

GIÁM SÁT CHẤT LƯỢNG DỮ LIỆU CỦA CƠ SỞ DỮ LIỆU KHOA HỌC VÀ CÔNG NGHỆ QUỐC TẾ

ThS Phan Huy Quế

Cục Thông tin KH&CN quốc gia

Tóm tắt: Phân tích các tiêu chí đánh giá chất lượng nội dung CSDL KH&CN quốc tế. Giới thiệu quy trình giám sát chất lượng dữ liệu của CSDL KH&CN quốc tế qua các công đoạn vận hành dữ liệu: thu thập, chọn lọc dữ liệu đầu vào; Xử lý, cập nhật, lưu trữ dữ liệu; Khai thác, sử dụng dữ liệu đầu ra. Đề xuất các giải pháp hỗ trợ đánh giá chất lượng nội dung CSDL KH&CN quốc tế.

Từ khóa: Chất lượng dữ liệu; chất lượng thông tin; nội dung thông tin; CSDL KH&CN quốc tế.

Monitoring data quality of the International S&T Database

Summary: Analyzes criteria for quality content evaluation of the international S&T database; presents process of monitoring data quality of the International S&T Database through stages of data operation: processing, updating and storing data, exploiting and utilizing output data; sets forth some solutions to supporting qualitative evaluation of international S&T database content.

Keywords: Data quality; information quality; information content; International S&T Database.

1. Đặt vấn đề

Một trong những nhiệm vụ chủ yếu của Đề án hội nhập quốc tế về khoa học và công nghệ (KH&CN) đến năm 2020 là tăng cường nguồn lực thông tin phục vụ hội nhập quốc tế về KH&CN nhằm khai thác kinh nghiệm và nguồn lực từ bên ngoài để phát triển năng lực nội sinh về KH&CN, tạo tiền đề cho sự phát triển bền vững trong thế kỷ XXI, đồng thời, từng bước phát triển năng lực hội nhập của quốc gia đối với những nội dung chủ yếu của hoạt động KH&CN quốc tế [1]. Để thực hiện nhiệm vụ nói trên, đề tài “Nghiên cứu xây dựng khung và hệ tiêu chí phát triển CSDL KH&CN quốc tế cho hoạt động hội nhập quốc tế về KH&CN của Việt Nam”, mã số KX06.02/11-15, do Cục Thông tin KH&CN quốc gia chủ trì thuộc Chương trình KH&CN trong điểm cấp nhà nước giai đoạn 2011-2015 “Nghiên cứu và phát triển hội nhập quốc tế về KH&CN”, đã xác định mục tiêu tổng thể là xây dựng bộ khung và các tiêu chí để phát triển CSDL KH&CN quốc tế phục vụ cho hoạt động hội nhập quốc tế về KH&CN của Việt Nam. Đây là một hệ thống thông tin KH&CN tích hợp gồm các khối dữ liệu về bốn đối tượng chủ yếu sau đây:

- Tổ chức KH&CN quốc tế;
- Chuyên gia KH&CN quốc tế;
- Chương trình, dự án hợp tác quốc tế và sự kiện KH&CN quốc tế;
- Công nghệ mũi nhọn (công nghệ chủ chốt).

Trong quá trình xây dựng, khai thác và phát triển CSDL KH&CN quốc tế, ngoài việc bảo đảm quá trình vận hành thông suốt của các khối dữ liệu nói trên, còn một việc không kém phần quan trọng là bảo đảm chất lượng dữ liệu thông qua hoạt động giám sát chất lượng dữ liệu. Cơ sở để giám sát chất lượng dữ liệu là các tiêu chí đánh giá chất lượng dữ liệu CSDL KH&CN quốc tế, được xây dựng dựa trên những nguyên tắc cơ bản của hoạt động thông tin KH&CN được quy định tại Khoản 1 và 2 Điều 4 Nghị định số 11/2014/NĐ-CP ngày 18 tháng 02 năm 2014 của Chính phủ về hoạt động thông tin KH&CN, là: chính xác, khách quan, đầy đủ, kịp thời và đáp ứng yêu cầu của tổ chức, cá nhân về chất lượng, khối lượng sản phẩm, dịch vụ thông tin KH&CN; hiệu quả trong việc khai thác, sử dụng và quản lý thông tin KH&CN phục vụ quản lý nhà nước, phát triển KT-XH, bảo đảm an ninh, quốc phòng và hội nhập quốc tế của đất nước [2]. Các nguyên

Nghiên cứu - Trao đổi

tác trên đều nhắm tới mục đích chủ yếu là bảo đảm chất lượng thông tin, phát huy cao nhất hiệu quả phục vụ thông tin KH&CN cho sự phát triển KT-XH của đất nước.

2. Tiêu chí đánh giá chất lượng dữ liệu CSDL KH&CN quốc tế

Đối với CSDL KH&CN quốc tế, chất lượng dữ liệu được phản ánh thông qua các tiêu chí chủ yếu sau đây:

a) *Tiêu chí chính xác*, gồm các nội dung sau:

- Tính đúng đắn, chính xác của thông tin miêu tả đối tượng so với nguồn dữ liệu gốc;
- Tính toàn vẹn và khách quan của thông tin miêu tả đối tượng, nghĩa là không có bất kỳ sự “áp đặt” hoặc “bóp méo thông tin” nào từ phía tổ chức, cá nhân quản trị CSDL. Trong trường hợp này, tiêu chí chính xác bảo đảm tính khách quan của thông tin trong CSDL.

Ngoài các nội dung trên, tiêu chí chính xác còn là việc thể hiện đúng đắn và trung thực nguồn thông tin được trích dẫn, không chỉ tạo sự tin cậy mà còn tạo điều kiện cho người sử dụng truy cập, khai thác nguồn dữ liệu gốc khi có nhu cầu.

b) *Tiêu chí đầy đủ*, gồm các nội dung sau:

- Độ đầy đủ của các trường dữ liệu đã xác định đối với mỗi đối tượng của các CSDL thành phần thuộc CSDL KH&CN quốc tế;
- Độ đầy đủ của các thông tin đã xác định đối với mỗi trường dữ liệu. Ngoài ra, độ đầy đủ của dữ liệu còn được đánh giá qua khả năng bao quát về *thời gian* và *không gian* của thông tin trong các trường dữ liệu.

c) *Tiêu chí cập nhật*, gồm các nội dung:

- Hoạt động cập nhật định kỳ hoặc thường xuyên thông tin về các đối tượng mới. Nói cách khác là tăng cường số lượng biểu ghi của CSDL;
- Cập nhật thay đổi của thông tin (nếu có) trong các trường dữ liệu về đối tượng của CSDL.

Tiêu chí cập nhật thể hiện tính mới và tính chất phát triển của CSDL.

d) *Tiêu chí hệ thống*, gồm các nội dung:

- Tính bao quát về thời gian và không gian của dữ liệu về đối tượng của CSDL;
- Mối liên kết chặt chẽ, lô-gíc về thông tin giữa trường dữ liệu về đối tượng trong cùng CSDL

cũng như với đối tượng của các CSDL thành phần khác. Ví dụ: từ thông tin về một chuyên gia KH&CN bất kỳ trong CSDL chuyên gia, có thể truy cập đến trường dữ liệu nơi chuyên gia làm việc (*CSDL tổ chức*), chương trình, dự án quốc tế chuyên gia hiện tham gia (*CSDL chương trình, dự án*), công nghệ mà chuyên gia là tác giả sáng chế (*CSDL công nghệ chủ chốt*) và ngược lại. Tiêu chí hệ thống giúp có cái nhìn toàn diện và sâu sắc hơn về đối tượng của CSDL. Để có được một cách hệ thống những thông tin về đối tượng, người quản trị CSDL phải có đầy đủ thông tin về đối tượng, thông tin phải được xử lý chính xác và phải được thường xuyên cập nhật. Tiêu chí hệ thống phản ánh chất lượng của dữ liệu qua việc tích hợp và cung cấp “bức tranh dữ liệu toàn cảnh” về đối tượng của CSDL. Như vậy, có thể nói, tiêu chí hệ thống là hệ quả của các *tiêu chí đầy đủ, tiêu chí chính xác* và *tiêu chí cập nhật*.

e) *Tiêu chí tiện dụng*, gồm các nội dung sau:

- Hình thức và nội dung thông tin được thể hiện đơn giản, hợp lý, bảo đảm thuận lợi cho quá trình khai thác, sử dụng của người dùng cũng như quá trình chia sẻ, trao đổi dữ liệu với các hệ thống khác;
- Các sản phẩm dịch vụ thông tin của CSDL được bao gói, trình bày phù hợp với sở thích của từng nhóm người dùng tin và được cung cấp theo phương thức phù hợp với yêu cầu của người dùng tin.

Ngoài ra, tiêu chí tiện dụng còn có những nội dung như: thuận tiện, nhanh chóng trong truy cập dữ liệu; đơn giản, tin cậy trong lưu trữ dữ liệu; an toàn dữ liệu...

3. Quy trình giám sát chất lượng dữ liệu CSDL KH&CN quốc tế

Căn cứ để giám sát chất lượng dữ liệu CSDL là các tiêu chí đánh giá chất lượng dữ liệu, được áp dụng đối với ba công đoạn chính của chu trình vận hành dữ liệu là:

- Thu thập, chọn lọc dữ liệu đầu vào;
- Xử lý, cập nhật, lưu trữ dữ liệu;
- Khai thác, sử dụng dữ liệu đầu ra.

Quy trình giám sát chất lượng dữ liệu CSDL KH&CN quốc tế được trình bày trong bảng dưới đây.

Nghiên cứu - Trao đổi

Bảng 1. Quy trình giám sát chất lượng dữ liệu CSDL KH&CN quốc tế

TT	Tiêu chí giám sát	Công đoạn/Nội dung giám sát		Phương pháp giám sát	Phương tiện hỗ trợ giám sát
		Thu thập, chọn lọc dữ liệu đầu vào	Xử lý, cập nhật, lưu trữ dữ liệu		
1.	Tiêu chí chính xác	+ <u>Giám sát xuất xứ của dữ liệu đầu vào:</u> - Xác định dữ liệu trong hay ngoài danh mục nguồn tin đã xác định của CSDL. - Trường hợp dữ liệu ngoài danh mục, cần làm rõ từ nguồn nào, độ tin cậy của nguồn tin, độ tin cậy của dữ liệu. + <u>Giám sát độ phù hợp của hình thức và nội dung dữ liệu đầu vào:</u> - Mức độ phù hợp của tiêu chuẩn định dạng dữ liệu với tiêu chuẩn quy định của CSDL; - Mức độ phù hợp của dữ liệu với các tiêu chí lựa chọn đối tượng của các CSDL thành phần thuộc CSDL KH&CN quốc tế.	+ <u>Giám sát cách thức tổ chức xử lý, cập nhật, lưu trữ dữ liệu:</u> - Mức độ hợp lý, khoa học của quy trình xử lý, cập nhật và lưu trữ dữ liệu; - Mức độ đúng đắn, chính xác việc thực hiện kỹ thuật xử lý, cập nhật và lưu trữ dữ liệu; - Mức độ tuân thủ và hiệu quả sử dụng các tiêu chuẩn, quy chuẩn kỹ thuật trong xử lý, cập nhật và lưu trữ dữ liệu; + <u>Giám sát tính chính xác, toàn vẹn và trung thực của dữ liệu:</u> - Mức độ đúng đắn, chính xác của thông tin miêu tả đối tượng so với nguồn dữ liệu gốc; - Tính toàn vẹn và khách quan của thông tin miêu tả đối tượng (không có áp đặt hoặc bóp méo	- Đối chiếu, so sánh. - Thao tác trên CSDL. - Xử lý thông tin phản hồi từ người dùng - Điều tra, khảo sát.	- Danh mục và địa chỉ website nguồn các CSDL thành phần của CSDL KH&CN quốc tế. - Danh mục và địa chỉ website các CSDL, Nhà xuất bản, Hệ thống thông tin trong nước và quốc tế. - Quy chế CSDL KH&CN quốc tế. - Bộ tiêu chí dữ liệu của các CSDL thành phần thuộc CSDL KH&CN quốc tế. - Tài liệu hướng dẫn về CSDL KH&CN quốc tế (nếu có). - Tiêu chuẩn, quy chuẩn kỹ thuật.

Nghiên cứu - Trao đổi

		thông tin từ phía người xử lý dữ liệu; - Thẻ hiện đúng đắn và trung thực nguồn thông tin được trích dẫn; - Tính chính xác và toàn vẹn của thông tin trong quá trình cập nhật; - Khả năng bảo đảm an toàn trong lưu trữ dữ liệu gốc.	bao gói; phương thức phục vụ...) - Khả năng liên kết chính xác của các trường dữ liệu (nếu có).		- Mỗi liên kết ngược (feedback) với người dùng tin. - Phiếu điều tra, khảo sát.
		+ <u>Giám sát mức độ đầy đủ của dữ liệu:</u> - Mức độ đầy đủ của các trường dữ liệu đã xác định đối với đối tượng của CSDL; - Mức độ đầy đủ của thông tin đã xác định đối với mỗi trường dữ liệu;	+ <u>Giám sát mức độ đầy đủ của thông tin trong các SP/DVT:</u> - Tỷ lệ thông tin phù hợp tìm thấy/tổng số thông tin phù hợp có trong CSDL đối với mỗi yêu cầu tìm tin; - Định lượng và định tính thông tin phù hợp của mỗi trường dữ liệu về đối tượng trong kết quả tìm tin. - Mức độ đầy đủ của các mối liên kết dữ liệu.	- Đối chiếu, so sánh. - Thao tác trên CSDL. - Xử lý thông tin phản hồi từ người dùng tin. - Điều tra, khảo sát.	- Tiêu chuẩn, quy chuẩn kỹ thuật. - Quy chế CSDL KH&CN quốc tế. - Bộ tiêu chí dữ liệu của các CSDL thành phần thuộc CSDL KH&CN quốc tế. - Tài liệu hướng dẫn về CSDL KH&CN quốc tế (nếu có). - Giáo trình, tài liệu hướng dẫn về xử lý, cập nhật và lưu trữ dữ liệu.
		+ <u>Giám sát mức độ đầy đủ của dữ liệu vào:</u> - Mức độ đầy đủ của dữ liệu so với yêu cầu đặt ra đối với nguồn dữ liệu đầu vào theo các khía cạnh: độ phủ lĩnh vực; thời gian; vị trí địa lý; quốc gia; v.v...; - Mức độ đầy đủ của tiêu chí dữ liệu đối với các đối tượng quản trị của CSDL.			
2.	Tiêu chí đầy đủ				

Nghiên cứu - Trao đổi

					- Mọi liên kết ngược (feedback) với người dùng tin. - Phiếu điều tra, khảo sát.
				- Tiêu chuẩn, quy chuẩn kỹ thuật. - Kết quả khảo sát nhu cầu thông tin của CSDL KH&CN quốc tế (nếu có). - Mọi liên kết ngược (feedback) với người dùng tin.	- Tài liệu hướng dẫn về CSDL KH&CN quốc tế (nếu có). - Mọi liên kết ngược (feedback) với người dùng tin.
			- Đối chiếu, so sánh. - Thao tác trên CSDL. - Xử lý thông tin phản hồi từ người dùng tin.	- Thao tác trên CSDL. - Xử lý thông tin phản hồi từ người dùng tin.	
			+ <u>Giám sát tính cập nhật của thông tin trong các SPDVTT:</u> - Mức độ phát triển về định lượng thông tin trong các SPDVTT; - Tỷ lệ thông tin mới/tổng số thông tin phù hợp tìm thấy đối với mỗi yêu cầu tin trong từng giai đoạn.	+ <u>Giám sát khả năng phân tích-tổng hợp thông tin trong các SPDVTT:</u> - Tính hệ thống và lô-gíc của thông tin phân tích-tổng hợp trong các SPDVTT; - Mức độ bao quát về không gian và thời gian của thông tin phân tích-tổng hợp trong các SPDVTT.	
			+ <u>Giám sát hoạt động cập nhật dữ liệu vào CSDL:</u> - Tính kịp thời và ổn định của hoạt động cập nhật dữ liệu mới vào CSDL; - Tính kịp thời và ổn định của hoạt động bổ sung, chỉnh sửa dữ liệu (nếu có) vào CSDL.	+ <u>Giám sát liên kết dữ liệu trong CSDL:</u> - Chất lượng tổ chức các mối liên kết dữ liệu trong cùng CSDL. - Chất lượng tổ chức các mối liên kết dữ liệu giữa các CSDL thành phần của CSDL KH&CN quốc tế.	
			+ <u>Giám sát tính mới của dữ liệu đầu vào:</u> - Tính cập nhật của nguồn sản sinh dữ liệu (xuất xứ của dữ liệu); - Tính ổn định/biến động của dữ liệu - Thời điểm hình thành dữ liệu.	-	
3.	Tiêu chí cấp nhật				
4.	Tiêu chí hệ thống				

Nghiên cứu - Trao đổi

5.	Tiêu chí tiện dụng	-	-	+ <u>Giám sát mức độ thuận lợi trong tìm kiếm và khai thác dữ liệu:</u> - Khả năng tìm kiếm dữ liệu theo các tiêu thức khác nhau đối với từng CSDL thành phần và CSDL tích hợp. - Mức độ thân thiện trong khai thác và sử dụng dữ liệu (download dữ liệu, định dạng và lưu trữ dữ liệu tại máy tính người dùng...). + <u>Giám sát mức độ thuận lợi trong chia sẻ, trao đổi dữ liệu:</u> - Khả năng và mức độ thuận tiện trong trao đổi, chia sẻ thông tin on-line và off-line của CSDL với các hệ thống khác trong phạm vi quốc gia. - Mức độ thân thiện trong trao đổi, chia sẻ thông tin on-line và off-line của CSDL với các hệ thống khác trong phạm vi quốc tế.	- Thao tác trên CSDL - Xử lý thông tin phản hồi từ người dùng tin. - Điều tra, khảo sát.	- Quy chế CSDL KH&CN quốc tế. - Tiêu chuẩn, quy chuẩn kỹ thuật. - Mối liên kết ngược (feedback) với người dùng tin. - Phiếu điều tra, khảo sát.
----	-----------------------	---	---	--	--	--

Nghiên cứu - Trao đổi

4. Kết luận

Giám sát chất lượng dữ liệu là một trong những hoạt động quan trọng và cần thiết để nâng cao chất lượng CSDL KH&CN quốc tế. Tuy nhiên, cần có những giải pháp hỗ trợ để triển khai thực hiện. Cụ thể như sau:

- Chọn lựa cơ quan nhà nước đủ tiêu chuẩn để giao trách nhiệm xây dựng và quản trị CSDL KH&CN quốc tế. Tiêu chuẩn cơ bản đối với cơ quan này là: Kinh nghiệm và năng lực xây dựng CSDL KH&CN; Kinh nghiệm về hoạt động thông tin KH&CN (do nội dung chủ yếu của hoạt động xây dựng và phát triển CSDL KH&CN quốc tế là thu thập, xử lý, lưu trữ và phục vụ thông tin KH&CN); Hạ tầng mạng thông tin hiện đại, tiên tiến; Đội ngũ nhân lực đủ sức xây dựng và quản trị CSDL KH&CN quốc tế, đặc biệt là đội ngũ nhân lực chất lượng cao về nghiệp vụ thông tin KH&CN;

- Xây dựng cơ chế về tạo nguồn dữ liệu đầu vào của CSDL KH&CN quốc tế, bao gồm

nguồn dữ liệu trong nước và nguồn dữ liệu từ nước ngoài. Trong điều kiện hiện nay, nguồn dữ liệu từ nước ngoài giữ vai trò chủ đạo trong cơ cấu dữ liệu đầu vào của CSDL. Cần tập trung đầu tư cho việc khai thác, chọn lọc, bảo đảm chất lượng nguồn dữ liệu đầu vào quan trọng này;

- Chú trọng hoạt động xây dựng và áp dụng tiêu chuẩn, quy chuẩn kỹ thuật, ưu tiên áp dụng các tiêu chuẩn hiện đại, tiên tiến trong xây dựng, duy trì và phát triển CSDL KH&CN quốc tế;

- Tăng cường mối quan hệ với người dùng tin của CSDL để thường xuyên tiếp thu ý kiến phản hồi về chất lượng SPDVT của CSDL nhằm cải tiến và nâng cao chất lượng phục vụ thông tin của CSDL;

- Tổ chức thực hiện hoạt động giám sát chất lượng dữ liệu KH&CN quốc tế với các phương tiện hỗ trợ cần thiết và đội ngũ nhân lực đáp ứng yêu cầu thực hiện hoạt động giám sát.

Tài liệu tham khảo

1. Quyết định số 735/QĐ-TTg ngày 18/5/2011 của Thủ tướng Chính phủ phê duyệt Đề án hội nhập quốc tế về KH&CN đến năm 2020.

2. Nghị định số 11/2014-NĐ-CP ngày 18/02/2014 của Chính phủ về hoạt động thông tin KH&CN.

3. Nguyễn Hữu Hùng (2005). Thông tin: từ lý luận tới thực tiễn.- H.: Văn hoá Thông tin, 835 tr.

4. Thông tin học-Cách tiếp cận dưới góc độ điều khiển học/Lê Trọng Hiến. - TC Thông tin và Tư liệu, 2010, số 1.- tr.11-16.

5. Nghiên cứu cơ sở lý luận và thực tiễn của việc hiện đại hóa Hệ thống thông tin KH&CN quốc gia Việt Nam trong bối cảnh hội nhập quốc tế.-Đề tài cấp bộ.-Chủ nhiệm đề tài: Nguyễn Hữu Hùng.- Cơ quan chủ trì: Trung tâm Thông tin Tư liệu KHCN Quốc gia.- Năm hoàn thành BC: 2005.

6. Nghiên cứu cơ sở khoa học và thực tiễn phát triển mạng thông tin tư liệu về khoa học và công nghệ ở Việt Nam. Đề tài cấp bộ.- Chủ nhiệm đề tài: Tạ Bá Hưng;Cơ quan chủ trì: Trung tâm Thông tin Tư liệu KHCN Quốc gia; Năm hoàn thành BC: .-2001.

7. Nghiên cứu áp dụng các chuẩn lưu trữ và trao đổi thông tin trong hệ thống thông tin khoa học và công nghệ quốc gia.-Đề tài cấp bộ.-Chủ nhiệm đề tài: Phan Huy Quế.- Cơ quan chủ trì: Trung tâm Thông tin KHCNQG.-Năm hoàn thành BC: 2003.

8. Bản thuyết minh đề tài nghiên cứu Khoa học xã hội và Nhân văn cấp nhà nước: "Nghiên cứu xây dựng khung và hệ tiêu chí phát triển cơ sở dữ liệu khoa học và công nghệ quốc tế cho hoạt động hội nhập quốc tế về khoa học và công nghệ"/CNĐT: Lê Xuân Định.-Cơ quan chủ trì: Cục Thông tin KH&CN quốc gia.

(Ngày Tòa soạn nhận được bài: 04-7-2014; Ngày phản biện đánh giá: 15-9-2014; Ngày chấp nhận đăng: 15-10-2014).