

Tổng luận số 10.2018

KHOA HỌC MỞ: CÁC XU HƯỚNG CHÍNH SÁCH GẦN ĐÂY



MỤC LỤC

I. TÍNH CẤP THIẾT VÀ TÁC ĐỘNG CỦA KHOA HỌC MỞ.....	5
1.1. Truy cập công bố khoa học.....	8
1.2. Truy cập dữ liệu nghiên cứu	11
II. TRUY CẬP MỞ CÔNG BỐ KHOA HỌC	13
2.1. Định nghĩa truy cập mở	13
2.2. Xuất bản truy cập mở và bảo vệ quyền sở hữu trí tuệ	19
2.3. Công bố truy cập mở và các tác động pháp lý của nó	24
III. DỮ LIỆU NGHIÊN CỨU MỞ.....	27
3.1. Nghiên cứu khoa học dựa vào dữ liệu	27
3.2. Định nghĩa dữ liệu mở	30
3.3. Chia sẻ dữ liệu: Thách thức và cơ hội	31
3.4. Các khung khổ bảo vệ dữ liệu mở ở các nước OECD.....	38
IV. QUẢN TRỊ KHOA HỌC MỞ: HOẠT ĐỘNG, XU HƯỚNG VÀ CHÍNH SÁCH...	42
4.1. Những nhân tố chủ chốt trong khoa học mở	42
4.2. Quản trị khoa học mở: Các xu hướng chính sách gần đây	43
KẾT LUẬN	52

GIỚI THIỆU

Khoa học mở thường đề cập đến những nỗ lực của các nhà nghiên cứu, các chính phủ, các cơ quan tài trợ nghiên cứu và cộng đồng khoa học để làm cho các đầu ra của nghiên cứu được tài trợ công được truy cập nhiều hơn ở định dạng số. Khoa học mở là sự giao thoa giữa tính mở có từ lâu đời trong khoa học và các công cụ công nghệ thông tin và truyền thông (CNTT-TT) và là công cụ để thúc đẩy nghiên cứu.

Một mặt, Internet và các nền tảng trực tuyến đang tạo ra nhiều cơ hội mới để tổ chức và phổ biến nội dung của những dự án nghiên cứu, công bố khoa học và bộ dữ liệu lớn, và làm cho chúng luôn sẵn sàng để các nhà khoa học và các nhà nghiên cứu khác cũng như cộng đồng doanh nghiệp và xã hội nói chung có thể truy cập được. Mặt khác, CNTT-TT cho phép thu thập một lượng lớn dữ liệu và thông tin, là cơ sở cho các thí nghiệm và nghiên cứu khoa học, góp phần làm cho khoa học ngày càng dựa vào dữ liệu. Kho dữ liệu và lưu trữ trực tuyến cung cấp khả năng lưu trữ, truy cập, sử dụng và tái sử dụng đầu vào và đầu ra của nghiên cứu khoa học (cả công bố khoa học và dữ liệu nghiên cứu), đẩy nhanh chuyển giao tri thức giữa các nhà nghiên cứu và các lĩnh vực khoa học, mở ra các phương thức mới cho hợp tác và nghiên cứu khoa học.

Các nhà kinh tế xem tri thức khoa học được tạo ra bởi nghiên cứu công như hàng hoá công, có nghĩa là mọi người đều có thể sử dụng tri thức mà không mất thêm chi phí khi nó được công khai, tạo ra lợi nhuận xã hội cao hơn. Theo đó những phát hiện cơ bản của khoa học được xem như một sản phẩm của cộng tác xã hội và được gán cho cộng đồng. Những đòi hỏi của các nhà khoa học về sở hữu trí tuệ thường nhằm đạt được sự công nhận và coi trọng. Cuộc đua để được là người công bố (được gọi là quy tắc ưu tiên) trong khoa học truyền thống là động lực mạnh mẽ để các nhà khoa học làm cho tri thức của họ được công khai.

Trong khi hệ thống dựa trên ý tưởng này đã hoạt động một phần thông qua hệ thống bình duyệt đồng nghiệp hiện tại và các công bố khoa học trên tạp chí thương mại, cuộc cách mạng CNTT-TT đã làm lung lay, nếu không phải là tất cả cơ chế đó thì ít nhất cũng là hệ thống công bố và phổ biến khoa học. Khoa học mở trong thời đại thông tin đồng thuận với quan điểm tri thức được tạo ra từ nghiên cứu công có những đặc điểm của hàng hoá công vượt ra ngoài khái niệm hàng được phát triển trong thế kỷ 18, khi CNTT-TT cho phép mở rộng khả năng làm phong phú thêm hàng hoá công và mở rộng phạm vi người sử dụng.

Trong những năm gần đây, khoa học mở đã trở thành chủ đề được quan tâm trong các chương trình nghị sự chính sách. Mặc dù công nhận khoa học mở là một khái niệm

rộng hơn sự truy cập mở đến dữ liệu nghiên cứu và các công bố khoa học ở tất cả các giai đoạn nghiên cứu, tổng luận **“Khoa học mở: Các xu hướng chính sách gần đây”** do Cục Thông tin khoa học và công nghệ quốc gia biên soạn cung cấp tổng quan về xu hướng chính sách khoa học mở gần đây, tập trung đặc biệt vào các sáng kiến thúc đẩy truy cập nhiều hơn các kết quả nghiên cứu được tài trợ công, bao gồm cả công bố khoa học và dữ liệu nghiên cứu.

Xin trân trọng giới thiệu!

**CỤC THÔNG TIN KHOA HỌC VÀ
CÔNG NGHỆ QUỐC GIA**

I. TÍNH CẤP THIẾT VÀ TÁC ĐỘNG CỦA KHOA HỌC MỞ

Có nhiều cơ sở luận giải khác nhau về chính sách cho khoa học mở và dữ liệu mở (Hộp 1.1) cũng như các tiêu chí đánh giá tác động của chúng là đa dạng. Một mặt, việc tiếp cận nhiều hơn đến đầu vào và đầu ra của nghiên cứu khoa học có thể nâng cao hiệu quả và năng suất của hệ thống khoa học và nghiên cứu nhờ giảm chi phí do trùng lặp trong việc thu thập, tạo lập, chuyển giao và tái sử dụng dữ liệu và tài liệu khoa học; cho phép triển khai nhiều nghiên cứu hơn từ cùng một dữ liệu; và tăng thêm cơ hội cho các nhà nghiên cứu tham gia vào quá trình nghiên cứu trong nước cũng như toàn cầu.

Hộp 1.1. Các luận cứ cho khoa học mở và dữ liệu mở cho nghiên cứu và đổi mới sáng tạo

Các yếu tố sau đây thường liên quan đến tính mở trong khoa học và nghiên cứu:

- **Nâng cao hiệu quả trong khoa học.** Nỗ lực phát triển khoa học mở có thể làm tăng hiệu quả và năng suất của hệ thống nghiên cứu nhờ: 1) giảm sự trùng lặp và chi phí trong tạo lập, chuyển giao và tái sử dụng dữ liệu; 2) cho phép triển khai nhiều nghiên cứu hơn từ cùng một dữ liệu; 3) tăng thêm cơ hội tham gia vào quá trình nghiên cứu trong nước cũng như toàn cầu.
- **Tăng tính minh bạch và chất lượng trong quá trình thẩm định nghiên cứu,** nhờ cho phép nhân rộng hơn và xác nhận các kết quả khoa học.
- **Đẩy nhanh chuyển giao tri thức.** Khoa học mở có thể làm giảm sự chậm trễ trong việc tái sử dụng các kết quả nghiên cứu khoa học nhờ các bài báo và bộ dữ liệu và thúc đẩy nhanh hơn quá trình ứng dụng kết quả nghiên cứu vào đổi mới sáng tạo.
- **Tăng cường sự lan toả tri thức đến nền kinh tế.** Tăng khả năng tiếp cận đến các kết quả nghiên cứu được tài trợ công có thể thúc đẩy hiệu ứng lan tỏa và thúc đẩy sự đổi mới sáng tạo trong nền kinh tế cũng như nâng cao nhận thức và lựa chọn có ý thức của người tiêu dùng.
- **Giải quyết hiệu quả hơn các thách thức toàn cầu.** Thách thức toàn cầu đòi hỏi các hành động phối hợp quốc tế. Các phương pháp tiếp cận khoa học mở và dữ liệu mở có thể thúc đẩy các nỗ lực hợp tác và chuyển giao tri thức nhanh hơn giúp hiểu biết tốt hơn về các thách thức như biến đổi khí hậu hoặc già hóa dân số và có thể giúp xác định các giải pháp.
- **Thúc đẩy sự tham gia của công chúng vào khoa học và nghiên cứu.** Các sáng kiến khoa học mở và dữ liệu mở có thể nâng cao nhận thức và sự tin tưởng vào khoa học của công chúng. Trong một số trường hợp, sự tham gia nhiều hơn của công chúng có thể dẫn đến sự tham gia tích cực hơn vào các thí nghiệm khoa học và thu thập dữ liệu.

Nguồn: OECD (2013a), *Background paper for the TIP workshop on Open Science and Open Data*

Mặt khác, khả năng tiếp cận kết quả nghiên cứu tăng lên (ở cả 2 dạng: công bố khoa học và dữ liệu) không chỉ thúc đẩy hiệu ứng lan tỏa đến các hệ thống khoa học mà còn thúc đẩy một cách mạnh mẽ hơn hệ thống đổi mới sáng tạo. Với quyền truy cập hạn chế đến công bố khoa học và dữ liệu, các công ty và cá nhân có thể sử dụng và tái sử dụng các kết quả khoa học để tạo ra các sản phẩm và dịch vụ mới (Hộp 1.2).

Hộp 1.2. Các cơ hội phát sinh từ khai phá văn bản và dữ liệu

Khai phá văn bản và dữ liệu (Text and data mining - TDM) là tập hợp các kỹ thuật khoa học máy tính để phân tích; trích xuất tri thức và thông tin từ các bộ dữ liệu số lớn (ví dụ: dữ liệu lớn) ra các xu hướng và mô hình mà mắt người không thể nhận thấy. Khai phá văn bản và dữ liệu được các nhà nghiên cứu sử dụng trong tất cả các lĩnh vực, từ những nhà sử gia đang quét các tài liệu và lưu trữ lịch sử, cho đến các chuyên gia y tế sử dụng để tìm các mẫu phổ biến trong hồ sơ y tế. Khai phá văn bản và dữ liệu là kỹ thuật được sử dụng rộng rãi trong các lĩnh vực như thiên văn học và di truyền học cũng như ở cả khu vực công và tư nhân. Thuật toán khai phá văn bản và dữ liệu khảo sát những tập dữ liệu quy mô lớn không chỉ chứa số liệu mà còn cả những dạng biểu ghi số khác, như văn bản, hình ảnh và file âm thanh. Khai phá văn bản và dữ liệu cho phép sử dụng tất cả các kỹ thuật phổ biến, tạo ra sự liên kết giữa các lĩnh vực nghiên cứu chưa được kết nối và tạo ra cơ hội lớn cho đổi mới sáng tạo.

Việc khai phá văn bản và dữ liệu có ảnh hưởng quan trọng trong cộng đồng học thuật. Với số lượng bài báo được công bố (và chưa được công bố) ngày càng tăng (ước tính khoảng 50 triệu vào năm 2010), các nhà khoa học và nhà nghiên cứu không thể truy cập, đọc và phân tích các công bố theo cách thủ công. Khai phá văn bản và dữ liệu cung cấp khả năng truy cập, quét và phân tích các công bố bằng máy tính. Ngành công nghiệp xuất bản đang phát triển các dịch vụ để làm cho cơ sở dữ liệu tạp chí khoa học ngày càng tương thích và chuẩn hóa thuật ngữ để các nhà nghiên cứu có thể áp dụng các kỹ thuật khai phá văn bản và dữ liệu dễ dàng hơn.

Nghiên cứu về những kỹ thuật này đã tiến bộ đáng kể trong những năm gần đây. Số lượng bài báo được công bố về chủ đề khai phá văn bản và dữ liệu từ đầu những năm 1990 cho thấy Hoa Kỳ chiếm đến 46,6% số các công bố khoa học liên quan đến khai phá văn bản và dữ liệu, tiếp đến là Anh (11,1%), Đài Loan (8,8%), Canada (5,7%) và Trung Quốc (4,6%).

Liệu hệ thống bản quyền hiện tại đang quảng bá hay cản trở khai phá văn bản và dữ liệu đang là một câu hỏi mở. Theo một báo cáo của Ủy ban Hệ thống thông tin chung (Joint Information Systems Committee - JISC) gần đây về giá trị và lợi ích của khai phá văn bản (JISC, 2012), những thỏa thuận cấp phép chính là rào cản quan trọng đối với việc sử dụng kỹ thuật khai phá văn bản trong cộng đồng nghiên cứu và giáo dục đại học ở Anh. Phân tích của OECD gần đây đã nêu bật bối cảnh trong đó các khuôn khổ về sở hữu trí tuệ đã thay đổi đáng kể như thế nào. Trong bối cảnh phát triển này, cách quy trình bản quyền khai phá văn bản và dữ liệu không phải lúc nào cũng rõ ràng trong tất cả các khía cạnh pháp lý. Cũng trong báo cáo này, có một số bằng chứng cho rằng các nhà nghiên cứu ở một số khu vực (như Liên minh châu Âu và Brazil) bị cấm tham gia vào khai phá văn bản và dữ liệu do sự lo ngại về vi phạm bản quyền trong quá trình này.

Nguồn: Clark, J. (2013), Text Mining and Scholarly Publishing, Publishing Research Consortium

Hộp 1.2. Các cơ hội phát sinh từ khai phá văn bản và dữ liệu

Khai phá văn bản và dữ liệu (Text and data mining - TDM) là tập hợp các kỹ thuật khoa học máy tính để phân tích; trích xuất tri thức và thông tin từ các bộ dữ liệu số lớn (ví dụ: dữ liệu lớn) ra các xu hướng và mô hình mà mắt người không thể nhận thấy. Khai phá văn bản và dữ liệu được các nhà nghiên cứu sử dụng trong tất cả các lĩnh vực, từ những nhà sử gia đang quét các tài liệu và lưu trữ lịch sử, cho đến các chuyên gia y tế sử dụng để tìm các mẫu phổ biến trong hồ sơ y tế. Khai phá văn bản và dữ liệu là kỹ thuật được sử dụng rộng rãi trong các lĩnh vực như thiên văn học và di truyền học cũng như ở cả khu vực công và tư nhân. Thuật toán khai phá văn bản và dữ liệu khảo sát những tập dữ liệu quy mô lớn không chỉ chứa số liệu mà còn cả những dạng biểu ghi số khác, như văn bản, hình ảnh và file âm thanh. Khai phá văn bản và dữ liệu cho phép sử dụng tất cả các kỹ thuật phổ biến, tạo ra sự liên kết giữa các lĩnh vực nghiên cứu chưa được kết nối và tạo ra cơ hội lớn cho đổi mới sáng tạo.

Việc sử dụng khai phá văn bản và dữ liệu có ảnh hưởng quan trọng trong cộng đồng học thuật. Với số lượng bài báo được công bố (và chưa được công bố) ngày càng tăng (ước tính khoảng 50 triệu vào năm 2010), các nhà khoa học và nhà nghiên cứu không thể truy cập, đọc và phân tích các công bố theo cách thủ công. Khai phá văn bản và dữ liệu cung cấp khả năng truy cập, quét và phân tích các công bố bằng máy tính. Ngành công nghiệp xuất bản đang phát triển các dịch vụ để làm cho cơ sở dữ liệu tạp chí khoa học ngày càng tương thích và chuẩn hóa thuật ngữ để các nhà nghiên cứu có thể áp dụng các kỹ thuật khai phá văn bản và dữ liệu dễ dàng hơn.

Nghiên cứu về những kỹ thuật này đã tiến bộ đáng kể trong những năm gần đây. Số lượng bài báo được công bố về chủ đề khai phá văn bản và dữ liệu từ đầu những năm 1990 cho thấy Hoa Kỳ chiếm đến 46,6% số các công bố khoa học liên quan đến khai phá văn bản và dữ liệu, tiếp đến là Anh (11,1%), Đài Loan (8,8%), Canada (5,7%) và Trung Quốc (4,6%).

Liệu hệ thống bản quyền hiện tại đang quảng bá hay cản trở khai phá văn bản và dữ liệu đang là một câu hỏi mở. Theo một báo cáo của Ủy ban Hệ thống thông tin chung (Joint Information Systems Committee - JISC) gần đây về giá trị và lợi ích của khai phá văn bản (JISC, 2012), những thỏa thuận cấp phép chính là rào cản quan trọng đối với việc sử dụng kỹ thuật khai phá văn bản trong cộng đồng nghiên cứu và giáo dục đại học ở Anh. Phân tích của OECD gần đây đã nêu bật bối cảnh trong đó các khuôn khổ về sở hữu trí tuệ đã thay đổi đáng kể như thế nào. Trong bối cảnh phát triển này, cách quy trình bản quyền khai phá văn bản và dữ liệu không phải lúc nào cũng rõ ràng trong tất cả các khía cạnh pháp lý. Cũng trong báo cáo này, có một số bằng chứng cho rằng các nhà nghiên cứu ở một số khu vực (như Liên minh châu Âu và Brazil) bị cấm tham gia vào khai phá văn bản và dữ liệu do sự lo ngại về vi phạm bản quyền trong quá trình này.

Nguồn: Clark, J. (2013), Text Mining and Scholarly Publishing, Publishing Research Consortium

Có bằng chứng cho thấy cả hai nhân tố là nghiên cứu và hệ thống đổi mới sáng tạo có thể gặp khó khăn trong việc tiếp cận tài liệu khoa học không được truy cập mở. Một số cuộc điều tra cho thấy những khó khăn trong việc tiếp cận tài liệu khoa học trong cộng đồng học thuật ở Hoa Kỳ và Châu Âu. Ví dụ, theo Ủy ban Phát triển kinh tế, 15% số học giả Hoa Kỳ và Canada từ tất cả các ngành cho rằng mức độ tiếp cận đến tài liệu của họ không thỏa đáng. Khảo sát của Ware và Monkman (2008) cho thấy chỉ 66% các nhà khoa học ở châu Âu và Trung Đông cho rằng có mức độ truy cập tốt hoặc tuyệt vời (85% ở Hoa Kỳ). Và những con số của các khu vực khác thậm chí còn thấp hơn. Các cuộc khảo sát của Rowlands và Nicholas (2005) và Sparks (2005) cũng chỉ ra rằng rào cản truy cập đến tài liệu khoa học cho các nhà nghiên cứu là do chi phí mua cao.

Các nước đang phát triển đặc biệt có thể được hưởng lợi từ truy cập mở đối với tài

liệu khoa học. Chan, Kirsop và Arunachalam (2005) cho thấy theo khảo sát của Tổ chức Y tế Thế giới (WHO), ở những nước có tổng sản lượng quốc gia (Gross National Product - GNP) bình quân đầu người dưới 1.000 USD, khoảng 56% các tổ chức y tế không mua được tạp chí; ở các nước có GNP bình quân đầu người từ 1.000 - 3.000 USD, tỷ lệ những tổ chức y tế không mua được tạp chí thấp hơn, nhưng vẫn lên đến 34%. Đây là lý do tại sao đã có nhiều sáng kiến cung cấp cho các nước đang phát triển khả năng tiếp cận đến tài liệu khoa học. Ví dụ, Chương trình Research4Life là chương trình đối tác công tư giữa ba cơ quan của Liên Hợp Quốc, hai trường đại học và các nhà xuất bản thương mại lớn cho phép thư viện có đủ tiêu chuẩn và người dùng của họ truy cập đến các tạp chí khoa học quốc tế có bình duyệt đồng nghiệp, sách và cơ sở dữ liệu (CSDL) miễn phí hoặc chỉ với chi phí nhỏ. Trong một số ngành, những tạp chí truy cập mở đã được xuất bản ngay tại các nước đang phát triển, chẳng hạn như tạp chí Khoa học y tế châu Phi.

Các nhà khoa học và học giả không phải là nhóm duy nhất được hưởng lợi từ những nỗ lực phát triển khoa học mở. Nhu cầu về truy cập kết quả nghiên cứu chủ yếu dưới dạng dữ liệu và công bố khoa học của cá nhân và khu vực doanh nghiệp là khá cao. Ví dụ, dữ liệu sử dụng từ PubMedCentral cho thấy 25% người dùng cá nhân hàng ngày là từ các trường đại học, 17% từ các công ty, 40% là công chúng và phần còn lại là từ chính phủ hoặc các nhóm khác. Một nghiên cứu gần đây về những doanh nghiệp vừa và nhỏ (DNVVN) chuyên sâu về nghiên cứu và phát triển (NC&PT) cho thấy 48% DNVVN cho rằng kết quả nghiên cứu rất quan trọng cho hoạt động kinh doanh của họ và hơn 2/3 cho biết có khó khăn trong việc tiếp cận tài liệu nghiên cứu. Ware (2009) đã tiến hành một cuộc khảo sát nhỏ tại Anh về những doanh nghiệp ở quy mô vừa, những người trả lời khảo sát cho biết họ không tiếp cận được 10-20% số bài báo. Cuối cùng, các nhà khoa học cũng lập luận rằng việc công khai dữ liệu nghiên cứu có thể nâng cao sự hiểu biết của công chúng về khoa học, thực hành dựa trên bằng chứng và các sáng kiến khoa học của công dân.

1.1. Truy cập công bố khoa học

Phân tích gần đây cho thấy trong thập kỷ qua, tỷ lệ bài báo truy cập mở liên tục tăng trong tổng số các bài báo khoa học. Những ước tính về tạp chí truy cập mở có thể khác nhau phụ thuộc vào định nghĩa về truy cập mở và các phương pháp được sử dụng trong phân tích, cũng như thời gian tiến hành phân tích. Việc tiếp cận được bài báo truy cập mở cũng có thể phụ thuộc vào các đường dẫn truy cập mở khác nhau. Nghiên cứu gần đây của Archambault et al. (2014) cho thấy, tính đến đầu năm 2014, có hơn 50% số bài báo khoa học được xuất bản từ năm 2007 - 2012 là truy cập được và tải xuống miễn phí trực tuyến.

Theo Laakso và Björk (2012), khoảng 17% số bài báo khoa học được xuất bản năm 2011 và được xử lý trong Scopus (đây được cho là CSDL toàn diện nhất về các bài báo khoa học) được các nhà xuất bản tạp chí cho truy cập mở (truy cập mở vàng). Hầu hết các bài báo đều có thể truy cập mở ngay lập tức (12%) trong khi 5% còn lại được cho truy cập mở sau 12 tháng kể từ khi công bố. Tổng số bài báo truy cập mở lại chiếm 0,7% trong tổng số bài báo được xuất bản năm 2011. Các bài báo truy cập mở có bao gồm chi phí xử lý bài báo (Article processing charge - APC) chiếm 49% trong tổng số bài báo truy cập mở vàng. Bằng chứng sơ bộ dường như cho thấy chi phí xử lý bài báo không có mối liên hệ chặt chẽ với hệ số tác động của tạp chí, đặc biệt là trong trường hợp truy cập mở lại.

Việc ước tính mức độ truy cập mở xanh là phức tạp hơn, vì các nhà nghiên cứu không chỉ lưu trữ bài viết trên kho chính thức mà còn trên các trang web cá nhân hoặc trên cơ sở hạ tầng số khác. Một số ước tính thận trọng hơn, chẳng hạn như những ước tính của Björk et al. (2013), cho rằng tỷ lệ bài báo truy cập mở xanh chiếm khoảng 12% tất cả các tài liệu được công bố vào thời điểm tiến hành phân tích. Các ước tính khác đưa ra con số này ở mức trên 20% vào năm 2011.

Lewis (2012) cho rằng truy cập mở vàng (tức là khi tác giả công bố trong những tạp chí truy cập mở trực tuyến) có thể chiếm 50% số lượng bài báo khoa học từ năm 2012 - 2017 và đến 90% số bài báo từ năm 2020 - 2025. Tuy nhiên, Miguel, Chichilla-Rodríguez và de Moya-Anegón (2011) chỉ ra rằng tỷ lệ các tạp chí truy cập mở xanh vượt trội hơn so với tỷ lệ công bố truy cập mở vàng. Ngoài ra, truy cập mở xanh (tức là khi tác giả tự lưu trữ bài viết trong kho lưu trữ trực tuyến) gần đây được lập luận là phương thức hiệu quả và chi phí phù hợp cho các nhà tài trợ, tổ chức và những bên liên quan khác.

Khối lượng tài liệu truy cập mở có thể khác nhau đáng kể giữa các lĩnh vực khoa học, thường do các nền văn hóa chia sẻ khác nhau trong các lĩnh vực khác nhau. Ví dụ, hành vi khoa học mở trong lĩnh vực vật lý năng lượng cao đã bắt đầu từ nhiều thập niên trước, khi các học giả gửi bản thảo tiền xuất bản (bản thảo các ấn phẩm của họ chưa xuất hiện trong các tạp chí bình duyệt đồng nghiệp) cho các đồng nghiệp trên khắp thế giới.

Björk et al. (2010) và UNESCO (2012) chỉ ra rằng truy cập mở xanh được đặc trưng bởi sự thay đổi đáng kể của các lĩnh vực khoa học. Theo những nghiên cứu này, truy cập mở xanh cao hơn ở lĩnh vực vật lý và thiên văn học; khoa học trái đất và môi trường; toán học; khoa học xã hội, nghệ thuật và nhân văn, so với lĩnh vực y học, hóa học hoặc sinh học và di truyền học. Họ cũng ước tính rằng hai trong số các kho được thiết lập tốt nhất - PubMedCentral cho y sinh học và arXiv cho vật lý và toán học, chứa 38% tất cả

các bản sao truy cập mở xanh và 94% toàn bộ bản sao trong các kho lưu trữ theo chủ đề. Như vậy, PubMedCentral chiếm ưu thế trong lĩnh vực khoa học sự sống và arXiv trong lĩnh vực vật lý và toán học. Theo Laakso (2014), thời gian cấm (thời gian tối thiểu tính đến khi đưa ra truy cập mở) không giống nhau giữa các ngành khác nhau. Một cuộc khảo sát gần đây của OECD cho thấy việc tiếp cận mở đã tiến triển nhưng sự khác nhau giữa các ngành vẫn còn đáng kể.

Đã có nhiều cuộc tranh luận trong giới học thuật về việc liệu các công bố truy cập mở có nhận được nhiều trích dẫn hơn những công bố không truy cập mở, điều này dẫn đến nỗ lực đo lường lợi thế trích dẫn truy cập mở (open access citation advantage). Hầu hết các nghiên cứu được tiến hành về chủ đề này có xu hướng cho rằng truy cập mở làm tăng tác động trích dẫn. Tuy nhiên, không có sự đồng thuận chung về cường độ của sự gia tăng đó. Mặc dù là số ít nhưng có những nghiên cứu không cho thấy bất kỳ lợi thế trích dẫn nào.

Cũng có những lập luận cho rằng lợi thế trích dẫn truy cập mở là do thành kiến chất lượng (quality bias): Các nhà nghiên cứu có xu hướng công bố thông qua truy cập mở những công trình tốt nhất và đây là lý do tại sao họ nhận được nhiều trích dẫn hơn. Tuy nhiên, Gargouri et al. (2010) cho thấy lợi thế trích dẫn không phải do thành kiến chất lượng mà do lợi thế chất lượng từ sự tự lựa chọn của người dùng để sử dụng và trích dẫn, mà không có bất kỳ cản trở nào liên quan đến khả năng truy cập có chọn lọc của người mua tạp chí.

Gentil-Beccot, Mele và Brooks (2009) đã phát hiện ra rằng việc phổ biến miễn phí và trực tuyến ngay lập tức các bản in bài tạp chí tạo ra lợi thế trích dẫn đáng kể trong lĩnh vực vật lý năng lượng cao. Điều này là bởi ở lĩnh vực vật lý năng lượng cao, các nhà nghiên cứu có xu hướng trích dẫn bản thảo tiền xuất bản. Ngoài ra, việc phân tích luồng nhấp chuột trên Internet trong các thư viện số hàng đầu cho thấy các nhà khoa học vật lý năng lượng cao có xu hướng thích tải xuống các bài báo từ kho lưu trữ hơn là từ các trang web tạp chí (các nhà khoa học vật lý năng lượng cao tải xuống nhiều hơn từ 4 - 8 lần một bài viết từ arXiv so với phiên bản được xuất bản chính thức của nó trên trang web tạp chí).

Kể từ khi Viện Y tế quốc gia Hoa Kỳ (NIH) thực hiện chính sách truy cập mở bắt buộc đối với những công bố của các nghiên cứu được tài trợ, số lượng bài báo truy cập được trên PubMed Central gia tăng đáng kể. Tính đến cuối năm 2014, PubMed Central có hơn 3,2 triệu bài báo toàn văn về y sinh học và các lĩnh vực liên quan. Số lượng bài báo được truy xuất tăng gấp đôi trong giai đoạn 2011 và 2014; trong năm 2011, khoảng

500.000 người dùng truy cập đã truy cập truy cập PubMed Central 1 lần/1 tuần, tải xuống 1 triệu bài viết. Đến năm 2014, số lượng người dùng truy cập đã đạt gần 1 triệu mỗi ngày và số lượng bài viết đã truy xuất vượt trên 2 triệu. Hardisty và Haaga (2008) đã tiến hành nghiên cứu về các bài báo y khoa và nhận thấy rằng các bác sĩ chuyên về sức khỏe tâm thần đã đọc những bài báo truy cập mở nhiều hơn hai lần bình thường.

Dự án Nghiên cứu về xuất bản truy cập mở (Study of Open Access Publishing - SOAP) đã tiến hành một cuộc khảo sát quy mô lớn về thái độ của các nhà nghiên cứu đối với xuất bản truy cập mở. Theo bản tóm tắt kết quả, các nhà nghiên cứu được khảo sát cho thấy thái độ tích cực đối với việc xuất bản truy cập mở; tỷ lệ này còn cao hơn trong lĩnh vực khoa học nhân văn (90% phản ứng tích cực) so với các ngành khoa học kỹ thuật, tự nhiên (khoảng 80% phản ứng tích cực). Lợi ích cho toàn bộ cộng đồng khoa học được coi là cao hơn so với lợi ích cá nhân của các nhà nghiên cứu. Hầu hết những người được hỏi xác định tài chính là rào cản chính đối với xuất bản truy cập mở, tiếp theo là chất lượng của tạp chí truy cập mở. Rào cản về kinh phí không đồng đều giữa các ngành: Với khoa học sinh học, khoa học nông nghiệp và khoa học liên quan đến y học, rào cản tài chính được coi là lớn hơn so với lĩnh vực kinh doanh và quản trị, thiên văn học và vũ trụ, hoặc khoa học xã hội. Do đó, theo kết quả của dự án này, các chính sách truy cập mở có nhiều khả năng có hiệu ứng cao trong các lĩnh vực mà thực tiễn truy cập mở vẫn chưa được phát triển, thay vì trong những lĩnh vực mà hầu hết mọi thứ đã sẵn sàng cho truy cập mở trực tuyến.

Việc ước tính giá trị kinh tế của các bài báo nghiên cứu và những đóng góp liên quan của chúng đến phát triển kinh tế là không đơn giản. Các ước tính được nêu trong nghiên cứu của Houghton và Sheehan (2009) phân tích tác động của việc tăng khả năng tiếp cận tới đầu ra của các nghiên cứu được tài trợ công ở Úc và ước tính khả năng tiếp cận tăng tạo ra doanh thu 9 tỷ đô la Úc trong vòng 20 năm. Houghton, Rasmussen và Sheehan (2010) ước tính rằng chính sách truy cập mở đối với các cơ quan nghiên cứu liên bang của Hoa Kỳ trong giai đoạn chuyển tiếp 30 năm có thể tạo ra trị giá khoảng 1,6 tỷ USD và lên tới 1,75 tỷ USD nếu không có thời gian cấm. Nền kinh tế Mỹ có thể hưởng lợi khoảng 1 tỷ USD một cách trực tiếp và số tiền còn lại sẽ chuyển thành ngoại tác kinh tế cho các nước khác. Theo các tác giả, những con số này sẽ cao hơn đáng kể so với chi phí ước tính thực hiện lưu trữ truy cập mở.

1.2. Truy cập dữ liệu nghiên cứu

Chia sẻ dữ liệu luôn được coi là hoạt động quan trọng cho nghiên cứu khoa học và được cộng đồng khoa học chấp nhận rộng rãi. Có một số bằng chứng cho thấy, liên quan

đến truy cập mở công bố khoa học, việc chia sẻ dữ liệu có thể làm tăng tỷ lệ trích dẫn của những bài báo khoa học và thúc đẩy hành vi khoa học lành mạnh. Trong khi chia sẻ dữ liệu cho phép sử dụng và tái sử dụng dữ liệu từ các nhà nghiên cứu và cá nhân khác; nó cũng giúp chống lại hành vi sai trái, gian lận trong khoa học và nghiên cứu, góp phần cải thiện việc thu thập và quản lý dữ liệu. Vì tất cả những lý do này, thực tiễn chia sẻ dữ liệu thường được cộng đồng nghiên cứu coi là tích cực.

Chia sẻ dữ liệu không chỉ cho phép xác minh các kết quả khoa học mà còn cho phép phân tích lại dữ liệu cho những mục đích khác với mục đích ban đầu. Điều này không chỉ tăng cường việc sử dụng dữ liệu, mà còn thúc đẩy sự cạnh tranh của các ý tưởng và nghiên cứu, thúc đẩy sự hợp tác. Ví dụ, Murray et al. (2009) nhận thấy rằng tiếp cận mở đến tài liệu nghiên cứu làm tăng động lực cho nghiên cứu bổ sung bằng cách khuyến khích việc thiết lập các hướng nghiên cứu mới.

Chia sẻ dữ liệu làm giảm sự trùng lặp của các nhà nghiên cứu khi họ cố gắng thu thập cùng một bộ dữ liệu. Lakhani et al. (2007) thấy rằng việc công khai thông tin về vấn đề nghiên cứu cho nhiều nhóm khác bên ngoài là phương tiện hiệu quả để giải quyết các vấn đề khoa học. Ngoài ra, việc công khai thông tin vấn đề nghiên cứu còn tạo điều kiện cho việc giải quyết vấn đề ở ranh giới hoặc ngoài lĩnh vực chuyên môn của họ, nhờ sự chuyển giao tri thức từ lĩnh vực này sang lĩnh vực khác. Williams (2010) đã phát hiện ra rằng các bài báo phân tích trình tự bộ gen được công bố công khai đã giúp tạo ra nhiều hơn 30% so với những bài báo phân tích các chuỗi được các quyền sở hữu trí tuệ bảo hộ. Ưu thế trong công bố khoa học và thương mại được tạo ra bởi các trình tự mở là đáng chú ý. Đối với những công bố khoa học, việc chia sẻ dữ liệu đặc biệt quan trọng đối với các nhà nghiên cứu ở các nước đang phát triển vì họ có ít khả năng thu thập dữ liệu do tốn kém và mất nhiều thời gian.

Năm 2007, Viện Y tế Quốc gia Hoa Kỳ (NIH) đã triển khai chính sách nghiên cứu của Hiệp hội về Gen mở rộng (Genome-Wide Association Studies - GWAS) và CSDL về kiểu gen và kiểu hình (dbGaP). Tính đến tháng 12 năm 2013, Viện Y tế quốc gia Hoa Kỳ đã nhận được hơn 17.500 yêu cầu về dữ liệu dbGaP từ 2.221 nhà nghiên cứu, dẫn đến 924 công bố khoa học. 25% các công bố này xuất hiện trên các tạp chí hàng đầu, bao gồm *PLoS One* và *Nature Genetics*. Những nghiên cứu có nguồn gốc từ dữ liệu dbGaP có tác động đáng kể đến những khám phá khoa học, nó cho phép có những khám phá mới trong nghiên cứu về bệnh Alzheimer và các nghiên cứu tâm thần liên quan. Với thành công này, năm 2014, Viện Y tế quốc gia Hoa Kỳ đã mở rộng chính sách chia sẻ dữ liệu di truyền GWAS sang chính sách chia sẻ dữ liệu được thu thập trong các loại nghiên cứu hệ gen khác.

Tại Hoa Kỳ, ước tính dữ liệu do Cục Thời tiết quốc gia Hoa Kỳ công bố đóng góp vào sự phát triển của thị trường khí tượng khu vực tư nhân khoảng 1,5 tỷ USD. Trong năm 2008, hình ảnh vệ tinh của NASA Landsat về môi trường bề mặt Trái đất được cung cấp miễn phí trên Internet. Việc sử dụng CSDL này tăng từ 19.000 hình ảnh mỗi năm (khi mỗi hình ảnh được bán với giá 600 USD) đến 2,1 triệu hình ảnh mỗi năm. Các công ty hàng đầu ở Silicon Valley như Google (cụ thể là Google Earth) sử dụng những hình ảnh này và được công bố mở ước tính đã tạo ra giá trị trực tiếp hơn 100 triệu USD mỗi năm cho nền kinh tế Hoa Kỳ. Theo ước tính gần đây về Sáng kiến Dữ liệu mở của Hoa Kỳ, dữ liệu mở có tiềm năng tạo ra hơn 3 nghìn tỷ USD mỗi năm với giá trị gia tăng trong các lĩnh vực như tài chính, sản phẩm tiêu dùng, y tế, năng lượng và giáo dục.

Một nghiên cứu của Học viện GovLab của Hoa Kỳ, một tổ chức sử dụng công nghệ và đổi mới sáng tạo để giải quyết các vấn đề chung, tìm hiểu cách các công ty Hoa Kỳ sử dụng dữ liệu mở của chính phủ thông qua dự án Open Data 500. Dự án này phân tích các công ty có trụ sở tại Hoa Kỳ (bao gồm cả các công ty quốc tế) đang sử dụng dữ liệu mở của Chính phủ - một nguồn lực quan trọng cho doanh nghiệp của họ. Hầu hết các công ty trong nghiên cứu đều thuộc các ngành công nghệ, tài chính và kinh doanh/dịch vụ pháp lý. Theo nghiên cứu này, dữ liệu được sử dụng rộng rãi nhất là dữ liệu của Bộ Thương mại, tiếp theo là Bộ Y tế và Dịch vụ nhân sinh và Ủy ban Chứng khoán và Giao dịch.

Các sáng kiến dữ liệu mở của Ủy ban châu Âu được kỳ vọng sẽ tạo ra thu nhập hàng năm, khoảng 140 tỷ Euro. Ngoài ra, một nghiên cứu của OECD (2013b) ước tính thị trường liên quan đến thông tin khu vực công cho khu vực OECD có thể đem lại giá trị khoảng 500 tỷ USD cộng thêm 200 tỷ USD nếu các rào cản sử dụng được gỡ bỏ, kỹ năng được nâng cao và hạ tầng dữ liệu được cải thiện.

JISC (2014) đã tiến hành một nghiên cứu về tác động kinh tế của ba trung tâm dữ liệu của Anh (Trung tâm Dữ liệu kinh tế và xã hội, Trung tâm Dữ liệu khảo cổ học và Trung tâm Dữ liệu khí quyển) ước tính rằng lợi nhuận trên đầu tư của mỗi trung tâm này tương đương khoảng 2 - 10 lần trong vòng 30 năm.

II. TRUY CẬP MỞ CÔNG BỐ KHOA HỌC

2.1. Định nghĩa truy cập mở

Thuật ngữ “truy cập mở” lần đầu tiên được định nghĩa chính thức tại một cuộc họp ở Budapest vào đầu tháng 12 năm 2001, mặc dù trước đó nó đã được đề cập trong những sáng kiến khác (ví dụ, về dữ liệu, Nguyên tắc Bermuda được công bố năm 1996 đã yêu cầu phải đưa ra công khai và nhanh chóng các dữ liệu về hệ gen). Trong cuộc họp được gọi là Sáng kiến truy cập mở Budapest, “truy cập mở” được định nghĩa là “*sự có sẵn*

miễn phí của các tài liệu khoa học trên Internet, cho phép bất kỳ người dùng nào đọc, tải xuống, sao chép, phân phối, in ấn, tìm kiếm hoặc liên kết tới toàn văn của các bài báo này, thu thập thông tin để lập chỉ mục, chuyển chúng thành dữ liệu cho phần mềm hoặc sử dụng chúng cho bất kỳ mục đích hợp pháp nào khác mà không có rào cản về tài chính, pháp lý hoặc kỹ thuật khác ngoại trừ những điều liên quan đến việc truy cập được Internet. Hạn chế duy nhất đối với việc sao chép, phân phối và vai trò duy nhất cho bản quyền trong lĩnh vực này là nên cho phép các tác giả kiểm soát tính toàn vẹn của công trình của họ và quyền được thừa nhận và trích dẫn đúng cách”.

Sau sáng kiến truy cập mở Budapest, Tuyên bố Bethesda được đưa ra tại cuộc họp của các nhà khoa học, cơ quan tài trợ, thư viện, cộng đồng khoa học và nhà xuất bản được tổ chức vào tháng 4 năm 2003. Vào tháng 10 cùng năm, Hội Max Planck Đức triệu tập một cuộc họp về “truy cập mở đối với tri thức trong lĩnh vực khoa học và nhân văn”. Cuộc họp này mở rộng các cuộc thảo luận sang cả lĩnh vực nhân văn và đưa ra Tuyên bố Berlin về truy cập mở đối với tri thức trong lĩnh vực khoa học và nhân văn.

Kết quả nghiên cứu truy cập mở có thể bao gồm kết quả nghiên cứu khoa học gốc, chẳng hạn như bài báo và tài liệu chuyên khảo, cũng như dữ liệu thô và siêu dữ liệu, tài liệu nguồn, các bản trình bày số của tài liệu hình ảnh và đồ họa; tài liệu đa phương tiện học thuật. Trên cơ sở những tuyên bố và sáng kiến ở Berlin và Budapest, có thể nêu ra ba đặc điểm cơ bản sau của truy cập mở gồm: Truy cập miễn phí, phân phối tiếp và lưu trữ thích hợp.

Trong các tài liệu nghiên cứu trên, “truy cập mở” được hiểu là miễn phí (gratis) tài liệu nghiên cứu có sẵn (công bố, dữ liệu nghiên cứu) trên Internet mà không có hạn chế kỹ thuật. Một số tuyên bố còn nhấn mạnh tầm quan trọng của việc gỡ bỏ các hạn chế pháp lý (Budapest, Bethesda, Berlin), bao gồm khả năng tái sử dụng hợp pháp không bị cản trở.

Truy cập mở vàng, truy cập mở xanh và truy cập mở lai

Một số phương pháp thực hiện truy cập mở cùng tồn tại với xuất bản số trong đó những hình thức phổ biến được đặt tên là truy cập mở vàng, truy cập mở xanh và truy cập mở lai. Những thuật ngữ dựa trên màu sắc này, ban đầu phổ biến ở Anh, nhưng sau đó đã được chấp nhận trên toàn thế giới.

Theo mô hình ***truy cập mở vàng (gold open access)***, tác giả gửi bài viết đến các tạp chí truy cập mở - tức là, những tạp chí trực tiếp cung cấp quyền truy cập mở miễn phí đến bài báo trực tuyến. Tạp chí truy cập mở thường được cấp phép theo một trong sáu giấy phép Creative Commons (CC) (Hộp 2.1 và 2.2). Trong số các tạp chí và CSDL truy cập

mở thành công là Thư viện Khoa học công cộng (Public Library of Science - PLOS), Biomed Central và truy cập mở được Springer Open Choice Publishing cung cấp. Chi phí xuất bản và doanh thu trong mô hình truy cập mở vàng thường được thu hồi thông qua phí xử lý bài báo (là phí xuất bản mà cơ quan của tác giả hoặc quỹ tài trợ nghiên cứu phải chi trả). Ngoài ra, một tạp chí truy cập mở có thể thu phí qua việc bán bản in còn phiên bản điện tử thì có thể truy cập mở. Tạp chí truy cập mở vàng có thể dựa vào các phương thức tài trợ khác (chẳng hạn như quảng cáo hoặc được các quỹ tài trợ) mà không tính phí cho tác giả hoặc độc giả.

Hộp 2.1. Các loại hình truy cập mở khác nhau: Truy cập mở tự do và truy cập mở miễn phí

Do cách tiếp cận khác nhau về hạn chế pháp lý đối với tái sử dụng và sự mơ hồ của từ “mở” (và “miễn phí”), nên thường có đề xuất phân biệt giữa *truy cập mở miễn phí (gratis)* và *truy cập mở tự do (libre)*. Thuật ngữ ***truy cập mở miễn phí (gratis open access)*** được sử dụng để biểu thị sự sẵn có của các công bố khoa học (và đôi khi cả dữ liệu nghiên cứu) mà không đòi hỏi trả tiền hoặc có hạn chế về mặt kỹ thuật. Thuật ngữ ***truy cập mở tự do (libre open access)*** bao gồm quyền truy cập miễn phí nói trên nhưng có thêm một điều khoản bổ sung rõ ràng rằng tài liệu không bị hạn chế về mặt pháp lý.

Sự phân biệt miễn phí/tự do dẫn đến câu hỏi: Có bao nhiêu hạn chế cần phải được loại bỏ để tài liệu đủ điều kiện là tài liệu “truy cập mở tự do”? Có thể có nhiều câu trả lời khác nhau cho câu hỏi này. Ví dụ, một số người cho rằng “có nhiều hơn một loại hoặc mức độ truy cập mở tự do” và dường như được coi là “truy cập mở tự do” khi loại bỏ một số hạn chế; do đó, các tài liệu theo giấy phép CC NC hoặc CC ND (Hộp 2.2) sẽ được coi là truy cập mở tự do. Một cách tiếp cận ngược lại có thể bắt nguồn từ cách hiểu biết về “tự do” được đề xuất trong các tác phẩm của Richard Stallman liên quan đến phần mềm tự do. Theo định nghĩa phần mềm tự do của Stallman, phần mềm tự do (libre) tồn tại khi người dùng có thể thực hiện tất cả bốn quyền tự do được định nghĩa trong đó: 1) Tự do chạy chương trình; 2) Tự do nghiên cứu chương trình; 3) Tự do phân phối lại; 4) Tự do phân phối các bản của các phiên bản sửa đổi, trong khi đó việc này bao gồm toàn bộ nội dung độc quyền theo pháp luật bản quyền. Cách hiểu về “tự do” theo đề xuất của Stallman đối với phần mềm đã được chấp nhận cho các tác phẩm vô hình khác trong định nghĩa về tác phẩm văn học tự do (Definition of Free Cultural Works - DFCW).

Ngay cả sự tự do theo định nghĩa của Stallman đối với phần mềm và được chuyển đổi sang các tác phẩm văn học trong DFCW không có nghĩa là không còn bất cứ sự hạn chế nào cả. Ví dụ, làm cho một nguồn tài nguyên trở thành miễn phí theo các định nghĩa này không có nghĩa là nó có thể được sử dụng theo cách thức cấu thành sự vi phạm quyền nhân thân hoặc quyền riêng tư. Một số nghĩa vụ của người dùng được nêu rõ ràng, chẳng hạn như nghĩa vụ thẩm quyền (điều BY - Hộp 2.2 và 2.3) hoặc các điều khoản copyleft (bảo lưu mọi quyền) theo đó cấm việc hạn chế quyền tự do của người khác (GPL hoặc CC BY-SA, Hộp 2.2 và 2.3).

Nguồn: Krzysztof Siewicz

Một loại hình đặc thù của tạp chí truy cập mở vàng được gọi là ***tạp chí lai (hybrid journal)***, ở đó một tạp chí thương mại cho phép một số bài báo cụ thể được truy cập mở, miễn là các chi phí xử lý bài báo được các tác giả hoặc cơ quan của họ chi trả. Các tạp chí lai có ưu điểm là tăng thêm nơi để các tác giả có thể công bố bài báo thông qua truy cập mở, vì ngày càng nhiều tạp chí thương mại cho phép hình thức truy cập mở lai này. Tuy nhiên, theo một số người, mô hình này có thể liên quan đến việc trả tiền hai lần cho cùng

một nội dung, một lần ở dạng chi phí xử lý bài báo và sau đó là tiền thông qua các đăng ký thuê bao của độc giả.

Hộp 2.2. Mô hình cấp phép Creative Commons (CC)

Hệ thống giấy phép Creative Commons (CC) có 4 yếu tố tùy chọn, gồm:

1. **BY** - Ghi nhận đóng góp (Attribution): Đây là yếu tố bắt buộc cho tất cả các giấy phép CC. Bạn cho phép người khác sao chép, phân phối, truyền đạt và trình diễn các tác phẩm thuộc quyền tác giả của bạn và các tác phẩm phái sinh từ tác phẩm này và thậm chí là sử dụng với mục đích thương mại theo cách thức được quy định bởi tác giả hoặc người cấp phép.
2. **NC** - Phi thương mại (NonCommercial): Bạn cho phép người khác sử dụng tác phẩm của bạn chỉ với mục đích phi thương mại.
3. **SA** - Chia sẻ với điều kiện như nhau (ShareAlike): Khi bạn sửa đổi, thay đổi hoặc xây dựng dựa trên tác phẩm theo điều kiện này, bạn có thể phân phối tác phẩm là kết quả của việc sử dụng tác phẩm gốc theo giấy phép giống hoặc tương tự giấy phép mà bạn được cấp từ tác phẩm gốc
4. **ND** - Không có phái sinh (NoDerivative): Người được bạn cho phép sử dụng tác phẩm của mình theo điều kiện này không được phép làm biến đổi tác phẩm của bạn cũng như không được tạo ra tác phẩm phái sinh từ việc sử dụng này.

Dựa vào 4 yếu tố tùy chọn, hệ thống Creative Commons có 6 giấy phép tiêu chuẩn, gồm:

1. **CC BY** (Giấy phép ghi nhận đóng góp - Creative Commons Attribution): "Giấy phép này cho phép những người khác có thể phổ biến, pha trộn (lấy 2 hoặc nhiều tài nguyên đang tồn tại và kết hợp chúng để tạo ra 1 tài nguyên mới), chỉnh sửa và xây dựng lại dựa trên công trình của bạn, thậm chí về mặt thương mại, miễn là họ phải ghi nhận đóng góp của bạn về sáng tạo gốc. Đây là giấy phép thích hợp được khuyến nghị cho mục đích phổ biến và sử dụng được tối đa các tài liệu cấp phép"

2. **CC BY-SA** (Giấy phép Ghi nhận đóng góp Chia sẻ với Điều kiện như nhau - Creative Commons Attribution ShareAlike): "Giấy phép này cho phép người dùng khác pha trộn, chỉnh sửa và xây dựng dựa trên công trình của bạn ngay cả với mục đích thương mại, miễn là họ ghi nhận đóng góp của bạn và cấp giấy phép cho những sáng tạo mới theo cùng các điều khoản của bản gốc. Giấy phép này thường được so sánh với giấy phép "bảo lưu mọi quyền" (copyleft) của phần mềm nguồn mở và tự do. Tất cả tác phẩm mới đều dựa trên tác phẩm của bạn sẽ cùng loại giấy phép, vì vậy, mọi dẫn xuất cũng sẽ cho phép sử dụng thương mại. Đây là giấy phép được Wikipedia sử dụng và được khuyến nghị cho những tài liệu có lợi từ việc tích hợp nội dung từ Wikipedia và các dự án cấp phép tương tự".

3. **CC BY-NC** (Giấy phép ghi nhận đóng góp phi thương mại - Creative Commons Attribution NonCommercial): "Giấy phép này cho phép người khác pha trộn, chỉnh sửa và xây dựng theo công trình của bạn cho mục đích phi thương mại và mặc dù các tác phẩm mới của họ vẫn phải ghi công và không mang tính thương mại, họ không bị bắt buộc cấp phép cho tác phẩm phái sinh của họ theo cùng điều khoản"

4. **CC BY-NC-SA** (Ghi nhận đóng góp Phi Thương mại Chia sẻ với Điều kiện Như nhau - Creative Commons Attribution NonCommercial ShareAlike): "Giấy phép này cho phép người khác pha trộn, tùy chỉnh, và xây dựng dựa trên tác phẩm của bạn với mục đích phi thương mại, với điều kiện chúng nâng cao uy tín của Bạn và cho phép cấp phép đối với các tác phẩm sáng tạo mới của họ theo các điều khoản giống hệt".

5. **CC BY-ND** (Ghi nhận đóng góp Không Phái sinh - Creative Commons Attribution NoDerivative): "Giấy phép này cho phép phân phối lại vì mục đích thương mại và phi thương mại, miễn là nó được chuyển giao và giữ nguyên vẹn (không đổi), cùng với ghi công tác giả"

6. **CC BY-NC-ND** (Ghi nhận đóng góp Phi Thương mại Không Phái sinh - Creative Commons Attribution NonCommercial NoDerivative): "Giấy phép này là loại giấy cấp phép hạn chế nhất trong số sáu loại Giấy phép chính của Creative Commons, cho phép người khác tái phân phối. Giấy phép này chỉ cho phép người khác tải về các tác phẩm của bạn và chia sẻ chúng cho người khác với điều kiện chúng phải làm tăng uy tín của bạn và người được cho phép không được thay đổi tác phẩm của bạn theo bất kì cách nào hoặc sử dụng chúng với mục đích thương mại".

Nguồn: www.creativecommons.org

Việc xuất bản truy cập mở đang thay đổi nhanh chóng và hiện tại người ta đang đặt ra câu hỏi: Phương pháp nào là hay nhất khi nói đến chi phí xử lý bài báo và tính bền vững của chúng? Theo Johnson (2015), cả các tổ chức và nhà xuất bản cần phải liên tục điều chỉnh các quy trình và hệ thống vì ngày càng có nhiều bài báo được xuất bản thông qua hình thức truy cập mở vàng. Ngoài ra, “mối quan hệ phức tạp giữa chi phí xử lý bài báo, cấp phép và thời hạn cấm vận vẫn là một chủ đề gây tranh luận”.

Một hình thức truy cập mở khác - được coi là bổ sung cho mô hình truy cập mở vàng trong khi thực tế có thể chồng chéo với nó - được gọi là **mô hình truy cập mở xanh (green open access)** và liên quan đến việc các tác giả tự lưu trữ bản thảo tiền xuất bản hoặc bản sau in của họ. Các tác giả cung cấp quyền truy cập vào các bài viết gốc hoặc đã xuất bản của họ bằng cách tự tạo ra các bản sao điện tử miễn phí cho tất cả mọi người. “Tự lưu trữ” đề cập đến việc cung cấp quyền truy cập mở đến một công bố bằng cách tải nó lên Internet, thường trong kho lưu trữ hoặc thông qua trang web của tác giả. Tự lưu trữ truy cập mở không phải là tự xuất bản; nó không phải là xuất bản trực tuyến mà không cần sự kiểm soát chất lượng (bình duyệt đồng nghiệp); và nó không phải là dành cho những tác phẩm mà tác giả mong muốn được trả tiền, chẳng hạn như sách hoặc các bài viết trên báo chí. Trong một số trường hợp cụ thể, bản sao xanh của một bài báo có thể được lưu trữ bởi nhà xuất bản thay vì tác giả.

Trong khi mô hình truy cập mở xanh đáp ứng các yêu cầu chính của truy cập mở - cụ thể là truy cập miễn phí, khả năng sao chép, sử dụng và phân phối tác phẩm và lưu trữ - thực tế là các công bố được xuất bản thông qua các kênh truyền thống, ở đó các tác giả chỉ giữ lại một số quyền nhất định đối với công bố/dữ liệu nghiên cứu của họ. Các bài báo tự lưu trữ thường đi kèm với văn bản giấy phép cho người dùng biết họ có thể làm gì và không được làm gì với bài báo đó. Ngược lại, việc phát hành kết quả nghiên cứu bằng mô hình truy cập vàng nói chung đảm bảo khả năng truy cập rộng hơn và ngay lập tức, khả năng tái sử dụng rõ ràng hơn, khả năng hiển thị và “khả năng tài trợ” của đầu ra của nghiên cứu trên Internet. Cho tới nay, vì truy cập mở vàng liên quan đến chi phí xử lý bài báo và áp dụng với một số tạp chí nhất định, truy cập mở vàng có thể hạn chế lựa chọn các nhà nghiên cứu công bố kết quả.

Các mô hình xuất bản truy cập mở vàng và truy cập mở xanh có những ưu điểm và nhược điểm khác nhau (Bảng 2.1). Ở phần lớn những quốc gia trả lời khảo sát của OECD, các chính sách truy cập mở xanh chiếm ưu thế, nhưng có nhiều quốc gia và tổ chức cấp kinh phí để trang trải chi phí truy cập mở vàng. Ví dụ, chính sách của chương trình Horizon 2020 của Ủy ban châu Âu cho phép tác giả lựa chọn từ cả hai kênh truy cập mở xanh và vàng. Chi phí xử lý bài báo phát sinh trong việc xuất bản truy cập mở có đủ

điều kiện để được hoàn trả từ các khoản tài trợ của Horizon 2020. Ngay cả khi kênh truy cập mở vàng được sử dụng, một số cơ quan tài trợ và tổ chức yêu cầu phải gửi bài báo vào một kho lưu trữ mở.

Bảng 2.1. Các sắc thái khác nhau của xuất bản truy cập mở

	Mô tả	Tính năng bổ sung	Ưu điểm	Nhược điểm
Truy cập mở vàng (vi lợi nhuận)	Bài báo có sẵn ngay tại thời điểm xuất bản, trong hầu hết các trường hợp, chi phí xuất bản được chi trả thông qua chi phí xử lý bài báo do tác giả hoặc nhà tài trợ trả	Những bài báo được xuất bản trên các tạp chí, trong đó tất cả bài báo đều có thể truy cập ngay lập tức Chỉ cung cấp truy cập mở đến bài báo, hoặc trong một số trường hợp tài liệu liên quan bổ sung như bộ dữ liệu, số liệu, hình ảnh hoặc video.	Truy cập mở vàng thường có quyền sử dụng lại đầy đủ theo Creative Commons (CC-BY) Truy cập ngay vào bài viết mà không có thời gian cấm vận Các nhà xuất bản ngày càng cung cấp dịch vụ sáng tạo về xuất bản truy cập mở vàng (ví dụ: Sage Open, Springer Open Choice). Một số nhà xuất bản như PLOS miễn phí xử lý cho các tác giả không được cơ quan tài trợ kinh phí	Các chi phí xử lý bài báo cần được chi trả bởi các nhà tài trợ hoặc cơ quan nghiên cứu Lựa chọn nơi công bố bị hạn chế
Truy cập mở vàng (không vi lợi nhuận)	Bài viết có sẵn ngay từ thời điểm xuất bản	Các bài báo được xuất bản trong các tạp chí đều có thể truy cập. Chỉ cung cấp truy cập mở đến bài viết, hoặc trong một số trường hợp, truy cập đến các tài liệu liên quan bổ sung như tập dữ liệu, số liệu, hình ảnh hoặc video	Truy cập mở vàng thường có quyền sử dụng lại đầy đủ theo giấy phép Creative Commons (CC-BY) Truy cập ngay vào bài báo mà không có thời gian cấm vận. Chi phí thấp hơn cho tác giả	Lựa chọn nơi công bố bị hạn chế
Truy cập mở lại	Tạp chí truy cập mở lại là tạp chí thương mại, trong đó một số bài báo là truy cập mở. Hình thức này thường yêu cầu thanh toán phí xử lý bài báo hoặc phí xuất bản cho nhà xuất bản Nếu nhận được thanh toán cho truy cập mở, khoản thanh toán này được bù trừ trong quá trình quy định của nhà xuất bản	Chi phí xử lý bài báo thay đổi đáng kể tùy thuộc vào tạp chí Truy cập mở ngay lập tức chỉ có thể được cung cấp cho bài báo hoặc, trong một số trường hợp, truy cập đến các tài liệu liên quan bổ sung như bộ dữ liệu, số liệu, hình ảnh hoặc video	Tác giả muốn công bố trên tạp chí truy cập mở không bị giới hạn trong số lượng ít tạp chí truy cập mở "đầy đủ"; họ có thể công bố trong các tạp chí truy cập mở lại của các nhà xuất bản chính. Nguy cơ cho các nhà xuất bản thấp hơn khi thiết lập tạp chí truy cập mở lại bởi vì họ vẫn nhận được thu nhập từ nguồn mua tạp chí	Số lượng tạp chí giới hạn (mặc dù đang tăng), nơi có thể xuất bản Giá của chi phí xử lý bài báo lại cao dẫn đến sự thu hút ít tùy chọn truy cập mở lại
Truy cập mở xanh (trường hợp các bản thảo trước in)	Bản thảo trước in của bài báo (tức là trước khi gửi cho một tạp chí để bình duyệt đồng nghiệp) có thể truy cập trực tuyến, thường là từ các trang web cá nhân hoặc của cơ quan, hoặc kho lưu trữ	Các bản thảo trước in của bài báo thường có sẵn trực tuyến, bao gồm cả bài đã xử lý hoặc bản không công bố của bài báo	Bài viết có thể được tải lên nhiều địa điểm: các kho lưu trữ cơ quan hoặc theo lĩnh vực, trang web cá nhân Không có thêm chi phí cho tác giả (không cần phải trả phí xử lý bài báo) Tác giả hoàn toàn tự do trong việc lựa chọn địa điểm xuất bản	Truy cập mở xanh thường không có quyền sử dụng lại đầy đủ theo giấy phép Creative Commons (CC-BY) Phiên bản của bài viết được gửi lưu trực tuyến không qua bình duyệt đồng nghiệp Chi phí bảo trì của kho lưu trữ
Truy cập mở xanh (bản thảo tác giả được chấp nhận)	Các phiên bản của bài báo (tức là sau khi trải qua bình duyệt đồng nghiệp và có sửa đổi theo yêu cầu để được tạp chí chấp nhận) có thể truy cập trực tuyến, thường ở kho lưu trữ cơ quan hoặc kho lưu trữ cá nhân hoặc trang web cá nhân.	Có thể có cấm vận tùy theo tạp chí	Không có thêm chi phí cho tác giả (không cần phải trả tiền xử lý bài báo) Tác giả có toàn quyền lựa chọn địa điểm công bố Phiên bản của bài báo có sẵn trực tuyến đã được bình duyệt đồng nghiệp	Truy cập mở xanh của bản thảo được chấp nhận thường liên quan đến thời gian cấm vận khai thác (thường đến 24 tháng). Truy cập mở xanh thường không có quyền sử dụng lại đầy đủ theo giấy phép Creative Commons (CC-BY) Chi phí bảo trì của kho lưu trữ

Cả hai mô hình truy cập mở vàng và truy cập mở xanh hiện đang được các chính phủ, cơ quan tài trợ, trường đại học và trung tâm nghiên cứu cũng như các bên liên quan đến khoa học mở khác ở các nước thành viên OECD phát huy. Trong khi mô hình truy cập mở xanh là mô hình mặc định cho truy cập mở cơ bản ở phần lớn các nước OECD, các biến thể của mô hình truy cập mở vàng đã nổi lên để đáp ứng mong muốn của tác giả được công bố trong các tạp chí hàng đầu và nỗ lực của các nhà xuất bản để phát triển các dịch vụ mới để các mô hình tìm kiếm lợi nhuận từ việc cạnh tranh. Một số quốc gia đã đưa ra các sáng kiến để thúc đẩy sự xuất hiện của các mô hình kinh doanh sáng tạo xung quanh quyền truy cập mở.

2.2. Xuất bản truy cập mở và bảo vệ quyền sở hữu trí tuệ

Bản quyền và các quyền sở hữu trí tuệ khác đóng vai trò quyết định trong cách thức công bố khoa học đang được cộng đồng khoa học phổ biến và sử dụng khi họ nhận mạnh các đòi hỏi cấp phép. Luật bản quyền và quyền sở hữu trí tuệ liên quan có thể hỗ trợ, cản trở hoặc là trung gian đối với việc thực hiện các nguyên tắc truy cập mở trong phổ biến các kết quả khoa học.

Mặc dù việc thực hiện những nguyên tắc truy cập mở được dựa trên các thỏa thuận hợp đồng giữa tác giả, nhà xuất bản và các trường đại học, khung pháp luật là yếu tố quyết định cách dàn xếp bản quyền được hình thành. Phương thức luật bản quyền quy định quyền và xác nhận giới hạn và ngoại lệ đối với các quyền này đóng vai trò then chốt trong thỏa thuận cấp phép.

Cách thức chính để xác định xem chế độ sở hữu trí tuệ nói chung và luật bản quyền nói riêng hỗ trợ truy cập mở trong khoa học như thế nào là xem xét phạm vi quyền được cấp cho các kết quả khoa học và các ngoại lệ có thể ở các nước khác nhau và cách thức mà cả các quyền và ngoại lệ hình thành cơ sở cho việc khai thác quyền của kết quả khoa học, hoặc bởi nhà nghiên cứu, tổ chức của nhà khoa học hoặc nhà xuất bản. Quyền càng mạnh (Hộp 2.1) thì khả năng thực hiện các quyền đó không qua cấp phép càng cao (Hộp 2.2 và 2.3), bằng cách đặt hạn chế sử dụng theo mô hình truyền thống hoặc bằng cách thúc đẩy tái sử dụng không hạn chế theo nguyên tắc truy cập mở. Các yếu tố khác - chẳng hạn như nguyên tắc về nguồn gốc (ví dụ như câu hỏi về những gì được quyền sở hữu trí tuệ bảo hộ hoặc không được bảo hộ), sở hữu quyền và thời hạn bảo hộ - là yếu tố quan trọng trong việc đánh giá khuynh hướng của chế độ sở hữu trí tuệ đối với mô hình truy cập mở, nhưng ít quyết định hơn.

Một đặc điểm quan trọng khác của luật bản quyền là sự cân bằng giữa những sử dụng được bảo lưu cho chủ bản quyền và sử dụng miễn phí - hoặc nói cách khác, sử dụng

mà không cần sự cho phép. Có thể có hai loại hình sử dụng: Loại hình thứ nhất là những sử dụng không thuộc phạm vi bảo hộ của luật bản quyền và sử dụng được bảo hộ nhưng được phép miễn trừ không cần xin phép, mặc dù đôi khi cần phải có “bồi thường hợp lý”. Loại hình thứ hai là “những ngoại lệ và hạn chế đối với bản quyền” thường cho phép sử dụng tác phẩm có bản quyền mà không cần sự cho phép của tác giả hoặc chủ sở hữu bản quyền. Các trường hợp ngoại lệ và giới hạn như vậy có thể được quy định trong luật thành văn hoặc án lệ, bao gồm sử dụng được bảo hộ bởi điều khoản “giao dịch hợp lý” hoặc “giao dịch công bằng”. Lập luận về chính sách cho những ngoại lệ và giới hạn có thể thay đổi tùy theo luật pháp của mỗi quốc gia. Chúng có thể gồm bảo vệ quyền hiến pháp hoặc quyền cơ bản, quy định trong thực tiễn và cạnh tranh của ngành công nghiệp, sự phổ biến kiến thức, hoặc những cân nhắc về thất bại thị trường. Ví dụ về loại thứ hai này có thể được thấy trong các trường hợp như trích dẫn; minh họa cho việc giảng dạy; bài báo về các chủ đề kinh tế, chính trị hoặc tôn giáo hiện tại; và sao chép tác phẩm với mục đích thông tin các sự kiện. Một lần nữa, với cách tiếp cận bảo hộ tối thiểu của các công ước, các bên ký kết được tự do ban hành những ngoại lệ và giới hạn khác, miễn là chúng chỉ áp dụng cho một số trường hợp đặc biệt không xung đột với việc khai phá bình thường tác phẩm và không xâm phạm một cách cố ý lợi ích hợp pháp của tác giả.

Mặc dù không thể đưa ra một sự thống nhất hoàn toàn, một số quốc gia tích cực khuyến khích việc tuân thủ các nguyên tắc truy cập mở đối với các công bố kết quả nghiên cứu được tài trợ công và dường như thúc đẩy cải cách vấn đề bản quyền theo hướng linh hoạt và thân thiện cho hoạt động nghiên cứu. Vương quốc Anh là một ví dụ điển hình về điều này: Trong khi Hội đồng Nghiên cứu đã áp dụng chính sách được gọi là “Con đường Vàng” (Golden Road) tới truy cập mở, trong đó yêu cầu các nhà nghiên cứu công bố kết quả theo Giấy phép Creative Commons Attribution 4.0, các nhà lập pháp cũng đã tiếp tục áp dụng các ngoại lệ mới về bản quyền, bao gồm một ngoại lệ cụ thể cho khai phá văn bản và dữ liệu. Hội đồng nghiên cứu Đức có thể chưa chính thức quy định nghĩa vụ đảm bảo truy cập mở của người tài trợ, nhưng các nhà lập pháp đã sửa đổi luật bảo vệ bản quyền tác giả để họ dễ dàng tuân thủ những thỏa thuận hợp đồng với nhà xuất bản.

Khung pháp lý ở châu Âu

Năm 2012, Ủy ban châu Âu đã công bố báo cáo của mình tới Nghị viện châu Âu và Hội đồng dưới nhan đề “Hướng tới việc tiếp cận tốt hơn với thông tin khoa học: Tăng cường lợi ích của đầu tư công trong nghiên cứu”. Theo các quan sát của Ủy ban, “các cuộc thảo luận về hệ thống phổ biến khoa học truyền thống tập trung vào việc tiếp cận ấn phẩm khoa học - tạp chí và tài liệu chuyên khảo. Tuy nhiên, việc cải thiện khả năng tiếp

cần đến dữ liệu nghiên cứu (kết quả thực nghiệm, quan sát và thông tin do máy tính tạo ra), tạo nền tảng cho phân tích định lượng làm cơ sở cho nhiều ấn phẩm khoa học ngày càng trở nên quan trọng”.

Hộp 2.3. Các mô hình và tùy chọn cấp phép

Khi thảo luận về các mô hình cấp phép và các tùy chọn, chúng ta cần chú ý đến sự phân biệt giữa 1) miễn phí và 2) truy cập mở tự do (Hộp 2.1). Cũng có thể có sự khác biệt song song về quan hệ pháp lý: 1) quan hệ giữa tác giả và nhà phân phối (nhà xuất bản, kho lưu trữ, v.v.); và 2) mối quan hệ giữa nhà phân phối (nhà xuất bản, kho lưu trữ, tác giả) và người dùng. Điều này dẫn đến 4 tình huống khác nhau:

1. Truy cập mở miễn phí trong quan hệ giữa tác giả và nhà phân phối

Vì quyền truy cập mở miễn phí không nhất thiết phải liên quan đến việc cấp giấy phép cho người dùng cuối, mối quan hệ pháp lý giữa tác giả và nhà phân phối có thể tuân theo mẫu chuẩn của hợp đồng xuất bản. Chỉ cần hợp đồng cho phép các nhà phân phối một cách rõ ràng làm cho những công bố có sẵn trên Internet có thể truy cập được

Nếu các bên mong muốn, hợp đồng có thể bao gồm một số điều khoản tùy chọn, ví dụ như nghĩa vụ rõ ràng đối với nhà phân phối để cung cấp quyền truy cập mở tài liệu, thời hạn cấm vận, v.v..

2. Truy cập mở tự do trong quan hệ giữa tác giả và nhà phân phối

Truy cập mở tự do liên quan đến việc cấp giấy phép miễn phí (CC BY, CC BY-SA) cho người dùng cuối, vì vậy điều này ngụ ý rằng người cấp phép phải có các quyền cần thiết hoặc là người giữ bản quyền. Phạm vi của các quyền nhận được phải tương đương hoặc nhiều hơn quyền được cấp trong giấy phép CC, nhưng các quyền không nhất thiết phải được chuyển cho người cấp phép CC tương lai. Giấy phép CC có thể được cấp dưới dạng giấy phép bổ sung từ người có giấy phép của chủ sở hữu bản quyền. Do đó, tác giả có thể giữ bản quyền đầy đủ và chỉ ủy quyền cho nhà phân phối cấp phép lại cho tác phẩm theo giấy phép CC. Hiện có cách thức dàn xếp pháp lý chẳng hạn như nhà phân phối nhận được từ tác giả giấy phép CC giống như bất kỳ ai khác. Một tùy chọn truyền thống hơn: Tác giả chuyển bản quyền cho nhà xuất bản để họ cấp giấy phép CC cho người dùng cuối. Hợp đồng giữa tác giả và nhà phân phối có thể bao gồm nhiều điều khoản phụ, như đã đề cập ở trên - như nghĩa vụ cung cấp quyền truy cập mở, thời gian cấm vận, v.v..

3. Truy cập mở miễn phí trong mối quan hệ giữa nhà phân phối và người dùng

Theo truy cập mở miễn phí, không cần có giấy phép cho người dùng cuối vì không cần thiết phải loại bỏ bất kỳ hạn chế bản quyền nào để đủ điều kiện là truy cập mở miễn phí. Nhưng có thể là tài liệu truy cập mở miễn phí được kèm theo một số giấy phép, chẳng hạn như giấy phép CC NC hoặc CC ND. Trong trường hợp này, lưu ý ở điểm 4 dưới đây sẽ được xem xét.

4. Truy cập mở tự do trong mối quan hệ giữa nhà phân phối và người dùng

Trong truy cập mở tự do, người dùng cuối được cấp giấy phép tự do mà nó loại bỏ bản quyền khỏi hạn chế có xung đột với quyền tự do của người dùng. Giấy phép có thể được coi là miễn phí nếu nó cho phép sử dụng không giới hạn, miễn phí, không độc quyền một tác phẩm và các dẫn xuất của nó. Tuy nhiên, nó có thể chứa một số nghĩa vụ không ảnh hưởng đến cốt lõi của tự do, chẳng hạn như công nhận quyền hoặc bảo lưu mọi quyền (**copyleft**)¹.

(1) **Copyleft** là quyền cho phép tác phẩm sáng tạo có thể được cộng đồng sửa đổi một cách tự do và yêu cầu tất cả các phiên bản được sửa đổi và mở rộng từ tác phẩm sáng tạo đó cũng phải tự do như vậy. Những người tin tưởng vào phong trào Copyleft đấu tranh để đưa ra lựa chọn thay thế cho chế độ hà khắc hiện hành về kiểm soát sở hữu trí tuệ. Phong trào này đưa ra khẩu hiệu châm biếm của mình “Bảo vệ những gì sai trái.

Nguồn: Krzysztof Siewicz

Báo cáo đánh dấu bước đi chính thức mới trên con đường thực hiện truy cập mở đến kết quả nghiên cứu được tài trợ công trong khoa học và nhân văn ở châu Âu. Các công bố khoa học không còn là các yếu tố duy nhất của chính sách truy cập mở: Các kết quả nghiên cứu mà từ đó tạo ra các công bố cũng phải được làm cho công chúng truy cập được. Để thực hiện chính sách này, Ủy ban châu Âu đã thiết lập sáng kiến thí điểm về truy cập mở với các bài báo nghiên cứu được bình duyệt đồng nghiệp trong Chương trình khung nghiên cứu FP7 (Seventh Research Framework Programme - FP7), hay còn gọi là dự án OpenAire, để đảm bảo rằng các kết quả từ nghiên cứu do quỹ tài trợ được phổ biến rộng rãi và hiệu quả nhất có thể cũng như đảm bảo khai thác và tác động tối đa cho các nhà nghiên cứu và những đối tượng khác.

Chính sách truy cập mở châu Âu không ràng buộc với các nước thành viên EU, những nước được tự do quyết định áp dụng chính sách phù hợp nhất với nhu cầu của cộng đồng khoa học của chính họ. Điều này dẫn đến một bức tranh không đồng nhất về chính sách truy cập mở trên khắp châu Âu, từ con đường vàng bắt buộc cho các công bố và dữ liệu nghiên cứu được Hội đồng Nghiên cứu Anh (RCUK) quy định, đến ưu tiên truy cập mở vàng ở Hà Lan, rồi đến con đường xanh cho công bố ở Đức. Ví dụ, trong những năm gần đây, Hội đồng nghiên cứu quốc gia của Anh và Hà Lan đã ban hành chính sách theo đó các khoản tài trợ nghiên cứu sẽ chỉ được trao cho những ứng viên cam kết xuất bản kết quả của họ, cả công bố và dữ liệu theo điều kiện truy cập mở.

Khung pháp lý ở Hoa Kỳ

Theo chỉ thị của Chính phủ Hoa Kỳ do Văn phòng Chính sách KH&CN (Chỉ thị Truy cập Công cộng) ban hành, tất cả các Bộ/cơ quan liên bang có chi phí nghiên cứu và phát triển hơn 100 triệu USD mỗi năm được yêu cầu xây dựng kế hoạch để làm cho kết quả nghiên cứu được công bố của các nghiên cứu do liên bang tài trợ có thể truy cập được sau một năm công bố. Ngoài ra, Dự luật Truy cập hợp lý đến nghiên cứu KH&CN (Fair Access to Science and Technology Research Act - FASTR) được trình lên Quốc hội Hoa Kỳ vào đầu năm 2013. Nếu được thông qua, dự thảo luật này sẽ hỗ trợ các mục tiêu của Chỉ thị với cấu trúc mạnh mẽ hơn của công cụ lập pháp. Nội dung dự luật này tương tự như Chỉ thị, với sự khác biệt nhỏ nhưng đáng kể về số lượng và các cơ quan chịu tác động, thời hạn cấm vận tối đa (sáu tháng thay cho một năm) và tham chiếu đến các công bố (cả hai) hoặc các dữ liệu nghiên cứu khác (Chỉ thị).

Những sửa đổi gần đây về luật bản quyền ở Đức...

Ở Đức, bổ sung gần đây cho Luật Bản quyền đề cập trực tiếp đến vấn đề cấp giấy phép các công bố khoa học được tạo ra nhờ tài trợ công như sau:

"Tác giả của đóng góp khoa học là kết quả của hoạt động nghiên cứu mà tài trợ công chiếm ít nhất 50% và xuất hiện trong tạp chí định kỳ được xuất bản ít nhất 2 lần mỗi năm, ngay cả khi họ đã trao cho nhà xuất bản hoặc người biên tập toàn quyền sử dụng được làm cho đóng góp của họ truy cập được một cách công cộng ở dạng bản thảo được chấp nhận sau 12 tháng kể từ lần xuất bản đầu tiên, trừ khi điều này phục vụ cho mục đích thương mại. Nguồn gốc của công bố đầu tiên phải được chỉ dẫn. Bất kỳ thỏa thuận sai trái nào liên quan đến những thiệt hại của tác giả đều không có hiệu lực".

Quy định này nhằm cho phép tác giả nộp công bố khoa học được tạo ra bởi nghiên cứu được tài trợ công (ít nhất 50%) và được xuất bản trong tạp chí định kỳ (ít nhất là 6 tháng/1 lần) được làm cho bản thảo được chấp nhận được truy cập công cộng cho các mục đích phi thương mại sau thời gian cấm vận 12 tháng. Quyền tái công bố không thể bị giới hạn bởi các thỏa thuận hợp đồng, có nghĩa là ngay cả khi tác giả đã cấp phép tất cả quyền độc quyền cho nhà xuất bản, tác giả vẫn sẽ được hưởng quyền tái công bố.

... và Khung pháp lý ở Anh

Gần đây, Anh đã thực hiện một số sửa đổi đối với khung bản quyền quốc gia nhằm tạo thuận lợi cho việc tiến hành nghiên cứu và phân tích khoa học. Phần 29A là đáng quan tâm được thể hiện như sau:

“29A. Bản sao cho phân tích văn bản và dữ liệu cho nghiên cứu phi thương mại

1. Việc tạo một bản sao tác phẩm của người có quyền tiếp cận hợp pháp đến với tác phẩm không vi phạm bản quyền trong tác phẩm miễn là đáp ứng yêu cầu:

(a) Bản sao được tạo ra để một người có quyền truy cập hợp pháp đến tác phẩm có thể tiến hành phân tích tính toán bất kỳ thứ gì được ghi lại trong tác phẩm cho mục đích duy nhất của nghiên cứu cho mục đích phi thương mại, và

(b) bản sao được kèm theo sự ghi nhận đóng góp đầy đủ (trừ khi điều này là không thể vì lý do thực tiễn hay cách khác).

2. Trong trường hợp bản sao tác phẩm đã được thực hiện theo phần này, bản quyền đối với tác phẩm bị vi phạm nếu:

(a) Bản sao được chuyển giao cho bất kỳ người nào khác, trừ khi việc chuyển giao được chủ sở hữu bản quyền cho phép hoặc

(b) Bản sao được sử dụng cho bất kỳ mục đích nào khác ngoài mục đích được đề cập trong mục (1) (a), trừ khi việc sử dụng được chủ sở hữu bản quyền cho phép.

3. Nếu bản sao được tạo ra theo phần này được xử lý tiếp theo

- (a) Nó được coi là bản sao vi phạm vì mục đích của giao dịch đó, và
- (b) Nếu giao dịch đó vi phạm bản quyền, nó sẽ được coi là bản sao vi phạm cho tất cả các mục đích tiếp theo.
4. Trong mục (3) "xử lý" có nghĩa là bán hoặc cho thuê, hoặc được đưa ra chào bán hoặc trưng bày để bán hoặc cho thuê.
5. Trong trường hợp một điều khoản của hợp đồng có mục đích ngăn chặn hoặc hạn chế việc tạo bản sao mà nó được điều chỉnh ở phần này, sẽ không vi phạm bản quyền, thì điều khoản đó là vô hiệu”.

Quy định này cơ bản làm rõ rằng việc sao chép tác phẩm cho mục đích khai phá văn bản và dữ liệu là không vi phạm bản quyền trong tác phẩm với điều kiện mục đích duy nhất là nghiên cứu cho mục đích phi thương mại. Quy định này cho thấy rõ rằng bất kỳ thỏa thuận hợp đồng nào nhằm hạn chế khả năng tạo bản sao theo quy định tại phần này là vô hiệu. Tuy nhiên, ngoại lệ không bao gồm Quyền CSDL riêng (*sui generis* database right - SGDR). Quan điểm của Chính phủ Anh là ngoại lệ giao dịch hợp lý quyền SGDR đối với các sử dụng khoa học phi thương mại cung cấp sự bảo vệ song song phù hợp cho nhu cầu hiện tại.

2.3. Công bố truy cập mở và tác động pháp lý của nó

Lựa chọn của tác giả về con đường công bố có thể được thúc đẩy bởi một số yếu tố, bao gồm cả sự mong đợi của họ về tác động mà công bố có thể có. Trong khi truy cập mở có thể làm tăng khả năng tiếp cận đến công bố, các nhà nghiên cứu phải đối đầu với thực tế rằng các tạp chí được xây dựng dựa trên các mô hình xuất bản truyền thống thường có uy tín lâu dài và mạng lưới đánh giá được thiết lập. Việc được tạp chí chấp nhận mang đến cho tác giả và công bố của họ một số danh tiếng mặc định được gắn liền với các bài báo đã được xuất bản trước đây trong tạp chí, bất kể các trích dẫn mà tài liệu có liên quan đến “chuẩn” cho nhan đề cụ thể đó. Trong phạm vi một số ưu đãi, như các quyết định khuyến mãi và tài trợ nghiên cứu, số liệu trắc lượng thư mục, quyết định xuất bản theo truy cập mở không thể được xem xét độc lập với những yếu tố như chất lượng và mức độ liên quan. Điều này cần được tính đến trong việc đánh giá mô hình và chính sách khác nhau.

Các tác giả khoa học mong muốn công bố tác phẩm qua các kênh truy cập mở cần xem xét ý nghĩa pháp lý của sự lựa chọn đó. Xuất bản truy cập mở (thông qua cả mô hình truy cập mở xanh và truy cập mở vàng) đòi hỏi sự nỗ lực hợp tác liên quan đến những bên liên quan khác nhau - đặc biệt là các tác giả khoa học, tổ chức nghiên cứu, cơ quan

tài trợ và nhà xuất bản. Các tác giả khoa học đặc biệt cần phải làm theo một quy trình nhiều bước:

- Xác định phạm vi quyền cần thiết để cung cấp quyền truy cập mở
- Có được các quyền này - hoặc, như thường xảy ra, tác giả phải ngăn chặn xử lý chúng ngay từ đầu
- Thực hiện các hoạt động nhất định để cho phép truy cập mở đến tác phẩm được công bố trong tạp chí (hoặc được chấp nhận để công bố trong tạp chí).

Phạm vi của các quyền cần thiết để truy cập mở tùy thuộc vào loại hình truy cập mở. Có nhiều quyền cần thiết hơn cho truy cập mở tự do vì nó ngụ ý việc cấp giấy phép miễn phí (CC BY hoặc CC BY-SA - Hộp 2.2 và 2.3), và không ai có thể cấp ít quyền hơn so với các quyền của tác giả. Đối với truy cập mở tự do, tác giả chỉ cần có quyền đưa công trình lên Internet là được. Các quyền khác có thể được chuyển giao cho nhà xuất bản hoặc bên thứ ba.

Tất cả quyền cần thiết cho truy cập mở được các tác giả giữ trong những trường hợp đơn giản, thường áp dụng cho tình huống điển hình của nghiên cứu được tài trợ công (ví dụ: tác giả là tác giả duy nhất, việc công bố không diễn ra trong quá trình được tuyển chọn và không có hợp đồng nào khác được ký kết liên quan đến công bố). Trong trường hợp đây là công trình chung, công trình được thực hiện trong khi được tuyển chọn, các tác phẩm phái sinh (ví dụ như bản dịch của các công bố khác), cũng như khi một hợp đồng với nhà xuất bản đã được ký kết, thường cần thiết phải có sự đồng ý của bên tương ứng để thực hiện truy cập mở. Cơ sở pháp lý chính xác cho sự đồng ý như vậy và hình thức phù hợp,... sẽ là khác nhau tùy theo hoàn cảnh và luật hiện hành. Bất cứ khi nào có nhu cầu về sự chấp thuận nó, cần được thông báo và rõ ràng - đặc biệt trong trường hợp truy cập mở tự do (bao gồm việc cấp phép miễn phí).

Những trường hợp phức tạp nảy sinh trong các tình huống mà khi những công trình là đối tượng của hợp đồng vì vấn đề tự do theo hợp đồng, trên thực tế không có giới hạn nào trong kết hợp các mối quan hệ pháp lý từ những hợp đồng đó. Điều này có nghĩa là có thể cần có tư vấn pháp lý cá nhân để xác định liệu hợp đồng có để tác giả có đủ quyền và nếu không, thì làm thế nào để thu xếp cho họ. Do đó, giải pháp đơn giản nhất là tránh các hợp đồng mà chúng có thể có những hạn chế và chỉ sử dụng giấy phép không độc quyền (chủ sở hữu bản quyền có thể cấp số lượng giấy phép không độc quyền không giới hạn, bao gồm cả giấy phép miễn phí). Ngoài ra, truy cập mở nên được thương lượng rõ ràng từ trước với các bên khác nữa. Mặc dù có một số ví dụ về những nhà xuất bản từ chối thương lượng hợp đồng chuyển giao quyền tác giả, mọi cuộc đàm phán như vậy

thường có thể thực hiện được và một sự thay đổi đơn giản của hợp đồng như vậy thành giấy phép không độc quyền sẽ không đòi hỏi tư vấn pháp lý phức tạp.

Trong một số trường hợp, các nhà xuất bản không hạn chế quyền truy cập mở trong hợp đồng biểu mẫu chuẩn của họ, hoặc thậm chí cung cấp rõ ràng quyền đó, khi họ thực hiện truy cập mở bằng con đường vàng hoặc xanh. Tuy nhiên, những hợp đồng chuẩn có thể có điều khoản liên quan đến chuyển giao bản quyền với việc cấp phép lại cho tác giả một số tùy chọn truy cập mở hạn chế (ví dụ như miễn phí nhưng không phải là tự do; sau thời hạn cấm vận; không phải bản được xuất bản cuối cùng mà chỉ có bản thảo đã được chấp nhận của tác giả).

Hiện tại, đối với một tác giả đơn lẻ muốn làm cho công bố của mình được truy cập mở, quy trình có thể phức tạp; ví dụ: các cuộc đàm phán cá nhân, có thể là trở ngại đối với tác giả. Các tổ chức như nhà tuyển dụng (ví dụ: các trường đại học), cũng như các cơ quan tài trợ có nhiều công cụ pháp lý khác nhau để giúp tác giả cá nhân nhờ có quan hệ với các nhà xuất bản. Thứ nhất, các tổ chức có thể ràng buộc về pháp lý tác giả phải tuân thủ chính sách truy cập mở nào đó. Thứ hai, các tổ chức cũng thường đóng vai trò là nhà xuất bản, nghĩa là họ có thể cung cấp giải pháp ngay lập tức cho các tác giả - những người không yêu cầu công bố lên tạp chí được đánh giá cao. Thứ ba, ngoài nghĩa vụ pháp lý, các tổ chức có thể cung cấp cho tác giả sự hỗ trợ có tổ chức trong mối quan hệ của họ với nhà xuất bản, cả về mặt pháp lý và kỹ thuật.

Một trong những bài học từ việc quan sát thực hiện thành công các chính sách truy cập mở ở cấp độ tổ chức để thúc đẩy truy cập mở là chính sách cần vượt ra ngoài sự hỗ trợ chung chung cho truy cập mở.

Chính sách cũng nên kiểm tra loại truy cập mở nào được theo đuổi và liệu nó là đòi hỏi hay đơn giản chỉ là khuyến nghị.

Bình duyệt đồng nghiệp mở

Mục đích chính của quá trình bình duyệt đồng nghiệp là duy trì chất lượng cao của kết quả nghiên cứu được công bố, và điều này được các nhà nghiên cứu đánh giá cao. Để tạo điều kiện thuận lợi và biểu hiện không thiên vị về ý kiến người đánh giá, việc bình duyệt theo truyền thống hoặc là ẩn danh đơn (người đánh giá biết thông tin về tác giả, nhưng tác giả không biết họ) hoặc ẩn danh kép (người đánh giá và tác giả ẩn danh). Bình duyệt ẩn danh kép đòi hỏi các tác giả phải chuẩn bị bản thảo ẩn danh hoặc thông tin tài trợ (không có tên và các tham chiếu đến các công trình trước đó, v.v..)

Tuy nhiên, việc bình duyệt truyền thống có một số nhược điểm. Thứ nhất, mô hình mang lại ít khuyến khích cho người đánh giá: họ không được ghi nhận khi họ dành thời gian và công sức để viết đánh giá. Thứ hai, quá trình này có thể không hoàn toàn minh bạch. Một số nhà phê bình cho rằng việc truy cập rộng rãi hơn vào dữ liệu có thể làm tăng cơ hội tránh việc công bố các bài báo có kết quả hoặc kết luận không chính xác. Đánh giá sau công bố được xem như là một thay thế, mặc dù nó cũng không phải là không có những thách thức.

Một số nghiên cứu nêu lên mối lo ngại về chất lượng của kết quả nghiên cứu được công bố trên các tạp chí khoa học. Mặc dù rất khó để ước tính số lượng bài báo khoa học được công bố có kết luận không chính xác, nhưng số lần trả lại bài viết có thể cung cấp thông tin về các vấn đề liên quan đến việc xác minh bằng kết quả khoa học bình duyệt đồng nghiệp truyền thống. Grieneisen và Zhang (2012) đã khảo sát 42 CSDL thư mục lớn nhất của một số lĩnh vực nghiên cứu chính và các trang web của nhà xuất bản. Họ nhận thấy rằng số lần trả lại bài đã tăng đáng kể sau năm 2001. Việc trả lại bài xảy ra nhiều hơn trong các lĩnh vực như y học, khoa học sự sống và hóa học so với các lĩnh vực như toán học, vật lý, kỹ thuật và khoa học xã hội. Theo nghiên cứu này, nguyên nhân chính của việc trả lại bài viết là do hành vi sai trái trong xuất bản (như đạo văn, tác quyền hoặc bản quyền), tiếp theo là sử dụng sai dữ liệu hoặc giải thích sai và nghiên cứu không đúng (chẳng hạn như sử dụng dữ liệu gian lận hoặc sửa đổi).

Nhằm giải quyết một phần các vấn đề nêu trên, những mô hình bình duyệt đồng nghiệp mở đang được đề xuất để bổ sung cho các mô hình bình duyệt truyền thống. Ví dụ, F1000Research, là tạp chí truy cập mở về khoa học sự sống, đã áp dụng mô hình trọng tài mở, trong đó các phản hồi của người bình duyệt đồng nghiệp đánh giá và tác giả định công bố công khai.

III. DỮ LIỆU NGHIÊN CỨU MỞ

3.1. Nghiên cứu khoa học dựa vào vào dữ liệu

Dữ liệu và đo lường luôn là vấn đề cơ bản đối với khoa học. Sự ra đời của các công cụ và phương pháp khai thác dữ liệu mới đã thúc đẩy một số đề xuất về sự xuất hiện của “khám phá khoa học dựa vào dữ liệu”, dựa trên cách sử dụng truyền thống của mô tả thực nghiệm, mô hình lý thuyết và mô phỏng các hiện tượng phức tạp. Điều này có ý nghĩa quan trọng cho cách phát hiện xảy ra như thế nào trong tất cả các lĩnh vực khoa học. Khoa học dựa vào dữ liệu cho phép phát triển các thí nghiệm khoa học cũng như mô phỏng thuật toán dựa trên máy tính, ngay cả trong những lĩnh vực truyền thống ít sử dụng dữ liệu hơn. OECD (2015) đề xuất coi dữ liệu là cơ sở hạ tầng (Hộp 3.1).

Hộp 3.1. Dữ liệu là cơ sở hạ tầng

Hầu hết dữ liệu (không phải tất cả) về nguyên tắc có thể được coi là tài nguyên cơ sở hạ tầng, vì chúng là “phương tiện được chia sẻ cho nhiều hướng” đáp ứng cả ba tiêu chí của tài nguyên cơ sở hạ tầng được Frischmann (2012) nhấn mạnh. Các tiêu chí này bao gồm:

Dữ liệu là hàng hóa không có tính cạnh tranh. Tính cạnh tranh hay tính không cạnh tranh về tiêu thụ mô tả mức độ nguồn tiêu thụ của một tài nguyên ảnh hưởng đến tiềm năng của nguồn tài nguyên đó trong đáp ứng nhu cầu của người khác. Dữ liệu là hàng hóa không có tính cạnh tranh bởi có thể được tiêu thụ chủ yếu không giới hạn số lần. Đặc tính này là nguồn gốc của sự lan tỏa đáng kể, liên quan đến sự gia tăng của năng suất tổng các yếu tố tổng hợp (TFP). Trong khi mọi người cho rằng phúc lợi xã hội được tối đa hóa khi một hàng hóa cạnh tranh thuần túy được tiêu thụ bởi người cần đánh giá nó cao nhất và cơ chế thị trường nói chung là phương tiện hiệu quả nhất cho phân phối hàng hóa đó và cho phân bổ các nguồn lực cần thiết để sản xuất hàng hóa, thì điều này không phải lúc nào cũng đúng đối với hàng hóa không cạnh tranh. Phúc lợi xã hội không được tối đa hóa khi hàng hóa chỉ được tiêu thụ bởi người đánh giá trả giá cao nhất mà cần được tiêu thụ bởi bất kỳ ai cần nó và coi trọng nó. Tối đa hóa việc tiếp cận đến hàng hoá không cạnh tranh về lý thuyết sẽ tối đa hóa phúc lợi xã hội, vì mọi lợi ích cá nhân bỏ sung đều không mất thêm chi phí.

Dữ liệu là tư liệu sản xuất. Dữ liệu không phải là hàng hoá tiêu dùng tốt hay hàng hóa bán thành phẩm. Trong hầu hết các trường hợp, dữ liệu có thể được phân loại là tư liệu sản xuất. Hệ thống tài khoản quốc gia (SNA) của Liên hợp quốc (2008) định nghĩa hàng hóa tiêu dùng hoặc dịch vụ tiêu dùng là “một sản phẩm được sử dụng [...] để đáp ứng trực tiếp nhu cầu; nhu cầu hoặc mong muốn cá nhân hoặc nhu cầu tập thể của các thành viên trong cộng đồng”. Ngược lại, hàng hóa bán thành phẩm và tư liệu sản xuất được sử dụng làm đầu vào để sản xuất hàng hóa khác. Tư liệu sản xuất, theo OECD, là “hàng hóa, trừ nguyên liệu đầu vào và nhiên liệu, được sử dụng để sản xuất hàng hóa và/hoặc dịch vụ khác”. Ngược lại với hàng hóa bán thành phẩm, chẳng hạn như nguyên liệu thô (ví dụ: dầu), tư liệu sản xuất không bị sử dụng hết, cạn kiệt hoặc biến đổi khi được sử dụng làm đầu vào để sản xuất hàng hóa khác. Trong hầu hết các trường hợp, dữ liệu thường được sử dụng làm đầu vào cho hàng hóa hoặc dịch vụ và điều này đặc biệt đúng đối với khối lượng dữ liệu lớn, có nghĩa là thay vì kết thúc vì chúng là công cụ chứ không phải là đầu cuối. Hơn nữa, dữ liệu cũng không phải là một lợi ích bán thành phẩm vì chúng không bị cạn kiệt khi được sử dụng, do tính chất không cạnh tranh của chúng. Cũng tương tự như với hầu hết các tư liệu sản xuất, dữ liệu có thể mất giá đặc biệt khi nó trở nên không phù hợp hơn cho mục đích sử dụng.

Dữ liệu là đầu vào vạn năng. Như Frischmann (2012) giải thích, “tài nguyên cơ sở hạ tầng cho phép nhiều hệ thống (thị trường và phi thị trường) hoạt động và đáp ứng nhu cầu từ nhiều loại người dùng khác nhau”. Chúng không phải là đầu vào đã được tối ưu hóa cho mục đích hạn chế cụ thể, nhưng “chúng cũng đa năng và cung cấp chức năng cơ bản và đa năng”. Dữ liệu thường có thể được thu thập cho mục đích cụ thể và trong trường hợp dữ liệu cá nhân, đặc điểm kỹ thuật định trước của mục đích. Tuy nhiên, về mặt lý thuyết không có giới hạn về việc dữ liệu có thể được sử dụng cho mục đích nào, và trên thực tế nhiều lợi ích của việc chia sẻ dữ liệu phát sinh từ việc tái sử dụng dữ liệu theo cách đã hoặc không thể dự kiến trước khi dữ liệu được thu thập. Ngoài ra, việc tái sử dụng dữ liệu được tạo trong một lĩnh vực có thể dẫn đến thông tin chi tiết hơn khi được áp dụng trong một lĩnh vực khác. Điều này là rõ ràng trong trường hợp dữ liệu khu vực công, nơi các bộ dữ liệu được sử dụng ban đầu cho các mục đích quản trị hành chính, được các doanh nhân sử dụng lại để tạo ra những dịch vụ mới chưa bao giờ được dự kiến khi dữ liệu ban đầu được tạo ra.

Nguồn: Source: OECD (2015), *Data-driven Innovation for Growth and Well-being*

Các công cụ phân tích dữ liệu (như máy học hoặc kỹ thuật nhận dạng mẫu) được các nhà khoa học sử dụng ngày càng nhiều để thu kiến thức về hiện tượng và kiểm tra hoặc xác nhận các mô hình. Các tập dữ liệu quy mô lớn cho phép thử nghiệm và mô phỏng trên máy tính, ngay cả trong những lĩnh vực mà các thí nghiệm truyền thống trong

phòng thí nghiệm là không thể hoặc quá khó để tổ chức. Với các tập dữ liệu đủ lớn, máy tính có thể phát hiện các phương thức và mối quan hệ phức tạp mà các nhà nghiên cứu không nhìn thấy được. Ngoài ra, khoa học dữ liệu và các nghiên cứu và thử nghiệm dựa trên thuật toán là cơ hội cho đổi mới sáng tạo và khám phá khoa học: nhiều lĩnh vực như khoa học máy tính hoặc khoa học dữ liệu hiện đang khai phá dữ liệu lớn nhằm tạo ra cơ hội để phát triển các thuật toán mới và hiệu quả hơn cho phân tích dữ liệu được sử dụng bởi những nhà nghiên cứu hoạt động trong các lĩnh vực khác nhau (trong cả khu vực công và tư nhân).

Các công cụ mới như các máy gia tốc hạt lớn hoặc kính viễn vọng và cả Internet như một công cụ thu thập dữ liệu, là những công cụ cho các phát triển mới trong khoa học, vì chúng đã thay đổi quy mô và độ chi tiết của dữ liệu được thu thập. Chẳng hạn, thiết bị Digital Sky Survey được vận hành năm 2000, trong tuần đầu tiên đã thu thập được dữ liệu nhiều hơn so với những gì thu được qua kính thiên văn trong lịch sử thiên văn học hoặc kính viễn vọng vô tuyến SKA mới có thể tạo ra tới 1 petabyte dữ liệu cứ 20 giây một lần. Hơn nữa, năng lực ngày càng tăng của phân tích dữ liệu đã làm cho nó có thể đem đến những hiểu biết sâu từ những bộ dữ liệu rất lớn một cách hợp lý và nhanh chóng. Ví dụ, trong di truyền học, máy giải trình tự gen ADN dựa trên các phân tích dữ liệu lớn giờ đây có thể đọc được khoảng 26 tỷ ký tự của mã di truyền của con người trong vài giây. Điều này đi đôi với việc giảm đáng kể chi phí của phân tích trình tự ADN trong 5 năm qua.

Những phát triển mới này, được mở rộng trên tất cả công cụ khoa học và các lĩnh vực khoa học, cho thấy tiềm năng phát triển khoa học mới, đề ra nhiều vấn đề mới cho chính sách khoa học. Những vấn đề này bao gồm từ kỹ năng mà các nhà khoa học và nhà nghiên cứu phải nắm vững với nhu cầu về khuôn khổ cho các kho dữ liệu tuân thủ theo tiêu chuẩn quốc tế về lưu giữ dữ liệu; đặt giao thức lưu giữ và siêu dữ liệu chung; bảo vệ tính toàn vẹn của dữ liệu; thiết lập các quy tắc cho các cấp truy cập khác nhau; và xác định quy tắc phổ biến giúp tạo điều kiện kết hợp những tập dữ liệu và cải thiện khả năng tương tác.

Tính đa dạng của dữ liệu khoa học

Dữ liệu nghiên cứu khoa học thay đổi rất lớn - về chủng loại và khối lượng, cũng như trong sử dụng và giá trị lâu dài. Có thể có bốn loại dữ liệu nghiên cứu đặc biệt quan trọng trong nghiên cứu:

Dữ liệu quan sát thu được từ các kính thiên văn, các vệ tinh, mạng cảm biến, khảo sát và các công cụ khác ghi lại thông tin lịch sử hoặc hiện tượng một lần (chẳng hạn như

dữ liệu thiên văn từ kính thiên văn Sloan Digital Sky Survey, SDSS) hoặc dữ liệu này cũng bao gồm dữ liệu nghiên cứu khoa học xã hội (như điều tra nhân khẩu học). Trong nhiều trường hợp, các dữ liệu này không thể được sao chép và nên được giữ lại.

Dữ liệu thực nghiệm có thể được thu thập từ máy có thông lượng cao (như máy gia tốc hạt), thông qua thử nghiệm lâm sàng; y sinh và dược phẩm, hoặc thông qua các thí nghiệm được kiểm soát khác. Việc lưu giữ đặc biệt quan trọng đối với dữ liệu thử nghiệm ở những nơi không khả thi hoặc đạo đức để nhân rộng việc thu thập dữ liệu. Điều này bao gồm một số dữ liệu liên quan đến con người và những loài đang bị đe dọa.

Dữ liệu tính toán được tạo ra từ các mô phỏng tính toán quy mô lớn. Mặc dù dữ liệu như vậy có thể được tái tạo bằng cách chạy lại mô phỏng nhưng có hai lý do tại sao dữ liệu tính toán có thể cần phải được bảo toàn trong trung hạn (ba năm trở lên). Thứ nhất, dữ liệu có thể được sử dụng làm nền tảng cho phân tích nội dung và trực quan hoá, sau đó, khai phá dữ liệu. Thứ hai, thời gian trên máy tính để tính toán bổ sung có thể không có sẵn trong khung thời gian ngắn. Điều này là phổ biến cho các tính toán quy mô rất lớn chạy trên các siêu máy tính được cộng đồng nghiên cứu chia sẻ, chẳng hạn như các máy tính tại các phòng thí nghiệm quốc gia của Bộ Năng lượng Hoa Kỳ và Quỹ Khoa học quốc gia (NSF).

Dữ liệu tham chiếu là dữ liệu được chỉnh lý biên tập cao và nhiều cộng đồng khoa học thường có nhu cầu cao đối với loại dữ liệu này. Những dữ liệu này được tạo ra cho nhiều mục đích, từ việc lập bản đồ bộ gen của người cho đến các dữ liệu theo thời gian dài về tình hình kinh tế và xã hội. Ngân hàng dữ liệu protein toàn cầu và nghiên cứu về động thái thu nhập là thí dụ về các bộ dữ liệu tham chiếu như vậy. Với tất cả những dữ liệu thống kê này, thường cần phải giữ gìn các tư liệu phụ trợ, chẳng hạn như hiệu chỉnh công cụ, thông số của thí nghiệm và sổ ghi chép trong phòng thí nghiệm. Trong khi hầu hết bộ sưu tập dữ liệu nghiên cứu lớn được sản xuất và sử dụng bởi các nhà nghiên cứu, chúng cũng có giá trị đối với chính sách công. Nhu cầu của chính sách công vượt ra ngoài khuôn khổ nhu cầu nghiên cứu và trở thành những vấn đề ưu tiên chung khi liên quan đến thông tin về khí hậu, địa chấn học, hải dương học, khảo sát lâm sàng và khảo sát nghiên cứu khoa học xã hội, các loài có nguy cơ tuyệt chủng, các địa điểm khảo cổ và vấn đề bảo mật nhạy cảm.

3.2. Định nghĩa dữ liệu mở

Tương tự như với truy cập mở, đã có nhiều nỗ lực khác nhau để xây dựng các định nghĩa về dữ liệu mở. Một điểm chung khá lý thú trong các cách tiếp cận là chúng nhấn mạnh đến quyền tái sử dụng để dữ liệu được gọi là mở. Tóm lại, dữ liệu mở là dữ liệu có

thể được sử dụng bởi bất kỳ ai mà không có bất cứ hạn chế về mặt kỹ thuật hoặc pháp lý nào. Việc sử dụng bao gồm cả truy cập và sử dụng lại. Liệu sử dụng mở như vậy tồn tại từ quan điểm pháp lý còn phụ thuộc vào khả năng áp dụng các hạn chế pháp lý có thể có hay không (nói cách khác, liệu các hạn chế đó có bị giấy phép tự do loại bỏ hay không).

Ví dụ, Quỹ Tri thức mở, phân biệt giữa: *sự mở về pháp lý* (khả năng để có được dữ liệu hợp pháp, tái sử dụng chúng, xây dựng và chia sẻ khác) và *sự mở về kỹ thuật* (có nghĩa là, không có rào cản kỹ thuật đối với sử dụng dữ liệu). Nói chung, dữ liệu mở đề cập đến những dữ liệu có sẵn không hạn chế và nó phải được đặc trưng bởi: i) tính có sẵn và truy cập: mọi người có thể lấy dữ liệu; ii) tái sử dụng và tái phân phối: Mọi người có thể tái sử dụng và chia sẻ dữ liệu; iii) sự tham gia toàn cầu: nghĩa là, bất kỳ ai cũng có thể sử dụng dữ liệu.

Năm 2004, các Bộ trưởng KH&CN của các nước OECD đã gặp nhau tại Pháp để thảo luận những hướng dẫn về truy cập đến dữ liệu nghiên cứu. Cuộc họp được tiếp nối vào năm 2007 với việc thông qua nguyên tắc với hướng dẫn của OECD về truy cập mở dữ liệu nghiên cứu được tài trợ công (Hộp 3.2) để tối đa hóa lợi ích phát sinh từ nghiên cứu được tài trợ công. Nguyên lý và nguyên tắc thừa nhận tầm quan trọng của truy cập mở đến dữ liệu nghiên cứu và tài liệu, nhưng họ cũng nhận ra sự cần thiết phải có sự phù hợp với khung khổ pháp lý của các quốc gia, chẳng hạn như luật bản quyền và bảo vệ quyền sở hữu trí tuệ.

3.3. Chia sẻ dữ liệu: Thách thức và cơ hội

Dữ liệu nghiên cứu, và dữ liệu nói chung, trong hầu hết các trường hợp là tài sản vô hình liên quan đến các tác nhân khác nhau và những bên liên quan trong các giai đoạn tạo dữ liệu, biên soạn và tái sử dụng dữ liệu khác nhau. Một số bên có thể đòi hỏi quyền sở hữu của mình đối với cùng bộ dữ liệu.

Dữ liệu, và đặc biệt là dữ liệu nghiên cứu, có thể được tạo ra hoặc thu thập bởi một cá nhân, sau đó được sử dụng bởi một bên khác, và sau đó thông tin có thể được biên dịch và làm mới bởi những người khác. Bên thứ ba tài trợ hoặc vận hành một số hoặc tất cả các hoạt động trên có thể yêu cầu quyền sở hữu dữ liệu. Ngoài ra, trong một số lĩnh vực khoa học, nhiều nhiệm vụ này có thể được thực hiện bằng máy hoặc tự động hoá.

Một yếu tố thiết yếu cho tính hữu ích của các nỗ lực chia sẻ dữ liệu là chất lượng của dữ liệu được công bố công khai. Khung Chất lượng và hướng dẫn của OECD cho hoạt động thống kê của OECD (2011) xác định bảy khía cạnh quan trọng của chất lượng dữ liệu (Hộp 3.3). Dự án Cơ hội trao đổi dữ liệu (Opportunity for Data Exchange - ODE), một dự án nghiên cứu do Ủy ban châu Âu tài trợ, đã phát triển một “Kim tự tháp dữ liệu”

(Hình 3.1) cho phép trực quan hóa các giai đoạn khác nhau của xử lý trên tập dữ liệu (data curation): từ dữ liệu thô chứa trong tệp máy tính cá nhân đến các dữ liệu được xử lý được liên kết với các công bố có chứa siêu dữ liệu hoặc thông tin chi tiết xung quanh dữ liệu.

Hộp 3.2. Nguyên lý và nguyên tắc của OECD về truy cập dữ liệu nghiên cứu từ tài trợ công

Tính mở (Openness). Khả năng truy cập mở đến dữ liệu nghiên cứu từ nguồn tài chính công phải dễ dàng, kịp thời, thân thiện với người dùng và tốt nhất là dựa trên Internet

Tính linh hoạt (Flexibility). Tính linh hoạt đòi hỏi phải được tính đến do những thay đổi nhanh chóng và thường xuyên không thể đoán trước của CNTT-TT bởi những đặc điểm của các lĩnh vực nghiên cứu khác nhau và bởi sự đa dạng của hệ thống nghiên cứu, khung pháp lý và văn hóa của mỗi quốc gia thành viên

Minh bạch (Transparency). Thông tin về dữ liệu nghiên cứu và tổ chức tạo ra dữ liệu, tư liệu về dữ liệu về tài liệu và các điều kiện cho sử dụng dữ liệu phải được quốc tế cung cấp một cách minh bạch, lý tưởng qua Internet.

Sự phù hợp về mặt pháp lý (Legal conformity). Các dàn xếp truy cập dữ liệu phải tôn trọng các quyền hợp pháp và lợi ích hợp pháp của tất cả các bên liên quan trong doanh nghiệp công. Hạn chế truy cập có thể được thiết lập vì lý do: an ninh quốc gia, quyền riêng tư và bảo mật, bí mật thương mại và quyền sở hữu trí tuệ, bảo vệ các loài quý hiếm, bị đe dọa hoặc có nguy cơ tuyệt chủng, quy trình pháp lý.

Bảo vệ sở hữu trí tuệ (Protection of intellectual property). Các dàn xếp truy cập dữ liệu cần xem xét việc áp dụng các quy định pháp luật về bản quyền và các luật sở hữu trí tuệ khác có liên quan đến cơ sở dữ liệu nghiên cứu được tài trợ công khai (như trong trường hợp quan hệ đối tác công-tư).

Trách nhiệm chính thức (Formal responsibility). Bố trí các dàn xếp truy cập nên cần thúc đẩy sự phát triển quy tắc và quy định liên quan đến trách nhiệm của các bên liên quan; nên phát triển với sự tham vấn của đại diện của tất cả các bên bị ảnh hưởng; phải đáp ứng các yếu tố như đặc điểm của dữ liệu, ví dụ: giá trị tiềm năng của chúng cho mục đích nghiên cứu. Các kế hoạch quản lý dữ liệu và tính bền vững lâu dài của dữ liệu cũng nên được xem xét.

Tính chuyên nghiệp (Professionalism). Các dàn xếp của tổ chức trong quản lý dữ liệu nghiên cứu phải dựa trên các tiêu chuẩn và giá trị chuyên nghiệp và có giá trị có liên quan thể hiện trong các quy tắc ứng xử của các cộng đồng khoa học có liên quan.

Tính tương tác (Interoperability). Việc dàn xếp truy cập cần xem xét các tiêu chuẩn tài liệu dữ liệu quốc tế có liên quan.

Chất lượng (Quality). Giá trị và tính hữu ích của dữ liệu phụ thuộc rất nhiều vào chất lượng của bản thân dữ liệu. Cần chú ý đặc biệt để đảm bảo tuân thủ các tiêu chuẩn chất lượng rõ ràng.

An ninh (Security). Cần chú ý hỗ trợ việc sử dụng các kỹ thuật và công cụ để đảm bảo tính toàn vẹn và bảo mật của dữ liệu nghiên cứu.

Hiệu quả (Efficiency). Một trong những mục tiêu trung tâm của việc thúc đẩy truy cập và chia sẻ dữ liệu là cải thiện hiệu quả của nghiên cứu khoa học được tài trợ công để tránh trùng lặp tốn kém và không cần thiết. Điều này cũng bao gồm cả phân tích chi phí và lợi ích để xác định các giao thức lưu giữ dữ liệu; sự tham gia của các tổ chức chuyên gia quản trị dữ liệu; và phát triển các cơ chế khuyến khích mới cho các nhà nghiên cứu và nhà sản xuất cơ sở dữ liệu

Trách nhiệm giải trình (Accountability). Việc thực hiện các sắp xếp truy cập dữ liệu phải được đánh giá định kỳ bởi nhóm người dùng, các tổ chức có trách nhiệm và cơ quan tài trợ nghiên cứu

Tính bền vững (Sustainability). Cần cân nhắc nghiêm túc đến tính bền vững của việc truy cập đến cận dữ liệu nghiên cứu được tài trợ công như một yếu tố quan trọng của cơ sở hạ tầng nghiên cứu

Nguồn: OECD (2007), *OECD Principles and Guidelines for Access to Research Data from Public Funding*.

Hộp 3.3. Khung chất lượng và hướng dẫn của OECD cho các hoạt động thống kê của OECD

Khung chất lượng và hướng dẫn của OECD cho các hoạt động thống kê của OECD xác định các khía cạnh sau:

1. Mức độ liên quan. “Được đặc trưng bởi mức độ mà dữ liệu được dùng để giải quyết các mục đích mà người dùng đang tìm kiếm. Nó phụ thuộc vào cả mức độ bao quát của các chủ đề được yêu cầu và việc sử dụng các khái niệm phù hợp

2. Độ chính xác. “Mức độ mà dữ liệu ước tính chính xác hoặc mô tả định lượng hoặc đặc tính mà chúng được thiết kế để đo lường

3. Độ tin cậy. “Độ tin cậy của các sản phẩm dữ liệu đề cập đến sự tin tưởng của người dùng đặt vào sản phẩm chỉ đơn giản là dựa trên hình ảnh của nhà sản xuất dữ liệu, tức là hình ảnh thương hiệu. Niềm tin của người dùng được xây dựng theo thời gian. Một khía cạnh quan trọng là tin tưởng vào tính khách quan của dữ liệu”

4. Tính kịp thời. “phản ánh khoảng thời gian giữa sự có sẵn và sự kiện hoặc hiện tượng mà chúng mô tả, nhưng được xem xét trong bối cảnh khoảng thời gian cho phép thông tin có giá trị và vẫn hoạt động

5. Khả năng truy cập. “phản ánh dữ liệu được truy cập và định vị dễ dàng như thế nào

6. Tính có thể giải thích. “Phản ánh sự dễ dàng mà người dùng có thể hiểu và sử dụng đúng cách và phân tích đúng cách dữ liệu”. Tính sẵn có của siêu dữ liệu đóng vai trò quan trọng ở đây, vì chúng cung cấp ví dụ “định nghĩa về khái niệm, tổng thể đối tượng mục tiêu, các nhân tố khả biến và thuật ngữ, liên quan đến cơ sở dữ liệu và thông tin mô tả những hạn chế của dữ liệu, nếu có”

7. Sự liên kết. “Phản ánh mức độ mà dữ liệu được kết nối hợp lý và nhất quán lẫn nhau. Sự liên kết ngụ ý rằng cùng một thuật ngữ không nên được sử dụng mà không cần giải thích cho các khái niệm khác nhau hoặc những mục dữ liệu khác nhau và không nên sử dụng những thuật ngữ khác mà không có giải thích cho cùng một khái niệm hoặc mục dữ liệu; và những biến thể trong phương pháp có thể ảnh hưởng đến giá trị dữ liệu, không nên được thực hiện mà không có giải thích.

Nguồn: OECD (2011), “Quality Framework and Guidelines for OECD Statistical Activities”

Hộp 3.3. Khung chất lượng và hướng dẫn của OECD cho các hoạt động thống kê của OECD

Khung chất lượng và hướng dẫn của OECD cho các hoạt động thống kê của OECD xác định các khía cạnh sau:

1. Mức độ liên quan. “Được đặc trưng bởi mức độ mà dữ liệu được dùng để giải quyết các mục đích mà người dùng đang tìm kiếm. Nó phụ thuộc vào cả mức độ bao quát của các chủ đề được yêu cầu và việc sử dụng các khái niệm phù hợp

2. Độ chính xác. “Mức độ mà dữ liệu ước tính chính xác hoặc mô tả định lượng hoặc đặc tính mà chúng được thiết kế để đo lường

3. Độ tin cậy. “Độ tin cậy của các sản phẩm dữ liệu đề cập đến sự tin tưởng của người dùng đặt vào sản phẩm chỉ đơn giản là dựa trên hình ảnh của nhà sản xuất dữ liệu, tức là hình ảnh thương hiệu. Niềm tin của người dùng được xây dựng theo thời gian. Một khía cạnh quan trọng là tin tưởng vào tính khách quan của dữ liệu”

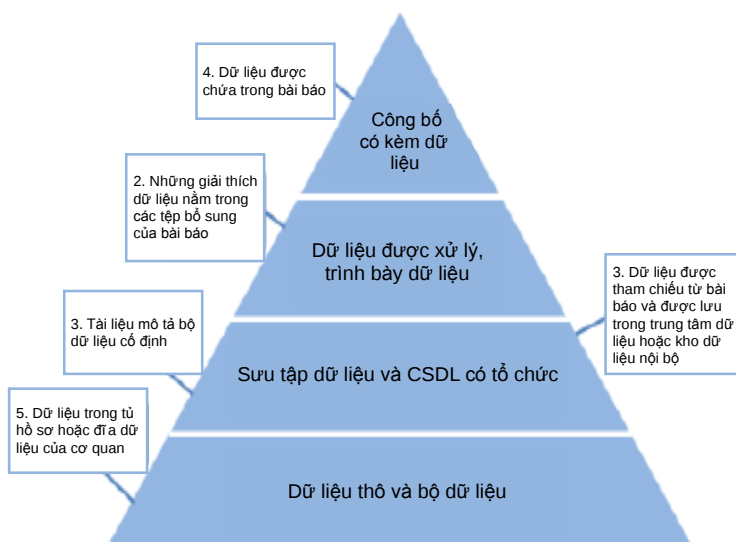
4. Tính kịp thời. “phản ánh khoảng thời gian giữa sự có sẵn và sự kiện hoặc hiện tượng mà chúng mô tả, nhưng được xem xét trong bối cảnh khoảng thời gian cho phép thông tin có giá trị và vẫn hoạt động

5. Khả năng truy cập. “phản ánh dữ liệu được truy cập và định vị dễ dàng như thế nào

6. Tính có thể giải thích. “Phản ánh sự dễ dàng mà người dùng có thể hiểu và sử dụng đúng cách và phân tích đúng cách dữ liệu”. Tính sẵn có của siêu dữ liệu đóng vai trò quan trọng ở đây, vì chúng cung cấp ví dụ “định nghĩa về khái niệm, tổng thể đối tượng mục tiêu, các nhân tố khả biến và thuật ngữ, liên quan đến cơ sở dữ liệu và thông tin mô tả những hạn chế của dữ liệu, nếu có”

7. Sự liên kết. “Phản ánh mức độ mà dữ liệu được kết nối hợp lý và nhất quán lẫn nhau. Sự liên kết ngụ ý rằng cùng một thuật ngữ không nên được sử dụng mà không cần giải thích cho các khái niệm khác nhau hoặc những mục dữ liệu khác nhau và không nên sử dụng những thuật ngữ khác mà không có giải thích cho cùng một khái niệm hoặc mục dữ liệu; và những biến thể trong phương pháp có thể ảnh hưởng đến giá trị dữ liệu, không nên được thực hiện mà không có giải thích.

Nguồn: OECD (2011), “Quality Framework and Guidelines for OECD Statistical Activities”



Hình 3.1. Kim tự tháp dữ liệu

Nguồn: Reilly, S. et al. (2011), *Report on Integration of Data and Publications*, ODE, *Opportunities for Data Exchange*

Nhiều cộng đồng khoa học vẫn chưa có quy định đánh giá chất lượng dữ liệu chuẩn như đối với các công bố khoa học. Như được nêu bật trong báo cáo của Royal Society (2012) và Dallmeier-Tiessen (2011), dữ liệu có ít giá trị nếu chất lượng của chúng không đáp ứng tiêu chí chất lượng tối thiểu. “Dữ liệu chất lượng tốt” ngụ ý không chỉ truy cập được (ví dụ có sẵn trên Internet) mà còn có thể hiểu được, có thể đánh giá, đáng tin cậy và, tất nhiên, có thể sử dụng lại được. Về mặt này, việc phát triển thông tin và siêu dữ liệu chia sẻ dữ liệu chi tiết là điều cần thiết cho việc sử dụng tiếp tục cùng một dữ liệu từ nhiều nhóm nghiên cứu.

Tuy nhiên, các nhà khoa học và nhà nghiên cứu không nhất thiết phải cần có những ưu đãi hay có kỹ năng để thực hiện những nhiệm vụ này, vì việc biên tập đúng đắn và phổ biến tập hợp dữ liệu là công việc tốn kém và tốn thời gian; thậm chí có thể được coi là loại hình đầu ra khoa học khác. Một giải pháp để giảm trở ngại là trích dẫn dữ liệu: Khả năng để các nhà nghiên cứu được công nhận làm công việc thu thập và biên tập/xử lý dữ liệu của họ thông qua các cơ chế tương tự như cơ chế cho trích dẫn đối với các bài báo nghiên cứu. Tuy nhiên, việc trích dẫn dữ liệu không nhất thiết là cách thức được chấp nhận rộng rãi trong cộng đồng học thuật. Một số nhà khoa học xem việc này như vấn đề giới hạn trích dẫn bài báo khoa học; các cơ quan tài trợ trong một số trường hợp đặt vấn đề về ý tưởng công nhận cá nhân là tác giả dữ liệu; và các chỉ số trích dẫn thư mục truyền thống chưa tính đến các trích dẫn không phải bài báo. Ngoài ra, có những rào cản kỹ thuật hạn chế sự phát triển của trích dẫn dữ liệu và các đo lường thư mục liên quan: thí dụ, sự không tương thích trong máy móc và phần mềm, cấu trúc tệp dữ liệu, lưu trữ và quản lý dữ liệu. Một số tổ chức đang tích cực tham gia khắc phục những rào cản này.

Theo Dự án nghiên cứu ODE trong chương trình khung FP7 của EU, trích dẫn dữ liệu có một số tính năng độc đáo do các thuộc tính cụ thể của tập hợp dữ liệu. Ví dụ, tập hợp dữ liệu có thể có kích thước rất khác nhau và không phải lúc nào cũng rõ ràng các yếu tố cụ thể bên trong tập dữ liệu mà các học giả đang đề cập đến - hoặc trong trường hợp các cập nhật đối với các tập dữ liệu, phiên bản này cần được trích dẫn. Theo dự án ODE, một số thực tiễn/thách thức tốt liên quan đến trích dẫn dữ liệu là:

- Trích dẫn của tập dữ liệu, với số nhận dạng, nên được liệt kê trong tài liệu tham khảo/thư mục để cho phép theo dõi và phát triển các cách đo lường trích dẫn.

- Nhà xuất bản cần cung cấp các hướng dẫn cho tác giả và người bình duyệt đồng nghiệp về trích dẫn dữ liệu.

- Không có sự thống nhất rõ ràng về các yêu cầu tính chính xác và lâu dài cho các bộ dữ liệu được coi là có thể trích dẫn hoặc được trích dẫn.

- Có sự thiếu rõ ràng và thống nhất về quyền tác giả đối với tập dữ liệu.
- Các nhà nghiên cứu cần nâng cao nhận thức trong cộng đồng của mình về lợi ích của việc trích dẫn dữ liệu và tuân thủ các hướng dẫn về trích dẫn dữ liệu đã thống nhất.

Các phương thức khác để công bố bộ dữ liệu là những tạp chí dữ liệu, đó là tập hợp các bài báo khoa học chuyên về công bố bài báo dữ liệu. Bài báo dữ liệu là bài báo có mục đích chính là mô tả tập dữ liệu hơn là báo cáo nghiên cứu và phân tích khoa học. Những bài báo dữ liệu chứa số liệu và mô tả về dữ liệu. Mục tiêu chính của bài báo đại diện là trở thành nguồn thông tin có thể trích dẫn về dữ liệu có uy tín đối với các học giả đã sản xuất và mô tả CSDL; công bố thông tin chi tiết về tập dữ liệu; và thông tin về sự tồn tại của dữ liệu đến cộng đồng khoa học. Các tạp chí dữ liệu có thể nhắm mục tiêu vào lĩnh vực khoa học rộng hơn cũng như lĩnh vực cụ thể, chẳng hạn như khoa học hệ thống Trái đất (khoa học địa lý).

Mặc dù khoa học mở có tác động tích cực đến doanh nghiệp khoa học, đổi mới sáng tạo và xã hội nói chung, vẫn còn một số lý do chính đáng để hạn chế tính mở của khoa học, đặc biệt là dữ liệu nghiên cứu, vượt ra ngoài các vấn đề kỹ thuật và không chỉ liên quan đến cộng đồng nghiên cứu mà còn liên quan đến xã hội nói chung. Chúng bao gồm nhiều vấn đề, như vấn đề liên quan đến quyền riêng tư của cá nhân và tổ chức hoặc an ninh quốc gia. Dữ liệu được thu thập trong quá trình nghiên cứu thường chứa thông tin cá nhân (ví dụ như hồ sơ y tế), khi mở các dữ liệu đó, quyền của các đối tượng dữ liệu phải được tôn trọng. Điều này không có nghĩa là dữ liệu không thể là dữ liệu mở, nhưng nó đòi hỏi thực hiện các thủ tục bảo vệ (Hộp 3.4).

Một thủ tục như vậy là ẩn danh, tuy nhiên việc này có thể dẫn đến việc không thể áp dụng toàn bộ chế độ bảo vệ dữ liệu cá nhân. Ngoài ra, không phải tất cả các kỹ thuật ẩn danh đều có hiệu quả. Một số quốc gia hiện đang quảng bá "quyền riêng tư theo thiết kế" cho dữ liệu sức khỏe của họ, thông qua việc sử dụng các công nghệ bảo mật tăng cường để đáp ứng cả nhu cầu sử dụng dữ liệu chăm sóc sức khỏe và bảo vệ quyền riêng tư.

Diễn đàn Khoa học toàn cầu của OECD đã xác định một loạt các thách thức liên quan đến khoa học định hướng dữ liệu và chia sẻ dữ liệu (Hộp 3.5). Theo Diễn đàn, những rào cản có thể liên quan đến nhiều vấn đề pháp lý; kỹ thuật hoặc thiếu kỹ năng, của cộng đồng khoa học và bên ngoài, để thực hiện các nhiệm vụ quản lý dữ liệu và chia sẻ dữ liệu.

Vì tất cả những lý do này (cụ thể là những thách thức xung quanh quyền sở hữu bộ dữ liệu, vấn đề thông tin cá nhân bảo mật và an ninh, các mức độ khác nhau về chất lượng của bộ dữ liệu, thiếu kỹ năng và ưu đãi trong cộng đồng các nhà nghiên cứu, v.v..)

thúc đẩy dữ liệu mở chắc chắn không dễ dàng như đối với việc thúc đẩy truy cập mở công bố khoa học. Mặc dù các chính sách đang phát triển nhanh chóng, nhiều chính sách và sáng kiến mới được phát triển gần đây mới thúc đẩy truy cập mở nhiều hơn là thúc đẩy dữ liệu mở.

Hộp 3.4. Phương tiện thực tiễn để ngăn chặn việc phát lộ thông tin

Các hoạt động phân tích dữ liệu trích xuất thông tin từ dữ liệu bằng cách tiết lộ bối cảnh trong đó dữ liệu được nhúng, tổ chức và cấu trúc của chúng. Có một số phương tiện thực tế để ngăn chặn hoặc tăng đáng kể chi phí trích xuất thông tin được nhúng trong dữ liệu thông qua phân tích dữ liệu, mặc dù chúng có thể ảnh hưởng tiêu cực đến tiện ích xử lý dữ liệu. Sau đây là một vài ví dụ:

Giảm dữ liệu (Data reduction). Giảm dữ liệu có thể được coi là phương tiện mạnh nhất để ngăn chặn việc trích xuất thông tin, bởi vì không thu thập dữ liệu, không có thông tin nào có thể được trích xuất. Các đối tượng dữ liệu có thể giữ lại hoặc từ chối cung cấp dữ liệu. Bộ điều khiển dữ liệu có thể thực hiện giảm thiểu dữ liệu. Như Pfizmann và Hansen (2010) đã nhấn mạnh, giảm thiểu dữ liệu “là chiến lược chung duy nhất (thông tin sai lệch hoặc đánh lạc hướng) để thực hiện tính không liên kết, vì tất cả dữ liệu cá nhân chính xác cung cấp một số liên kết.

Mật mã hoá (Cryptography). Mật mã hoá thực tế là cách làm “bao gồm nguyên tắc, phương tiện và phương pháp chuyển đổi dữ liệu nhằm che giấu nội dung thông tin, thiết lập tính xác thực, ngăn chặn sự sửa đổi không bị phát hiện, ngăn chặn sự thoái thác hoặc sử dụng trái phép”. Nó là một phương tiện công nghệ chính cung cấp bảo mật dữ liệu trong hệ thống thông tin và truyền thông. Mật mã có thể được sử dụng để bảo vệ tính bảo mật của dữ liệu, chẳng hạn như dữ liệu tài chính hoặc dữ liệu cá nhân khi được lưu trữ hay đang truyền đi. Mật mã cũng có thể được sử dụng để xác minh tính toàn vẹn của dữ liệu bằng cách phát hiện dữ liệu đã bị thay đổi hay chưa và định danh người hoặc thiết bị đã gửi chúng.

Nhận diện xoá định danh (De-identification). Bao gồm các quy trình khác nhau, từ ẩn danh đến sử dụng tên giả (bút danh). Những hình thức này có chung mục đích ngăn chặn việc trích xuất các thuộc tính nhận dạng (tức là nhận dạng lại) hoặc ít nhất là tăng đáng kể chi phí nhận dạng lại. Ẩn danh hoá là quá trình trong đó thông tin nhận dạng của một thực thể bị loại trừ hoặc che giấu để danh tính của đối tượng không thể nhận ra, hoặc trở nên quá tốn kém để được xây dựng lại. Một số nghiên cứu cho thấy khi được liên kết với dữ liệu khác, hầu hết các dữ liệu ẩn danh có thể lại được bỏ ẩn danh; có nghĩa là, thông tin nhận diện có thể được xây dựng lại. Tuy nhiên, nhiều ứng dụng yêu cầu một số loại định danh và việc ẩn danh hoàn toàn sẽ ngăn chặn bất kỳ giao tiếp hoặc giao dịch hai chiều hữu ích nào. Do đó, bút danh được sử dụng, mà ở đó hầu hết thuộc tính nhận dạng (như là số nhận dạng) trong một bản ghi dữ liệu được thay thế bằng số nhận dạng nhân tạo duy nhất (tức là bút danh).

Tính không liên kết và phân tách chức năng tách (Unlinkability and functional separation). Tính không liên kết được tạo ra từ các quá trình để đảm bảo rằng bộ xử lý dữ liệu không thể phân biệt liệu các mục quan tâm có liên kết với nhau hay không. Theo ISO Pfizmann và Hansen (2010), tính không liên kết “đảm bảo rằng người dùng có thể sử dụng nhiều nguồn hoặc lần dịch vụ hoặc tài nguyên mà không người nào khác có thể liên kết các ứng dụng này với nhau”. Xoá nhận diện là để tạo ra tính không liên kết, nhưng nó không thể đảm bảo tính không liên kết. Phương tiện kỹ thuật khác có thể là tách chức năng và phân phối (phân cấp).

Bổ sung nhiễu thông tin giả (Noise addition and disinformation). Việc bổ sung “nhiều” vào bộ dữ liệu có thể làm cho việc phân tích dựa trên bộ dữ liệu hoàn chỉnh trở nên quan trọng trong khi che giấu các thuộc tính dữ liệu nhạy cảm. Tìm sự cân bằng phù hợp để bảo vệ sự riêng tư trong khi giảm thiểu chi phí cho tiện ích xử lý dữ liệu là một thách thức. Đưa thông tin giả là việc đưa thông tin sai không chính xác một cách cố ý để gây hiểu nhầm. Các kỹ thuật bổ sung nhiễu được coi là đầy hứa hẹn để giúp bảo vệ quyền riêng tư và bảo mật trong cơ sở dữ liệu, trong khi vẫn giữ tất cả các bộ dữ liệu thống kê gần với bản gốc. Công trình “Bảo mật phân biệt” là một ví dụ.

Nguồn: OECD (2015), Data-Driven Innovation for Growth and Well-Being

Hộp 3.5. 9 thách thức đối với nghiên cứu dựa vào dữ liệu và nghiên cứu dựa trên bằng chứng

Diễn đàn Khoa học Toàn cầu của OECD gần đây đã xác định 9 thách thức đối với nghiên cứu dựa vào dữ liệu và nghiên cứu dựa trên bằng chứng như sau:

1. Một lượng lớn dữ liệu số đang được tạo ra ở quy mô chưa từng thấy, một phần nhờ sự phát triển của CNTT-TT. Độ tin cậy, hiệu lực thống kê và tính khái quát của các dạng dữ liệu mới chưa được hiểu đầy đủ.
2. Trong khi dữ liệu hành chính, điều tra và tổng điều tra dân số được thu thập rộng rãi bởi các cơ quan thống kê quốc gia và cơ quan chính phủ, nhiều biểu ghi dữ liệu vi mô có sẵn ở mức độ thấp hơn nhiều.
3. Những dạng dữ liệu cá nhân mới, chẳng hạn như dữ liệu mạng xã hội, ngày càng được tạo ra và thu thập ngày càng nhiều. Việc sử dụng dữ liệu đó có thể tạo ra rủi ro cho quyền riêng tư cá nhân.
4. Quyền pháp lý, văn hóa, ngôn ngữ và quyền sở hữu của các rào cản truy cập cản trở sự hợp tác xuyên quốc gia và khai thác dữ liệu quốc tế, đặc biệt là trong các ngành khoa học xã hội.
5. Các chương trình nghiên cứu toàn cầu đòi hỏi sự phối hợp liên ngành và quốc tế ngày càng tăng.
6. Việc hợp tác và chia sẻ kinh nghiệm giữa các quốc gia trong việc phát triển tài nguyên dữ liệu có thể so sánh được là cần thiết để để khai thác đầy đủ các tiềm năng của bộ dữ liệu.
7. Các nhà nghiên cứu thường thiếu nguồn tài nguyên hoặc kỹ năng để đảm bảo rằng dữ liệu họ sử dụng, thu thập và sản xuất là có sẵn để tái sử dụng.
8. Đầu tư quốc gia vào kỹ năng và cơ sở hạ tầng liên quan đến việc tạo và quản lý dữ liệu là cần thiết để tránh nguy cơ mất dữ liệu hoặc suy thoái.
9. Các nhà nghiên cứu cần có tập hợp những khuyến khích phù hợp để đảm bảo chia sẻ dữ liệu hiệu quả.

Nguồn: Adapted from OECD (2013b), "New data for understanding the human condition - International perspectives"

3.4. Các khung bảo vệ dữ liệu mở ở các nước OECD

Việc mở rộng chính sách truy cập mở đối với dữ liệu nghiên cứu được tài trợ công làm nảy sinh một số vấn đề pháp lý và chính sách khác với những vấn đề liên quan đến việc công bố khoa học và chuyên khảo. Vì vấn đề truy cập mở dữ liệu nghiên cứu - không giống với công bố khoa học - là một mục tiêu chính sách tương đối mới, những tính năng cụ thể của dữ liệu nghiên cứu ít được chú ý.

Trên phạm vi quốc tế, sự bảo hộ dành cho các CSDL (như bộ sưu tập dữ liệu hoặc nhiều yếu tố khác) đã được thiết lập - hoặc được xác nhận - bởi cả điều 10 (2) của Hiệp định TRIPS và trong điều 5 của WCT: "*Sưu tập dữ liệu hoặc tài liệu khác, cho dù ở dạng máy tính có thể đọc được hay dạng khác, do lý do lựa chọn hoặc sắp xếp nội dung của chúng tạo thành trí tuệ sẽ được bảo vệ như vậy ...*"

Trong khi những công bố khoa học hầu như luôn thu hút sự bảo hộ bản quyền theo luật bản quyền của các nước thành viên OECD, dữ liệu nghiên cứu riêng lẻ và bộ dữ liệu chứa chúng có thể không dễ dàng nằm trong chế độ bản quyền. Tuy nhiên, một số quốc gia OECD cũng như Liên minh châu Âu đã áp dụng các khung pháp lý có chứa nhiều nội

dung cho dữ liệu nghiên cứu. Ở những quốc gia khác, đáng chú ý là Hoa Kỳ, bộ dữ liệu không có sự bảo hộ quyền sở hữu trí tuệ cụ thể.

CSDL là đối tượng cụ thể được bảo vệ bởi bản quyền trong một số trường hợp, nhưng ở một số khu vực - ví dụ ở Liên minh châu Âu, Nhật Bản, Hàn Quốc - cũng được bảo vệ bởi Quyền CSDL riêng. Lớp bảo vệ bổ sung này được áp dụng ở một số quốc gia và được dành cho các CSDL không phụ thuộc vào việc có hay không có sáng tạo trí tuệ (tức là "lựa chọn hoặc sắp xếp"). Thay vào đó, những thứ được bảo vệ chính là sự đầu tư vào việc tạo CSDL, tức là trong việc thu thập, xác minh hoặc trình bày dữ liệu. Loại quyền này là điển hình của Chỉ thị CSDL của EU và luật pháp của một số quốc gia khác. Cần lưu ý rằng mặc dù việc bảo vệ dành cho các CSDL gốc tập trung vào việc sắp xếp hoặc lựa chọn mà không mở rộng đến nội dung của CSDL, Quyền CSDL riêng cung cấp sự bảo vệ chống lại sao chép các phần quan trọng của CSDL, ít nhất là ở mức độ nào đó, với chính dữ liệu.

Sự phức tạp về tình trạng quyền của dữ liệu nghiên cứu ở châu Âu và các khu vực khác có khả năng ảnh hưởng xấu đến cơ hội tái sử dụng các bộ sưu tập dữ liệu khoa học do khó khăn (cho cả với những tổ chức nghiên cứu đưa CSDL trở thành truy cập được cho người dùng tiềm năng) trong xác định liệu một CSDL nào đó có được bao quát bởi Quyền CSDL riêng hay không và mức độ việc tái sử dụng và khai thác CSDL có thể diễn ra một cách tự do. Cho dù việc sử dụng các bộ sưu tập hay CSDL cho mục đích nghiên cứu và học tập cá nhân nói chung, việc khai phá văn bản và dữ liệu nói riêng, được điều chỉnh bởi những ngoại lệ phù hợp liên quan đến bản quyền hoặc Quyền CSDL là không rõ ràng. Việc sử dụng giấy phép Creative Commons 4.0 (Hộp 2.3 và 2.4) có thể làm giảm bớt sự không rõ ràng, bằng cách nêu rõ những gì có thể và không thể thực hiện được với tài liệu được cấp phép.

Khung pháp lý trong Liên minh châu Âu

Ở châu Âu, để đủ điều kiện được bảo vệ bản quyền, bộ sưu tập dữ liệu, bảng số liệu phải thể hiện đủ mức độ nguyên gốc trong lựa chọn hoặc sắp xếp chúng - tức là thông qua lựa chọn này hoặc sắp xếp đó, tác giả có thể thể hiện sự sáng tạo và lựa chọn tự do của mình.

Việc liệu bộ sưu tập dữ liệu nghiên cứu khoa học có đáp ứng những tiêu chí nguyên gốc vẫn là một câu hỏi thực tế và được xác định trên cơ sở từng trường hợp cụ thể. Tuy nhiên, nếu việc lựa chọn hoặc sắp xếp nội dung của CSDL khoa học được thực hiện bởi các yếu tố kỹ thuật hoặc các lệnh chính xác và toàn diện, thì tác giả có thể sẽ ít hoặc không có tác động hoặc không có sự sáng tạo đáng kể trong lựa chọn, tránh hoặc kết hợp

dữ liệu trong bộ sưu tập dữ liệu. Do đó, hầu hết CSDL khoa học, trong nhiều trường hợp không đáp ứng yêu cầu ngưỡng bảo vệ bản quyền.

Chỉ thị Xã hội Thông tin chứa một ngoại lệ về bản quyền có thể áp dụng trong một số trường hợp. Điều 5 (3) (a) của Chỉ thị này cho phép các nước thành viên cung cấp ngoại lệ trong trường hợp duy nhất “sử dụng cho mục đích minh họa duy nhất cho giảng dạy hoặc nghiên cứu khoa học, miễn là nguồn tin bao gồm cả tên tác giả, được chỉ ra, trừ khi điều này không thể và được biện minh bởi mục đích phi thương mại để đạt được”. Ngoại lệ này là tùy chọn; Các nước thành viên có thể quyết định có thực hiện hay không. Kết quả là, các quốc gia thành viên có những quy tắc và quy định khác nhau trong bối cảnh này và một số quốc gia không chấp nhận ngoại lệ nghiên cứu nào (ví dụ như Hà Lan và Tây Ban Nha). Kết quả là, các ngoại lệ nghiên cứu thường được thực hiện mơ hồ và không đồng đều ở cấp quốc gia, điều này có thể bất lợi cho một số nhà nghiên cứu.

Bộ sưu tập các công trình, dữ liệu, hoặc các tài liệu khoa học khác được sắp xếp một cách có hệ thống hoặc có phương pháp và có thể truy cập riêng lẻ, điện tử hoặc bằng các phương tiện khác, có thể được bảo vệ theo Quyền CSDL riêng. Thông qua Điều 7 của Chỉ thị CSDL, như được thực hiện trong luật pháp của các Nước thành viên, thì nhà sản xuất CSDL nên chứng minh được đầu tư đáng kể (đánh giá chất lượng hoặc định lượng) trong việc thu thập, xác minh hoặc trình bày nội dung của CSDL có toàn quyền ngăn chặn việc khai phá và tái sử dụng toàn bộ hoặc một phần đáng kể, dữ liệu đánh giá chất lượng hoặc định lượng về nội dung của CSDL đó. Giống như bảo vệ bản quyền, Quyền CSDL riêng phát sinh tự động, mà không cần có bất kỳ một yêu cầu hình thức nào, tại thời điểm hoàn thành CSDL hoặc đưa nó ra cho công chúng.

Trong trường hợp việc "thu thập, xác minh hoặc trình bày" các bộ dữ liệu nghiên cứu thể hiện sự đầu tư đáng kể thực sự cần thiết để đủ điều kiện bảo hộ, Quyền CSDL riêng dẫn đến hai quyền có thể chuyển nhượng được cho nhà sản xuất CSDL: quyền trích xuất và quyền sử dụng lại đáng kể các phần của CSDL và được định nghĩa tương ứng như sau: “(a) 'trích xuất' nghĩa là việc chuyển vĩnh viễn hoặc tạm thời toàn bộ hoặc một phần đáng kể nội dung của CSDL sang môi trường khác bằng bất kỳ phương tiện nào hoặc dưới bất kỳ hình thức nào; (b) "tái sử dụng" có nghĩa là bất kỳ hình thức nào làm cho dữ liệu trở thành có sẵn cho công chúng tất cả hoặc một phần đáng kể những nội dung của CSDL bằng cách phân phối bản sao, cho thuê, trực tuyến hoặc các hình thức truyền tải khác”.

Việc bảo hộ theo Quyền CSDL riêng phải kéo dài trong 15 năm kể từ ngày đầu tiên của tháng 1 năm sau ngày CSDL được hoàn thành. Thời hạn bảo vệ cho CSDL có thể bắt

đầu lại với hai điều kiện đều liên quan đến thuật ngữ "đáng kể". Điều kiện thứ nhất là sự sửa đổi đáng kể nội dung của CSDL, được đánh giá định lượng hoặc định tính, có thể bao gồm bổ sung, xóa hoặc thay đổi (bao gồm sắp xếp lại nội dung). Điều kiện thứ hai, sửa đổi đáng kể nhờ một đầu tư đáng kể, việc đánh giá định lượng hoặc định tính. Đây là một trong những điều khoản gây tranh cãi nhất của Chỉ thị vì theo một số người, nó dường như cung cấp cơ sở cho việc bảo vệ vĩnh viễn quyền của CSDL.

Hà Lan

Hà Lan cho đến nay là nước thành viên duy nhất đã quy định rõ ràng việc thực hiện Quyền CSDL riêng của các cơ quan khu vực công. Điều 8 của Luật CSDL của Hà Lan không cho một cơ quan công quyền có quyền thực thi các Quyền CSDL độc quyền trừ khi quyền được bảo lưu một cách rõ ràng bằng một điều khoản chung trong một luật, nghị định hoặc pháp lệnh, hoặc trong một trường hợp cụ thể được thông báo trên bản thân CSDL hoặc trong khi CSDL được cung cấp cho công chúng.

Hàn Quốc

Ở Hàn Quốc, Chương IV Luật Bản quyền bảo hộ các bộ sưu tập là các tác phẩm nguyên gốc của tác giả nếu chúng có tính sáng tạo trong lựa chọn, sắp xếp hoặc sáng tác nội dung. CSDL được định nghĩa là các bộ sưu tập mà nội dung của chúng được sắp xếp hoặc tạo ra theo cách mà bất kỳ ai cũng có thể truy cập hoặc tìm kiếm được các nội dung trong CSDL. Do đó các CSDL nguyên gốc được bảo hộ bản quyền khi chúng đáp ứng các điều kiện trên. CSDL không nguyên gốc thuộc đối tượng bảo hộ theo Chương IV có hiệu lực từ năm 2003. Các nhà sản xuất CSDL đầu tư đáng kể vào nguồn nhân lực hoặc vật chất để tạo lập một CSDL, hoặc đổi mới, xác thực hoặc bổ sung nội dung có quyền sao chép, phân phối CSDL này. Một quốc tịch khác có thể là người thụ hưởng Quyền bảo vệ CSDL riêng với điều kiện tương nhượng, nếu quốc gia đó được bảo hộ theo một hiệp ước mà Hàn Quốc đã tham gia. Các quyền của nhà sản xuất CSDL bị giới hạn nhiều hơn bản quyền, do những hạn chế và ngoại lệ đối với bản quyền một mặt có thể áp dụng thay thế cho các quyền của nhà sản xuất CSDL, mặt khác, việc sử dụng toàn bộ hoặc một phần đáng kể CSDL được cho phép cho mục đích giáo dục, học tập hoặc nghiên cứu, hoặc để báo cáo các sự kiện hiện tại. Thời hạn bảo hộ được tái thiết lập sau 5 năm.

IV. QUẢN TRỊ KHOA HỌC MỞ: HOẠT ĐỘNG, XU HƯỚNG VÀ CHÍNH SÁCH

4.1. Những nhân tố chủ chốt trong khoa học mở

Có một số nhân tố trong hệ thống đổi mới sáng tạo cục bộ, quốc gia và toàn cầu tham gia vào những nỗ lực khoa học mở:

Các nhà nghiên cứu là những người đứng đầu trong nỗ lực thúc đẩy khoa học mở. Những động lực đối với nhà khoa học tham gia khoa học mở có thể trải rộng từ giá trị văn hóa có trong khoa học (như tính cởi mở, sự sẵn lòng tham gia vào xã hội) đến sự cần thiết, như phát triển hạ tầng cơ sở công nghệ cho phép sự cộng tác. Các nhà nghiên cứu đáp lại những khuyến khích từ các tổ chức tài trợ, các trường đại học và các tổ chức nghiên cứu công. Sự căng thẳng có thể tồn tại giữa mô hình “xuất bản hoặc hư hỏng” cạnh tranh và mối quan tâm đến chia sẻ dữ liệu và cộng tác.

Các bộ/cơ quan chính phủ đã phát triển những chiến lược quốc gia cho khoa học mở, hoặc đó là những chiến lược độc lập hoặc như là một phần của chương trình nghị sự mở quốc gia lớn hơn. Những chương trình nghị sự này xác định những ưu tiên cấp quốc gia cần được chuyển đổi thành những sáng kiến cụ thể được thực hiện bởi các tác nhân của hệ thống đổi mới sáng tạo.

Các cơ quan tài trợ nghiên cứu là những tác nhân chủ chốt trong thúc đẩy các nỗ lực khoa học mở, bởi họ là những tác nhân chịu trách nhiệm xác định cơ chế và yêu cầu đối với những bên hưởng lợi từ các tài trợ và kinh phí nghiên cứu. Ở một số nước trong vài năm gần đây, các cơ quan tài trợ đã tăng cường thực hiện các quy định và cơ chế thúc đẩy khoa học mở và trong một số trường hợp là bắt buộc thực hiện thông qua các hỗ trợ tài chính cho chi phí thực hiện truy cập mở hoặc chi trả cho các chi phí liên quan đến công bố dữ liệu hoặc các tài liệu nghiên cứu khác.

Ở một số nước OECD, **các trường đại học và viện nghiên cứu công lập** có mức độ tự chủ và có trách nhiệm xây dựng các chính sách riêng của mình để hỗ trợ khoa học mở và thực hiện các chính sách của các hội đồng hoặc cơ quan tài trợ. Ngoài ra, các trường đại học và các tổ chức giáo dục đại học có thể đóng vai trò trong việc đào tạo sinh viên và các nhà nghiên cứu để phát triển các kỹ năng cần thiết để cho phép thực hành khoa học mở - từ các kỹ năng cơ bản liên quan đến việc sử dụng kho lưu trữ trực tuyến đến kỹ năng cần thiết để thực hiện làm sạch, biên tập và quản lý dữ liệu.

Thư viện, kho lưu trữ và trung tâm dữ liệu là những nhân tố chủ chốt và là những tác nhân hỗ trợ cho khoa học mở. Các thư viện đã chấp nhận vai trò của mình và hiện đang tích cực hoạt động trong việc bảo tồn, quản lý, xuất bản và phổ biến các tài liệu

khoa học số, dưới dạng công bố khoa học, dữ liệu và các nội dung liên quan đến nghiên cứu khác. Thư viện và kho lưu trữ tạo thành cơ sở hạ tầng vật lý cho phép các nhà khoa học chia sẻ việc sử dụng và tái sử dụng kết quả công việc của họ, và chúng có vai trò thiết yếu trong việc tạo ra phong trào khoa học mở.

Các tổ chức và quỹ phi lợi nhuận tư nhân có thể đóng một vai trò quan trọng trong việc phát triển, nâng cao nhận thức và khuyến khích văn hóa khoa học mở. Họ không chỉ tài trợ cho nghiên cứu mở và đưa ra các yêu cầu trong thỏa thuận tài trợ, mà còn phát triển và tạo thuận lợi cho việc hình thành mạng lưới các bên liên quan trên toàn thế giới.

Các nhà xuất bản khoa học tư nhân cung cấp một loạt các hỗ trợ cho truy cập mở (ví dụ thông qua con đường truy cập mở vàng hoặc xuất bản trong các tạp chí lai) và các dịch vụ liên quan như duy trì kho dữ liệu số và các bộ dữ liệu hoặc tài liệu khoa học khác, hoặc phát triển các công cụ khai phá văn bản và dữ liệu (Text and Data Mining hoặc TDM).

Các doanh nghiệp là một phần của nhu cầu về các công bố và dữ liệu truy cập mở mà họ sử dụng để phát triển các sản phẩm và dịch vụ mới. Ví dụ, các doanh nghiệp như các công ty dược phẩm cũng hỗ trợ khoa học mở thông qua quan hệ đối tác công-tư với các trường đại học hoặc cung cấp tài chính cho các thử nghiệm lâm sàng mở.

Cuối cùng, **các thực thể siêu quốc gia** đóng một vai trò quan trọng trong việc định nghĩa các thỏa thuận hoặc các hướng dẫn phối hợp quốc tế nhằm giải quyết các vấn đề khoa học mở với quan điểm quốc tế và toàn cầu. Các tổ chức liên chính phủ (IGOs) đóng một vai trò quan trọng trong việc thúc đẩy phối hợp liên chính phủ ở cấp quốc tế và trong việc định hình chương trình nghị sự chính trị thông qua việc xây dựng các hướng dẫn và nguyên tắc xung quanh các chủ đề đặc thù mà sau đó chúng được các nước thành viên thông qua và thực hiện. Các tổ chức liên chính phủ như OECD, UNESCO, EU và Ngân hàng Thế giới trong những năm gần đây đã tích cực thúc đẩy các nỗ lực khoa học mở của các nước thành viên và (trong một số trường hợp) các nước không thành viên.

4.2. Quản trị dữ liệu mở: Những xu hướng chính sách gần đây

Các nước trên thế giới đang phát triển ngày càng nhiều các khung pháp lý và chính sách, các hướng dẫn và sáng kiến để khuyến khích tính mở nhiều hơn trong khoa học. Các nước cũng đang thúc đẩy cơ chế khuyến khích và tài trợ.

Tuy nhiên, cần lưu ý rằng chính sách truy cập mở (chính sách thúc đẩy truy cập tới các bài báo khoa học) hoàn thiện hơn các chính sách dữ liệu nghiên cứu mở. Điều này là do việc xác định các bộ dữ liệu không dễ như một bài báo nghiên cứu cuối cùng. Ngoài

ra, để được sử dụng lại, dữ liệu cần được làm sạch và được liên kết với siêu dữ liệu. Việc công bố dữ liệu cũng có thể làm tăng mối quan ngại về quyền riêng tư và bảo mật cũng như quyền sở hữu và các vấn đề sở hữu trí tuệ liên quan.

Theo một cuộc khảo sát gần đây được Ủy ban châu Âu tiến hành, các cơ quan tài trợ có ít chính sách dữ liệu mở hơn các chính sách truy cập mở: 48 nhà tài trợ chính sách truy cập mở được ghi nhận tại ROARMAP (Cơ quan đăng ký các chính sách) ở châu Âu, Brazil, Canada và Hoa Kỳ, 23% trong số đó không công bố dữ liệu theo yêu cầu và 38% không đề cập đến dữ liệu trong mô tả chính sách. Tuy nhiên, 29% trong số các chính sách chỉ thị chia sẻ dữ liệu nghiên cứu và 10% khuyến khích chia sẻ dữ liệu mà không có chỉ thị. Các tổ chức nghiên cứu (chẳng hạn như trường đại học hoặc trung tâm nghiên cứu công) có các kết quả tương tự, với 42% có chính sách truy cập mở nhưng chỉ 11% báo cáo có chính sách dữ liệu mở. Kết quả này cũng có thể liên quan đến thực tế là một vài kho lưu trữ của các tổ chức được dành riêng cho dữ liệu nghiên cứu, mặc dù cơ sở hạ tầng cho dữ liệu nghiên cứu mở có vẻ phát triển hơn so với khung chính sách cho dữ liệu nghiên cứu mở (36% người trả lời cho biết tổ chức của họ đã phát triển kho lưu trữ dữ liệu nghiên cứu trong khi 11% cho biết họ đang áp dụng các chính sách dữ liệu nghiên cứu mở).

Các nỗ lực và quản trị khoa học mở có thể được hưởng lợi từ các phương pháp tiếp cận tư vấn đa bên và được tổ chức thành mạng lưới, bao gồm tất cả nhân tố có liên quan: chính phủ, viện nghiên cứu và khu vực học thuật, xã hội dân sự và khu vực doanh nghiệp. Nhiều biện pháp chính sách để thúc đẩy khoa học mở có thể được phát triển và chấp nhận bởi các tác nhân đa dạng ở cấp quốc gia cũng như ở cấp tổ chức, chẳng hạn như trong trường hợp của các trường hợp đại học và viện nghiên cứu công. Các biện pháp chính sách có thể bao gồm nhiều hành động và sáng kiến khác nhau, chẳng hạn như quy tắc bắt buộc, cơ chế khuyến khích hoặc người sáng tạo. Các biện pháp thuộc ba loại hành động này có thể được thực hiện cùng nhau để thúc đẩy truy cập mở bằng những phương pháp tiếp cận tích hợp và đa diện. Tuy nhiên, các xu hướng chính sách gần đây cho thấy phần lớn sáng kiến liên quan đến quy tắc và yêu cầu bắt buộc và việc phát triển cơ sở hạ tầng cho khoa học mở. Cho đến nay, có ít sáng kiến liên quan đến định nghĩa về cơ chế khuyến khích phi tiền tệ: thông thường, các cơ chế khuyến khích ở dạng tài trợ bao gồm cả chi phí xuất bản truy cập mở, trong khi định nghĩa tiêu chí tài trợ mới cho các nhà nghiên cứu tham gia vào truy cập mở và các hoạt động về dữ liệu mở ít phổ biến hơn, mặc dù chúng hiện đang được thảo luận ở một số quốc gia.

Các quy định và yêu cầu bắt buộc

Ở một số nước OECD, các cơ quan tài trợ lớn bắt buộc truy cập tự do tới các kết quả nghiên cứu mà họ tài trợ. Ở hầu hết các quốc gia, yêu cầu này được giới hạn ở cấp độ truy cập miễn phí hay truy cập mở tới kết quả nghiên cứu được tài trợ công và các nhà nghiên cứu tự do lựa chọn cách họ muốn tiết lộ bài báo nghiên cứu - có nghĩa là hoặc đi theo đường vàng (thông qua các tạp chí truy cập mở) hoặc tự lưu trữ bài báo trên các kho lưu trữ trực tuyến (con đường xanh). Chính sách truy cập mở ở Anh và Hà Lan không giống như vậy. Tại Anh, theo các khuyến cáo của báo cáo Finch, con đường vàng đã được ưu tiên hơn con đường xanh như một cách để ủy bắt buộc và có được quyền truy cập mở vào kết quả nghiên cứu một cách hiệu quả hơn. Ở Hà Lan, chiến lược quốc gia về truy cập mở ưu tiên xuất bản truy cập mở vàng, mặc dù chấp nhận truy cập mở xanh.

Ở một số quốc gia và một số tổ chức, các nhà nghiên cứu được yêu cầu công bố nghiên cứu ngay tại ngày thời điểm bản thảo được chấp nhận. Trong các trường hợp khác, thời hạn cấm vận được cho phép. Thời hạn cấm vận có thể khác nhau, nhưng thường không quá 12 tháng kể từ khi xuất bản. Trong một số trường hợp, bài báo được yêu cầu rõ ràng ở định dạng có thể đọc được bằng máy.

Yêu cầu về việc tiết lộ các bộ dữ liệu mở đã được các quốc gia triển khai ít hơn và mức độ thực hiện chính sách ở khía cạnh này ở giai đoạn kém hoàn thiện hơn, trong đó Dự án thí điểm Dữ liệu mở trong các chương trình của EU là một ví dụ. Khi những chính sách bắt buộc đối với việc công bố dữ liệu được chấp nhận, những yêu cầu thường bao gồm việc phát triển kế hoạch quản trị dữ liệu (xem xét các vấn đề như quản trị dữ liệu, bảo trì, bảo quản) và phát triển siêu dữ liệu.

Các yêu cầu có thể ít nhiều cụ thể. Chúng liên quan đến lộ trình truy cập mở cụ thể, loại giấy phép bản quyền hợp pháp, loại kho lưu trữ (thể chế hoặc các quy định cụ thể). Những yêu cầu khác có thể liên quan đến việc xuất bản dữ liệu ở định dạng rõ ràng để có thể tương thích hoặc phát triển siêu dữ liệu. Các quy định bắt buộc đối với cả các công bố và dữ liệu thường cho phép ngoại lệ vì những lý do như bảo mật, tính riêng tư. Ngoài ra, các quy định bắt buộc được thiết kế phù hợp với nguyên tắc đạo đức của quốc gia và khung bản quyền.

Ví dụ về sáng kiến quốc gia áp dụng các yêu cầu bắt buộc đã được phát triển ở một số nước trên thế giới. Tại Bỉ, quy định của cơ quan tài trợ khu vực Liên bang Wallonia-Brussels yêu cầu các nhà nghiên cứu lưu giữ các kết quả nghiên cứu trong kho lưu giữ nội bộ. Kế hoạch quốc gia về nghiên cứu và đổi mới KH&CN của Tây Ban Nha yêu cầu sử dụng kho lưu trữ nội bộ hay kho lưu trữ theo chủ đề, không quá 12 tháng sau khi xuất bản. Biên bản ghi nhớ của Văn phòng Chính sách KH&CN Hoa Kỳ (OSTP) quy định

việc xây dựng kế hoạch để tăng cường sự tiếp cận công cộng tới các ấn phẩm và dữ liệu số từ nghiên cứu được tài trợ, bao gồm các vấn đề bảo quản lâu dài. Hội đồng Nghiên cứu Anh đã phát triển các hướng dẫn nêu rõ rằng nghiên cứu được tài trợ công phải có sẵn, tốt hơn là bằng mô hình truy cập mở vàng, nhưng truy cập mở xanh là một lựa chọn có thể chấp nhận được. Các hướng dẫn của Hội đồng Nghiên cứu Anh cũng bao gồm các chỉ dẫn về giấy phép bản quyền để sử dụng trong trường hợp truy cập mở vàng.

Ở Phần Lan, Sáng kiến Nghiên cứu và Khoa học mở (ATT) do Bộ Giáo dục và Văn hóa triển khai xem xét việc áp dụng các quy tắc bắt buộc để thúc đẩy khoa học mở. Ngoài ra, Viện Hàn lâm Phần Lan đề xuất xuất bản truy cập mở bất cứ khi nào có thể. Quỹ Khoa học quốc gia Thụy Sĩ yêu cầu người được tài trợ đưa bài báo hoặc sách vào trong kho lưu trữ theo chủ đề hoặc các kho lưu trữ nội bộ. Cả Cơ quan tài trợ Phần Lan và Thụy Sĩ đều chấp nhận xuất bản truy cập mở xanh hoặc vàng.

Cơ quan KH&CN Nhật Bản khuyến cáo việc xuất bản truy cập mở các đầu ra của nghiên cứu mà họ tài trợ, thông qua kho lưu trữ nội bộ (truy cập mở xanh). Bộ Giáo dục, Văn hóa, Thể thao, Khoa học và Công nghệ Nhật Bản (MEXT) yêu cầu xuất bản truy cập mở các luận án tiến sĩ, theo nghị định của Bộ.

Cơ chế tài trợ và khuyến khích

Khoa học mở có thể được hỗ trợ bằng cách xác định cơ chế khuyến khích phù hợp để thúc đẩy hành vi mở trong khoa học và nghiên cứu. Các cơ chế khuyến khích có thể nhắm vào các khía cạnh khác nhau của quá trình nghiên cứu, bao gồm ví dụ hỗ trợ tài chính cho các nỗ lực khoa học mở, thừa nhận và tài trợ cho các nhà nghiên cứu thực hiện các hoạt động khoa học mở hoặc sử dụng các chỉ số đánh giá mới và rộng hơn có tính đến khoa học mở và tác động của nó. Cho đến nay, hỗ trợ tài chính cho các nỗ lực khoa học mở đã được thực hiện thường xuyên hơn các loại khuyến khích khác.

Thừa nhận các hoạt động liên quan đến khoa học mở trong việc đánh giá các nhà nghiên cứu, ví dụ về phân bổ tài trợ hoặc quy trình thăng tiến nghề nghiệp, có thể là cách tốt để thúc đẩy các nỗ lực khoa học mở. Ở hầu hết các quốc gia, khung chính sách hiện tại không khuyến khích các nỗ lực chia sẻ, đặc biệt là đối với kết quả, bộ dữ liệu hoặc tài liệu nghiên cứu khác ở giai đoạn trước xuất bản. Khoa học là một quá trình cạnh tranh và các nhà nghiên cứu có thể lo sợ hành vi phi đạo đức hoặc đơn giản là có thể không tận dụng được các cơ hội khoa học mở nếu không thực sự thừa nhận và chính thức công nhận. Các cuộc điều tra liên quan đến thái độ và hành vi của các nhà nghiên cứu có thể giúp thiết kế các cơ chế khuyến khích hiệu quả hơn. Ví dụ, ở Chile, chính sách truy cập mở quốc gia được giám sát bởi một ủy ban thẩm định gồm các chuyên gia quốc tế xem

xét nhận thức của sinh viên đại học, học giả, nhà nghiên cứu và biên tập viên... về chính sách.

Một vài khuyến khích để thúc đẩy chia sẻ dữ liệu thậm chí còn rõ rệt hơn. Tác động về mặt học thuật đối với cá nhân hoặc tổ chức có ảnh hưởng lớn đến việc phân bổ kinh phí nghiên cứu (nghĩa là thành công hay thất bại của các đề xuất tài trợ nghiên cứu); nó cũng ảnh hưởng đến sự tiến bộ của sự nghiệp học tập của cá nhân và tình hình của tổ chức. Do đó, những đo đạc tác động của học thuật được sử dụng bởi các cơ quan tài trợ nghiên cứu đang định hình mạnh mẽ hành động và quyết định của các nhà nghiên cứu, khuyến khích một số hành vi nhất định.

Hiện nay, những đo lường tác động của học thuật tập trung vào những công bố trên tạp chí khoa học, trong khi chia sẻ (công bố) dữ liệu thường không được xem xét. Các nhà nghiên cứu được tài trợ khi công bố các kết quả thu được từ việc phân tích bộ dữ liệu, thay vì công bố các bộ dữ liệu đó. Điều này có thể tạo ra các tình huống trong đó các nhà nghiên cứu sẵn sàng bảo hộ các bộ dữ liệu của họ chứ không phải công bố chúng. Mặt khác, một số tạp chí uy tín (như Nature và Science) đòi hỏi sự sẵn có của tất cả các bộ dữ liệu cần để hiểu, đánh giá, nhân rộng và xây dựng dựa trên kết quả nghiên cứu được công bố.

Hơn nữa, thu thập dữ liệu, làm sạch, thẩm định và biên tập là nhiệm vụ tiêu tốn thời gian và các nhà nghiên cứu thường muốn khai thác triệt để những cơ hội phát sinh từ dữ liệu mà họ thu thập và tạo ra để tối đa hóa kết quả mà họ có được từ một bộ dữ liệu. Nếu không có sự thừa nhận những hoạt động này, thì có rất ít khuyến khích cho các nhà nghiên cứu chia sẻ dữ liệu hoặc làm cho nó luôn sẵn có cho cộng đồng nghiên cứu truy cập được bằng cách phát triển siêu dữ liệu và các thông tin cần thiết khác để tái sử dụng. Ngoài việc công nhận những nhiệm vụ này quy trình thăng tiến nghề nghiệp hoặc phân bổ tài trợ, cộng đồng nghiên cứu đã đề xuất phát triển công cụ trích dẫn dữ liệu để ghi nhận tác giả bộ của các dữ liệu và siêu dữ liệu.

Ví dụ về các sáng kiến được phát triển để tạo ra những khuyến khích cho các nỗ lực khoa học mở bao gồm những điều sau đây. Sáng kiến quốc gia khoa học mở tại Phần Lan (ATT) sẽ khuyến khích sự tham chiếu đến dữ liệu và các phương pháp cũng như trao thưởng cho các nhóm và nhà nghiên cứu thực hiện có những nỗ lực thúc đẩy khoa học mở. Tại Bỉ, cơ quan tài trợ ở Wallonia đã triển khai mô hình Đại học Liège sử dụng kho lưu trữ truy cập mở xanh để đánh giá và trao cho các nhà nghiên cứu liên kết với tổ chức đó. Ở Tây Ban Nha, việc đưa ra tiêu chí trong các quy trình đánh giá của các nhà nghiên cứu có tính đến các nỗ lực thúc đẩy khoa học mở hiện đang được thảo luận nhưng chưa

được thực thi. Tại Hoa Kỳ, số liệu thống kê về số lượng công bố và tải xuống từ kho lưu trữ PubMed Central cho thấy số lượng bài báo nghiên cứu có thể được truy cập công cộng. Ngoài ra, Viện Y tế quốc gia Hoa Kỳ thường xuyên gán số nhận dạng duy nhất cho bộ dữ liệu để chúng có thể được trích dẫn chính xác.

Cơ sở hạ tầng, kỹ năng cho khoa học mở, khung pháp lý

Cơ sở hạ tầng

Nhiều nước đã đầu tư vào cơ sở hạ tầng cần thiết để thúc đẩy khoa học mở, bao gồm kho lưu trữ trực tuyến, CSDL, thư viện kỹ thuật số và lưu trữ và các nền tảng chứa thông tin về các dự án NC&PT và lý lịch của các nhà nghiên cứu. Phần Lan đã triển khai một lộ trình phát triển cơ sở hạ tầng để thúc đẩy khoa học mở. Trung Quốc đã phát triển các nền tảng trực tuyến để lưu giữ dữ liệu và công bố. Argentina đã phát triển CSDL SICyTAR với thông tin về lý lịch, các công bố và các hợp tác nghiên cứu của các nhà nghiên cứu. Cả Chile và Mexico đã đầu tư vào một số kho quốc gia để chia sẻ các bài báo khoa học và dữ liệu. Tại Tây Ban Nha, RECOLECTA là kho lưu trữ quốc gia và cơ sở hạ tầng chính cho phép các nhà nghiên cứu và các bên liên quan khác tự do lưu trữ và truy cập các công bố khoa học. Tại Ba Lan, một số cổng trực tuyến và thư viện ảo đã được phát triển để tạo điều kiện lưu trữ truy cập mở. Tại Anh, Hội đồng Lãnh đạo cơ sở hạ tầng điện tử là cơ quan điều phối duy nhất chịu trách nhiệm về chiến lược cơ sở hạ tầng điện tử của Anh và tư vấn cho Bộ Kinh doanh, Đổi mới và Kỹ năng. Ủy ban châu Âu cũng đã tích cực trong việc thúc đẩy sự phát triển của kho lưu trữ và các nền tảng của các quốc gia thành viên EU. Cơ sở hạ tầng này không chỉ có thể chia sẻ tài liệu khoa học mà còn thúc đẩy sự hợp tác cởi mở và phát triển một nền văn hóa khoa học mở. Cơ quan Khoa học và Công nghệ Nhật Bản đã phát triển J-STAGE, một nền tảng trực tuyến mở cho các nhà xuất bản, nơi có thể gửi các bài báo, quản lý bình duyệt đồng nghiệp và xuất bản các bài báo trực tuyến.

Kỹ năng

Yếu tố hỗ trợ bổ sung liên quan đến kỹ năng cần thiết cho phép những nhân tố và các bên liên quan khác nhau sử dụng và tạo ra công cụ khoa học mở. Một số nghiên cứu đã gợi ý rằng sự thiếu hụt các kỹ năng liên quan đến dữ liệu tạo thành một trong những rào cản chính trong việc sử dụng phân tích dữ liệu. Những kỹ năng liên quan đến khoa học dữ liệu được phân phối không đồng đều giữa các quốc gia, như được gợi ý bởi một số chỉ số liên quan đến khả năng giải quyết vấn đề trong môi trường giàu công nghệ. Kết quả là, kỹ năng cho phép các nỗ lực khoa học mở cần được xác định và tăng cường trong chương trình giáo dục. Ngoài ra, đào tạo và phát triển kỹ năng có thể nhắm mục tiêu các

nhà khoa học và nhà nghiên cứu cùng với tất cả những chu kỳ giáo dục và làm việc. Không chỉ sinh viên mà còn cả nhiều nhà nghiên cứu và nhà khoa học trưởng thành cần có hoặc có được các kỹ năng cần thiết để sử dụng kho lưu trữ mở và lưu trữ, học cách phát triển siêu dữ liệu, và làm mới và duy trì dữ liệu mà chúng sản xuất.

Các sáng kiến gần đây nhằm vào việc phát triển kỹ năng cho khoa học mở bao gồm Hướng dẫn quản lý dữ liệu do Bộ Giáo dục và Văn hóa Phần Lan phát triển; Hướng dẫn vấn đề liên quan đến quản lý dữ liệu và mô tả dịch vụ hiện có để quản lý dữ liệu có sẵn trong nước. Ở Phần Lan, nhiều cơ sở giáo dục đại học cung cấp các khóa đào tạo cho các nhà nghiên cứu và sinh viên về quản lý dữ liệu và sở hữu dữ liệu. Ở Ba Lan, trung tâm nghiên cứu mới về dữ liệu lớn, OCEAN, được phát triển bởi Trung tâm liên ngành về Mô hình toán học và tính toán của Đại học Warszawa và được đồng tài trợ bởi Trung tâm Nghiên cứu và Phát triển Quốc gia, với mục tiêu không chỉ cung cấp cơ sở hạ tầng điện tử để lưu trữ dữ liệu mà còn cả là chuyên môn và đào tạo để phân tích dữ liệu lớn. Ngoài ra, Ba Lan là một trong những đối tác trong một dự án mới bắt đầu, FOSTER, nhằm hỗ trợ các nhà nghiên cứu trẻ trong việc áp dụng cách tiếp cận mở và tuân thủ các chính sách tiếp cận mở. Dự án dự định thiết lập một chương trình đào tạo toàn châu Âu về tiếp cận mở và dữ liệu mở, củng cố các hoạt động đào tạo trên khắp châu Âu. Dự án cũng bao gồm các đối tác từ Đan Mạch, Pháp, Đức, Hà Lan, Bồ Đào Nha, Tây Ban Nha và Anh.

Pháp gần đây đã thành lập một số sáng kiến để phát triển các kỹ năng liên quan đến khoa học mở. Thư viện khoa học kỹ thuật số quốc gia đã tạo ra một nhóm làm việc tập trung vào các ngành nghề mới xung quanh việc truy cập mở và dữ liệu mở. Trung tâm Truyền thông Khoa học Trực tiếp (CCSD) cung cấp đào tạo trong nền tảng truy cập mở quốc gia (HAL). Cuối cùng, bảy đơn vị đào tạo khu vực cung cấp các khóa học cho các nhà nghiên cứu và sinh viên tiến sĩ về truy cập mở và nghiên cứu dữ liệu mở.

Tại Hoa Kỳ, bản ghi nhớ OSTP nêu bật tầm quan trọng của việc hỗ trợ đào tạo, giáo dục và phát triển lực lượng lao động liên quan đến quản lý dữ liệu khoa học, phân tích, lưu trữ, bảo quản và quản lý. Chỉ riêng NIH, việc phát triển các kỹ năng phân tích dữ liệu là một trong bốn trụ cột chính của dữ liệu lớn cho sáng kiến Kiến thức, có mục tiêu phát triển thêm các kỹ năng trong khoa học về dữ liệu lớn ngoài việc phát triển các kỹ năng sử dụng dữ liệu và phân tích trong lĩnh vực y sinh học. Ngoài ra, một số Viện và Trung tâm NIH cung cấp kinh phí đào tạo về tin - y sinh học.

Theo Kế hoạch hành động về Chính phủ mở 2.0, Chính phủ Canada sẽ tối đa hóa quyền truy cập vào nghiên cứu khoa học được liên bang tài trợ, khuyến khích cộng tác và tham gia rộng hơn với cộng đồng khoa học, khu vực tư nhân và công chúng. Nội dung

được hoàn thành vào năm 2016 bao gồm: xây dựng hồ sơ về năng lực kỹ năng số của Canada và tăng hiểu biết về mối quan hệ giữa kỹ năng số và thị trường lao động và kết quả xã hội; phát triển công cụ trực tuyến và tài liệu đào tạo để cải thiện kỹ năng số của cá nhân; tài trợ cho các sáng kiến của khu vực tư nhân và xã hội dân sự nhằm cải thiện kỹ năng số.

Ở Anh, Chiến lược năng lực dữ liệu tập trung vào (ngoài các vấn đề khác) vốn nhân lực và phát triển kỹ năng để phân tích dữ liệu cũng như khả năng truy cập dữ liệu và kỹ năng chia sẻ dữ liệu ở người tiêu dùng, doanh nghiệp và học viện. Ngoài ra, việc thành lập trung tâm đào tạo tiên sĩ về dữ liệu lớn đã được công bố ở một số trường đại học và các tổ chức giáo dục đại học trong nước. Với những mục tiêu tương tự, Viện dữ liệu mở đã được thành lập gần đây tại London - Anh. Trung tâm Tin học Quốc gia Ấn Độ đã tổ chức một số hội thảo để nâng cao nhận thức khoa học mở cho các học giả.

Tại Đức, Hiệp hội Helmholtz, quan hệ đối tác của 18 trung tâm nghiên cứu khoa học kỹ thuật và sinh học y tế, thường xuyên đào tạo và tổ chức các lớp về bảo trì và quản lý dữ liệu nghiên cứu để thúc đẩy nền văn hóa khoa học mở cho các nhà nghiên cứu.

Khung pháp lý

Khung pháp lý thân thiện với khoa học mở là phương tiện bổ sung để thúc đẩy các nỗ lực phát triển khoa học mở. Một số quốc gia hiện đang thảo luận về việc sửa đổi các quy định về sở hữu trí tuệ đối với nghiên cứu hoặc miễn thuế. Ví dụ, Australia và Phần Lan hiện đang thảo luận về các sửa đổi khung pháp lý hiện hành về công bố kết quả nghiên cứu được tài trợ công khai để luật bản quyền ngày càng thân thiện với khoa học mở. Ở Đức, Luật bản quyền quốc gia được sửa đổi vào năm 2013 cho phép các nhà khoa học và nhà nghiên cứu được tài trợ công có quyền hợp pháp tải các công bố của họ trực tuyến, ngay cả khi họ đã chuyển nhượng quyền khai thác chúng cho nhà xuất bản, sau thời gian cấm lên tới 12 tháng. Gần đây, Vương quốc Anh đã thông qua một loạt sửa đổi khung pháp lý về bản quyền (có hiệu lực vào năm 2014), bao gồm quyền tự do sử dụng lại tài liệu được sao chép hoặc ghi lại cho mục đích giáo dục và phi thương mại.

Khoa học mở toàn cầu

Quản trị khoa học mở trên phương diện quốc tế và trong một số trường hợp, trên phương diện toàn cầu được các tổ chức chính phủ quốc tế (IGO) tạo điều kiện thuận lợi. IGO đóng vai trò quan trọng trong việc thúc đẩy phối hợp giữa Chính phủ ở cấp quốc tế và trong việc định hình chương trình nghị sự chính trị thông qua việc xây dựng những nguyên tắc xung quanh các chủ đề cụ thể, sau đó được các nước thành viên chấp nhận và thực hiện.

Các IGO như OECD, UNESCO, EU và Ngân hàng Thế giới đã hoạt động trong những năm gần đây trong việc thúc đẩy nỗ lực khoa học mở của thành viên và, trong một số trường hợp, các nước không phải là thành viên. Vào tháng 6 năm 2013, Bộ trưởng Khoa học G8 đã đưa ra một tuyên bố thừa nhận tầm quan trọng của dữ liệu nghiên cứu mở và truy cập mở đối với các bài báo nghiên cứu được tài trợ công.

OECD đã chủ động xây dựng các hướng dẫn và nguyên tắc về chủ đề liên quan đến khoa học mở, bao gồm việc tiếp cận dữ liệu nghiên cứu (Hộp 3.2 và 3.3) hoặc thông tin khu vực công. OECD, cùng với WIPO và các tổ chức quốc tế khác, có vai trò cơ bản trong việc phát triển giấy phép Creative Commons mới cho IGO (Hộp 2.4). Nhiều ủy ban OECD và các bên công tác đã làm việc về vấn đề liên quan đến dữ liệu mở và khoa học mở. Ví dụ, tiềm năng của khoa học mở và nỗ lực mở rộng nghiên cứu để giải quyết bệnh Alzheimer và chứng mất trí nhớ đã được nhấn mạnh gần đây bởi tư vấn chuyên gia OECD “Mở rộng hợp tác toàn cầu để thúc đẩy đổi mới bệnh Alzheimer và mất trí nhớ” cũng như ở các lĩnh vực y tế khác. Liên minh châu Âu đã áp dụng và thúc đẩy các nỗ lực khoa học mở trong Chương trình khung nghiên cứu và đổi mới gần đây nhất; đặc biệt, Horizon 2020 yêu cầu truy cập mở vào các ấn phẩm nghiên cứu được tài trợ thông qua chương trình này.

UNESCO cam kết thúc đẩy tiếp cận mở với tài liệu khoa học của các tổ chức nghiên cứu được tài trợ công. UNESCO tiến hành hoạt động nâng cao nhận thức và thúc đẩy thực hành tiếp cận mở giữa các nhà hoạch định chính sách, nhà nghiên cứu và người quản lý tri thức. Nhiều hành động được thúc đẩy trong hợp tác với mạng lưới toàn cầu các văn phòng khu vực và địa phương của tổ chức. UNESCO dành sự chú ý đặc biệt đến những lợi ích phát sinh từ việc tiếp cận mở với các nước châu Phi và các nước đang phát triển khác, nơi những nỗ lực tiếp cận mở ít phát triển hơn. UNESCO đã biên soạn một loạt hướng dẫn chính sách về việc áp dụng và thực hiện nỗ lực tiếp cận mở, nhằm cung cấp thông tin cho các nước thành viên UNESCO; để giúp chọn chính sách truy cập mở phù hợp nhất với bối cảnh cụ thể của họ; và thúc đẩy việc áp dụng chính sách truy cập mở.

Cuối cùng, Ngân hàng Thế giới đã thông qua chính sách nội bộ truy cập mở hoàn toàn cho các ấn phẩm mà nó tạo ra, và đả cải thiện đáng kể việc cung cấp dữ liệu công cho tất cả người dùng và các bên liên quan có thể.

KẾT LUẬN

Qua những phân tích ở trên có thể rút ra những kết luận sau:

Khoa học mở là một phương tiện và không phải là mục đích. Các chiến lược và chính sách khoa học mở là một phương tiện để hỗ trợ chất lượng khoa học tốt hơn, tăng cường cộng tác và sự gắn kết giữa nghiên cứu và xã hội mà chúng có thể dẫn đến tác động kinh tế và xã hội cao hơn tác động của nghiên cứu được tài trợ công.

Khoa học mở rộng hơn truy cập mở đến công bố khoa học hoặc dữ liệu; nó bao gồm nhiều khía cạnh và các giai đoạn của quá trình nghiên cứu. Điều quan trọng là phải nhớ rằng khoa học mở là một khái niệm rộng hơn, bao gồm khả năng tương tác của cơ sở hạ tầng khoa học, các phương pháp nghiên cứu mở và chia sẻ (như ứng dụng mở và mã nguồn mở) và các công cụ thân thiện với máy móc cho phép, ví dụ, khai phá văn bản và dữ liệu.

Các chính sách thúc đẩy dữ liệu mở vẫn còn chưa đủ so với các chính sách thúc đẩy truy cập mở đến các công bố khoa học. Mặc dù các nguyên tắc tiếp cận mở với dữ liệu khoa học được thiết lập tương đối tốt ở các nước OECD, nhưng phạm vi tiếp cận này là khác nhau rất lớn giữa các quốc gia. Điều này là do thực tế rằng các bộ dữ liệu không dễ dàng được xác định và được định nghĩa như với các bài báo nghiên cứu học thuật. Sự đa dạng của dữ liệu khoa học, các truyền thống và tiêu chuẩn khác nhau trong xử lý dữ liệu cũng là vấn đề. Một số thách thức bổ sung liên quan đến các tập dữ liệu có thể là định nghĩa về quyền sở hữu các tập dữ liệu quy mô lớn có thể được thu thập bởi máy móc hoặc các nhà cung cấp phần mềm; quyền riêng tư; bảo mật; và thậm chí cả các vấn đề an ninh quốc gia. Thêm vào đó, một số loại dữ liệu nhất định, chẳng hạn như hồ sơ y tế, là đặc biệt nhạy cảm với các vấn đề về quyền riêng tư.

Các chính sách khoa học mở cần có tính nguyên tắc nhưng phải điều chỉnh phù hợp với thực tế địa phương (quốc gia). Các chính sách khoa học mở đòi hỏi đa dạng các phương pháp tiếp cận có tính đến nhu cầu của các tác nhân khác nhau tham gia vào các dự án nghiên cứu. Ví dụ, nếu một dự án nghiên cứu liên quan đến các đối tác khu vực kinh doanh và lợi ích thương mại, các yêu cầu chia sẻ kết quả nghiên cứu có thể sẽ khác với trường hợp chỉ có các tổ chức công tham gia dự án. Trong các trường hợp khác, các vấn đề về quyền riêng tư hoặc bảo mật có thể phải xem xét khi xử lý từng lớp dữ liệu nhất định.

Cần có các cơ chế khuyến khích tốt hơn để thúc đẩy chia sẻ dữ liệu giữa các nhà nghiên cứu. Trong khi tất cả các nhà nghiên cứu khu vực công quan tâm đến tối đa hóa việc chia sẻ các bài báo nghiên cứu đã công bố, điều này là chưa chắc đúng cho các bộ dữ liệu nghiên cứu, đặc biệt ở giai đoạn tiền xuất bản. Ngoài ra, công tác làm sạch và biên

tập dữ liệu (curation) (ví dụ, bổ sung siêu dữ liệu) là một hoạt động tốn thời gian nhưng ít khi được thừa nhận trong các cơ chế đánh giá hoặc các thủ tục phân bổ tài trợ. Hầu hết các đánh giá của các trường đại học và các nhà nghiên cứu là gần như dựa hoàn toàn vào các chỉ số giảng dạy và trách lượng thư mục, mà chúng đóng góp ít giá trị cho việc đánh giá chia sẻ các đầu vào tiền xuất bản và kết quả sau công bố, chẳng hạn như dữ liệu và các thông tin liên quan khác. Việc mở rộng cơ chế trích dẫn đến các bộ dữ liệu có thể giải quyết một phần vấn đề này.

Cần phát triển kỹ năng liên quan đến dữ liệu. Các kỹ năng của các nhà nghiên cứu cần có để chia sẻ các công bố khoa học hoặc tập dữ liệu một cách mở trên mạng là không đồng đều như nhau. Một số ngành như khoa học máy tính hoặc vật lý có thể có truyền thống hơn trong việc tải lên tài liệu nghiên cứu lên các kho và trong biên tập và duy trì các tập dữ liệu lớn. Tuy nhiên, các nhà nghiên cứu trong một số ngành khác có thể cần được đào tạo thêm để phát triển các kỹ năng cần thiết để làm cho khoa học mở được triển khai. Ngoài ra, các sinh viên và công chúng cũng cần có được các kỹ năng để tận dụng, sử dụng và tái sử dụng tập hợp dữ liệu được cộng đồng nghiên cứu chia sẻ. Một số quốc gia hiện đang xây dựng các chương trình giảng dạy về khoa học dữ liệu để giải quyết vấn đề này.

Cần đào tạo và nâng cao nhận thức cho các nhà nghiên cứu về phát triển văn hóa khoa học mở. Các cuộc khảo sát gần đây về hành vi của các nhà khoa học cho thấy không phải tất cả các nhà nghiên cứu đều nhận thức được các khả năng mà khoa học mở cung cấp. Ở một số quốc gia, nhiều tổ chức khác nhau thường xuyên tổ chức các hội thảo và các khóa đào tạo cho các nhà nghiên cứu để nhận thức được những khả năng này.

Kho lưu trữ và nền tảng trực tuyến sẽ không có tác động nếu thông tin mà chúng chứa không có chất lượng tốt. Nếu các kho lưu trữ không thân thiện với người dùng và các bộ dữ liệu chúng chứa chưa được làm sạch và biên tập/xử lý đúng đắn hoặc các siêu dữ liệu chưa được phát triển đầy đủ, việc sử dụng chúng có thể sẽ khó tối đa hóa.

Chi phí bảo quản dài hạn các kết quả nghiên cứu được truy cập mở cần được xem xét. Mở truy cập không phải là không có chi phí. Nhiều chính phủ và tổ chức nghiên cứu hiện đang chịu các chi phí để cung cấp truy cập mở đến các bài báo và dữ liệu, cũng như chi phí lưu trữ và bảo quản các tập dữ liệu trên mạng. Với khối lượng dữ liệu tăng lên nhanh chóng, các tổ chức công sẽ gặp khó khăn để tìm nguồn tài trợ bền vững và đưa ra mô hình kinh doanh bền vững. Quan hệ đối tác công-tư với các nhà cung cấp dịch vụ tư nhân có thể cung cấp những giải pháp sáng tạo.

Cần có khung pháp lý rõ ràng cho việc chia sẻ các công bố khoa học và tái sử dụng các tập dữ liệu ở cấp độ quốc gia và quốc tế. Sự thiếu rõ ràng về việc diễn giải khung pháp lý quốc gia và quốc tế có thể cản trở việc chia sẻ hoặc tái sử dụng các kết quả

ngiên cứu. Ngoài ra, cần có các hướng dẫn rõ ràng về khai phá văn bản và dữ liệu, vì công cụ này sẽ được các nhà nghiên cứu sử dụng ngày càng nhiều trong tương lai. Gần đây, một số quốc gia OECD đang thảo luận hoặc đã sửa đổi khung pháp lý quốc gia để làm cho chúng ngày càng thân thiện với khoa học mở.

Các phương pháp tư vấn liên quan đến tất cả các nhân tố cho khoa học mở là một thành phần chủ chốt của các chiến lược khoa học mở thành công. Các nỗ lực khoa học mở bao gồm các cộng đồng và các nhân tố khác nhau: Các nhà nghiên cứu, các viện và trung tâm nghiên cứu, thư viện và trung tâm dữ liệu, các tổ chức phi lợi nhuận tư nhân, các tổ chức kinh doanh bao gồm các nhà xuất bản hàn lâm, các thực thể siêu quốc gia, công dân, v.v.. Những nhân tố này không nhất thiết phải có cùng các khuyến khích, mục tiêu hoặc kỳ vọng. Một chiến lược thành công cần phải tính đến sự đa dạng này và phản ứng một cách phù hợp.

Cần có sự hợp tác quốc tế trong lĩnh vực khoa học mở để giải quyết các thách thức toàn cầu. Hợp tác quốc tế ngày càng trở nên quan trọng hơn bao giờ hết bởi các công bố khoa học và dữ liệu ở dạng số có thể di chuyển xuyên biên giới quốc gia. Cơ sở hạ tầng được chia sẻ và tương thích là cần thiết để phổ biến các kết quả nghiên cứu và thúc đẩy hợp tác khoa học. Những nỗ lực này có thể giúp tránh sự trùng lặp trong các nỗ lực, cũng như giúp chia sẻ các rủi ro hoặc các khoản đầu tư liên quan. Ngoài ra, các nước BRIC như Brazil, Trung Quốc và Ấn Độ cũng đang triển khai các chính sách khoa học mở và lộ trình cơ sở hạ tầng dữ liệu. Hợp tác quốc tế và hợp tác trong lĩnh vực khoa học mở sẽ trở nên quan trọng hơn khi sự sản sinh tri thức và hoạt động NC&PT toàn cầu đang hướng nhiều hơn tới các nền kinh tế đang nổi lên. Hơn nữa, việc giải quyết những thách thức toàn cầu sẽ đòi hỏi sự tiếp cận và chia sẻ công cộng nhiều hơn các dữ liệu nghiên cứu quốc gia, và do đó, đòi hỏi sự hợp tác lớn hơn ở cấp độ toàn cầu.

Các nhà hoạch định chính sách cần thúc đẩy khoa học mở nhưng đồng thời bảo tồn cạnh tranh. Cạnh tranh là một khía cạnh quan trọng của doanh nghiệp khoa học: thúc đẩy truy cập mở và dữ liệu mở quá sớm có thể phản tác dụng trong một số trường hợp.

Biên soạn: Nguyễn Lê Hằng

Đào Thanh Vân

Tài liệu tham khảo chính: OECD (2015), Making open science a reality, OECD Publishing, Paris.