

MỘT SỐ NỘI DUNG PHÁT TRIỂN HỆ THỐNG CÁC CSDL TẠI VIỆN KHOA HỌC XÃ HỘI VIỆT NAM

Trần Mạnh Tuấn

Tại Viện Thông tin Khoa học Xã hội hiện có 15 CSDLTM đang được tạo lập, cập nhật. Ngoài ra, có 1 CSDLTM được tạo lập trên cơ sở tích hợp các CSDLTM sách của các thư viện tại các viện nghiên cứu chuyên ngành trực thuộc Viện Khoa học Xã hội Việt Nam. Tại hầu hết các Viện nghiên cứu chuyên ngành trực thuộc Viện KHXHVN, ngoài các loại CSDL thư mục phản ánh nguồn tài liệu là sách, tạp chí khoa học, hầu hết đều có một số CSDL đặc thù. Ví dụ, đó là các CSDL thư mục phản ánh các luận án khoa học (thạc sĩ và tiến sĩ), các CSDL phản ánh nguồn tư liệu khảo sát, điều tra... Có thể nói, hệ thống các loại CSDL được tạo lập và phát triển tại Viện KHXHVN là rất đa dạng, phong phú.

Cho đến nay đã xuất hiện nhiều công trình nghiên cứu, các luận án, luận văn khoa học thực hiện công việc này với quan điểm xem xét chúng là các sản phẩm thông tin được sử dụng để kiểm soát, tìm kiếm, truy cập các nguồn tài liệu hiện được quản lý tại các cơ quan thông tin thư viện thuộc Viện Khoa học Xã hội Việt Nam. Trong báo cáo này, cách tiếp cận như trên không được áp dụng. Ở đây, chúng tôi tập trung vào việc nghiên cứu đối chiếu giữa **chúng** với 4 nhóm CSDL được tạo lập theo thiết kế dịch vụ biên mục tích hợp (Connexion) do OCLC triển khai. Dịch vụ này được tạo lập từ 1998 và phát triển mạnh dưới hình thức Dự án CORC (Co-operative Online Resource Catalog). Bốn nhóm CSDL này bao gồm: a) CSDL (thường là CSDLTM) trực tiếp phản ánh nguồn thông tin (bộ sưu tập thông tin) – Resource DB; b) CSDL có chức năng quản lý người dùng và việc khai thác, sử dụng thông tin (Authority File); c) CSDL có chức năng cung cấp danh mục và kết nối trực tiếp đến các nguồn tin trên mạng (Path-Finder File); d) CSDL quản lý DDC và LCSH (WebDewey). Cách khảo sát và đối chiếu như thế cho đến nay chưa được thực hiện.

Viện Thông tin Khoa học Xã hội, cũng như tất cả các cơ quan thành viên của Hệ thống thông tin thư viện khoa học của nước ta, mới tập trung vào việc phát triển một loại CSDL (bao gồm các CSDL biệt lập và các OPAC) tương ứng với nhóm

¹ Viện Thông tin Khoa học xã hội

các CSDL phản ánh nguồn tin (CSDL thuộc nhóm a) nêu trên) trong Dự án CORC. Tại một số cơ quan thông tin lớn, như Trung tâm Thông tin Khoa học và Công nghệ Quốc gia, Thư viện Quốc gia ..., đã quan tâm tới việc xây dựng CSDL để quản lý người đọc, song mới dừng ở mức là các CSDL mang tính chất nội bộ, được khai thác một cách tự trị và biệt lập với hệ thống các CSDL khác (đây có thể xem là các CSDL thuộc nhóm b) trong Dự án CORC ở mức giản lược hơn). Còn lại, các CSDL thuộc nhóm c) và d) chưa hề được quan tâm tạo lập và phát triển. Đây là một hạn chế lớn khi xem xét từ quan điểm và quyền lợi của người dùng tin, từ phía nhà cung cấp và môi giới dịch vụ thông tin, cũng như từ phía người xây dựng chính sách, người làm nhiệm vụ quản lý Nhà nước đối với hoạt động thông tin thư viện.

Hệ thống các CSDL thuộc nhóm c) sẽ là công cụ quan trọng trợ giúp người dùng trực tiếp khai thác các nguồn thông tin trên mạng. Và đó đó, chúng cũng có vai trò quan trọng trợ giúp tích cực đối với các chuyên gia tìm kiếm, thu thập và phát triển nguồn tin trên cơ sở các nguồn thông tin khổng lồ trên web. Đây là hệ thống các CSDL siêu dữ liệu về các nguồn tài nguyên trên mạng, giúp các chuyên gia thông tin thư viện định hướng và bảo trì việc tích hợp nguồn tin mà mình tạo ra với nguồn tin bên ngoài. Về thực chất, nhóm CSDL này là một *tập hợp danh sách liên kết đến các nguồn tin đã lựa chọn trên mạng và danh sách này được biên tập lại để phù hợp với mục đích cụ thể* (theo Eric Childress, chuyên gia của OCLC). Ở giai đoạn hiện nay, khi tập trung phát triển CSDL thuộc nhóm pathfinder, ta cần chú ý tới sự xuất hiện và khả năng ứng dụng hệ thống chỉ số DOI (Digital Object Identifier), một công cụ xác định tài liệu trên mạng rất hữu dụng¹.

CSDL WebDewey (nhóm d)) *trong khuôn khổ dự án CORC* có 2 ý nghĩa quan trọng và chủ yếu: đối với chuyên gia thông tin thư viện trong quá trình biên mục, và vì thế nó là công cụ hạt nhân của dịch vụ biên mục tích hợp Connexion mà OCLC cung cấp; đối với người dùng tin trong quá trình khai thác các nguồn thông tin trên mạng.

¹ DOI là tên của một tài liệu/đề mục trên mạng. DOI chính là một hệ thống nhận dạng liên tục và động (mang tính chất của loại ngôn ngữ hậu kết hợp), và là hệ thống thực hiện chức năng trao đổi tương tác các thông tin số hoá đã được quản lý trên mạng. Tham khảo thêm: <http://www.doi.org/>.

Ở nước ta, do chưa xây dựng CSDL thuộc nhóm này, nên vấn đề tự động hoá trong quá trình xử lý thông tin chưa thể được triển khai, đồng thời cũng chưa thể nói đến việc tạo lập hoặc kiểm soát được sự tương thích giữa các quá trình xử lý thông tin với nhau (việc xử lý thông tin tại các cơ quan thông tin thư viện khác nhau), cũng như giữa quá trình xử lý thông tin với quá trình tìm tin sau này. Điều đó, đến lượt mình, làm hạn chế đáng kể hiệu quả và khả năng khai thác các nguồn hay các hệ thống thông tin ở nước ta nói chung và tại Viện Khoa học Xã hội Việt Nam nói riêng. Vấn đề này đã từng được nêu lên¹, song hầu như chưa nhận được sự quan tâm của các đồng nghiệp. Tuy nhiên, sự khác biệt khi đề cập tới vấn đề trên chính là trong khuôn khổ dự án CORC, WebDewey là CSDL quản trị DDC và LCSH, các ngôn ngữ tư liệu mang tính chất ngôn ngữ tiền kết hợp (pre-coordinated), còn trong các đề tài nghiên cứu mà chúng tôi đã thực hiện, CSDL quản trị từ khoá địa danh Việt Nam được tạo ra là công cụ để kiểm soát một loại thesaurus đặc biệt và là loại ngôn ngữ hậu kết hợp (post-coordinated). Vì lí do đó và vì các đơn vị từ vựng của ngôn ngữ này được xây dựng nên từ ngôn ngữ tự nhiên của người dùng tin, mà ý nghĩa và tác động của CSDL này không chỉ chủ yếu đối với người xử lý thông tin, mà còn đối với người tìm tin sau này, là công cụ quan trọng kiểm soát và bảo trì sự tương tác giữa các quá trình *xử lý thông tin* và *tìm tin*.

Từ hệ thống các CSDL thư mục được tạo lập và khai thác tại Viện Khoa học Xã hội Việt Nam, nhận thấy sự trùng lặp trong phần các CSDL thư mục quản lí tài liệu dạng sách là rất lớn. Điều này dẫn đến sự tổn hao về nguồn lực, làm giảm đáng kể hiệu quả xây dựng các nguồn tin dạng số. Viện Thông tin Khoa học Xã hội, với tư cách là cơ quan giúp Chủ tịch Viện Khoa học Xã hội Việt điều phối hoạt động thông tin thư viện KHXH cần sớm tìm ra các giải pháp khắc phục thực trạng trên. Xây dựng và đề xuất một quy hoạch cho phát triển hệ thống CSDL thư mục tại đây là một việc làm nhằm tránh được sự trùng lặp trong quá trình xây dựng các loại

¹ Nghiên cứu, xây dựng phương pháp sử dụng từ khoá địa danh Việt Nam trong các CSDL tại Viện Thông tin KHXH: Đề tài nghiên cứu cấp Viện.- H.: Viện TTKHXH. 2005. 98 tr. Một số kết quả nghiên cứu và khuyến nghị đã được công bố trên các diễn đàn như Tạp chí *Thông tin & Tư liệu*, *Thông tin Khoa học xã hội*...(Cùng tác giả).

CSDL tại đây, góp phần nâng cao hiệu quả đầu tư cho hoạt động thông tin-thư viện KHXH.

Cũng cần lưu ý là các nguồn tin cần được kiểm soát trong lĩnh vực KHXH là rất lớn, đa dạng và phân tán ở mức cao (được tạo ra vào mọi lúc, ở mọi nơi, bởi mọi chủ thể...). Do đó, hệ thống các CSDL phản ánh chúng – các CSDL thuộc nhóm a) – là rất lớn, công kênh, và gồm nhiều loại khác nhau, chứ không chỉ đơn giản là các CSDL thư mục phản ánh nguồn tài liệu dạng sách, tài liệu trong một số ít tạp chí khoa học như hiện có. Tại Viện Khoa học Xã hội Việt Nam hiện còn rất nhiều loại nguồn thông tin đặc thù khác chưa được kiểm soát bởi các CSDL tương ứng. Ví dụ, đó là các bộ phận khác nhau tạo nên nguồn tin khoa học nội sinh. Điều này đã được phản ánh trong một đề tài nghiên cứu vào năm 2006¹.

Mỗi loại cơ quan thông tin-thư viện, trên cơ sở chức năng, nhiệm vụ của mình mà có ưu thế trong việc quản lý và cung cấp các dịch vụ liên quan tới một loại nguồn thông tin xác định². Chính vì thế, việc tạo lập các CSDL ở những cơ quan thông tin-thư viện khác nhau cũng có những khác biệt nhau. Mục đích của sự khác biệt ấy chính là làm sao để việc khai thác nguồn thông tin tại mỗi cơ quan này đạt được ở mức cao nhất có thể. Và cũng chính vì thế mà khả năng chia sẻ nguồn tin giữa các cơ quan thông tin-thư viện được thực hiện một cách lâu dài, bền vững. Điều đó cho phép một số cơ quan khác nhau, trong quá trình liên kết và chia sẻ nguồn lực với nhau, sẽ giúp mỗi một có thể khai thác được đầy đủ nguồn tài nguyên đã được tạo lập từ các thành viên khác và ngược lại.

Phát triển hệ thống các CSDL thư mục phản ánh nguồn tin KHXH cần chú ý tới việc phát triển các loại CSDL thư mục phản ánh các chuyên đề đang thu hút sự quan tâm của giới nghiên cứu, ví dụ các chuyên đề *văn hoá và phát triển, toàn cầu hoá, nghiên cứu Việt Nam...* Về vấn đề này, có thể tham khảo thêm các nguyên lý xây dựng hệ thống chuyên đề trong quá trình triển khai *dịch vụ cung cấp thông tin theo chuyên đề*, một loại dịch vụ thông tin đang thu hút sự quan tâm của đông đảo giới dùng tin là các nhà khoa học. Ngoài ra, trong quá trình phát triển hệ thống

¹ *Thực trạng và các biện pháp nâng cao hiệu quả quản lý, khai thác nguồn tin khoa học nội sinh tại Viện Khoa học Xã hội Việt Nam*, Đề tài nghiên cứu cấp Bộ.- H.: Viện Khoa học Xã hội Việt Nam. 2006. 180tr.

² Xem Guichat, Menou General Introduction to Information Works.- Paris: UNESCO. 1988. 353 p.

CSDL thư mục chuyên đề, cần đặc biệt quan tâm tới sự liên kết, phối hợp chặt chẽ giữa Viện Thông tin KHXH và các cơ quan thông tin-thư viện chuyên ngành thuộc các tổ chức nghiên cứu, đào tạo về KHXH, trước hết là các Đại học đa ngành lớn của Quốc gia như Đại học Quốc gia Hà Nội, Tp. Hồ Chí Minh, các đại học vùng...

Để tiến tới xây dựng một quy hoạch phát triển các CSDL thư mục phản ánh nguồn thông tin tài liệu về KHXH của nước ta cần nhận diện đầy đủ sự phân bố của nguồn tin này theo nhiều khía cạnh khác nhau.

Vấn đề xây dựng và phát triển hệ thống các CSDL thư mục phản ánh nguồn tin KHXH trong nước cũng hết sức phức tạp, và cần phải được nghiên cứu, khảo sát từ nhiều khía cạnh khác nhau. Viện KHXH Việt Nam là tổ chức khoa học lớn nhất có chức năng tạo ra nguồn thông tin về KHXH của quốc gia, một bộ phận lớn và quan trọng trong nguồn thông tin khoa học *nội sinh* của quốc gia. Chính vì thế, quan điểm và kế hoạch phát triển các CSDL thư mục với tư cách công cụ thực hiện chức năng kiểm soát thư mục nguồn thông tin KHXH của quốc gia của Viện KHXH Việt Nam có một ý nghĩa và vị trí đặc biệt quan trọng trong quá trình xây dựng và phát triển ngân hàng thông tin khoa học của quốc gia.

Viện Thông tin Khoa học Xã hội cần nghiên cứu, xây dựng và đề xuất hướng phát triển các CSDL thư mục *bao quát các bộ phận khác nhau trong nguồn tin khoa học nội sinh* tại Viện Khoa học Xã hội Việt Nam. Ví dụ đó là CSDL thư mục phản ánh các luận án/luận văn khoa học, báo cáo kết quả nghiên cứu, nguồn học liệu được tạo lập và sử dụng trong hoạt động đào tạo, thuyết minh đề tài nghiên cứu, các dự án điều tra cơ bản, hệ thống tư liệu điền dã của các ngành KHXH, CSDL dữ kiện về các tổ chức, cá nhân hoạt động trong lĩnh vực KHXH, CSDL về các CSDL trong lĩnh vực KHXH (CSDL siêu dữ liệu)... Trước mắt, đối với CSDL thư mục quản lý tài liệu là sách và tạp chí trong lĩnh vực KHXH của nước ta, có thể được tạo nên từ các CSDL thư mục quốc gia do Thư viện Quốc gia Việt Nam xây dựng và phổ biến. Tạo lập và phát triển hệ thống CSDL theo hướng đó cũng có nghĩa là tiến hành các chuẩn bị cần thiết để thực hiện việc chia sẻ nguồn lực giữa các thành viên trong Hệ thống thông tin thư viện trên phạm vi quốc gia.

Nếu việc tạo ra nguồn tin KHXH được thực hiện bởi tất cả các tổ chức nghiên cứu và đào tạo của Việt Nam, thì rõ ràng quan điểm và kế hoạch tạo lập, phát triển hệ thống CSDL thư mục về KHXH cũng phải được xây dựng trên cơ sở thực tiễn này. Điều đó cho thấy, trong quá trình xây dựng và phát triển các CSDL

thư mục về KHXH ở nước ta, cần xác định rõ và hợp lý vai trò, vị trí của Viện Thông tin khoa học xã hội, các cơ quan thông tin thư viện chuyên ngành KHXH và các tổ chức nghiên cứu-đào tạo hữu quan khác của cả nước. Điều này cũng hoàn toàn có thể thực hiện được bởi các điều kiện và giải pháp về công nghệ hiện tại: Sự ra đời của Internet và các chuẩn trong xử lý, bao gói thông tin đã cho phép thông tin được phát triển và quản lý một cách phân tán và việc khai thác, truy cập thông tin được thực hiện một cách thuận lợi mà không bị lệ thuộc vào các yếu tố không gian và thời gian.

Tựu chung, quy hoạch xây dựng hệ thống CSDL thư mục tại Viện KHXH Việt Nam cần được xây dựng trên nguyên tắc:

- Tập trung vào việc xây dựng hệ thống CSDL thư mục phản ánh nguồn tin được tạo ra – chính là những thông tin phản ánh các loại hình kết quả hoạt động khoa học - tại Viện KHXH Việt Nam.

- Xây dựng một số quy định / chuẩn thống nhất cho việc trình bày và xử lý đối với mỗi loại nguồn thông tin được tạo ra tương ứng với mỗi loại kết quả hoạt động khoa học: nghiên cứu khoa học; quản lý khoa học; điều tra cơ bản, điền dã; đào tạo; hợp tác quốc tế; hội thảo, hội nghị khoa học...

- CSDL thư mục sẽ được tạo lập tại chính nơi mà nguồn tin mà chúng phản ánh được tạo ra.

Cần xây dựng một số CSDL có chức năng quản trị các công cụ xử lý thông tin đang được sử dụng phổ biến trong Hệ thống thông tin thư viện quốc gia – các CSDL tương ứng với nhóm c) của Dự án CORC. Ví dụ, đó là các CSDL quản trị Bảng phân loại DDC đã được dịch sang tiếng Việt (WebDewey cho phiên bản tiếng Việt), CSDL quản trị khung đề mục chủ đề do Thư viện Quốc hội Mỹ xây dựng và phổ biến (LCSH - với điều kiện khi hệ thống này được dịch sang tiếng Việt!) và các CSDL quản trị các loại ngôn ngữ tư liệu phổ biến khác trong các cơ quan thông tin thư viện KHXH: Hệ thống từ khoá có kiểm soát nói chung (thesaurus trong các ngành KHXH), các hệ thống từ khoá địa danh, từ khoá thời gian...

Nếu như, kế hoạch phát triển các CSDL tương ứng với nhóm a) trong Dự án CORC được thực hiện chủ yếu bởi các cơ quan thông tin thư viện chuyên ngành (nơi mà nguồn tin khoa học nội sinh được tạo ra), thì nhiệm vụ chủ yếu trong việc tạo lập và phát triển các CSDL tương ứng với nhóm c) trong Dự án CORC lại thuộc về Viện Thông tin Khoa học Xã hội – cơ quan có chức năng điều phối và đóng vai

trò cơ quan đầu ngành của Hệ thống thông tin thư viện KHXXH, một cơ quan có chức năng nghiên cứu và phát triển trong lĩnh vực thông tin-thư viện của nước ta.

Tính chất khó khăn và phức tạp trong việc thực hiện nhiệm vụ này là ở chỗ: Việc xây dựng các CSDL quản trị các loại ngôn ngữ tư liệu chưa hề được giải quyết trong thực tiễn công tác thông tin-thư viện nước ta; hơn nữa, còn là vì bản thân tính chất phức tạp của các *bài toán cần giải quyết*. Ví dụ CSDL quản trị từ khoá địa danh Việt Nam cần xử lý được các *tình huống* rất phổ biến là sự thay đổi địa danh hành chính (Việt Nam) qua thời gian, hoặc quan hệ rất phức tạp giữa *địa danh hành chính* và *địa danh phi hành chính*... thông qua việc xác lập quan hệ giữa các loại dữ liệu có liên quan một cách phù hợp.... Hơn nữa, để có thể thực sự là một công cụ hữu hiệu và tiện lợi cho người dùng, chúng lại cần phải được tích hợp vào hệ thống các loại CSDL khác, trước hết là tương ứng với nhóm a) trong dự án CORC. Chỉ đến khi đó, chúng mới trở thành các công cụ trợ giúp người dùng theo đúng ý nghĩa, mục đích và đòi hỏi của điều đó.

Trước mắt, Viện Thông tin Khoa học Xã hội cần sớm nghiên cứu, thiết kế và xây dựng CSDL quản trị từ khoá địa danh Việt Nam, và xem đó là một công cụ quan trọng trợ giúp cho các quá trình *xử lý thông tin* (định từ khoá địa lí) cũng như *tìm tin theo dấu hiệu địa danh Việt Nam*.

Các công việc và đòi hỏi đối với vấn đề quy hoạch và phát triển hệ thống CSDL như đã nêu trên đây là rất phức tạp, ngay khi chúng mới chỉ dừng ở các nghiên cứu thử nghiệm. Nhưng nếu không triển khai, nếu không xây dựng và duy trì được sự phối hợp bền vững giữa các cơ quan khá nhau, nếu không tận dụng công sức, trí tuệ hợp lý cả ở trong và ngoài nước, thì không thể tạo nên được các đổi mới và các bước phát triển về chất trong hoạt động thông tin-thư viện trong lĩnh vực KHXXH nói riêng và khoa học và công nghệ nói chung ở nước ta. Đây vừa là nhiệm vụ của những chuyên gia nghiên cứu cũng như những người làm công tác tại các cơ quan thông tin thư viện, đồng thời cũng là nhiệm vụ của những người làm công tác *quản lí nhà nước* về hoạt động thông tin-thư viện, về ứng dụng công nghệ thông tin và truyền thông nói chung. Bởi một điều đơn giản: Để có thể giải quyết các vấn đề phức tạp như trên không chỉ đòi hỏi sự hợp sức và liên kết giữa nghiên cứu và triển khai, mà còn cần một chính sách đầu tư hợp lí, đồng bộ và ổn định - *nhiệm vụ của quản lí Nhà nước*.

Thay phần kết luận

Có thể dễ nhận thấy, trong phạm vi Hệ thống thông tin thư viện quốc gia, cũng như trong phạm vi hoạt động nghiên cứu và đào tạo, các vấn đề như đã nêu không thể và không chỉ được xem là nhiệm vụ của bất kỳ cơ quan thông tin thư viện khoa học nào, mà cần được xem là nhiệm vụ lâu dài và chung của mọi tổ chức thông tin thư viện của quốc gia. Hệ thống các cơ quan thông tin thư viện KHXH cũng như hệ thống các thư viện đại học ở nước ta cần đặc biệt quan tâm tới vấn đề này bởi nhiều lí do khác nhau. Trước hết là bởi vì sự liên kết về mặt nhận thức và các hoạt động khoa học cụ thể giữa các hệ thống này với nhau, giữa các thành viên của mỗi hệ thống nêu trên đều chưa trở thành một nền nếp, và do đó, hiệu quả của điều đó còn rất hạn chế. Trong khi đó, sự liên kết giữa nghiên cứu và đào tạo giữa các trường đại học, Viện nghiên cứu đang trở nên rất chặt chẽ - điều cho thấy nhu cầu kết nối giữa các chủ thể trên từ phía người dùng tin là rất rõ rệt. Có nhiều hoạt động phản ánh sự liên kết phối hợp giữa các thư viện đại học với các thành viên của Hệ thống thông tin thư viện KHXH, trong đó, phối hợp cùng nhau phát triển các nguồn thông tin dạng số, phát triển hệ thống các loại CSDL thành phần theo mô hình của dự án CORC là các nội có tác động và ý nghĩa rất đặc sắc. Hy vọng rằng điều đó sẽ sớm trở thành hiện thực.

TÀI LIỆU THAM KHẢO

1. Durbin R., Nalen J., Pitre N . E-Book collection development and management: The quandary of establishing policies and guidelines for academic library collections .// *Advances in Library Administration and Organization*, Volume 19, 2002, pp. 59-84.
2. Hanrong Wang, Guiling Niu and William J. Hubbard. Current status of reference services in academic libraries in mainland China: A web analysis.// *The International Information & Library Review*, Volume 36, Issue 2, June 2004, pp. 105-110.
3. Hoàng Lê Minh. Dự án *Hệ thống thông tin thư viện điện tử liên kết các trường đại học* và việc tăng cường tiềm lực khoa học-công nghệ cho phát triển kinh tế-xã hội tại Thành phố Hồ Chí Minh. 2003.
4. Introduction to CORC, <http://staff.oclc.org/oclc/research/projects/core/>
5. King D.W. Some Thoughts on Academic Library Collections.// *The Journal of Academic Librarianship*, Volume 30, Issue 4, July 2004, pp. 261-264.
6. McDonald A.. Planning academic library buildings for a new age: Some principles, trends, and developments in the United Kingdom .// *Advances in Librarianship*, Volume 24, 2000, pp. 51-79.