dữ liệu (gọi chung là **tác phẩm**) hoặc đã hết thời hạn bảo hộ của luật, chẳng hạn như Luật sở hữu trí tuệ, hoặc tác giả/người nắm giữ các quyền của tác phẩm đó khước từ các quyền của mình và hiến tặng tác phẩm đó vào phạm vi công cộng.
- theo Luật sở hữu trí tuệ của Việt Nam, Điều 27. Thời hạn bảo hộ quyền tác giả[2], tác phẩm của một tác giả hoặc đồng tác giả sẽ **tự động nằm trong phạm vi công cộng** sau khi tác giả hoặc đồng tác giả cuối cùng chết được 50 năm; nếu giả thiết tuổi thọ trung bình của người Việt Nam là 70 tuổi, thì hầu như tất cả các tác phẩm do các tác giả Việt Nam sáng tạo ra từ năm 1900 trở về trước trong suốt chiều dài vài ngàn năm lịch sử của Việt Nam hiện đang nằm trong phạm vi công cộng.
- một khi tác phẩm nằm trong phạm vi công cộng, thì không chỉ 100 triệu người dân Việt Nam, mà cả hơn 8 tỷ người trên thế giới có quyền truy cập **miễn phí** tới các tác phẩm đó, bất kể nó ở dạng (kỹ thuật) số hay không; nếu tác phẩm nằm trong phạm vi công cộng ở dạng (kỹ thuật) số, bất kỳ ai cũng có thể sử dụng nó tùy theo ý của mình, miễn là nên có một **đường liên kết ngược về địa chỉ web của tác phẩm** (kỹ thuật) số đó.
- kinh nghiệm thực tế trên thế giới cho thấy, cụ thể là tại các quốc gia thuộc Liên minh châu Âu[3]: *“Số hóa nội dung thuộc phạm vi công cộng không tạo ra các quyền mới đối với nó: các tác phẩm nằm trong phạm vi công cộng ở dạng tuần tự (analogue form) tiếp tục nằm trong phạm vi công cộng một khi chúng được số hóa.”*

hoặc:

b) **có bản quyền** và **được cấp phép theo một giấy phép mở** cho phép truy cập, sử dụng lại, tái mục đích, tùy chỉnh và phân phối theo các điều kiện nhất định, được cung cấp cho **tất cả các tác nhân** ngay lập tức hoặc nhanh nhất có thể,... và **miễn phí**.

- **khẳng định**: dữ liệu nghiên cứu mở là **có bản quyền** một khi nó được **cấp phép mở**
- giấy phép mở được gắn vào trong dữ liệu nghiên cứu mở phải đảm bảo cho phép **bất kỳ ai** cũng có thể **miễn phí** truy cập, sử dụng lại, tái mục đích, **tùy chỉnh** và phân phối nó theo các điều kiện nhất định, **ngay lập tức hoặc nhanh nhất có thể**;
- cụm từ ‘**tùy chỉnh**’ (**Adapt**) nhấn mạnh rằng dữ liệu nghiên cứu mở phải là dữ liệu, khi được **cấp phép mở**, ví dụ như các giấy phép CC BY[4]

hoặc CC BY-SA[5], mà bất kỳ ai cũng có quyền pha trộn, biến đổi, và xây dựng dựa vào nó vì bất kỳ mục đích gì, kể cả thương mại hóa.

và

2. Dữ liệu nghiên cứu mở bao gồm cả dữ liệu số và tuân tự, dữ liệu thô và đã qua xử lý, cùng siêu dữ liệu kèm theo dữ liệu đó, mà:

- a) người sử dụng **phải tuân thủ thừa nhận ghi công** cho tác giả và/hoặc người nắm giữ bản quyền của dữ liệu nghiên cứu mở; điều này yêu cầu người sử dụng phải tuân thủ các điều khoản và điều kiện của giấy phép mở được gắn vào dữ liệu nghiên cứu mở nhằm tôn trọng các quyền bản quyền của (các) tác giả và/hoặc người nắm giữ các quyền bản quyền của dữ liệu nghiên cứu mở đó; việc không tuân thủ thừa nhận ghi công sẽ được hiểu là người sử dụng biến dữ liệu nghiên cứu mở của người khác thành dữ liệu của mình và vì vậy được coi như là hành động “ăn cắp”.


Three “Layers” Of Licenses



Hình 2: Ba “lớp” các giấy phép Creative Commons: (1) Legal Code - Mã pháp lý; (2) Human Readable - Người đọc được; (3) Machine Readable - Máy đọc được

- b) không phải chỉ người đọc được, mà **máy cũng phải đọc được**; một trong những yêu cầu để máy đọc được là bản thân giấy phép được gắn vào dữ liệu nghiên cứu mở cũng phải thỏa mãn yêu cầu này. Hệ thống các giấy phép Creative Commons có 3 lớp giấy phép: (1) Mã pháp lý (Legal Code) dành cho các luật sư; (2) Chứng thư Chung (Commons Deed) hay còn gọi là lớp giấy phép “Người đọc được” (Human Readable); và (3) Máy đọc được (Machine Readable) bằng việc sử dụng Ngôn ngữ Thể hiện Quyền của Creative Commons - CC REL[6] (Creative Commons Rights Expression Language); điều này giúp cho các giấy phép Creative Commons được sử

dụng rất phổ biến cho các tài nguyên (kỹ thuật) số, bao gồm dữ liệu nghiên cứu mở, được cấp phép mở; nói cách khác, nhờ có CC REL, máy hiểu được các quyền và bốn phần của giấy phép được gắn vào các tài nguyên số; và ngược lại, máy sẽ không thể hiểu được các ngôn ngữ thông thường có trong bất kỳ văn bản quy phạm pháp quy nào, dạng như: “*Dữ liệu mở của cơ quan nhà nước: là dữ liệu được cơ quan nhà nước có thẩm quyền công bố rộng rãi cho cơ quan, tổ chức, cá nhân tự do sử dụng, tái sử dụng, chia sẻ*”[7]; trong thực tế, việc sử dụng trình chọn giấy phép Creative Commons (CC License Chooser[8]) để gắn giấy phép CC cho tài nguyên số sẽ cho phép người sử dụng chọn lựa chọn ‘người đọc được’ và/hoặc ‘máy đọc được’.

 <p>This work is licensed under a Creative Commons Attribution 4.0 International License.</p>	<pre>
This work is licensed under a Creative Commons Attribution 4.0 International License.</pre>
<p>Cấp phép mở CC Người đọc được!</p>	<p>Cấp phép mở CC theo Ngôn ngữ Thể hiện Quyền - REL (Rights Expression Language): Máy đọc được!</p>

Hình 3: So sánh việc cấp phép mở CC để ‘Người đọc được’ và ‘Máy đọc được’

- c) cuối cùng, rất quan trọng: dữ liệu nghiên cứu mở phải là Tìm thấy được, Truy cập được, Tương hợp được, Sử dụng lại được - FAIR (Findable, Accessible, Interoperable, Reusable); ngắn gọn, dữ liệu nghiên cứu mở phải tuân thủ các nguyên tắc quản lý dữ liệu FAIR.

3. FAIR là gì?

Như được nêu ở trên, FAIR là các ký tự đầu của các từ tiếng Anh gồm Findable, Accessible, Interoperable, Reusable, có nghĩa là Tìm thấy được, Truy cập được, Tương hợp được, và Sử dụng lại được.

Theo FORCE11[9], một cộng đồng gồm các học giả, thủ thư, nhân viên lưu trữ, nhà xuất bản và nhà cấp vốn nghiên cứu được thành lập để giúp tạo điều kiện cho sự thay đổi hướng tới cải thiện việc tạo lập và chia sẻ kiến thức, biến đổi truyền thông học thuật truyền thống trở nên hiện đại, phù hợp với kỷ nguyên số thông qua việc sử dụng hiệu quả công nghệ thông tin, đã đưa ra chỉ dẫn các nguyên tắc xuất bản dữ liệu theo FAIR như sau[10]:

“Một trong những thách thức của khoa học tăng cường dữ liệu là tạo thuận lợi để phát hiện tri thức bằng việc **hỗ trợ cho con người và máy** nhằm phát hiện,

truy cập tới, tích hợp và phân tích, thực hiện các tác vụ đúng thích hợp đối với dữ liệu khoa học và liên kết tới các thuật toán và tiến trình công việc của chúng. Vì vậy dữ liệu, bao gồm cả siêu dữ liệu khoa học, cần phải là *Tìm thấy được, Truy cập được, Tương hợp được và Sử dụng lại được*”.

Để tìm thấy được (Findable):

- F1: (siêu) dữ liệu chỉ định một mã thường trực toàn cầu duy nhất và vĩnh viễn
- F2: dữ liệu được mô tả với siêu dữ liệu phong phú (rich metadata).
- F3: (siêu) dữ liệu được đăng ký hoặc được đánh chỉ mục trong tài nguyên có khả năng tìm kiếm được.
- F4: siêu dữ liệu chỉ định mã dữ liệu.

Để Truy cập được (Accessible):

- A1: (siêu) dữ liệu có khả năng truy xuất được bằng mã của chúng khi sử dụng giao thức truyền thông được tiêu chuẩn hóa.
- A1.1: giao thức đó là mở, tự do, và triển khai được vạn năng.
- A1.2: giao thức đó cho phép thủ tục xác thực và ủy quyền, ở các nơi cần thiết.
- A2: siêu dữ liệu là truy cập được, thậm chí khi dữ liệu không còn tồn tại nữa.

Để có khả năng tương hợp được (Interoperable):

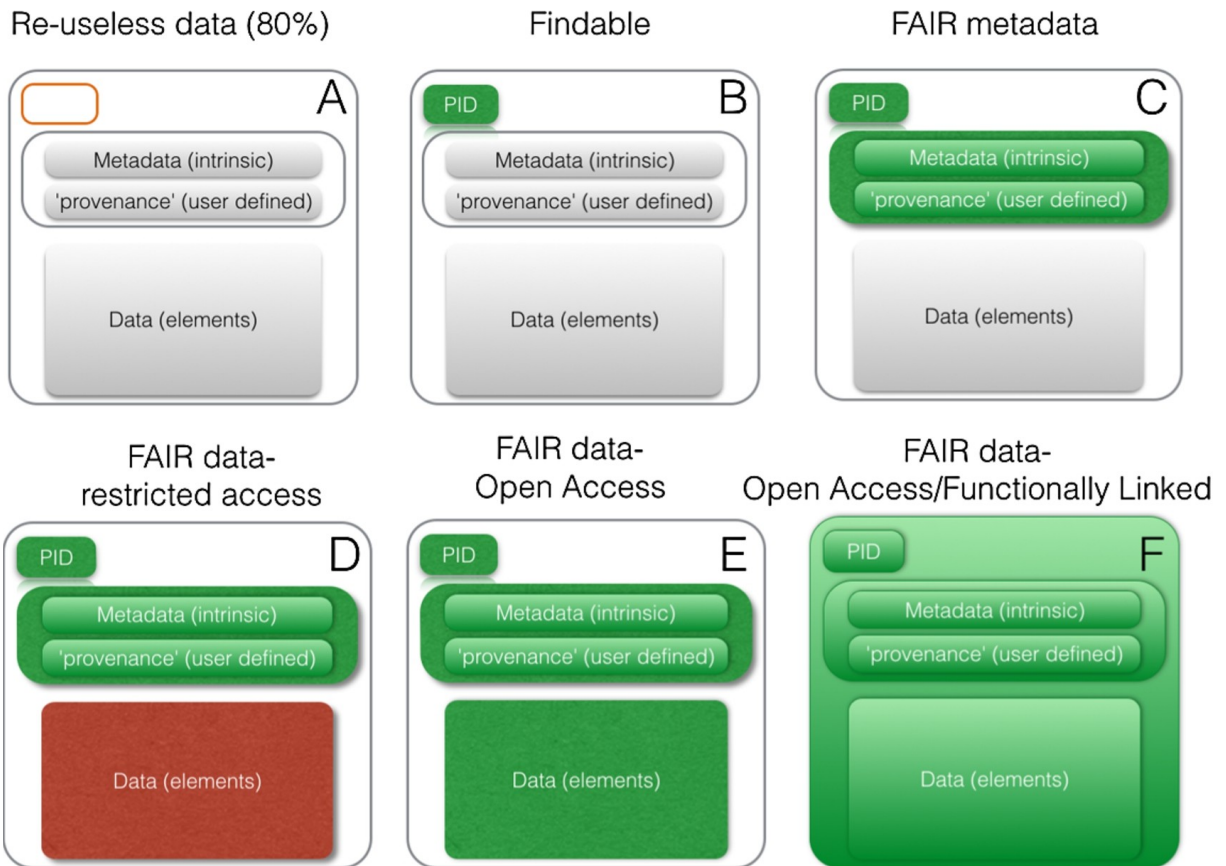
- I1: (siêu) dữ liệu sử dụng ngôn ngữ chính thống, truy cập được, được chia sẻ và áp dụng được rộng rãi để trình bày tri thức.
- I2: (siêu) dữ liệu sử dụng các từ vựng tuân theo các nguyên tắc FAIR.
- I3. (siêu) dữ liệu gồm các tham chiếu đủ điều kiện tới (siêu) dữ liệu khác.

Để sử dụng lại được (Re-usable):

- R1: (siêu) dữ liệu có nhiều thuộc tính chính xác và thích hợp.
- R1.1: (siêu) dữ liệu được phát hành với giấy phép sử dụng dữ liệu rõ ràng và truy cập được.
- R1.2: (siêu) dữ liệu có liên kết tới nguồn gốc xuất xứ của chúng.
- R1.3: (siêu) dữ liệu đáp ứng được các tiêu chuẩn cộng đồng phù hợp lĩnh vực.

Không chỉ dừng lại ở việc liệt kê các thành tố của từng trong số các nguyên tắc FAIR, FORCE11 còn đưa ra *‘Chỉ dẫn các nguyên tắc xuất bản dữ liệu để Tìm thấy được, Truy cập được, Tương hợp được và Sử dụng lại được phiên bản b1.0’* [11], chi tiết hóa rất nhiều các công việc cần phải làm để (siêu) dữ liệu khoa học đáp ứng được FAIR, để nhận thấy nhất trong số đó là việc sử dụng các mã nhận diện thường trực - PID (Persistent IDentifier). Lý do chính cho việc sử dụng PID ngày càng gia tăng vì không có PID, dữ liệu hầu như không sử dụng lại được, điều hoàn toàn trái ngược với FAIR.

Data as increasingly FAIR Digital Objects

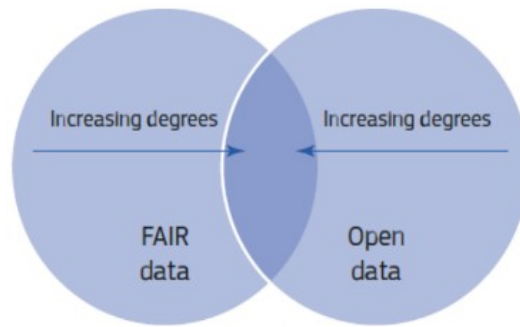


Hình 4: Dữ liệu như các đối tượng số FAIR ngày càng gia tăng

Điều có thể sơ bộ rút ra được từ các nguyên tắc FAIR và chỉ dẫn của FORCE11 về các nguyên tắc đó cho thấy đề (siêu) dữ liệu khoa học và nghiên cứu cùng với siêu dữ liệu đi cùng với nó đạt được FAIR là rất không dễ.

4. Môi quan hệ giữa Dữ liệu FAIR và Dữ liệu (nghiên cứu) Mở

Như ở phần trên đã nêu, tài liệu Khuyến nghị Khoa học Mở của UNESCO 2021 định nghĩa ‘Dữ liệu nghiên cứu mở’ là phù hợp với FAIR, mặc dù trong thực tế “Dữ liệu có thể là FAIR hoặc Mở, vừa là FAIR vừa là Mở, hoặc không là FAIR không là Mở. Các lợi ích lớn nhất khi dữ liệu vừa là FAIR vừa là Mở, vì khi không có các hạn chế sẽ hỗ trợ được rộng lớn nhất có thể cho sử dụng lại, và sử dụng lại ở phạm vi rộng. Để tối đa hóa những lợi ích của việc biến dữ liệu FAIR thành hiện thực, và **trong ngữ cảnh của các sáng kiến Khoa học Mở**, các nguyên tắc FAIR nên được triển khai kết hợp với yêu cầu chính sách rằng **dữ liệu nghiên cứu nên là Mở mặc định** - đó là, Mở trừ phi có lý do tốt để hạn chế truy cập hoặc sử dụng lại”[12]. Nói một cách khác, mức độ FAIR của dữ liệu càng gia tăng, thì mức độ tính mở sẽ càng cao, và ngược lại, như được minh họa trên **Hình 5**.



Hình 5: Mối quan hệ giữa dữ liệu FAIR và dữ liệu Mở

Tài liệu ‘Chi phí của việc không có dữ liệu nghiên cứu FAIR - Phân tích chi phí - lợi ích đối với dữ liệu nghiên cứu FAIR’ [13] có đoạn: “**Dữ liệu sẽ không thực sự sử dụng lại được** trừ phi nó là **MỞ**, nghĩa là, sẵn sàng theo một **giấy phép mở** và với các chi phí tượng trưng (trong hầu hết các trường hợp chi phí bằng 0), và **tính mở thường đi cùng với sự triển khai các nguyên tắc FAIR**”. Lưu ý này là hoàn toàn phù hợp với định nghĩa ‘Dữ liệu nghiên cứu mở’ của UNESCO được nêu ở trên.

5. Kết luận và gợi ý

Để dữ liệu nghiên cứu mở tuân thủ các nguyên tắc dữ liệu FAIR, dữ liệu đó vừa phải tuân thủ với định nghĩa dữ liệu nghiên cứu mở như được nêu trong Khuyến nghị Khoa học Mở của UNESCO, cùng lúc với các nguyên lý dữ liệu FAIR. Rõ ràng, điều này là rất không dễ thực hiện trong ngắn hạn đối với bất kỳ nhà nghiên cứu khoa học nào.

Để có thể có được dữ liệu nghiên cứu mở, trước tiên, nhà nghiên cứu khoa học cần có năng lực cấp phép mở đúng cho dữ liệu nghiên cứu của mình, có thể và thường bằng việc sử dụng công cụ chọn giấy phép của CC (CC License Chooser), cũng như có năng lực để làm việc với dữ liệu nghiên cứu mở của các nhà nghiên cứu khác. Để làm được những việc đó, tới lượt nó, các nhà nghiên cứu khoa học cần có năng lực để phân biệt các dữ liệu nghiên cứu mở với dữ liệu nghiên cứu không mở, cũng như phân biệt các dữ liệu nghiên cứu mở nằm trong phạm vi công cộng với các dữ liệu được cấp phép mở đúng; các năng lực để tìm kiếm, sử dụng/sử dụng lại, tạo lập (bao gồm việc sửa lại, pha trộn, cùng tạo lập) và chia sẻ dữ liệu nghiên cứu mở. Các năng lực được nêu ở đây cũng đều nằm trong khung năng lực tài nguyên giáo dục mở (TNGDM).

Một trong vô số các công việc để có thể có dữ liệu nghiên cứu mở tuân thủ với các nguyên tắc dữ liệu FAIR một nhà nghiên cứu khoa học có thể bắt đầu là có khả năng làm việc với và chỉ định các mã nhận diện thường trực - PID (Persistent Identifier) duy nhất, ví dụ như mã nhận diện đối tượng số - DOI (Digital Object Identifier) cho dữ liệu, siêu dữ liệu, hay nhiều dạng kết quả nghiên cứu khác như được minh họa trên **Hình 6**; hay mã nhận diện nhà nghiên cứu và người đóng góp mở - ORCID (Open Researcher and Contributor Identifier). Các công việc này đều nằm trong các kỹ năng khoa học mở cốt lõi, được cơ quan hàng không vũ trụ Mỹ NASA gợi ý cho các nhà

ngiên cứu bắt đầu với khoa học mở[14], và rộng hơn, là nằm trong đề xuất của Tim Berner - Lee về web dữ liệu với dữ liệu liên kết (Linked Data)[15] để thay thế cho web tài liệu siêu văn bản với việc sử dụng mã nhận diện tài nguyên thống nhất – URI (Uniform Resource Identifier) thay cho bộ định vị tài nguyên thống nhất – URL (Uniform Resource Locator), một cách tương ứng.

Type	Type
<input type="checkbox"/> Publication (1816828) - <ul style="list-style-type: none"> <input type="checkbox"/> Article (1152581) <input type="checkbox"/> Taxonomictreatment (471803) <input type="checkbox"/> Conferencepaper (81474) <input type="checkbox"/> Report (20274) <input type="checkbox"/> Other (18921) <input type="checkbox"/> Book (14820) <input type="checkbox"/> Deliverable (14422) <input type="checkbox"/> Section (13492) <input type="checkbox"/> Thesis (9008) <input type="checkbox"/> Preprint (7769) <input type="checkbox"/> Workingpaper (5931) <input type="checkbox"/> Technicalnote (2825) <input type="checkbox"/> Proposal (784) <input type="checkbox"/> Softwaredocumentation (657) <input type="checkbox"/> Milestone (581) <input type="checkbox"/> Datamanagementplan (471) <input type="checkbox"/> Peerreview (459) <input type="checkbox"/> Patent (301) <input type="checkbox"/> Annotationcollection (255) 	<input type="checkbox"/> Image (808195) - <ul style="list-style-type: none"> <input type="checkbox"/> Figure (520873) <input type="checkbox"/> Photo (280895) <input type="checkbox"/> Drawing (4638) <input type="checkbox"/> Other (1176) <input type="checkbox"/> Plot (308) <input type="checkbox"/> Diagram (305) <input type="checkbox"/> Dataset (200178) <input type="checkbox"/> Software (106041) <input type="checkbox"/> Presentation (41341) <input type="checkbox"/> Other (20401) <input type="checkbox"/> Poster (17202) <input type="checkbox"/> Video (8283) <input type="checkbox"/> Lesson (4825) <input type="checkbox"/> Physicalobject (1127)

Hình 6: Các dạng dữ liệu có thể được gắn mã nhận diện đối tượng số (DOI)[16]

Tài liệu tham khảo

- [1] UNESCO, 23/11/2021: *Recommendation on Open Science*: <https://unesdoc.unesco.org/ark:/48223/pf0000379949>. Bản dịch sang tiếng Việt: https://www.dropbox.com/s/l3q04t99nil5mgo/379949eng_Vi-25112021.pdf?dl=0
- [2] Cổng thông tin điện tử Quốc hội: *VĂN BẢN HỢP NHẤT SỐ 11/VBHN-VPQH LUẬT SỞ HỮU TRÍ TUỆ*: <https://quochoi.vn/tintuc/Pages/tin-hoat-dong-cua-quoc-hoi.aspx?ItemID=72721>
- [3] Lê Trung Nghĩa (2023): *Chuyển đổi số và vấn đề bản quyền trong hoạt động thư viện - kinh nghiệm của Europeana*: <https://giaoducmo.avnuc.vn/bai-viet-toan-van/chuyen-doi-so-va-van-de-ban-quyen-trong-hoat-dong-thu-vien-kinh-nghiem-cua-europeana-974.html>
- [4] Creative Commons: *CC BY 4.0*: <https://creativecommons.org/licenses/by/4.0/>
- [5] Creative Commons: *CC BY-SA 4.0*: <https://creativecommons.org/licenses/by-sa/4.0/>
- [6] CC Wiki: *CC REL*: https://wiki.creativecommons.org/wiki/CC_REL
- [7] Trang tin điện tử Chính phủ: *Nghị định số 47/2020/NĐ-CP của Chính phủ: Quản lý, kết nối và chia sẻ dữ liệu số của cơ quan nhà nước*: <https://vanban.chinhphu.vn/?pageid=27160&docid=199754>
- [8] Creative Commons: *License Chooser*: <https://creativecommons.org/choose/>

- [9] [2] FORCE11: *About FORCE11*: <https://force11.org/info/about-force11/>
- [10] [3] FORCE11: *The FAIR Data Principles*:
<https://www.force11.org/group/fairgroup/fairprinciples>. Bản dịch sang tiếng Việt:
<https://vnfoss.blogspot.com/2017/07/cac-nguyen-tac-du-lieu-fair.html>
- [11] FORCE11: *Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1.0*: <https://force11.org/info/guiding-principles-for-findable-accessible-interoperable-and-re-usable-data-publishing-version-b1-0/>. Bản dịch sang tiếng Việt: <https://vnfoss.blogspot.com/2017/07/chi-dan-cac-nguyen-tac-xuat-ban-du-lieu.html>
- [12] European Commission, 2018: *Turning FAIR into Reality - Final Report and Action Plan from the European Commission Expert Group on FAIR Data*:
<https://op.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283>. Bản dịch sang tiếng Việt:
https://www.dropbox.com/s/wtiraiu8svilgei/turning_fair_into_reality_1-Vi-30042019.pdf?dl=0, phần 2.3 FAIR và dữ liệu mở, tr. 34-35.
- [13] European Commission, 2018: *Cost of not having FAIR research data. Cost-Benefit analysis for FAIR research data*: http://publications.europa.eu/resource/cellar/d375368c-1a0a-11e9-8d04-01aa75ed71a1.0001.01/DOC_1. Bản dịch sang tiếng Việt:
https://www.dropbox.com/s/10kblenwlhca1jx/Cost-of-not-having-FAIR-research-data_Vi-02042023.pdf?dl=0, tr.14
- [14] NASA (2023): *Guide for Your Open Science Journey*: https://nasa.github.io/Transform-to-Open-Science-Book/Open_Science_Cookbook/Your_Open_Science_Journey.html.
Bản dịch sang tiếng Việt: <https://giaoducmo.avnuc.vn/khoa-hoc-mo/huong-dan-lo-trinh-khoa-hoc-mo-cho-ban-888.html>
- [15] Tim Berners-Lee (2026): *Linked Data*:
<https://www.w3.org/DesignIssues/LinkedData.html>
- [16] Ảnh chụp và tùy chỉnh từ màn hình của trang Zenodo: <https://zenodo.org/search?page=1&size=20&q=>. Truy cập ngày 14/07/2023.



Giấy phép nội dung: [CC BY 4.0 Quốc tế](https://creativecommons.org/licenses/by/4.0/)



Lê Trung Nghĩa, ORCID iD: <https://orcid.org/0009-0007-7683-7703>

Viện Nghiên cứu, Đào tạo và Phát triển Tài nguyên Giáo dục Mở

Hiệp hội các trường đại học cao đẳng Việt Nam.

(Bài viết cho Hội thảo: Dữ liệu nghiên cứu khoa học mở và lưu trữ dữ liệu nghiên cứu khoa học mở do Trung tâm Công nghệ thông tin - Bộ Khoa học và Công nghệ tổ chức tại Trường Đại học Kinh tế TP. Hồ Chí Minh ngày 13/09/2023).