

Thư viện nguồn mở: một tất yếu đối với khoa học hiện đại?



Song song với phương thức thông tin khoa học truyền thống, in trên giấy hay điện tử, muốn đọc phải trả tiền hoặc miễn phí là một phương thức thông tin khoa học mới gọi là “thông tin khoa học trực tiếp”, dựa trên sự tự lưu trữ, bởi chính các nhà nghiên cứu, các sản phẩm của họ trong các “thư viện nguồn mở”, từ đó ra đời khái niệm tự do công bố.

Vậy lợi ích của các thư viện mở được các nhà nghiên cứu sử dụng là gì? Đây là cơ hội và thách thức của nó?

Một phương thức thông tin khoa học mới

Các tài liệu lưu trữ mở là các cơ sở dữ liệu ở đó các tác giả có thể tự công bố trên internet các kết quả nghiên cứu khoa học của mình, giúp

các đồng nghiệp của họ trên toàn thế giới có thể tiếp cận chúng trực tiếp và miễn phí. Người ta còn gọi đây là phương thức “thông tin khoa học trực tiếp” để nhấn mạnh rằng, ngược với các tạp chí khoa học truyền thống, không có một trung gian nào tham gia vào, không có ban biên tập cũng chẳng có nhà xuất bản thương mại. Đúng là các quản trị mạng phải thực hiện việc kiểm tra sơ đẳng nội dung trước khi đưa lên mạng, nhưng sự kiểm tra này chỉ kéo dài vài phút đối với mỗi bài; đơn giản chỉ là nhằm mục đích loại các bài rõ ràng không phù hợp.

Không giống như việc đưa lên mạng một văn bản trên các trang cá nhân hay các blog, các tài liệu lưu trữ mở đòi hỏi một sự tổ chức nhất định được thực hiện ở cấp độ cơ quan (trường đại học, viện nghiên cứu, phòng thí nghiệm...). Những cơ quan này dành riêng một máy chủ để lưu trữ các kết quả khoa học của các thành viên của mình, thuộc tất cả các chuyên ngành – đó là cách được ưa chuộng của một số chuyên gia Anh-Mỹ. Đôi khi các thông tin được lưu trữ theo chủ đề, như *ArXiv* chẳng hạn, trang này tập trung vào vật lý và toán học. Ở Pháp, nhiều cơ quan đã quyết định nhóm nhau lại và thành lập một dữ liệu mở chung. Năm 2004, INRIA đã nối với thư viện nguồn mở được triển khai bởi CNRS năm 2001. Mới đây, rất nhiều cơ quan khác (các trường ĐH, các trường lớn, các tổ chức nghiên cứu như INRA và INSERM) đã ký một thỏa thuận chung và đã bắt đầu xây dựng một mã nguồn mở phỏng theo *ArXiv*, nhưng chặt chẽ hơn và đặc biệt là rộng hơn về số các chuyên ngành: tại đây có các kết quả nghiên cứu không chỉ trong lĩnh vực khoa học tự nhiên (toán, vật lý, hóa học...), mà còn cả trong các lĩnh vực khoa học xã hội và nhân văn. Cơ sở của các tài liệu tương tác, được gọi tắt là HAL (<http://hal.archives-ouvertes.fr>), đã nhận được khoảng 2000 văn bản mỗi tháng. Sáng kiến này đã được Viện hàn lâm khoa học Pháp chính thức ủng hộ ngay từ năm 2005. Ở một số nước khác trên thế giới, các cơ quan như NIH (Viện quốc gia về sức khỏe) ở Mỹ, Hội đồng nghiên cứu Anh, các trường đại học, các hiệp hội (Wellcome Trust) cũng ủng hộ chính sách thư viện điện tử nguồn mở.



Nhà toán học Nga Grigory Perelman, người đã chứng minh được giả thuyết thế kỷ của

Các cơ sở dữ liệu này bao gồm: Mọi loại văn bản mà nhà nghiên cứu muốn cho cộng đồng khoa học biết: các bản thảo của các bài báo đang trong quá trình biên tập để được đăng tải, góp ý cho một tác phẩm khoa học, luận văn hay luận án, báo cáo khoa học hoặc các bài báo đã được đăng trên tạp chí truyền thống (với điều kiện tạp chí này đồng ý) và cả dữ liệu liên quan, như tên của tác giả, cơ quan công tác, ngày, các tài liệu tham khảo... (điều này làm cho việc tìm kiếm tài liệu tham khảo và lập danh mục các tác phẩm đã công bố trở nên dễ dàng hơn).

Ưu điểm và nhược điểm

Việc triển khai và duy trì các dữ liệu mở có một chi phí rất thấp so với số tiền mà các cơ sở nghiên cứu chi cho việc mua tài liệu khoa học tăng rất nhanh trong vòng hai mươi năm trở lại đây và làm cho rất nhiều phòng thí nghiệm và tổ chức nghiên cứu lo lắng. Tại Pháp, các chi tiêu cho việc mua tạp chí khoa học lên tới vài trăm triệu euro mỗi năm. Trong khi đó, chi phí cho thư viện điện tử nguồn mở thấp hơn từ 500 đến 1000 lần, vì chỉ phải trả

Pointcaré ("Nếu một đa tạp ba chiều compact không có biên là đơn liên, thì nó đồng phôi với mặt cầu ba chiều"), đã không công bố các công trình nghiên cứu của mình trên các tạp chí quốc tế, như thói quen từ lâu của các nhà khoa học. Ông đã đưa các kết quả của mình lên mạng Internet trên một địa chỉ được thành lập năm 1991 theo sáng kiến của nhà vật lý học người Mỹ Paul Ginsparg mà tất cả mọi người đều có thể truy cập miễn phí (<http://arxiv.org>).

lương cho một vài kỹ sư tin học.

Với nhà khoa học, việc công bố công trình nghiên cứu trên thư viện nguồn mở nhanh hơn rất nhiều so với đôi khi mất hơn một năm của việc công bố một bài báo trong một tạp chí khoa học. Còn đối với độc giả thì một tập hợp các thông tin trong một số lĩnh vực khoa học ở một lưu trữ duy nhất, giúp việc tìm kiếm các tài liệu trở nên dễ dàng hơn.

Thư viện điện tử nguồn mở còn cung cấp một truy cập miễn phí cho tất cả các nhà khoa học trên toàn thế giới, (sau khi Liên Xô tan rã, rất nhiều thành viên của trường phái vật lý lý thuyết xuất sắc của Nga đã có thể tiếp tục con đường trí tuệ trong chuyên ngành của mình bằng cách truy cập miễn phí vào các công trình mới nhất nhờ có *ArXiv*). Hơn nữa, việc truy cập tự do và từ xa sẽ đi theo hướng dân chủ hóa tri thức, trong chừng mực nó làm cho việc tham khảo các bài báo khoa học chuyên ngành trở nên dễ dàng hơn đối với tất cả người dân.

Ngoài ra, thư viện điện tử nguồn mở còn cho phép giải phóng khỏi khái niệm tài liệu cố định mãi mãi trong một công bố khoa học trên giấy. Trên

thực tế, một tác giả có thể đặt bài trong cùng một "hộp" các phiên bản kế tiếp nhau – có ghi ngày tháng cập nhật – của cùng một bài. Như vậy người ta sẽ có thể thấy trên thư viện điện tử nguồn mở phiên bản của tác giả đã đăng trên tạp chí truyền thống, phiên bản được sửa đổi và chấp nhận (với đường dẫn tới bài báo đã được đăng tải), và cả các phiên bản trước ở đó, chẳng hạn, các lỗi chính tả trong các dữ liệu hay các phương trình được sửa lại cho đúng. Điều này đặc biệt hữu ích cho các tác giả của các bài báo tổng hợp, họ có thể thực hiện việc cập nhật sau khi đăng tải, tính đến đồng thời các phản ứng của độc giả và các phát triển khoa học mới nhất. Cuối cùng, trong trường hợp có sai sót, tác giả có thể rút lại bài báo và nói rõ lý do. Tất cả những điều này không thể xảy ra đối với bài báo khoa học đăng trên tạp chí giấy.

Một khó khăn về kỹ thuật của thư viện nguồn mở đó là đảm bảo sự tồn tại lâu dài của các dữ liệu lưu trữ. Các địa chỉ điện tử của các tài liệu phải rất ổn định để các tài liệu này có thể dùng làm quy chiếu trong các nghiên cứu khoa học sau. Các di chuyển phần mềm và máy móc phải được thực hiện rất nhanh để tập dữ liệu lưu trữ vẫn có khả năng truy cập được trong nhiều thập kỷ sau. Trong các điều kiện này, người ta có thể hy vọng rằng tuổi thọ của các tài liệu điện tử chỉ ít cũng dài như tuổi thọ của các tài liệu in trên giấy (thường là vài thế kỷ). Nhưng để đạt được mục đích này, cần phải thiết lập một sự quản lý tập trung tài liệu lưu trữ ở tầm quốc gia.

Ngoài ra việc duyệt một bài báo bởi một ban biên tập gồm các chuyên gia là một quá trình lâu dài và tốn kém. Do vậy việc đặt một bài báo trong thư viện điện tử nguồn mở thường xảy ra đồng thời với việc bài báo đó được các tạp chí khoa học chấp nhận đăng tải nó; người đặt bài khi đó ghi rõ nguồn của bài báo đã công bố, tạo thành một dạng đảm bảo khoa học.

Thư viện nguồn mở và tạp chí truyền thống: chung sống hòa bình

Các thư viện điện tử nguồn mở liệu có làm phá sản các nhà xuất bản tạp chí truyền thống không?

Kinh nghiệm với *ArXiv*, hiện đã 15 tuổi, chứng tỏ rằng không. Trên thực tế, hai phương thức thông tin này không loại trừ nhau. Người ta thậm chí còn nhận thấy một hiệu ứng hợp tác giữa các hệ lưu trữ mở và hoạt động của các tạp chí truyền thống. Một số tác giả, trước khi gửi bài đăng trên một tạp chí, thường bắt đầu bằng việc đưa lên một thư viện nguồn mở, để tận dụng các phản ứng tức thì của các đồng nghiệp; các phản ứng này đôi khi giúp tác giả chỉnh sửa nội dung của bài báo, cho ra đời một phiên bản thứ hai – và chính phiên bản tốt hơn này sẽ được gửi đăng trên tạp chí giấy.

Trên một khía cạnh khác, nhà xuất bản khoa học có thể tra cứu trên thư viện nguồn mở tập hợp các tác phẩm của tác giả, để có một ý tưởng về con người khoa học của anh ta. Sự lựa chọn những người viết tóm tắt bài báo khoa học cũng trở nên dễ dàng hơn vì thư viện nguồn mở giúp ban biên tập dễ dàng biết nhà khoa học nào hiện đang quan tâm đến lĩnh vực của bài báo, và như vậy có khả năng chấp nhận viết tóm tắt bài báo.

Phần lớn các báo khoa học hiện nay đã chấp nhận việc đưa phiên bản một bài báo của tác giả lên thư viện nguồn mở của một cơ quan nghiên cứu khoa học nào đó, với các phương thức đa dạng. Tuy nhiên, một số nhà xuất bản lại qui định chỉ được đưa lên thư viện nguồn mở bài báo khoa học đã được đăng ở tạp chí từ vài tháng

đến một năm.

Rõ ràng là thư viện nguồn mở không phải là một dạng mới của tạp chí truyền thống, hay đơn giản là một sự kết tụ vào trong cùng một cổng điện tử nội dung của rất nhiều tạp chí. Đó là một công cụ nghiên cứu mới, với các quy luật trí tuệ riêng của nó. Với công cụ này, sự tham khảo duy nhất trong thông tin giữa các nhà khoa học không còn là một bài báo được ấn bản nữa, và tài liệu khoa học có thể có một đời sống thú vị trước và sau khi được công bố.

Các thư viện nguồn mở đưa chúng ta trở lại một truyền thống cũ, truyền thống về trao đổi trực tiếp các thư từ khoa học giữa các nhà bác học. Phương thức thông tin đầu tiên này đã kéo dài vài thế kỷ. Nó đã không mất đi với sự ra đời ngay từ thế kỷ XVII, dưới sự che chở của các viện hàn lâm và các hội bác học, các ấn bản như *Philosophical Transactions of the Royal Society* ở Anh hay *Journal des Sçavants* ở Pháp, và các xuất bản khoa học như chúng ta biết ngày nay kể từ cuối thế kỷ XIX. Xuất bản khoa học, đến lượt nó, cũng sẽ không biến mất, mà sẽ chung sống với các thư viện nguồn mở. Nhưng đâu sẽ là điểm cân bằng? Tương lai sẽ nói cho chúng ta biết.

Quốc Chiên Theo *Pour la Science* số 352