

THƯ VIỆN CÓ THỂ NẮM GIỮ VAI TRÒ TRUNG TÂM TRONG QUẢN LÝ DỮ LIỆU NGHIÊN CỨU

Chia sẻ và lưu trữ dài hạn dữ liệu nghiên cứu đang ngày càng được chú trọng trong quá trình nghiên cứu, do góp phần thúc đẩy tiến trình khoa học và tối đa hóa khoản thu lại cho nhà đầu tư nghiên cứu. Dù mức độ nhận thức về chia sẻ dữ liệu giữa các ngành khoa học có sự khác biệt, một điều không thể phủ nhận là chia sẻ dữ liệu nghiên cứu đang thu hút được sự quan tâm của tất cả các ngành khoa học.

Tuy nhiên, chúng ta vẫn đang ở trong giai đoạn chuyển từ khoa học trọng tâm tài liệu sang trọng tâm dữ liệu và cơ sở hạ tầng cho việc chia sẻ và tái sử dụng dữ liệu khoa học vẫn chưa hoàn thiện. Số lượng tăng nhanh của các công cụ và cảm biến ngày càng tinh vi cho thấy dữ liệu khoa học sẵn có để chia sẻ đang phát triển nhanh chóng, tuy nhiên vẫn còn có nhiều việc cần làm để có thể trích xuất dữ liệu này từ ổ cứng của nhà nghiên cứu cũng như đảm bảo dữ liệu đó có thể được truy cập được về lâu dài.

Sự nổi lên của xuất bản điện tử đã và đang làm phá vỡ các vai trò thông tin truyền thống và thư viện có vị thế ra sao trong khoa học trọng tâm dữ liệu cũng chưa sáng tỏ. Rõ ràng đang có nhiều cơ hội mở ra để thư viện phát huy vị thế trung tâm của mình nhưng đồng thời nếu thư viện không tự mình có những thay đổi sớm thì sẽ có các tổ chức khác chiếm lấy vị thế này.

Hệ sinh thái phức hợp

Hệ thống xuất bản khoa học hiện đại đang ngày càng trở nên phức tạp hơn. Quy trình xuất bản truyền thống có thể được hiểu một cách cơ bản như một vòng lặp đơn giản qua việc các bài báo được chuyển từ các nhà nghiên cứu tới các nhà xuất bản,

và tiếp tục tới các thư viện rồi quay trở lại với nhà nghiên cứu. Ngược lại, hệ thống xuất bản hiện đại có thể được hình dung như một mạng lưới - với những mối quan hệ mới và những chủ thể có vai trò chồng lấn đan xen ngày càng nhiều. Ngày nay, các bài báo công bố không chỉ đơn thuần được đưa tới các tạp chí, mà còn được chuyển tới các kho lưu trữ nội bộ hoặc kho lưu trữ theo chủ thể. Trong đó, các kho lưu trữ nội bộ thường do các đơn vị cung cấp dịch vụ thư viện quản lý và chứa đựng các tài liệu xám cũng như là các ấn phẩm chính thức. Thêm vào đó, xu hướng kết hợp các tạp chí điện tử của các nhà xuất bản cũng đang làm giảm vai trò thu mua của thư viện cùng lúc với việc các tạp chí truy cập mở có thể phá vỡ vai trò phân phối truyền thống của các nhà xuất bản.

Tất nhiên, các mô hình truyền thống và hiện đại nêu trên đều được mô tả theo một cách đơn giản hóa và mô hình tạp chí truyền thống phức tạp hơn nhiều so với mô hình vòng lặp đề cập ở đây. Tuy nhiên, điều có thể thấy rõ ràng là vai trò tổ chức đang ngày càng bớt tính cứng nhắc trong hoạt động xuất bản hàn lâm. Chính trong hệ sinh thái thiên biến này mà nhu cầu về các dịch vụ trọng tâm dữ liệu ngày càng gia tăng.

Mối quan tâm ngày càng tăng về dữ liệu khoa học và nhu cầu về các dịch vụ trọng tâm dữ liệu đang mang đến nhiều cơ hội cho thư viện để tái định hình vai trò trung tâm của mình trong các cơ sở nghiên cứu, nhưng sự trùng lặp vai trò với các tổ chức khác cũng cho thấy các tổ chức này có thể nhanh chóng khẳng định được vị thế trong các lĩnh vực mà thư viện đã luôn coi là thuộc quyền hạn của mình.

Vòng đời dữ liệu

Từ mô hình vòng đời dữ liệu, ta có thể xem xét được các cơ hội dành cho các dịch vụ thư viện trong việc chia sẻ dữ liệu nghiên cứu. Kho lưu trữ dữ liệu Vương quốc Anh đã phân chia vòng đời dữ liệu theo sáu giai đoạn: xử lý dữ liệu, phân tích dữ liệu, bảo quản dữ liệu; cấp quyền truy cập dữ liệu và tái sử dụng dữ liệu. Nhiều giai đoạn này có thể được hưởng lợi nhiều từ những kỹ năng của cộng đồng thư viện.

Quản lý dữ liệu hiệu quả bắt đầu ngay từ khi tiến hành nghiên cứu chứ không phải là trong giai đoạn sau nghiên cứu. Các thư viện nghiên cứu cần ở vị trí tư vấn về cấu trúc, cách thức lưu trữ và siêu dữ liệu đối với dữ liệu nghiên cứu.

Bảo quản dữ liệu dài hạn đòi hỏi cách thức định dạng, lưu trữ và nhu cầu siêu dữ liệu khác so với dữ liệu trong quá trình tạo lập dữ liệu. Một lần nữa, các thư viện nghiên cứu cần ở vị trí tư vấn trong hoạt động này. Quan trọng hơn cả, việc lưu trữ dữ liệu của dự án nghiên cứu cần mở rộng phạm vi ra khỏi khuôn khổ của một dự án đơn lẻ và cần được lưu trong một kho lưu trữ phù hợp.

Cấp quyền truy cập dữ liệu là một lĩnh vực nữa mà thư viện có thể thể hiện vai trò của mình. Cấp quyền truy cập không chỉ đòi hỏi dữ liệu cần sẵn có mà còn cần có thể tìm thấy được và đi kèm với các quyền cho phép tái sử dụng. Về việc này, cộng đồng thư viện có kinh nghiệm trong việc thiết lập các hệ thống phân loại và vấn đề bản quyền.

Trong giai đoạn cuối cùng của vòng đời dữ liệu, việc tái sử dụng dữ liệu đặt ra yêu cầu tìm thấy được dữ liệu và đảm bảo rằng có đủ thông tin hiệu quả để dữ liệu được tái sử dụng. Tất nhiên là có một sự khác biệt giữa các cơ hội sẵn có và các cơ hội đang được thực hiện và trong phần lớn trường hợp, các thư viện vẫn đang tiếp tục trọng tâm tư liệu.

Mặc dù có nhiều thư viện đã thiết lập các kho lưu trữ nội bộ, các kho lưu trữ này vẫn tập trung chủ yếu vào tài liệu hơn là dữ liệu. Trong 2,727 kho dữ liệu được liệt kê tại OpenDOAR, Thư mục các kho truy cập mở, chỉ có 131 kho là được đưa vào danh sách kho chứa đựng các bộ dữ liệu (4.8%). Điều này không khác mấy so với kết quả từ truy vấn tương tự được tiến hành từ đầu năm 2011 là 4.1%.

Sự phân biệt giữa kho dữ liệu nội bộ và các bộ dữ liệu chuyên ngành cho thấy một bức tranh tối màu hơn đối với các kho lưu trữ nội bộ khi chỉ có 4% các kho lưu trữ nội bộ được liệt kê là có chứa các bộ dữ liệu so với 11.1 % các kho lưu trữ chuyên ngành.

Các kho lưu trữ chuyên môn có tiềm năng cung cấp nhiều giao diện tương tác có tính đổi mới hơn dành riêng cho các loại dữ liệu đặc thù hơn là kho dữ liệu nội bộ thông thường có thể cung cấp, mặc dù khi dữ liệu tồn tại lâu hơn dự án, câu hỏi còn lại là ai sẽ là người chịu trách nhiệm cho những công trình này trong thời gian dài nếu một kho lưu trữ đóng cửa hoặc nhu cầu dữ liệu/siêu dữ liệu cần được cập nhật.

Nếu thư viện không thể cung cấp các dịch vụ dữ liệu đổi mới và hiệu quả thì các đơn vị khác sẽ chiếm lấy công việc này. Một ví dụ điển hình cho việc này là các tạp chí dữ liệu mới.

Tái đóng gói dữ liệu

Hai tạp chí truy cập mở mới được dựa trên dữ liệu nghiên cứu thay vì các phát hiện nghiên cứu và bắt đầu được xuất bản từ 2014 là Tạp chí Dữ liệu Khoa học (Scientific Data Journal) của Tập đoàn Xuất bản Nature và Tạp chí Dữ liệu Địa khoa học (Geoscience Data Journal) của Nhà xuất bản Wiley. Hai ấn phẩm này cung cấp thông tin mô tả chi tiết về cách thức và lý do mà bộ dữ liệu được thu thập đồng thời kết nối tới chính bộ dữ liệu tại một trong số các kho dữ liệu được cho phép.

Những ấn phẩm dạng này không những đem tới nhiều lợi ích cho cá nhân nhà nghiên cứu và ngành khoa học nói chung

mà còn khuyến khích nhà nghiên cứu chia sẻ dữ liệu nhiều hơn. Vì đây là những ấn phẩm được bình duyệt và qua việc chia sẻ dữ liệu nghiên cứu tại những ấn phẩm này, nhà khoa học vẫn nhận được sự công nhận cho công trình nghiên cứu của mình mà không nhất thiết phải có những nhận định hoặc phát hiện mới trong đó. Ngoài ra, những ấn phẩm dạng này cũng giúp nhà nghiên cứu tìm kiếm và tái sử dụng các bộ dữ liệu sẵn có.

Có rất nhiều ưu điểm của các tạp chí dữ liệu mới này có thể đạt được bởi cộng đồng thư viện mà không cần đến các nhà xuất bản. Cũng có nhiều quan ngại về việc các nhà xuất bản mở rộng phạm vi hoạt động của họ khi có một số người cáo buộc họ lạm dụng vị thế trên thị trường để tăng cao giá thành tạp chí. Nhưng nếu các thư viện vẫn muốn tiếp tục thể hiện được vai trò liên quan của mình trong lĩnh vực xuất bản thì họ cần quyết tâm thích ứng và học từ sự quyết tâm đổi mới dữ liệu của các nhà xuất bản.

Thư viện có thể cung cấp các dịch vụ gì?

Iain Hrynaszkiewicz, trưởng bộ phận dữ liệu và xuất bản HSS trong nghiên cứu mở tại Tập đoàn xuất bản Nature/Palgrave Macmillan cho rằng, vai trò của cán bộ thư viện sẽ tiếp tục thay đổi với trọng tâm là kỹ năng về dữ liệu “Trong quá khứ các cán bộ thư viện đã được tham gia đào tạo về kỹ năng thông tin theo mọi hình thức, và dữ liệu nghiên cứu có vai trò ngày càng ngang bằng với các kết quả đầu ra nghiên cứu khác (như là bài báo nghiên cứu) trong hoạt động đánh giá và tài trợ nghiên cứu. Việc cung cấp các khóa đào tạo về đánh giá, lưu trữ, xuất bản và quản lý dữ liệu theo đó là điều tất yếu dành cho cán bộ thư viện”.

“Những người mô tả dữ liệu (data descriptors) như là những người được Tạp chí dữ liệu khoa học công bố, đóng vai trò quan trọng trong việc phát hiện, tái sử dụng dữ liệu cũng như đáp ứng các yêu cầu về chia sẻ dữ liệu của nhà tài trợ và các cơ sở nghiên cứu. Họ có thể được coi là một phần

của thực tiễn tốt nhất về quản lý dữ liệu nghiên cứu và lên kế hoạch xuất bản của bất kì nghiên cứu nào. Vì lý do này, chúng tôi mong muốn được phối hợp với các nhà quản lý thư viện và chuyên gia thông tin trong việc nâng cao kỹ năng và hiểu biết của nhà nghiên cứu về tầm quan trọng của quản lý dữ liệu”.

Nhu cầu về sự chuyên nghiệp trong quản lý dữ liệu tại chỗ đảm bảo rằng ngành thông tin - thư viện sẽ có chỗ đứng trong tương lai quản lý dữ liệu, mặc dù chưa thể rõ khả năng có thể đáp ứng trước các nhu cầu này của hầu hết các thư viện và cán bộ thư viện là ra sao.

Thay đổi mang tính kỹ thuật này đòi hỏi cần có các bộ kỹ năng mà có thể bị hạn chế trong lĩnh vực thông tin do trọng tâm đào tạo của các trường đào tạo về thư viện. Các chuyên gia thông tin có thể tiếp tục là một phần của công việc nhưng luôn có khả năng rằng, những người lao động này sẽ được phân bổ vào các dự án đặc thù.

Kết luận

Thiên hướng tự chuyển mình trong các dịch vụ thư viện đã trở thành chủ đề thường thấy trong tài liệu ngành trong nhiều năm. Thông thường thì các công trình này bao gồm các dự đoán rằng, nếu không bị diệt vong thì vai trò của thư viện sẽ thay đổi vượt xa mọi sự thừa nhận. Nhu cầu ngày càng nhiều về các dịch vụ trọng tâm dữ liệu mang lại nhiều cơ hội lớn nằm trong phạm vi thư viện nhưng cũng có nhiều đối thủ cạnh tranh tiềm tàng khác.

Nếu thư viện tiếp tục làm tròn vai trò truyền thống cốt lõi của mình thì cần có thêm các cách tiếp cận dữ liệu sáng tạo hơn. Không đổi mới thành công sẽ khiến cho vai trò của thư viện và cán bộ thư viện trở nên mờ nhạt hơn.

Nguyễn Minh Hiền (lược dịch)

Nguồn: <https://www.researchinformation.info/feature/libraries-could-play-key-role-managing-research-data>