

CHUYỂN ĐỔI SỐ NGÀNH THÔNG TIN KHOA HỌC VÀ CÔNG NGHỆ ĐỂ HỘI NHẬP QUỐC TẾ

Tạ Tuấn Anh

Công ty cổ phần Phát triển nguồn mở và Dịch vụ FDS

Tóm tắt: Bài viết tổng quát thực trạng phát triển hệ thống thông tin KH&CN tại Việt Nam, chiến lược và nền tảng yêu cầu để quản lý thông tin KH&CN quốc gia và từ đó đưa ra các giải pháp chuyển đổi số để hội nhập quốc tế.

Từ khóa: Hệ thống thông tin; khoa học và công nghệ; chuyển đổi số; cơ sở dữ liệu

Digital transformation in science and technology information for international integration

Abstract: The article summarizes the development of science and technology information system in Vietnam, the strategy and platform required to manage national S&T information and then providing digital transformation solutions for international integration.

Keywords: Information system; science and technology; digital transformation; database

Thông tin là động lực thúc đẩy phát triển KH&CN ở tất cả các quốc gia. Trong kỷ nguyên chuyển đổi số, phương thức quản lý thông tin KH&CN sẽ thay đổi theo hướng mở và hội nhập chung vào mạng lưới toàn cầu dựa trên nền tảng của internet.

1. Thực trạng phát triển thông tin khoa học và công nghệ

Thông tin KH&CN có ý nghĩa rất lớn trong sự phát triển nền kinh tế, xã hội của đất nước. Có nhiều loại hình ứng dụng khác nhau của thông tin KH&CN dành cho các đối tượng sử dụng khác nhau gồm nhà quản lý, nhà khoa học, người dân, doanh nghiệp và các tổ chức. Thông tin KH&CN thường được xây dựng dưới hình thức của các cơ sở dữ liệu cho phép người dùng tra cứu, tìm kiếm thông tin tư liệu phục vụ nghiên cứu và phát triển, như: thư viện điện tử; atlas điện tử; tạp chí điện tử; thông tin sở hữu công nghiệp; tiêu chuẩn kỹ thuật; kho dữ liệu mở.

Việc phát triển các hệ thống thông tin KH&CN đang rất được quan tâm trên thế giới. Bên cạnh các thư viện truyền thống, các loại hình ứng dụng khác được dùng phổ biến cho thông tin KH&CN có thể kể ra gồm: i) các CSDL xuất bản điện tử, trong đó bao gồm các CSDL miễn phí như các tạp chí truy cập mở hoặc CSDL có thu phí

như Proquest, ScienceDirect, Springerlink, IEEE,...; ii) các CSDL chỉ mục và thông tin trích dẫn như ISI, Scopus, PubMed, Google Scholar, OpenCitation,...; iii) các kho lưu trữ truy cập mở như Datacite, arXiv, OpenAIRE,...; vi) các hệ thống quản lý đăng ký định danh cho tài liệu xuất bản và các cán bộ nghiên cứu, như: Crossref (DOI), ORCID, ISNI,...; vi) các hệ thống thông tin nghiên cứu chuyên ngành của các tổ chức như PubChem (hóa học), GlobalChange (biến đổi khí hậu), InspireHEP (vật lý năng lượng cao),...

Tại Việt Nam, các hệ thống thông tin KH&CN được phát triển theo hệ thống phân cấp quản lý của nhà nước, gồm có: i) hệ thống thông tin ở cấp độ quốc gia do Cục Thông tin KH&CN quốc gia thuộc Bộ KH&CN quản lý; ii) hệ thống thông tin do các bộ, ngành quản lý theo lĩnh vực; iii) hệ thống thông tin do các tỉnh thành quản lý theo địa bàn hành chính; iv) hệ thống thông tin phục vụ nghiên cứu và phát triển của các tổ chức tại cơ sở (viện, trường). Những ứng dụng được khai thác nhiều nhất hiện nay ở trong nước là thư viện số của các viện, trường và CSDL quản lý thông tin đề tài, nhiệm vụ nghiên cứu ở các cấp. Nhiều tạp chí khoa học ở trong nước cũng đang chuyển dịch để thực hiện xuất bản điện tử

và cung cấp truy cập mở cho người dùng. Chúng ta cũng đã bước đầu xây dựng được một CSDL thông tin trích dẫn cho các tạp chí xuất bản trong nước (V-CitationGate do Đại học Quốc gia Hà Nội phát triển).

Sự tồn tại của nhiều hệ thống thông tin KH&CN khác nhau ở các cấp mang lại sự đa dạng của nguồn tin nhưng lại dẫn đến sự chồng chéo, thiếu tính liên kết và không đầy đủ thông tin trong các CSDL. Nghiên cứu phát triển nguồn tin KH&CN trong thời đại công nghệ số [1] đã đưa ra một phân tích chi tiết về thực trạng phát triển các nguồn tin KH&CN của nước ta hiện nay. Các hạn chế chính trong phát triển các hệ thống thông tin KH&CN tại Việt Nam có thể tổng kết như sau:

- Một là, có nhiều CSDL khác nhau được phát triển nhưng chủ yếu chỉ dùng cho các nhu cầu phục vụ cục bộ, thiếu sự liên kết, chia sẻ thông tin giữa các CSDL. Mới chỉ bước đầu hình thành được mạng liên kết trên phạm vi quy mô hẹp (trong địa phương, theo lĩnh vực) và vẫn còn manh mún.

- Hai là, chưa phát triển được các hệ thống thông tin KH&CN ở quy mô rộng, chứa đầy đủ thông tin trong các lĩnh vực quản lý; Chưa khai thác được thế mạnh của các nền tảng công nghệ đám mây cho phép tích hợp, tập hợp thông tin từ nhiều CSDL.

- Ba là, thiếu nền tảng dùng chung, nhất quán trong phạm vi cả nước để tạo cơ sở hình thành một mạng lưới liên kết các hệ thống thông tin KH&CN ở quy mô quốc gia. Hiện có một số trường đại học trong nước đang cùng khai thác một nền tảng công nghệ chung của nhà cung cấp thương mại (ví dụ, tailieu.vn) để xây dựng hệ thống thư viện số. Tuy nhiên, nền tảng này chưa đáp ứng được đầy đủ các nhu cầu về quản lý thông tin KH&CN trong thực tế.

2. Chiến lược chuyển đổi số thông tin khoa học và công nghệ

Chuyển đổi số là một yêu cầu cấp thiết trong bối cảnh hội nhập quốc tế để phát triển ngành thông tin KH&CN. Cách tiếp cận trong chuyển đổi số là xây dựng hệ thống thông tin KH&CN quốc gia như là một nền tảng thay cho cách tiếp cận hướng tới các ứng dụng như hiện nay.

Nền tảng thông tin KH&CN không nên được hiểu theo nghĩa hẹp là ứng dụng CNTT vào quản lý thông tin KH&CN. Theo nghĩa rộng, nền tảng là sự đổi mới toàn diện cơ chế hoạt động của thông tin KH&CN dựa trên dữ liệu số trong kỷ nguyên công nghiệp 4.0. Căn cứ vào chính sách KH&CN quốc gia, nền tảng số bao gồm một chiến lược chuyển đổi số trên cơ sở khai thác các nền tảng CNTT tiên tiến để quản lý thông tin KH&CN.

Chiến lược chuyển đổi số được coi như bản kế hoạch tổng thể (dài hạn) hướng tới sự đổi mới toàn diện hệ thống vận hành về thông tin KH&CN của quốc gia để đạt các mục tiêu chính như sau:

i) Hình thành mạng lưới liên kết dữ liệu về thông tin KH&CN trên toàn quốc và hội nhập với quốc tế;

ii) Khuyến khích sự tham gia của các tổ chức phi chính phủ, doanh nghiệp tư nhân cùng đóng góp phát triển mạng lưới thông tin KH&CN;

iii) Quy hoạch phân luồng phát triển các CSDL thông tin KH&CN theo cả chiều rộng và chiều sâu: các CSDL theo hướng lĩnh vực hẹp, cung cấp dữ liệu chuyên sâu phục vụ cho công tác nghiên cứu của các nhà khoa học, còn các CSDL theo chiều rộng chứa thông tin toàn diện, đa ngành, có tính tổng hợp cao phục vụ cho các cơ quan nhà nước, người dân và doanh nghiệp;

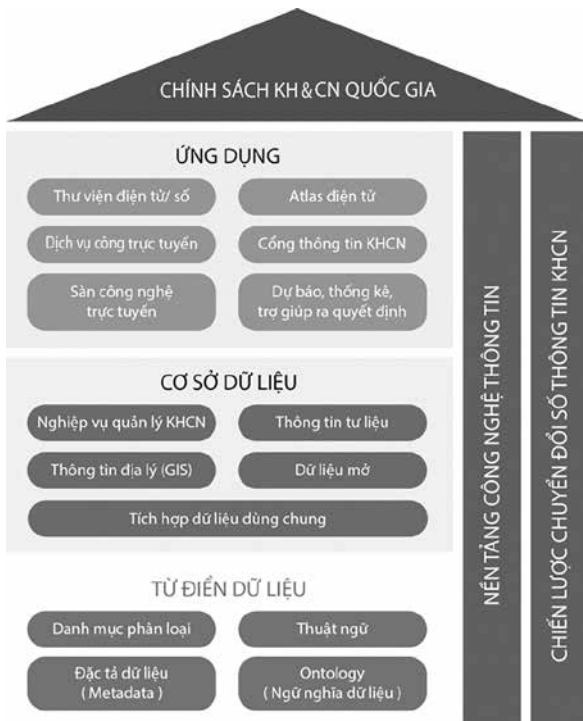
iv) Ứng dụng các công nghệ tiên tiến của kỷ nguyên số, trong đó có trí tuệ nhân tạo (AI), dữ liệu lớn (Bigdata), internet kết nối vạn vật (IoT) vào thu thập, lưu trữ, xử lý, phân tích dữ liệu tự động để gợi ý thông tin hoặc hỗ trợ người dùng ra quyết định;

v) Đổi mới chính sách theo hướng khoa học mở để thúc đẩy truy cập mở đối với kết quả của các công trình nghiên cứu và phát triển.

3. Xây dựng nền tảng quản lý thông tin khoa học và công nghệ quốc gia

Nền tảng quản lý thông tin KH&CN dựa trên CNTT được xây dựng trên cơ sở của ba tầng kiến trúc là: từ điển dữ liệu, cơ sở dữ liệu và ứng dụng (Hình 1). Các ứng dụng phải được phát triển trên một nền tảng chung gồm: từ điển dữ liệu, các mã định danh và các CSDL.

3.1. Từ điển dữ liệu



Hình 1. Khung kiến trúc nền tảng số thông tin KH&CN quốc gia

Từ điển dữ liệu quy định các khái niệm và lược đồ dữ liệu được sử dụng thống nhất để phục vụ chia sẻ dữ liệu trong toàn hệ thống thông tin KH&CN. Từ điển được xây dựng ở các mức độ trừu tượng từ thấp đến cao gồm: i) danh mục thuật ngữ (thesaurus); ii) danh mục phân loại (taxonomy); iii) đặc tả siêu dữ liệu (metadata); iv) đặc tả ngữ nghĩa dữ liệu (ontology).

Thuật ngữ là các từ khóa thường dùng để đánh chỉ mục cho các tài liệu nhằm mục đích tìm kiếm theo yêu cầu của người dùng. Các thuật ngữ được thu thập từ các văn bản quản lý nhà nước hoặc từ các từ điển theo lĩnh vực chuyên ngành (y tế, tài chính, ngân hàng,...).

Danh mục phân loại là loại dữ liệu dùng để tham chiếu trong các CSDL. Đây là loại dữ liệu dùng chung để thống nhất khi khai báo và tạo lập dữ liệu trong các hệ thống thông tin. Danh mục phân loại có thể được quy định trong các văn bản quản lý nhà nước, ví dụ như quyết định số 12/2008/QĐ-BKHCN quy định phân loại các lĩnh vực nghiên cứu,

các mục tiêu kinh tế-xã hội, và các dạng của hoạt động KH&CN.

Danh mục cũng có thể là các bảng phân loại theo chuẩn quốc tế như bảng phân loại về sáng chế quốc tế (IPC); kiểu dáng công nghiệp (bảng phân loại locarno); các yếu tố hình của nhãn hiệu trong đăng ký nhãn hiệu (bảng phân loại Vienne); hàng hóa/dịch vụ trong đăng ký nhãn hiệu (bảng phân loại Nice); bảng phân loại lĩnh vực tiêu chuẩn kỹ thuật.

Đặc tả siêu dữ liệu là định nghĩa tập hợp các thuộc tính mô tả được sử dụng khi lập chỉ mục cho các loại thông tin như thư viện tài liệu, bộ dữ liệu và các đối tượng khác. Trang web của DDC (Digital Curation Center) liệt kê một danh sách các chuẩn đặc tả siêu dữ liệu được dùng trên thế giới. Các chuẩn thường dùng là Dublin core trong lĩnh vực TT-TV; DIF (Directory Interchange Format) trong lĩnh vực khoa học tự nhiên; Darwin Core trong lĩnh vực sinh học; DDI (Data Documentation Initiative) trong lĩnh vực khoa học xã hội và lưu trữ; TEI (Text Encoding Initiative) trong lĩnh vực khoa học xã hội, ngôn ngữ và con người; ISO 19115 trong lĩnh vực thông tin địa lý [2].

Trong kỷ nguyên số, dữ liệu dùng cho thông tin KH&CN sinh ra không chỉ để cho con người tìm kiếm sử dụng mà phải có khả năng để cho máy tính xử lý hoàn toàn tự động. Chính vì vậy, dữ liệu được tạo ra cần kèm theo đặc tả ngữ nghĩa. Hầu hết các CSDL thông tin KH&CN phổ biến trên thế giới hiện nay đều đã và đang được chuyển đổi để đáp ứng khả năng cung cấp ngữ nghĩa dưới dạng của dữ liệu liên kết (linked data). Trong lĩnh vực thông tin KH&CN, một số ontology sau đây thường dùng để đặc tả ngữ nghĩa của dữ liệu: BIBO (Bibliographic Ontology) [3] dùng để mô tả ngữ nghĩa các chỉ mục tài liệu; ontology được ánh xạ với các lược đồ dữ liệu cũ như Bibtext, MARC, ScienceDirect; SPAR (Semantic Publishing and Referencing Ontologies) [4] là một tập các ontology mô tả ngữ nghĩa của dữ liệu dùng trong lĩnh vực xuất bản và tham chiếu trích dẫn khoa học; GCIS (Global Change Information System) [5] được sử dụng để xây dựng CSDL của chương trình nghiên cứu biến đổi khí hậu toàn cầu do Chính phủ Mỹ tài trợ; Datacite Schema [6] mô tả ngữ

nghĩa cho các bộ dữ liệu mở được chia sẻ trên internet; Springer Nature SciGraph Ontology [7] mô tả ngữ nghĩa cho dữ liệu của nhà xuất bản Springer; UN Document Ontology [8] dùng để mô tả ngữ nghĩa cho tất cả các tài liệu được sử dụng trong các tổ chức của Liên Hợp Quốc.

3.2. Các cơ sở dữ liệu

Chuyển đổi số thông tin KH&CN sẽ đặt trọng tâm vào việc xây dựng các nền tảng CSDL mà trên đó có thể xây dựng nhiều loại ứng dụng khác nhau để cung cấp dịch vụ cho người dùng đầu cuối. Sử dụng chung nền tảng CSDL sẽ tránh được hiện trạng đầu tư chồng chéo, dữ liệu không đầy đủ và thiếu toàn diện tại nhiều hệ thống khác nhau. Các CSDL nằm trong nền tảng được phân vào 5 nhóm chính dưới đây:

- Loại thứ nhất, các CSDL nghiệp vụ được tạo ra từ các ứng dụng phục vụ công tác quản lý nhà nước về KH&CN dưới dạng các dịch vụ công như quản lý đề tài, nhiệm vụ KH&CN; quản lý đăng ký hoạt động của các tổ chức, doanh nghiệp KH&CN; quản lý nhà nước trong các lĩnh vực sở hữu công nghiệp, tiêu chuẩn đo lường chất lượng, v.v.

- Loại thứ hai là các CSDL thông tin tư liệu cung cấp kho lưu trữ, dịch vụ tìm kiếm thông tin tài liệu bao gồm chỉ mục và nội dung toàn văn. CSDL này là nền tảng của các ứng dụng thư viện điện tử, thư viện số, tạp chí truy cập mở. Một số CSDL lớn đang khai thác ở Việt Nam hiện nay là CSDL toàn văn các công bố trong nước CSDL nhiệm vụ KH&CN, do Cục Thông tin KH&CN quốc gia quản lý; các CSDL tài liệu nội sinh, tạp chí truy cập mở của các viện nghiên cứu và trường đại học.

- Loại thứ ba là CSDL thông tin địa lý (GIS), cung cấp nền tảng cho các ứng dụng atlas điện tử thường dùng trong một số ngành khoa học tự nhiên như địa lý, trái đất, tài nguyên và môi trường, sinh học,... Atlas điện tử có thể được xây dựng theo nhóm chuyên đề trên phạm vi cả nước hoặc được tổng hợp cho một địa bàn hành chính (ví dụ atlas điện tử tổng hợp vùng Tây Nguyên, atlas điện tử của các tỉnh thành).

- Loại thứ tư là các kho lưu trữ chia sẻ dữ liệu mở. Đây là nền tảng cung cấp các bộ dữ liệu có thể dùng cho nghiên cứu chuyên

sâu của các nhà khoa học. Dữ liệu mở có thể được tạo ra từ việc kết xuất một CSDL sẵn có hoặc là kết quả của một nghiên cứu đã thực hiện. Người dùng có thể khai thác các bộ dữ liệu mở theo một trong hai hình thức là ngoại tuyến (tải về toàn bộ các bản ghi và sử dụng) hoặc trực tuyến (truy vấn lấy các bản ghi dữ liệu theo yêu cầu).

- Cuối cùng, CSDL tích hợp là nền tảng cung cấp thông tin tổng hợp được thu thập từ nhiều nguồn CSDL khác nhau dựa trên một cấu trúc lược đồ thống nhất. Điểm khác biệt của CSDL tích hợp dùng chung là tập trung vào ngữ nghĩa và sự liên kết của dữ liệu ở các nhóm thông tin khác nhau. Trong CSDL tích hợp thường không chứa các loại dữ liệu chi tiết như nội dung toàn văn hoặc thông tin thuộc về quy trình xử lý nghiệp vụ. CSDL trích dẫn khoa học là một ví dụ điển hình của dạng này. Nó tập trung vào việc thu thập thông tin bài báo từ nhiều nguồn xuất bản để từ đó phân tích ra mối quan hệ trích dẫn giữa các bài báo với nhau.

3.3. Các mã định danh trường tồn

Trên thế giới hiện nay, xu thế chung của các hệ thống thông tin KH&CN là dùng hệ thống các mã định danh trường tồn (Persistent Identification) để định danh các tài nguyên. Một mã PID sẽ tồn tại suốt đời và không thay đổi cho một đối tượng dữ liệu được mô tả. Một trong những đặc điểm sử dụng mã PID là người dùng có thể truy xuất đầy đủ thông tin của đối tượng được mô tả trên Web mà không cần biết hệ thống CNTT phía sau của nó. Ví dụ như người dùng có thể truy cập tới nơi xuất bản gốc của các công bố có mã định danh số cho đối tượng (DOI) tại địa chỉ: <https://doi.org/<doi>>.

Các loại mã PID có thể được dùng để định danh cho 3 nhóm đối tượng chính trong thông tin KH&CN. Thứ nhất là các mã định danh sử dụng cho các loại tài nguyên số như DOI (Digital Object Identifier), Handle, ARK (Archival Resource Key). Thứ hai là các loại mã dùng để định danh cho cá nhân như Scopus ID, ResearcherID, ORCID ID, ResearchGate, Linkedin. Cuối cùng là các loại mã dùng để định danh cho tổ chức như ISNI (ISO 27729), PSI (Publisher Solutions International), Ringgold, LEI (Legal Entity Identifier).

4. Giải pháp chuyển đổi số để hội nhập quốc tế

Trong thời đại công nghệ số, việc phát triển các tài nguyên thông tin KH&CN nội sinh được coi là nhiệm vụ công việc trọng tâm [1]. Tuy nhiên, việc phát triển các nguồn thông tin nội sinh phải bảo đảm tính mở và chuẩn hóa để có thể hội nhập với quốc tế. Trước tiên, các công bố trong nước cần chuyển đổi để đáp ứng tiêu chí xuất bản điện tử theo chuẩn mực của thế giới. Các tạp chí và kỷ yếu của hội nghị khoa học phải được xuất bản kèm theo mã DOI để người dùng có thể truy cập tới các bài báo gốc đăng tải trên internet. Thư viện số được xây dựng tại các viện, trường và cơ quan quản lý khoa học để lưu trữ các kết quả trong nghiên cứu và đào tạo. Kết quả có thể là các tài liệu, phần mềm, bộ dữ liệu được quản lý để đáp ứng các chuẩn của truy cập mở. Người dùng có thể tìm kiếm và khai thác đầy đủ nội dung điện tử nếu dữ liệu đã được cấp phép mở (tùy chọn một giấy phép phù hợp theo Creative Commons).

Dạng thông tin nội sinh thứ hai cần được phát triển để phục vụ công tác quản lý nghiên cứu và phát triển tại các bộ, ngành, địa phương. Đây là những CSDL cung cấp các thông tin KH&CN đầy đủ về nguồn nhân lực KH&CN, tổ chức KH&CN, các nhiệm vụ KH&CN và các kết quả đã đạt được. Dữ liệu lưu trữ trong hệ thống cần được chuẩn hóa để có thể dễ dàng chia sẻ, trao đổi trong mạng lưới thông tin KH&CN. Tiêu chuẩn dữ liệu được sử dụng phổ biến nhất hiện nay cho các hệ thống thông tin KH&CN là CERIF [9] do tổ chức EuroCRIS chủ trì xây dựng. CERIF được khuyến cáo sử dụng cho tất cả các hệ thống thông tin nghiên cứu ở các cấp độ từ viện, trường, vùng, quốc gia, cho tới cả Liên minh châu Âu. Tiêu chuẩn này đã thực hiện mô hình hóa toàn bộ thông tin KH&CN thành các nhóm đối tượng thông tin phải quản lý gồm cá nhân, tổ chức hoạt động khoa học, dự án nghiên cứu, kết quả công bố, sản phẩm khoa học, sáng chế, tài chính,...

Ở cấp độ quốc gia, một CSDL thông tin KH&CN được xây dựng trên cơ sở tích hợp dữ liệu từ tất cả các nguồn cung cấp tin nội sinh ở trong nước. CSDL quốc gia thực

hiện chuẩn hóa và liên kết dữ liệu từ nhiều nguồn khác nhau để cung cấp “một cửa” tìm kiếm thông tin KH&CN trong cả nước. Hệ thống này sẽ hoạt động như một kho dữ liệu tập trung theo mô hình của hệ thống OpenAIRE [10] phục vụ phát triển nghiên cứu mở tại châu Âu.

Giao thức lưu trữ mở OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) [11] sẽ được lựa chọn làm giao thức tiêu chuẩn để thu thập dữ liệu thông tin KH&CN đa dạng từ nhiều nguồn. Đây là tiêu chuẩn toàn cầu được áp dụng trong các ứng dụng liên quan đến quản lý, lưu trữ tài liệu và thông tin tư liệu. OAI-PMH được xây dựng dựa trên nền tảng công nghệ web (API Restful). Giao thức này có tính tổng quát và độ linh hoạt cao, cho phép thu thập các bộ dữ liệu theo nhiều kiểu mẫu với metadata mô tả khác nhau. Tất cả các CSDL KH&CN trong nước và của quốc gia cần được nâng cấp, chuyển đổi để hỗ trợ chuẩn giao thức trao đổi dữ liệu OAI-PMH.

Chỉ số trích dẫn các công bố được xem là một thông tin quan trọng để đo lường hiệu quả của các nghiên cứu hiện nay. V-CitationGate [12] là CSDL thư mục đầu tiên tại Việt Nam đã phân tích được chỉ số trích dẫn trong cho các công bố khoa học. V-CitationGate đã kết nối thành công các ấn phẩm khoa học xuất bản trên 50 Tạp chí khoa học của Việt Nam có thông tin trích dẫn trong nước và trên thế giới. Tuy nhiên, hiện nay CSDL mới chỉ tập trung vào phân tích chỉ số cho các bài báo và tạp chí trong nước. Cần phải mở rộng CSDL để có được thông tin chỉ số trích dẫn của các tác giả là người Việt có công bố khoa học ở trong và ngoài nước. Hệ thống các chỉ số đánh giá trong CSDL cũng cần được liên kết với các hệ thống chỉ số đánh giá thông dụng khác trên thế giới như ISI, Scopus, Altmetric,...

Để thông tin KH&CN trong nước có thể hội nhập quốc tế, cần áp dụng các loại mã định danh trường tồn (PID) khi tạo lập dữ liệu. Các mã định danh này bảo đảm tính duy nhất để giúp dễ dàng liên kết các thông tin trong phân tích dữ liệu. Cụ thể, các công bố khi xuất bản cần được gắn với mã định danh DOI; lưu trữ các tài nguyên số có thể gắn với các mã định danh Handle, ARK; các cá nhân hoạt động KH&CN cần được

khuyến khích tham gia vào hệ thống đăng ký định danh quốc tế như ORCID.

Với sự phát triển của internet, các nền tảng dựa trên nguồn lực đám đông đang tạo ra một động lực mới, rất hiệu quả trong phát triển thông tin KH&CN. Mô hình này sử dụng các đóng góp của chính người dùng để xây dựng nội dung thông tin. Hệ thống điển hình đầu tiên ra đời theo mô hình này chính là trang thông tin Wikipedia. Trong lĩnh vực thông tin KH&CN, các hệ thống hoạt động dựa trên đám đông có thể kể đến như: arXiv.org, cho phép người dùng lưu trữ, chia sẻ các báo cáo hoặc bài báo trước khi xuất bản; academia.edu là một mạng xã hội cho phép các cán bộ nghiên cứu và phát triển chia sẻ và theo dõi kết quả nghiên cứu của các cá nhân.

Chuyển đổi số thông tin KH&CN ở Việt Nam cũng cần ưu tiên phát triển một nền tảng tương tự như arXiv.org hoặc academia.edu. Đây là nơi người dùng tự do chia sẻ các kết quả, công trình nghiên cứu của cá nhân. Nó là mô hình giúp tập hợp tri thức từ cộng đồng một cách nhanh chóng, đầy đủ và hiệu quả nhất so với tất cả các mô hình khác. Tuy nhiên, mô hình này chỉ đảm bảo tính bền vững khi nó vận hành dựa trên một nền tảng kinh doanh mang được lợi ích cho người tham gia đóng góp về mặt danh tiếng, cơ hội phát triển nghề nghiệp hoặc thậm chí là tài chính thu được. Với một kho tư liệu khổng lồ do người dùng đóng góp (bao gồm cả các tổ chức), chúng ta có thể dễ dàng xây dựng được nhiều tiện ích hỗ trợ người dùng dựa trên các công nghệ tiên tiến như AI, học máy. Ví dụ hệ thống có thể phát hiện một cách nhanh chóng về khả năng “đạo văn” của một tài liệu tiếng Việt bất kỳ.

Kết luận

Chuyển đổi số ngành thông tin KH&CN sẽ dựa trên việc phát triển các nền tảng công nghệ. Quá trình chuyển đổi số của ngành được gắn kết với sự phát triển chung của quốc gia về chính phủ điện tử. Nhà nước không nên đầu tư vào xây dựng các ứng dụng mà tập trung vào việc tạo lập hạ tầng, nền tảng số, trong đó có nền tảng dịch vụ chia sẻ và trao đổi dữ liệu về thông tin KH&CN theo chuẩn mực quốc tế; nền tảng quản lý CSDL quốc gia thống

nhất, tập trung được các nguồn tin trong cả nước. Nhà nước cần tạo cơ chế để khuyến khích xã hội hóa phát triển các ứng dụng trong lĩnh vực thông tin KH&CN như thư viện điện tử, thư viện số, mạng chia sẻ dữ liệu mở, sàn giao dịch kết nối cung cầu,....

Xu thế chung trong hội nhập quốc tế hiện nay là theo đuổi các chính sách về khoa học mở, dữ liệu mở và truy cập mở. Đầu tư xây dựng các hệ thống công nghệ cần tránh theo hướng tự thiết kế (adhoc) mà cần dựa trên các tiêu chuẩn phổ quát trên thế giới trong lĩnh vực thông tin KH&CN. Trong quá trình chuyển đổi số, các nguyên tắc FAIR (Findable - Accessible - Interoperable - Reusable): Tìm thấy được- Truy cập được- Tương hợp được- Sử dụng lại được [13] luôn cần được áp dụng cho thiết kế dữ liệu của các hệ thống thông tin KH&CN.

TÀI LIỆU THAM KHẢO

- Đào Mạnh Thắng, Trần Thị Hải Yến (2017). “Phát triển nguồn tin khoa học và công nghệ trong thời đại công nghệ số”, Tạp chí Thư viện Việt Nam - 2017 - Số 1 - Tr. 12-21
 - List of Metadata Standards, <http://www.dcc.ac.uk/resources/metadata-standards/list>
 - Bibliographic Ontology Specification, <http://bibliontology.com>
 - Semantic Publishing and Referencing Ontologies, <http://www.sparontologies.net/>
 - GCIS Ontology, <https://data.globalchange.gov/gcis.owl>
 - DataCite Metadata Schema, <http://schema.datacite.org/>
 - SN SciGraph, A Linked Open Data platform for the scholarly domain, <https://scigraph.springernature.com/explorer/ontology/>
 - The United Nations System Document Ontology, <https://unsceb-hlcm.github.io/onto-undo/>
 - Carlos Sousa Pinto, Cláudia Simões, Luis Amaral. “CERIF – Is the Standard Helping to Improve CRIS?”, *Procedia Computer Science*, Vol 33, 2014, Pages 80-85, ISSN 1877-0509.
 - OpenAIRE. European project supporting Open Science, <https://www.openaire.eu/>
 - OAI-PMH. Open Archives Initiative Protocol for Metadata Harvesting, <https://www.openarchives.org/pmh/>
 - V-CitationGate. Cơ sở dữ liệu thư mục Vietnam Citation Gateway, <https://vcgate.vnu.edu.vn/>
 - Wilkinson, M. D. et al. “The FAIR Guiding Principles for scientific data management and stewardship”, *Scientific Data*. Vol 3, 2016
- (Ngày Tòa soạn nhận được bài: 10-5-2019; Ngày phản biện đánh giá: 20-9-2019; Ngày chấp nhận đăng: 15-10-2019).