

## ỨNG DỤNG DỮ LIỆU MỞ LIÊN KẾT NHẪM NÂNG CAO CHẤT LƯỢNG CỦA CÁC KHO TRI THỨC SỐ

ThS Nguyễn Danh Minh Trí

Trường Đại học KHXH&NV, Đại học Quốc gia Tp.HCM

**Tóm tắt:** Kho tri thức số là một phương thức lưu trữ hiệu quả và đang trở nên ngày càng phổ biến trên thế giới. Thông qua việc xây dựng kho tri thức số, các cơ sở giáo dục đang góp phần tạo nên một xã hội học tập rộng khắp và nâng tầm chất lượng giáo dục. Trong các kho tri thức số, nhiều nguồn tài nguyên đã được xây dựng trên nền tảng dữ liệu mở liên kết và có tiềm năng phát triển lâu dài. Bài báo trình bày khái quát về dữ liệu mở liên kết và giải pháp ứng dụng dữ liệu mở liên kết trong việc nâng cao chất lượng của các kho tri thức số.

**Từ khóa:** Dữ liệu mở liên kết; kho tri thức số.

### APPLYING LINKED OPEN DATA TO IMPROVE THE QUALITY OF DIGITAL KNOWLEDGE HUB

**Abstract:** Digital knowledge hub is an effective storage method and is becoming more popular around the world. Through building digital knowledge hubs, educational institutions are contributing to create a broad learning society and improving the quality of education. In digital knowledge hubs, many resources have been built on linked open data platforms and have long-term growth potential. The article presents an overview of linked open data and linked open data application solutions in improving the quality of digital knowledge hubs.

**Keywords:** Linked Open Data; digital knowledge hub.

## MỞ ĐẦU

Dữ liệu mở đã trở thành một lựa chọn ưu tiên trong việc xây dựng các nguồn tài nguyên thông tin. Nhiều quốc gia và tổ chức giáo dục đã và đang chuyển hướng sang sử dụng dữ liệu mở như một phần tất yếu của quá trình phát triển. Dữ liệu mở thường được liên kết với nhau trong một mạng lưới rộng khắp và còn được gọi là dữ liệu mở liên kết. Với lợi thế được xây dựng từ các cộng đồng đóng góp đa dạng, dữ liệu mở liên kết hứa hẹn sẽ tạo ra những kho tri thức số chất lượng, góp phần đem lại sự phát triển giáo dục bền vững cho nhân loại trong tương lai.

## 1. KHÁI QUÁT VỀ DỮ LIỆU MỞ LIÊN KẾT VÀ KHO TRI THỨC SỐ

### 1.1. Khái quát về dữ liệu mở

Dữ liệu mở là một khái niệm không mới nhưng chứa đựng rất nhiều ý nghĩa khi tiếp cận dưới nhiều góc độ khác nhau. Một định nghĩa chính thức về dữ liệu mở vẫn còn là một điều khá mới mẻ [Nguyễn Danh Minh Trí, 2019]. Phần lớn định nghĩa cho rằng: “Dữ liệu mở là dữ liệu mà bất cứ ai cũng có thể truy cập, sử dụng hoặc chia sẻ theo một giấy phép mở”, bao gồm một phiên bản

định nghĩa ngắn gọn có thể truy cập được, và tham chiếu đến định nghĩa chính thức. Theo đó, dữ liệu mở là các dữ liệu điện tử hoặc dữ liệu số phải được tự do sẵn có cho mọi người sử dụng và tái xuất bản theo ý muốn mà không bị hạn chế bởi bản quyền, bằng sáng chế hoặc các cơ chế kiểm soát khác.

Theo xu hướng phát triển của các kho tri thức số, các nguồn dữ liệu mở sẽ phát sinh ngày càng lớn và đa dạng. Dữ liệu mở có thể là dữ liệu văn bản thuần túy, dữ liệu đa phương tiện, dữ liệu từ các thí nghiệm khoa học,... được phát hành theo giấy phép mở, thường là giấy phép chia sẻ tương tự. Các tính năng chính của dữ liệu mở bao gồm [Trần Minh, 2017]:

- *Tính sẵn sàng và sự truy cập:* dữ liệu mở phải sẵn sàng một cách tổng thể và không nhiều hơn chi phí tái sinh hợp lý, được tải về qua internet. Dữ liệu đó cũng phải là sẵn sàng ở dạng thuận tiện và có khả năng tùy biến được.

- *Tính tái sử dụng và phân phối lại:* dữ liệu mở phải được cung cấp theo các điều khoản cho phép sử dụng lại và phân phối lại, bao gồm cả việc trộn lẫn với các tập hợp dữ liệu khác.

- *Tính cộng đồng*: đề cập đến sự tham gia toàn cầu của người dùng. Mọi người phải có khả năng sử dụng, sử dụng lại và phân phối lại. Không có sự phân biệt giữa các lĩnh vực của đời sống hoặc giữa con người hoặc các nhóm người.

Dữ liệu mở có một tầm ảnh hưởng nhất định trong giáo dục, đặc biệt là giáo dục mở. Trong đó, tính mở của thông tin và dữ liệu có những ảnh hưởng lớn đến giáo dục cũng như chất lượng đào tạo [Nguyễn Danh Minh Trí, 2017]. Tính mở của các nguồn dữ liệu không chỉ tác động đến việc dạy và học mà còn ảnh hưởng đến quá trình nghiên cứu và chính sách của các tổ chức giáo dục.

Trên thế giới đã có rất nhiều nguồn tài nguyên giáo dục mở được xây dựng dựa trên nền tảng dữ liệu mở [Atenas & Havemann, 2015] và phần lớn được đóng góp bởi các cộng đồng tri thức thông qua môi trường web [Dhanarajan & Porter, 2013]. Một trong những cộng đồng đóng góp lớn đến từ các nhà nghiên cứu, nhà giáo dục, giáo viên và người học đã và đang góp phần xây dựng các nguồn tài nguyên giáo dục mở chất lượng [Butcher, 2015].

Rất nhiều hệ thống khác nhau đã và đang tham gia vào quá trình lưu trữ và phân phối dữ liệu mở [Nguyễn Thế Hùng, 2019]. Điều này đã tạo ra dòng chảy dữ liệu mở đang gia tăng một cách nhanh chóng trên toàn cầu. Việc triển khai và duy trì cơ sở hạ tầng để lưu trữ các nguồn dữ liệu mở cũng là một yếu tố then chốt. Dữ liệu mở thường được liên kết và chia sẻ thông qua mạng internet và các ứng dụng phần mềm hỗ trợ. Quá trình liên kết dữ liệu mở diễn ra thường xuyên và liên tục sẽ giúp cho các nguồn tài nguyên và kho tri thức số được xây dựng ngày càng phong phú và chất lượng hơn.

## 1.2. Dữ liệu mở liên kết và kho tri thức số

Kho tri thức số là nơi lưu trữ các tài liệu đến từ nhiều ngành và lĩnh vực khác nhau. Xây dựng kho tri thức số không thể tách rời việc xây dựng những nguồn dữ liệu chất lượng, trong đó có dữ liệu mở. Về cơ bản,

dữ liệu mở không cần liên kết đến dữ liệu khác. Tuy nhiên, dữ liệu mở sẽ phát huy tối đa tiềm năng khi được liên kết trong một mạng lưới cộng đồng đóng góp đa dạng và có tính bền vững cao.

Dữ liệu được lưu trữ trong các hệ thống hiện nay có khối lượng rất lớn và đa dạng [Nguyễn Danh Minh Trí, 2018]. Tuy nhiên, không phải tất cả các dữ liệu đều ở trạng thái mở và sẵn sàng cho người dùng sử dụng. Để tận dụng tiềm năng kết nối của các nguồn dữ liệu mở, đã có nhiều tổ chức tham gia vào quá trình liên kết dữ liệu, trong đó có W3C. W3C là một tổ chức cộng đồng đã có nhiều nỗ lực trong việc bảo vệ tính mở của dữ liệu và hướng tới việc làm giàu đám mây dữ liệu mở liên kết.

Khi đề cập đến dữ liệu mở liên kết, không thể không nhắc đến khái niệm dữ liệu liên kết. Dữ liệu liên kết (Linked Data-LD) là tập hợp các nguyên tắc thiết kế để chia sẻ các dữ liệu được kết nối với nhau trên web mà máy tính có thể đọc được [Ontotext]. Các liên kết này có được thông qua kết nối mạng và được các ứng dụng hỗ trợ đọc tự động [Wikipedia]. Dữ liệu mở liên kết (Linked Open Data-LOD) là sự pha trộn của dữ liệu liên kết và dữ liệu mở. Do vậy, có thể nói LOD là tập các dữ liệu vừa được liên kết với nhau và được xây dựng dựa trên các nguồn dữ liệu mở.

Để có được những nguồn LOD chất lượng sẽ cần có sự đầu tư bài bản và khả năng truy cập hiệu quả về mặt công nghệ, nhất là truy cập mở. Xây dựng chính sách truy cập mở là điều kiện cần thiết để phát triển các nguồn tài nguyên giáo dục mở [Lê Trung Nghĩa, 2018]. Sự phát triển mạnh mẽ của công nghệ web, đặc biệt là công nghệ web ngữ nghĩa ngày nay đã tạo điều kiện lớn cho việc truy cập mở nhằm xây dựng các nguồn LOD chất lượng ở nhiều lĩnh vực.

Công nghệ web ngữ nghĩa (Semantic Web-SW) đã và đang thay đổi cách thức lưu trữ và phổ biến thông tin, trong số đó có LOD [Piedra et al., 2013]. LOD sẽ tạo nên

các kết nối dữ liệu có cấu trúc trên web, từ đó hỗ trợ quá trình xuất bản thông tin đạt hiệu quả cao. Các kết nối này sẽ ngày càng được mở rộng tùy theo quy mô và định hướng phát triển các nguồn dữ liệu mở khác nhau. Một ví dụ nổi bật của LOD chính là cơ sở dữ liệu Dbpedia. Dbpedia là một dự án tận dụng nguồn lực cộng đồng để trích xuất thông tin có cấu trúc từ Wikipedia và làm cho các thông tin này sẵn sàng trên web để cung cấp cho các ứng dụng khác.

LOD có nhiều lợi ích và được ứng dụng rộng rãi trên thực tế. Được xây dựng trên nền tảng là các tiêu chuẩn phổ biến và chính sách minh bạch của dữ liệu mở, LOD là hữu dụng cho các tổ chức cũng như xã hội. Một trong các lợi ích phổ biến của LOD chính là góp phần xây dựng các nguồn tài nguyên giáo dục mở và các kho tri thức số chất lượng [Kaltenböck et al., 2016]. LOD giúp phá vỡ các “cấu trúc ống thông tin đóng” đang tồn tại giữa các định dạng và loại bỏ các hàng rào giữa các nguồn dữ liệu khác nhau. Từ đó, sử dụng các tiêu chuẩn của LOD sẽ tạo ra khả năng tích hợp và dễ dàng duyệt qua các nguồn dữ liệu phức tạp. Các chỉ dẫn đó cũng cho phép dễ dàng cập nhật và mở rộng các mô hình dữ liệu mở.

LOD có thể được sử dụng để trình bày dữ liệu theo cách thức được liên kết và theo một tập hợp các nguyên tắc toàn cầu. Điều này sẽ củng cố và làm gia tăng chất lượng dữ liệu. Hơn nữa, cơ sở dữ liệu đồ thị ngữ nghĩa cho việc trình bày LOD sẽ tạo ra các liên kết ngữ nghĩa giữa các nguồn và định dạng rời rạc phân tán khác nhau và suy diễn ra tri thức mới vượt ra khỏi các tri thức có sẵn. Đây chính là nguyên tắc chính để xây dựng các kho tri thức số dựa trên LOD.

LOD cũng góp phần tạo ra nhiều ứng dụng thông minh để phục vụ cuộc sống như: ứng dụng thông minh phát hiện các hàng xóm tốt trong khu vực dựa trên các dữ liệu về trường học, giao thông, các tòa nhà văn phòng và các câu lạc bộ/các vườn hoa trong khu vực đó chẳng hạn; hoặc là

ứng dụng thông minh hỗ trợ đưa ra các giải pháp cải thiện tính sáng tạo và đổi mới cho tất cả những người phát triển, các công dân và các doanh nghiệp. Họ có thể sử dụng tất cả các tập hợp dữ liệu đó để đặt mọi thứ vào ngữ cảnh và tạo ra tri thức cũng như các ứng dụng hữu ích mới.

### **1.3. Khái quát về các kho tri thức số phổ biến được cộng đồng đóng góp**

#### ***Wikipedia***

Wikipedia (bách khoa toàn thư mở) là một website cung cấp thông tin dạng bách khoa toàn thư tự do. Ứng dụng này là kết quả của sự cộng tác giữa những người đọc từ khắp nơi trên thế giới. Nội dung của Wikipedia rất đa dạng, gần như bao trùm tất cả mọi khía cạnh của đời sống con người, từ chính trị, kinh tế, khoa học, công nghệ đến môi trường, đời sống, thiên nhiên, xã hội,... và tồn tại dưới nhiều ngôn ngữ khác nhau. Kho tri thức số được Wikipedia lưu trữ trực tuyến là một trong những kho tri thức số khổng lồ của nhân loại tính đến thời điểm hiện tại và lượng dữ liệu vẫn đang không ngừng được tăng lên hàng ngày.

Điểm hạn chế lớn nhất của Wikipedia là sự quá đa dạng, nên sẽ gặp khó khăn trong việc quản lý, đảm bảo tính đúng đắn cho tất cả nội dung được chia sẻ. Bên cạnh đó, việc cho phép người dùng có thể chỉnh sửa tự do bất cứ nội dung nào mà họ muốn sẽ dẫn đến nhiều hành động phá hoại nội dung hoặc chen các thông tin quảng cáo. Mặc dù đội ngũ quản trị của Wikipedia đã có cam kết là sẽ đảm bảo tính đúng đắn, vẫn sẽ tồn tại một lượng lớn nội dung và dữ liệu chưa hợp lý cần được chỉnh sửa.

#### ***Wikihow***

Ra mắt vào năm 2005 tại Palo Alto, California-Hoa Kỳ, WikiHow đã dần trở thành một trang web chia sẻ cách thức làm mọi thứ rất nổi tiếng và phổ biến trên thế giới [Nguyễn Danh Minh Trí, 2018]. Là một website cộng đồng trực tuyến mang tính chất bách khoa, Wikihow giúp người dùng

có thể tìm ra giải pháp cho những vấn đề gặp phải trong cuộc sống một cách dễ dàng và nhanh chóng. Ngoài ra, website này còn bao gồm một cơ sở dữ liệu rộng lớn về các hướng dẫn cách làm mọi việc rất chi tiết. Cũng giống như Wikipedia, người dùng sẽ tham gia đóng góp bài viết cho Wikihow nhằm làm phong phú dần các chủ đề của ứng dụng này. Những phương pháp xây dựng nội dung được sử dụng phổ biến nhất bao gồm: dịch và tạo bài viết, chỉnh sửa bài viết, tạo mới chủ đề. Phần lớn bài viết trên Wikihow đều được chia sẻ theo giấy phép xuất bản mở (Creative Commons-CC). Lượng dữ liệu mở liên kết được Wikihow lưu trữ vẫn đang gia tăng một cách nhanh chóng và ngày càng hỗ trợ tốt hơn nhu cầu tìm giải pháp của người dùng.

Wikihow có thể được xem như một công cụ hiệu quả trong việc xây dựng các nguồn tài nguyên giáo dục mở (Open Educational Resources-OER) [Nguyễn Danh Minh Trí, 2018]. Mặc dù Wikihow tiếng Việt chưa được phổ biến, cộng đồng người dùng ở Việt Nam vẫn đang có nhiều nỗ lực xây dựng và khuyến khích sự phát triển của website hữu ích này. Tính đến năm 2018, WikiHow được khai thác trong 18 ngôn ngữ với hơn 200.000 bài báo bằng tiếng Anh, hơn 3.000 bài viết bằng tiếng Việt và đang được tiếp tục phát triển với số lượng nhiều hơn.

### **VOER**

Chương trình Tài nguyên Giáo dục Mở Việt Nam (Vietnam Open Educational Resources-VOER) được hỗ trợ bởi Quỹ Việt Nam (The Vietnam Foundation) với mục tiêu xây dựng kho tài nguyên giáo dục mở cho người Việt sử dụng và truy nhập miễn phí nguồn tài liệu hữu ích phục vụ giảng dạy, học tập và nghiên cứu, cũng như phục vụ cho toàn xã hội. Về cơ bản, nội dung trong VOER sẽ được lưu trữ dưới hai định dạng: một là các module, là một chủ đề nhỏ hoặc một phần hoàn chỉnh của chủ đề lớn; hai là collection, là tập hợp các module được sắp xếp theo một trình tự nhất định để tạo

thành một cuốn sách hoặc giáo trình. Cách tổ chức nội dung theo module và collection làm cho việc chia sẻ, sử dụng và tái sử dụng nội dung mềm dẻo và dễ dàng hơn bao giờ hết. Bất kỳ tác giả nào khi đăng ký một tài khoản trên hệ thống đều có thể đóng góp nội dung.

VOER xây dựng các nguồn tài nguyên dựa trên cơ chế huy động đóng góp từ cộng đồng người dùng là các giảng viên trong các trường đại học [Nguyễn Danh Minh Trí, 2017], các viện nghiên cứu... Người dùng có thể đóng góp nội dung dưới dạng các module nhỏ [Nguyễn Danh Minh Trí, 2018], từ đó các nguồn dữ liệu mở sẽ được liên kết lại để có một kho tri thức đồ sộ, phủ kín các lĩnh vực và sẵn sàng cho việc tạo ra các giáo trình phục vụ cho việc học tập, giảng dạy và nghiên cứu một cách hiệu quả.

### **Hệ tri thức Việt số hóa**

Hệ tri thức Việt số hóa là một hệ thống tổng hợp các tài liệu được số hóa, lưu trữ và phổ biến tri thức cho toàn xã hội, trong đó đối tượng chủ yếu là thanh thiếu nhi. Hệ thống đã thu hút một lượng lớn người dùng tham gia với vai trò vừa khai thác, vừa đóng góp để làm giàu các tài nguyên tri thức số hóa của Việt Nam.

Kho dữ liệu mở thuộc đề án Hệ tri thức Việt số hóa là nguồn lực chiến lược quốc gia, với sự tham gia đóng góp bởi các cơ quan nhà nước, tổ chức, doanh nghiệp và mỗi người dân. Hệ tri thức Việt số hóa được tổng hợp từ 02 nguồn:

- Nguồn tri thức cơ bản có sẵn được xây dựng từ nhiều chuyên gia giáo dục đến từ các lĩnh vực như khoa học, chính trị, kỹ thuật, công nghệ...
- Nguồn tri thức cộng đồng được thu thập và liên tục cập nhập gồm các tri thức khoa học thường thức trong đời sống xã hội như: chăm sóc sức khỏe, y tế, phòng chống dịch bệnh, trồng trọt, chăn nuôi, thủy lợi, công nghệ bảo quản, chế biến và khoa học đời sống.

## 2. ỨNG DỤNG DỮ LIỆU MỞ LIÊN KẾT NHẪM NÂNG CAO CHẤT LƯỢNG CÁC KHO TRI THỨC SỐ

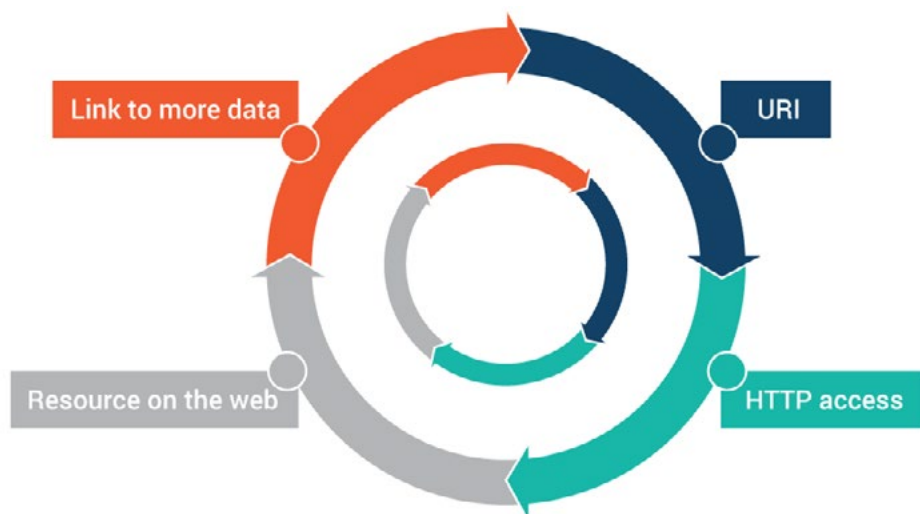
### 2.1. Xây dựng các nguồn dữ liệu mở liên kết

Như đã trình bày ở mục 1.2, dữ liệu mở liên kết (LOD) là một trong những khái niệm và trụ cột chính của web ngữ nghĩa (SW). SW còn được biết đến như là web của dữ liệu [Dicheva & Dichev, 2013] và chủ yếu được phát triển dựa trên những mối liên kết giữa các tập hợp dữ liệu mà cả con người và máy tính có thể hiểu được. LOD là một thành phần không thể thiếu của SW nhằm hiện thực hóa các hướng đi tốt để tạo ra những mối liên kết đó.

SW hỗ trợ lưu trữ nhiều dạng dữ liệu tùy theo các loại đối tượng khác nhau trên thực tế, từ các khái niệm và đối tượng, con

người, các vị trí được kết nối cùng nhau và chứa một lượng lớn dữ liệu được kết nối. Tuy nhiên, để liên kết, pha trộn và tích hợp các tập dữ liệu khổng lồ từ các nguồn dữ liệu thô rải rác phân tán, SW dựa trên LOD cần các chỉ dẫn cơ bản để liên kết dữ liệu nhằm tạo ra các nguồn dữ liệu ý nghĩa hơn.

Tim Berners-Lee, người phát minh ra World Wide Web, đã nêu lên một hướng đi mới thông qua việc thiết kế LOD vào năm 2006. Ông cũng đồng thời cung cấp các chỉ dẫn về việc sử dụng các công nghệ web được chuẩn hóa để thiết lập các liên kết ở cấp độ dữ liệu giữa các nguồn dữ liệu khác nhau. Trên cơ sở đó, các nguồn lực cộng đồng có thể được tận dụng để làm phong phú thêm kho tri thức số đang được lưu trữ trong hệ thống hiện tại.



**Hình 1.** Mối liên hệ của bốn nguyên tắc thiết kế dữ liệu liên kết đề xuất bởi Tim Berners - Lee vào năm 2006 [Ontotext]

Để xây dựng được web của LOD, cần có các điều kiện tiên quyết như: dữ liệu có thể được tham chiếu trên web, người dùng có thể cung cấp các đường liên kết web tới dữ liệu này từ bên trong dữ liệu của riêng họ. Theo Tim Berners-Lee, có 4 nguyên tắc cần tuân thủ khi thiết kế LOD, cụ thể là [Bizer, Heath, & Berners-Lee, 2009]:

- Sử dụng URI như là tên của các mô tả dữ liệu

Mã nhận diện tài nguyên thống nhất (Uniform Resource Identifier - URI), còn gọi là mã nhận diện mở, là nhận diện duy nhất của tất cả tài nguyên được kết nối sao cho có thể phân biệt và tích hợp mà không bị nhầm lẫn với nhau.

URI là một hệ thống nhận dạng toàn cầu duy nhất [Ngomo et al., 2014], được sử dụng để đặt tên duy nhất cho các nguồn tài nguyên (ví dụ: những nội dung kỹ thuật số

trên web, các đối tượng trong thế giới thực, các khái niệm trừu tượng,...). Theo đó, URI hỗ trợ phân biệt giữa những tài nguyên hoặc các đối tượng từ các tập dữ liệu khác nhau.

- Sử dụng các HTTP URI sao cho người dùng có thể tra cứu được các tên đó

Giao thức HTTP cung cấp một cơ chế đơn giản để truy xuất tài nguyên, khi các nguồn dữ liệu mở có thể được xác định bởi URI kết hợp với giao thức này. Điều này sẽ làm cho các nguồn dữ liệu trở nên dễ tìm hơn. Kết quả của quá trình này sẽ cho phép người dùng xuất bản bất kỳ loại dữ liệu nào và bổ sung cho không gian dữ liệu toàn cầu.

- Khi ai đó tra cứu một URI, ứng dụng sẽ cung cấp thông tin hữu ích bằng việc sử dụng các tiêu chuẩn (RDF, SPARQL).

Để có thể sử dụng URI một cách hiệu quả, việc truy vấn sử dụng RDF hoặc SPARQL đã được đề xuất [d'Aquin et al., 2013]. Cơ sở dữ liệu đồ thị ngữ nghĩa là công nghệ được phát triển để lưu trữ các dữ liệu được kết nối với nhau và tạo ý nghĩa cho các dữ liệu được kết nối đó thông qua việc làm giàu về ngữ nghĩa cho các tập hợp dữ liệu.

Khung Mô tả Tài nguyên (Resource Description Framework - RDF) là mô hình tiêu chuẩn, một định dạng biểu diễn dựa trên đồ thị cho việc xuất bản và trao đổi dữ liệu trên web được tổ chức W3C phát triển. RDF là tiêu chuẩn được sử dụng trong cơ sở dữ liệu đồ thị ngữ nghĩa, còn được tham chiếu tới như là bộ 3 RDF (RDF triplestore). Không giống như cơ sở dữ liệu quan hệ, bộ ba này là một công nghệ được phát triển để lưu trữ dữ liệu được kết nối với nhau và có khả năng hỗ trợ việc suy ra các dữ kiện mới từ những dữ liệu hiện có.

Bên cạnh đó, SPARQL là ngôn ngữ truy vấn được W3C chuẩn hóa để lấy và thao tác dữ liệu được lưu trữ ở định dạng RDF. Điều này cho phép người dùng có thể tìm kiếm trên web dữ liệu (hoặc bất kỳ cơ sở dữ liệu nào) và tìm ra các mối quan hệ dữ liệu mới, từ đó có được khả năng làm giàu các nguồn dữ liệu hiện có của mình.

- Đưa vào các liên kết tới các URI khác sao cho có thể phát hiện được nhiều mối liên hệ ngữ nghĩa hơn.

Tương tự với web siêu văn bản, các liên kết đến các URI khác làm cho dữ liệu được kết nối với nhau và cho phép người dùng tìm thấy những đối tượng đang cần. Bằng cách liên kết thông tin mới với các tài nguyên hiện có, người dùng có thể tối đa hóa việc tái sử dụng và liên kết lẫn nhau giữa các đối tượng dữ liệu nhằm tạo ra một mạng lưới liên kết phong phú có ý nghĩa và được hỗ trợ xử lý trong các hệ thống máy tính.

Sau khi có được web của LOD làm cơ sở để cộng đồng tham gia đóng góp, bước tiếp theo là cần nâng cao chất lượng của các nguồn dữ liệu mở. Đây là một trong những yếu tố then chốt trong sự phát triển của LOD.

## 2.2. Nâng cao chất lượng các nguồn dữ liệu mở

Để có thể nâng cao chất lượng các kho tri thức số được xây dựng từ các nguồn LOD, một yếu tố khác mà người dùng cần quan tâm là nâng cao khả năng xây dựng và triển khai hạ tầng cho các nguồn dữ liệu mở. Quá trình này không thể tách rời việc thu thập và chỉnh sửa các nguồn dữ liệu mở đã có sẵn cũng như đang phát sinh mới. Bước tiếp theo là tập trung phát triển khả năng kết nối giữa các nguồn dữ liệu này.

Theo đề xuất của Tim Berners-Lee, người dùng có thể hiện thực hóa hạ tầng dữ liệu mở theo 5 bước đáp ứng các tiêu chí của bộ quy tắc FAIR (Findable-Accessible-Interoperable-Reusable: Tìm thấy - Truy cập - Tương hợp - Tái sử dụng) như sau [Trần Minh, 2017]:

Bước 1 - Thực hiện cấp phép mở (Open License): Chia sẻ dữ liệu (dưới định dạng bất kỳ kể cả dùng định dạng pdf hoặc html) để có thể truy cập được trên internet và được cấp giấy phép truy cập mở.

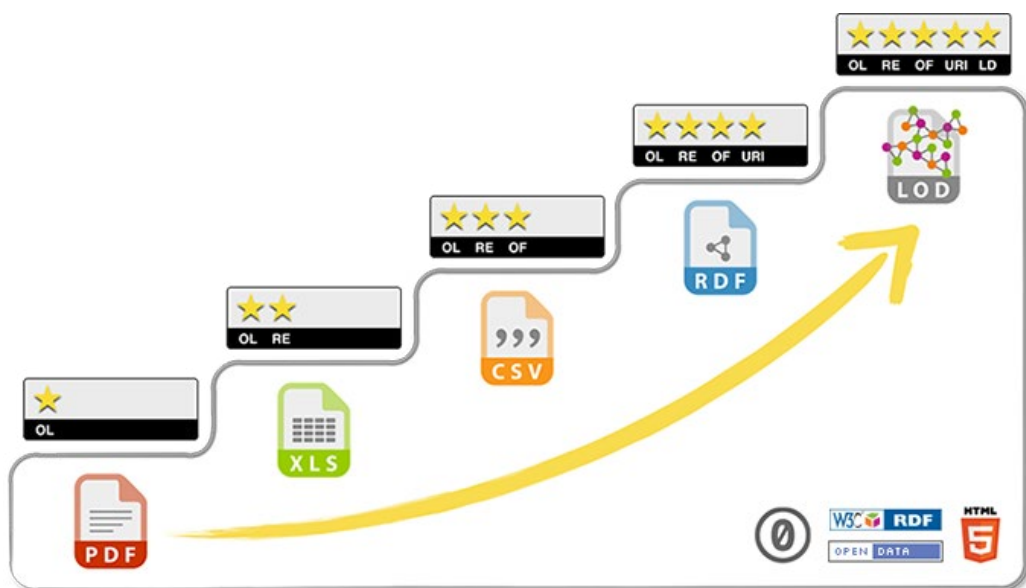
Bước 2 - Đảm bảo khả năng máy đọc được (Machine Readable): Dữ liệu được chia sẻ dưới định dạng mà máy có thể đọc

và xử lý được nội dung chứa đựng bên trong các nguồn dữ liệu này.

**Bước 3 - Lưu trữ và xuất bản theo định dạng mở (Open Format):** Dữ liệu được chia sẻ dưới các định dạng theo tiêu chuẩn mở (không bị phụ thuộc vào chỉ một nhà cung cấp dịch vụ phần mềm ứng dụng).

**Bước 4 - Sử dụng mã định danh URI:** Sử dụng các mã định danh toàn cầu URI để mô tả dữ liệu và siêu dữ liệu. Trong trường hợp này dữ liệu cần phải được mô hình hóa theo một chuẩn, được khuyến cáo bởi tổ chức W3C.

**Bước 5 - Nâng cao chất lượng các nguồn dữ liệu liên kết:** là cấp độ cao nhất thỏa mãn đủ các tiêu chí của tiêu chuẩn FAIR, cho phép các bộ dữ liệu có thể tham chiếu lẫn nhau thông qua các thuật ngữ dùng chung được định nghĩa dưới dạng của một từ điển dữ liệu. Từ điển dữ liệu này sẽ được phát triển dần theo thời gian và hỗ trợ việc tham chiếu cũng như liên kết dữ liệu giữa nhiều tổ chức khác nhau nhằm làm phong phú thêm các nguồn dữ liệu mở.



**Hình 2. Mô hình 5 bước phát triển hạ tầng dữ liệu mở [Nguồn 5stardata.info]**

Sau khi đã triển khai được hạ tầng cho các nguồn dữ liệu mở theo 5 bước trên, kết hợp với web của LOD đã được định hình, các tổ chức và người dùng đã có nền tảng vững chắc để nâng cao chất lượng của các kho tri thức số. Vấn đề còn lại nằm ở chiến lược xây dựng và duy trì cộng đồng đóng góp nhằm tối ưu hóa phương thức thu thập và xử lý các dữ liệu đã được thay đổi hoặc mới phát sinh. Quá trình xây dựng này cần nhiều thời gian và những bước đi đúng đắn, phù hợp với mục tiêu phát triển của từng kho tri thức số khác nhau.

## KẾT LUẬN

Kho tri thức số đã trở thành một trong

những kho tri thức lớn của nhân loại, hỗ trợ mạng lưới giáo dục trên toàn cầu thông qua việc lưu trữ nhiều dạng tài nguyên học thuật khác nhau. Trong các dạng tài nguyên đó thì dữ liệu mở đã xuất hiện như một xu thế tất yếu. Dữ liệu mở thường được liên kết với nhau trong các mạng lưới nên còn được gọi là dữ liệu mở liên kết. Việc ứng dụng các nguồn dữ liệu mở liên kết dựa trên cơ sở đóng góp từ cộng đồng là một trong những phương pháp hiệu quả nhằm xây dựng và nâng cao chất lượng của các kho tri thức số. Nhận thức được xu hướng phát triển và những lợi thế mà dữ liệu mở liên kết mang lại, nhiều cơ quan tổ chức và quốc gia đã và đang có những sự đầu tư cần thiết để tạo

ra nhiều cơ hội học tập hơn cho xã hội. Tuy nhiên, để việc ứng dụng dữ liệu mở liên kết được hiệu quả thì cần có những sự đầu tư dài hạn về mặt chiến lược, công nghệ cũng như phát triển các cộng đồng người dùng đóng góp đa dạng nhằm tạo ra các kho tri thức số ngày càng phong phú và chất lượng hơn trong tương lai.

### TÀI LIỆU THAM KHẢO

- Atenas, J. and Havemann, L. (Eds.). (2015). Open Data as Open Educational Resources: Case studies of emerging practice. London: Open Knowledge, Open Education Working.
- Bizer, C., Heath, T. and Berners-Lee, T. (2009). Linked data - The story so far. International Journal on Semantic Web and Information Systems (IJSWIS), 5, 1-22.
- Butcher, N. (2015). A basic guide to open educational resources (OER). UNESCO and COL.
- Dhanarajan, G. and Porter, D. (2013). Open educational resources: An Asian perspective. Vancouver: Commonwealth of Learning and OER Asia.
- Dicheva, D. and Dichev, C. (2013). Leveraging domain specificity to improve findability in OER repositories. Research and Advanced Technology for Digital Libraries, 466-469.
- d'Aquin, M., Adamou, A. and Dietze, S. (2013). Assessing the educational linked data landscape. Proceedings of the 5th Annual ACM Web Science Conference. New York, USA.
- Kaltenböck, M., Bauer, M. and Blumauer, M. (2016). Linked open data: The essentials, a quick start guide for decision makers, edition mono/monochrom. Vienna, Austria.
- Ngomo, A.C.N. et al. (2014). Introduction to linked data and its lifecycle on the Web. In Koubarakis M. et al. (eds), Reasoning web. Reasoning on the web in the big data era. Reasoning web 2014. Lecture Notes in Computer Science, 8714. Cham: Springer.
- Piedra et al. (2013). Using linked open data to improve the search of open educational resources for engineering students, 2013 IEEE Frontiers in Education Conference (FIE).
- Lê Trung Nghĩa (2018). Giáo dục mở ở Việt Nam: Thách thức và giải pháp. Tạp chí Tia sáng, số ra ngày 14/4/2018.
- Lê Trung Nghĩa (2019). Hai điều kiện tiên quyết cho dữ liệu mở. Tạp chí Tia sáng, 16, 14-17.
- Nguyễn Danh Minh Trí (2019). Vai trò và cơ hội của các thư viện trong việc lưu trữ và quản lý dữ liệu mở phục vụ cộng đồng. Tạp chí Thư viện Việt Nam, 1, 24-29.
- Nguyễn Danh Minh Trí (2017). Vai trò của tài nguyên giáo dục mở và truy cập mở trong việc nâng cao chất lượng giáo dục đại học Việt Nam. Tạp chí Thư viện Việt Nam, 1, 48-53.
- Nguyễn Danh Minh Trí (2017). Tổng quan về ảnh hưởng của tính mở trong giáo dục đại học. Tạp chí Thư viện Việt Nam, 4, 13-19 & 44.
- Nguyễn Danh Minh Trí (2018). Tổng quan về Khoa học dữ liệu. Tạp chí Thông tin và Tư liệu, 6, 16-23.
- Nguyễn Danh Minh Trí (2018). Phân tích các kỹ năng cốt lõi nhằm sử dụng và phát triển tài nguyên giáo dục mở. Tạp chí Thư viện Việt Nam, 4, 8-12.
- Nguyễn Danh Minh Trí (2018). Xây dựng các nguồn tài nguyên giáo dục mở dựa trên nền tảng Wikihow tiếng Việt. Tạp chí Thư viện Việt Nam, 2, 22-27.
- Nguyễn Thế Hùng (2019). Giá trị của Dữ liệu mở là gì, vì sao Chính phủ nên mở dữ liệu? Truy cập ngày 9/8/2020 từ <https://viettimes.vn/gia-tri-cua-du-lieu-mo-la-gi-vi-sao-chinh-phu-nen-mo-du-lieu-369019.html>.
- Trần Minh. (2017). Báo cáo tổng quan về dữ liệu mở, Viện Công nghiệp phần mềm và nội dung số Việt Nam. Bộ Thông tin và Truyền thông.
- West, P. and Victor, L. (2011). Background and action paper on OER. The William and Flora Hewlett Foundation.
- Wikipedia. Truy cập ngày 9/8/2020 từ: <https://www.wikipedia.org/>, [https://vi.wikipedia.org/wiki/Dữ\\_liệu\\_liên\\_kết](https://vi.wikipedia.org/wiki/Dữ_liệu_liên_kết), [https://en.wikipedia.org/wiki/Linked\\_data](https://en.wikipedia.org/wiki/Linked_data), [https://en.wikipedia.org/wiki/Open\\_data](https://en.wikipedia.org/wiki/Open_data).

(Ngày Tòa soạn nhận được bài: 15-7-2022;  
Ngày phản biện đánh giá: 10-9-2022; Ngày  
chấp nhận đăng: 15-11-2022).