

Lưu trữ tư liệu khoa học trong thời đại khoa học mở tại Việt Nam: THỰC TRẠNG VÀ KHUYẾN NGHỊ CHÍNH SÁCH

Chu Thị Ngân

Trung tâm Dữ liệu và Thông tin Khoa học - Viện Hàn lâm Khoa học và Công nghệ Việt Nam

“

Khoa học mở đang trở thành xu hướng toàn cầu, thúc đẩy chia sẻ, minh bạch và hợp tác trong nghiên cứu. Trong đó, lưu trữ dữ liệu khoa học đóng vai trò then chốt, bảo đảm khả năng truy cập mở, tính minh bạch và tái sử dụng dữ liệu - nền tảng cho đổi mới sáng tạo. Bài viết phân tích hiện trạng lưu trữ dữ liệu nghiên cứu tại Việt Nam trong mối liên hệ với yêu cầu của khoa học mở, thông qua đánh giá chính sách hiện hành, kinh nghiệm quốc tế và thực tiễn trong nước. Trên cơ sở đó, bài viết đề xuất các khuyến nghị nhằm xây dựng hệ thống lưu trữ dữ liệu nghiên cứu thống nhất, bền vững, đáp ứng tiêu chuẩn quốc tế, góp phần nâng cao chất lượng nghiên cứu và tăng cường hội nhập quốc tế trong lĩnh vực khoa học và công nghệ tại Việt Nam.

”



Mô hình khoa học mở theo định hướng toàn cầu của UNESCO. Nguồn: unesdoc.unesco.org.



Sinh viên tra cứu tại Trung tâm Thông tin - Thư viện, Đại học Quốc gia Hà Nội. Nguồn: Đại học Quốc gia Hà Nội.

Khoa học mở - tổng quan một số mô hình quốc tế

Khoa học mở (Open Science) đang trở thành mô hình phát triển khoa học mới trong bối cảnh chuyển đổi số toàn cầu, với nền tảng là sự minh bạch, chia sẻ và cộng tác. Khác với phương thức truyền thống, khoa học mở khuyến khích công khai dữ liệu, tài liệu, phần mềm và bài báo khoa học để tăng cường khả năng tiếp cận và tái sử dụng kết quả nghiên cứu. Theo UNESCO (2021), khoa học mở giúp tri thức khoa học trở nên dễ tiếp cận và minh bạch hơn vì lợi ích toàn xã hội. Khoa học mở không chỉ thu hẹp khoảng cách số mà còn gia tăng khả năng tiếp cận tri thức, đặc biệt là ở các quốc gia đang phát triển. Trong khu vực ASEAN, Thái Lan, Indonesia và Malaysia đã xây dựng chiến lược khoa học mở. Do đó, Việt Nam cũng cần xây dựng một hệ sinh thái khoa học mở, trong đó lưu trữ và truy cập tư liệu khoa học là yếu tố quan trọng.

Bên cạnh các nền tảng lưu trữ dữ liệu quy mô quốc tế, nhiều trường đại học trên thế giới cũng đã xây dựng kho lưu trữ riêng để bảo vệ, chia sẻ và nâng cao giá trị của kết quả nghiên cứu. Những hệ thống này, thường do thư viện hoặc trung tâm dữ liệu nghiên cứu quản lý, đóng vai trò hỗ trợ các nhà khoa học tuân thủ yêu cầu chia sẻ dữ liệu của các cơ quan tài trợ. Một ví dụ tiêu biểu là Harvard Dataverse, nền tảng do Đại học Harvard (Hoa Kỳ) phát triển, cho phép lưu trữ, chia sẻ và công bố dữ liệu nghiên cứu theo tiêu chuẩn quốc tế, đồng thời cung cấp các công cụ tổ chức, gắn thẻ (metadata) và cấp phát DOI cho các bộ dữ liệu. Liên minh châu Âu là một trong những khu vực tiên phong trong việc thúc đẩy chính sách khoa học mở, đặc biệt là thông qua việc yêu cầu các nhà nghiên cứu nộp dữ liệu vào các kho lưu trữ mở như Zenodo khi nhận tài trợ từ các chương trình Horizon 2020 và Horizon

Europe. Zenodo là một nền tảng lưu trữ dữ liệu nghiên cứu mở, được phát triển bởi Tổ chức Nghiên cứu Hạt nhân châu Âu (CERN), cung cấp không gian miễn phí để lưu trữ và chia sẻ dữ liệu khoa học, phần mềm, bài báo và các tài nguyên học thuật khác. Các nhà nghiên cứu được khuyến khích, thậm chí bắt buộc, phải chia sẻ dữ liệu nghiên cứu của mình nhằm thúc đẩy tính minh bạch, khả năng truy cập, chia sẻ và tái sử dụng dữ liệu trong cộng đồng khoa học quốc tế. Mô hình này thể hiện sự chuyển dịch mạnh mẽ từ các chính sách khuyến khích sang các quy định bắt buộc về chia sẻ dữ liệu, góp phần xây dựng hệ sinh thái khoa học mở toàn diện.

Tại Nhật Bản, nền tảng JST (Japan Science and Technology Agency) Data Platform đóng vai trò quan trọng trong việc quản lý và chia sẻ dữ liệu của các dự án nghiên cứu công. JST Data Platform hỗ trợ nhiều loại dữ liệu, bao gồm dữ liệu khoa học, kỹ thuật và công nghệ, giúp các nhà nghiên cứu và tổ chức dễ dàng truy cập, khai thác thông tin phục vụ cho các nghiên cứu tiếp theo. Hệ thống này không chỉ phục vụ các nhà nghiên cứu Nhật Bản, mà còn giúp tăng cường hợp tác quốc tế trong nghiên cứu khoa học.

Đại học Mahidol (Thái Lan) đã triển khai kho lưu trữ dữ liệu nghiên cứu, yêu cầu giảng viên và nghiên cứu viên nộp dữ liệu đầu ra vào kho này. Việc chia sẻ dữ liệu nghiên cứu không chỉ hỗ trợ phát triển khoa học mở, mà còn thúc đẩy một môi trường minh bạch, cởi mở trong cộng đồng học thuật, góp phần nâng cao hợp tác quốc tế trong nghiên cứu.

Khoa học mở không đơn thuần là việc công khai kết quả nghiên cứu, mà còn là một sự chuyển dịch toàn diện trong cách thức lưu trữ, quản lý và chia sẻ dữ liệu khoa học. Theo đó, dữ liệu nghiên cứu bao gồm dữ liệu gốc, dữ liệu phân tích, mã nguồn và phần mềm, cần được tổ chức theo hướng dễ tiếp cận, dễ tái sử dụng và minh bạch trong toàn bộ vòng đời nghiên cứu. Đây cũng chính là tinh thần của nguyên tắc FAIR (Findable - Accessible - Interoperable - Reusable), là tiêu chuẩn quốc tế trong quản trị dữ liệu khoa học mở.

Để đáp ứng nguyên tắc FAIR, mỗi bộ dữ liệu phải có mã định danh duy nhất (DOI hoặc Handle), đi kèm siêu dữ liệu (metadata) mô tả đầy đủ nội dung, cấu trúc và phương pháp thu thập, cũng như giấy phép sử dụng rõ ràng. Dữ liệu phải được lưu trữ trên các nền tảng hỗ trợ liên thông và tương tác dữ liệu, cho phép công cụ tự động tìm kiếm và truy xuất, đồng thời đảm bảo quyền riêng tư và bảo mật khi cần.

Thực trạng và khuyến nghị chính sách cho Việt Nam

Tại Việt Nam, khoa học mở bước đầu đã được các cơ quan quản lý và cơ sở nghiên cứu quan tâm, song việc triển khai vẫn còn rời rạc, thiếu hệ thống và mới dừng ở mức thăm dò. Một trong những hạn chế chính là chưa có hệ thống lưu trữ dữ liệu nghiên cứu quốc gia thống nhất; dữ liệu thường phân tán theo từng đề tài, thiếu chuẩn hóa siêu dữ liệu và định danh, gây khó khăn cho việc truy xuất và tái sử dụng.

Hiện chỉ một số trường đại học lớn như Đại học Quốc gia Hà Nội, Đại học Bách khoa Hà Nội hay Đại học Tôn Đức Thắng xây dựng thư viện số, nhưng chủ yếu tập trung vào bài báo và luận văn, chưa bao phủ dữ liệu nghiên cứu gốc hay mã nguồn phần mềm. Việt Nam cũng chưa có quy định pháp lý bắt buộc lưu trữ, công bố hoặc công nhận dữ liệu nghiên cứu là sản phẩm khoa học chính thức, khiến động lực chia sẻ trong cộng đồng nghiên cứu còn thấp. Bên cạnh đó, hạ tầng công nghệ chưa được đầu tư đầy đủ và chưa có sự tích hợp với các kho dữ liệu quốc tế như Zenodo, Dryad hay Figshare.

Một trong những thách thức lớn trong việc triển khai khoa học mở là sự thiếu vắng một kho dữ liệu nghiên cứu quốc gia tập trung. Dữ liệu từ các dự án khoa học tại các trường đại học và viện nghiên cứu hiện vẫn phân tán, không được lưu trữ hệ thống, gây khó khăn trong việc chia sẻ và tái sử dụng cho các nghiên cứu tiếp theo. Bên cạnh đó, hiện chưa có quy định bắt buộc nộp và lưu trữ dữ liệu đầu ra từ các dự án nghiên cứu. Việc thiếu trách nhiệm và cơ chế chia sẻ sau nghiên cứu làm giảm tính minh bạch và

hạn chế khả năng tiếp cận dữ liệu trong cộng đồng khoa học. Mặc dù khoa học mở đã được công nhận là xu thế toàn cầu, Việt Nam vẫn thiếu các quy định và hướng dẫn cụ thể về lưu trữ và chia sẻ dữ liệu. Điều này gây khó khăn trong việc thiết lập một hệ sinh thái nghiên cứu minh bạch, đồng bộ. Hiện tại, một số kho tài liệu số như thư viện của các trường đại học đã được xây dựng, nhưng chủ yếu phục vụ tài liệu học thuật như sách, bài giảng, chưa đáp ứng được nhu cầu lưu trữ và chia sẻ dữ liệu nghiên cứu khoa học, vốn đòi hỏi các định dạng đặc thù và yêu cầu cao về bảo mật, truy xuất.

Rào cản lớn khác là nhận thức hạn chế về khoa học mở. Nhiều nhà nghiên cứu, giảng viên và cán bộ thư viện chưa thấy rõ tầm quan trọng của việc chia sẻ dữ liệu, dẫn đến thiếu động lực tham gia. Hạ tầng kỹ thuật phục vụ khoa học mở cũng còn yếu, thiếu hệ thống máy chủ, phần mềm phù hợp và nhân lực có chuyên môn. Ngoài ra, sự thiếu vắng các chính sách thúc đẩy từ cơ quan quản lý cũng là trở ngại lớn. Việc chưa có cơ chế khuyến khích, đánh giá và công nhận đóng góp trong chia sẻ dữ liệu khiến các nhà khoa học ít quan tâm đến việc lưu trữ và công khai dữ liệu nghiên cứu.

Để thúc đẩy khoa học mở và nâng cao hiệu quả lưu trữ, chia sẻ dữ liệu nghiên cứu tại Việt Nam, cần xây dựng một kho dữ liệu quốc gia mạnh, bền vững, kết nối các nhà khoa học, viện nghiên cứu, tổ chức trong nước và quốc tế. Kho dữ liệu này không chỉ lưu trữ thông tin nghiên cứu mà còn tạo môi trường minh bạch, hỗ trợ chia sẻ dữ liệu khoa học. Hệ thống cần tuân thủ các chuẩn quốc tế như Dublin Core và DataCite để đảm bảo bảo mật và khả năng truy cập cao.

Các nhà nghiên cứu, trường đại học và viện nghiên cứu cần có trách nhiệm nộp dữ liệu đầu ra vào kho dữ liệu quốc gia hoặc các kho quốc tế ngay sau khi hoàn thành đề tài, nhằm tạo nguồn dữ liệu phong phú, phục vụ nghiên cứu tiếp theo và tăng cường minh bạch khoa học.

Chuẩn hóa mô tả dữ liệu là yếu tố quan trọng, giúp cải thiện khả năng chia sẻ và nâng cao chất lượng nghiên cứu. Đồng thời, cần đào tạo đội ngũ cán bộ thư viện, thông tin và chuyên gia trong lĩnh vực quản lý, chia sẻ dữ liệu để hỗ trợ các nhà khoa học.

Trong khi kho dữ liệu quốc gia chưa hoàn thiện nên khuyến khích sử dụng các hệ thống lưu trữ mở quốc tế như Zenodo, Dryad, Figshare. Bên cạnh đó, có thể triển khai thử nghiệm chính sách lưu trữ dữ liệu tại các đơn vị tiên phong như Viện Hàn lâm Khoa học và Công nghệ Việt Nam, Đại học Quốc gia Hà Nội và Đại học Quốc gia TP Hồ Chí Minh trước khi áp dụng rộng rãi.

*
* *

Lưu trữ tư liệu khoa học không chỉ là một bước kỹ thuật trong nghiên cứu, mà là yếu tố then chốt thúc đẩy khoa học mở tại Việt Nam. Trong bối cảnh toàn cầu hóa khoa học, việc chia sẻ và tái sử dụng dữ liệu góp phần tăng cường minh bạch, mở ra cơ hội hợp tác và sáng tạo. Do đó, cần thiết lập một chính sách lưu trữ dữ liệu nghiên cứu đồng bộ, phù hợp với chuẩn quốc tế và nhu cầu thực tiễn.

Chính sách này cần ưu tiên xây dựng kho dữ liệu quốc gia, chuẩn hóa mô tả dữ liệu và thúc đẩy chia sẻ qua các hệ thống lưu trữ mở. Điều này không chỉ nâng cao chất lượng nghiên cứu mà còn giúp kết quả khoa học được tiếp tục khai thác hiệu quả. Thành công trong xây dựng hệ thống lưu trữ phụ thuộc vào sự phối hợp giữa cơ quan quản lý, viện nghiên cứu, trường đại học và cộng đồng học thuật. Các cơ quan cần tạo hành lang pháp lý thuận lợi, còn các cơ sở nghiên cứu cần nghiêm túc áp dụng quy định về lưu trữ và chia sẻ dữ liệu. Đồng thời, cộng đồng học thuật cũng cần nhận thức rõ vai trò của lưu trữ trong nâng cao chất lượng nghiên cứu ✍