

# SƠ LƯỢC VỀ OAI-PMH

(Theo “*Using the Open Archives Initiative Protocol for Metadata Harvesting*” của Timothy W. Cole và Muriel Foulonneau)

LẠI TRẦN THANH SƠN  
& VŨ NGUYỄN ANH  
Chuyên viên Thư viện số  
Phòng Công tác Kỹ thuật  
Thư viện ĐH Khoa học Tự nhiên  
Đại học Quốc gia TP. HCM

## Định nghĩa:

OAI-PMH hay *Sáng kiến lưu trữ mở cho giao thức gặt hái siêu dữ liệu – Open Archives Initiative - Protocol for Metadata Harvesting* được thiết kế nhằm giúp cho khả năng tương tác giữa các thư viện số tốt hơn và tạo thuận lợi cho việc phổ biến thông tin hiệu quả hơn. Đặc biệt, nó cung cấp các quy tắc và khung sườn cho việc chia sẻ các siêu dữ liệu mô tả, cả việc tạo siêu dữ liệu sẵn có và việc thu thập ghi chép siêu dữ liệu một khi chúng được tạo ra. Từ góc độ kỹ thuật, OAI-PMH được xem là tương đối đơn giản so với các giao thức khác cùng loại hoặc có liên quan.

Dù kỹ thuật không phức tạp, OAI-PMH đã chứng minh nó khá mạnh và hữu ích. Mặc dù các khái niệm kỹ thuật, cách mà giao thức sử dụng để thể hiện có thể phức tạp, nhưng khi đưa vào ứng dụng, nó lại khá dễ dàng đối với những người không thuộc lĩnh vực tin học. Giao

thức này dễ dàng trong việc đưa vào ứng dụng trong lĩnh vực lưu trữ thông tin.

## Lịch sử:

Tháng 7 năm 1999, Paul Ginsparg, Rick Luce, và Herbert Van de Sompel của Los Alamos National Laboratory (LANL), một thư viện ở bang New Mexico đã gửi thư mời về một buổi hội thảo về việc thúc đẩy công việc hướng tới “Dịch vụ phổ cập cho việc tự lưu trữ các tài liệu học thuật”. Buổi hội thảo đã diễn ra từ ngày 21 đến 22/10/1999 ở Santa Fe, New Mexico.

Kết quả của buổi hội thảo là sự ra đời của tổ chức Open Archives Initiative (OAI) và đồng thời cũng ban hành một khung tổng quát và các thiết lập về mặt kỹ thuật được biết đến với tên gọi “Hiệp ước Santa Fe”.

Buổi hội thảo thứ hai của OAI được tổ chức vào ngày 3 tháng 6 năm 2000 ở

SanAntonio, Texas, nhằm bổ sung và thúc đẩy các mục tiêu đã đề ra trong hội nghị đã diễn ra tại Santa Fe. Hiệp ước đã tận dụng các kỹ thuật và phương pháp được sử dụng trong nguyên mẫu của Universal Preprint Service (UPS), được phát triển bởi Herbert Van de Sompel và các cộng sự và mùa hè năm 1999. Những thành công ban đầu của các mô hình UPS đã thúc đẩy việc đưa ra một giao thức mới để cập nhật và chia sẻ dữ liệu cũng như tương tác giữa các thư viện số. Mô hình UPS đã thành công trong việc chứng minh sự hiệu quả trong việc tìm kiếm qua kho lưu trữ và thông tin tài nguyên trên một số kho bản in điện tử (e print). Trong các mô hình UPS và hiệp ước Santa Fe đã giả định khá cụ thể về các chức năng nội bộ và cách thức tổ chức của nhà cung cấp dữ liệu và nhà cung cấp dịch vụ. Trong khi những nghiên cứu đưa ra về mong muốn cho việc gặt hái và chia sẻ dữ liệu. Hiệp ước Santa Fe cũng được ràng buộc chặt chẽ theo tên của các tập hợp con của giao thức. Một điều rõ ràng là việc gặt hái và chia sẻ tài liệu đã vượt qua ngoài tầm của nhóm nghiên cứu ban đầu, hiệp ước Santa Fe cho thấy sự cần thiết của việc tổng quát hóa, đưa ra mô hình về tổ chức cụ thể hơn.

Điều này dẫn đến sự thành lập ban chỉ đạo của OAI vào giữa năm 2000. Ủy ban về kiểm tra, kỹ thuật để cải thiện hiệp ước Santa Fe cũng được thành lập trong thời gian này.

Cuộc họp đầu tiên của tổ chức OAI-PMH được tổ chức vào ngày 7 - 8/09/2000, tại đại học Cornell. Hội nghị đã đưa ra các quy trình chi tiết về việc

đưa dữ liệu lên và tải dữ liệu xuống trong việc gặt hái siêu dữ liệu, các mô hình trước đây cũng được hoàn thiện hơn.

Một hội thảo khác của OAI cũng được tổ chức ở Châu Âu từ ngày 18 – 20/09/2000, kết hợp với hội nghị thư viện số ở Lisbon, Bồ Đào Nha. Một cuộc gặp gỡ ngắn của Ủy ban kỹ thuật cũng được diễn ra vào ngày 25/11/2000, tại CERN, ở Geneva, Switzerland.

Đầu tháng 11/2000, một bản thử nghiệm của OAI-PMH được thử nghiệm bởi Ủy ban kỹ thuật OAI trong nhóm thử nghiệm của liên hiệp. Để so sánh với hiệp ước Santa Fe, bản thử nghiệm alpha của OAI-PMH đã vượt qua được các tiêu chuẩn của Eprint và tập trung nhiều hơn cho việc truyền tải các siêu dữ liệu từ nhà cung cấp dữ liệu đến nhà cung cấp dịch vụ. Trong mô hình mới này, thì chuẩn XML cũng được sử dụng chính thức, xác định ngữ nghĩa và cấu trúc của OAI-PMH. Một vấn đề được chú ý là việc giữ cho kỹ thuật của giao thức phải đơn giản, đặc biệt là đối với nhà cung cấp siêu dữ liệu cho OAI. Một lưu ý là sự thay đổi từ yêu cầu ban đầu là các nhà cung cấp dữ liệu phải hỗ trợ chuẩn eprint giao thức cụ thể. Thay vào đó, nhà cung cấp dữ liệu được yêu cầu là cung cấp các siêu dữ liệu theo chuẩn đơn giản là Dublin Core (DC).

Nhu cầu sử dụng DC như là “một ngôn ngữ chung” hay một chuẩn siêu dữ liệu cấp thấp vẫn là một vấn đề đang được tranh cãi cho tới ngày nay. Trong khi Ủy ban điều hành, Ủy ban kỹ thuật của OAI đều đồng tình rằng nhu cầu sử dụng DC, một dạng thức của siêu dữ

liệu, sẽ trở nên phổ biến rộng rãi hơn so với dạng thức Dienst hay E print, đa số đều cho rằng tất cả các nguồn tài nguyên đều có thể mô tả đầy đủ bằng định dạng DC đơn giản.

Bản thử nghiệm của OAI-PMH được thử nghiệm trong tháng 11 tới 12/2000. Ngày 1/11/2000, giao thức đã được cập nhật nhiều lần trong thời gian chạy thử nghiệm để giải quyết các vấn đề phát sinh.

Được sự đề nghị của OAI và Ủy ban kỹ thuật của OAI trong năm 2000, sự phát triển của OAI được quan tâm từ khi bắt đầu đến khi đưa ra quốc tế. Ngay cả khi OAI đang trong quá trình thử nghiệm, thì các kế hoạch đưa OAI ra ngoài hoạt động đã được tiến hành ở Hoa Kỳ và Châu Âu. Buổi giới thiệu trang trọng OAI-PMH được tổ chức tại Washington DC vào ngày 23/01/2001, được gọi là "OAI OPEN DAY" của Hoa Kỳ. "OAI OPEN DAY" của Châu Âu tổ chức vào 26/02/2001 tại Berlin, Đức. Cả hai sự kiện này đều giới thiệu về phiên bản 1.0 của OAI. Những người tham dự các sự kiện trên được thử nghiệm phiên bản 1.0 cũng như phản hồi ý kiến của mình về phiên bản này của OAI. Ủy ban kỹ thuật của OAI thông qua các người thử nghiệm đã nhận thấy cần phải cải tiến phiên bản 1.0 để làm cho giao thức trở nên mạnh mẽ và ổn định hơn cho các ứng dụng thực tiễn ở quy mô lớn. Vào ngày 2 tháng 6/2001, một sửa đổi nhỏ trong giao thức đã cho ra đời bản 1.1, được thực hiện để giữ cho giao thức đồng bộ với thay đổi trong những đặc điểm của chuẩn W3C XML. Ngoài bản

1.1 ra, sau này không còn thêm bản 1.x nào được thực hiện nữa.

Trong năm 2001, ủy ban kỹ thuật đã phân tích các vấn đề đã được đưa ra bởi những người dùng thử bản beta, hầu hết là về các vấn đề về phạm vi. Đầu năm 2002, Ủy ban kỹ thuật đã biên soạn 1 danh sách gồm 18 tiêu chuẩn kỹ thuật. Mỗi nhóm nhỏ của Ủy ban kỹ thuật viết một bản tóm tắt, mô tả từng tiêu chuẩn kỹ thuật trên, các vấn đề này được thảo luận qua các cuộc hội thảo qua điện thoại. Mọi nỗ lực đều được sử dụng để đáp ứng mong muốn của các nhà thử nghiệm bản beta.

Những thay đổi từ bản 1.1 lên bản 2.0 phản ánh sự ảnh hưởng của những đặc điểm kỹ thuật. Ngày 14/06/2002, phiên bản 2.0 của OAI-PMH đã được phát hành, đây có thể coi là một bản ổn định của OAI, có thể xây dựng các ứng dụng quan trọng trên những môi trường khác nhau. Đến thời gian viết bài này (2010) thì bản 2.0 vẫn là bản mới nhất.

Những thay đổi cho bản 1.1 dẫn đến việc ra đời bản 2.0, phản ánh nỗ lực để làm sáng tỏ các chi tiết đặc điểm kỹ thuật, sửa các lỗi nhỏ, và thêm vài chức năng nâng cao. Ngoài việc cung cấp những hướng dẫn hữu ích nó còn cho thấy các nguyên tắc phụ trợ và đặc điểm kỹ thuật cho mục đích sử dụng đặc biệt của OAI-PMH. Phiên bản 2.0 không có những thay đổi lớn về mặt cấu trúc, thiết kế cơ bản, nhưng nó lại không thể tương thích ngược được. Bằng chứng, là sau 4 năm ra đời, một số dữ liệu cũ của nhà cung cấp vẫn chưa được truyền tải. Kết quả là, một số nhà cung cấp dịch vụ buộc phải

quay lại sử dụng phiên bản 1.1 để gỡ hái dữ liệu từ nhà cung cấp dữ liệu.

### **Các đặc điểm của OAI-PMH:**

#### **Phạm vi:**

OAI-PMH làm việc với dữ liệu cấu trúc, cụ thể với các dữ liệu thể hiện bằng cách sử dụng ngôn ngữ XML (Extensible Markup Language). Người dùng cấp cao của giao thức đang bắt đầu thử nghiệm với những cách để mở rộng và phát triển phạm vi này bao gồm các lớp khác nhau của siêu dữ liệu và thậm chí cả nội dung đầy đủ, nhưng các nhà phát triển của OAI-PMH đã tập trung vào siêu dữ liệu mô tả khi chúng tạo ra các giao thức. Phạm vi OAI-PMH cũng được xác định tập trung ngầm vào siêu dữ liệu mà nó mô tả rời rạc, kỹ thuật số. OAI-PMH có thể được sử dụng cho các siêu dữ liệu mô tả các loại tài nguyên thông tin.

Đối với các mục tiêu của đề tài này, xác định mục đích chung của siêu dữ liệu là "dữ liệu về dữ liệu" là một điểm khởi đầu tốt. Tuy nhiên, định nghĩa này vẫn còn bao gồm nhiều, đôi khi trùng lặp các lớp của siêu dữ liệu, Ví dụ, siêu dữ liệu kỹ thuật, siêu dữ liệu cấu trúc, siêu dữ liệu bảo quản, siêu dữ liệu mô tả, siêu dữ liệu hành chính, và siêu dữ liệu quyền... Tất cả như là một lớp con của siêu dữ liệu nói chung, sự gỡ hái các bản ghi siêu dữ liệu mô tả là một sự sưu tầm của các nguồn tài nguyên thông tin số giống như các phiếu biên mục sách tại các thư viện địa phương. Bản ghi một siêu dữ liệu mô

tả tóm tắt những thuộc tính của nguồn tài nguyên thông tin số, hỗ trợ cho việc phát hiện, định vị, phân loại, phân nhóm, liên quan, phiên dịch, và xác định nguồn tài nguyên. Các thuộc tính điển hình chứa trong một bản ghi siêu dữ liệu mô tả bao gồm các thuộc tính quen thuộc như nhan đề, tên tác giả, nhà xuất bản, ngày xuất bản, và các tiêu đề đề mục. Các lĩnh vực mô tả vật lý của việc ghi mục lục phiếu được tăng cường đối với mô tả nội dung số với các thuộc tính như định dạng đối tượng số, mức độ (ví dụ, kích thước tập tin), và loại tài nguyên. Một kí hiệu xếp giá của thư viện có thể được thay thế bằng một URL (Universal Resource Locator, địa chỉ web của các tài nguyên kỹ thuật số) hay một số nhận diện thích hợp khác một cách tương tự như DOI (Digital Object Identifier). Bất kì thuộc tính của một nguồn tài nguyên số mà nó có thể giúp người dùng tìm thấy một nguồn tài nguyên và xác định xem liệu tài nguyên đó có thể hữu ích để đáp ứng nhu cầu thông tin trực tiếp của mình thì có thể xem như là siêu dữ liệu mô tả. Siêu dữ liệu mô tả được tìm hiểu trong các lĩnh vực của thư viện số là một thuật ngữ mở rộng, nhưng nó tương đối dễ dàng cho cán bộ thư viện được đào tạo trong biên mục truyền thống để hiểu và đánh giá.

OAI-PMH đã được tạo ra đặc biệt để cho phép việc phục vụ phát triển qua những nguồn phân phối riêng biệt, các đối tượng thông tin tương đối độc lập ( đa dạng về kích thước, thể loại) và tương đối phức tạp. Điều

này bao gồm các tài nguyên thông tin như các bài báo trực tuyến, hình ảnh số hóa, video clip, các đoạn âm thanh và các trang web cá nhân.

### **Mục tiêu:**

Mục đích chính của giao thức là truyền tải dữ liệu từ vị trí này sang vị trí khác trong không gian thông tin ảo World Wide Web, nhằm để chia sẻ và tập hợp các thông tin hữu ích. OAI-PMH chia sẻ thông tin trực tuyến với các nhà cung cấp dữ liệu OAI và các nhà cung cấp dịch vụ. Trong thuật ngữ của mạng máy tính, thì OAI-PMH là một ví dụ về kiến trúc đơn giản của Client – server.

Trong 6 năm đầu hoạt động , OAI-PMH thông thường được sử dụng là để tổng hợp nhiều nguồn tài nguyên lại với nhau trong một dịch vụ tìm kiếm và cung cấp truy cập thông tin tích hợp.

Trong thời đại “Mua sắm tại chỗ” hiện nay, nhu cầu tích hợp các nguồn tài nguyên thông tin vào một chỗ theo mô hình OAI-PMH là điều cần thiết. Tìm kiếm giữa các kho “siêu dữ liệu” được thực hiện trên các hệ thống được thiết kế tối ưu hóa cho từng mục đích riêng biệt phục vụ nhu cầu của độc giả. Các tập hợp siêu dữ liệu có thể được phân tích để tìm ra mối quan hệ giữa các nội dung của các siêu dữ liệu với nhau được tổ chức bởi các nhà cung cấp.

OAI cho phép cộng tác giữa các nhà cung cấp dữ liệu và cung cấp dịch vụ. Các dịch vụ được cung cấp

bởi các nhà cung cấp dịch vụ bị phụ thuộc vào chất lượng của các dữ liệu được thu vào. Những lợi ích mà nhà cung cấp dữ liệu mang lại lại phụ thuộc vào chất lượng dịch vụ đang được cung cấp. Sự hợp tác có thể công khai hay bí mật, những nhà cung cấp dữ liệu và dịch vụ có thể bàn bạc với nhau, tham khảo để thống nhất về chuẩn hóa và các dịch vụ kèm theo.

### **OAI-PMH- Cái gì không thuộc về nó**

Vẫn còn một vài quan niệm sai lầm thường gặp như là OAI-PMH là gì và nó được sử dụng cho mục đích gì. Những quan niệm sai một phần bởi vì theo tên của nó và một phần bởi vì sự liên kết chặt chẽ của nó với một số các công nghệ mà nó sử dụng. Một cách để giúp làm rõ một số quan niệm sai lầm này là định nghĩa OAI-PMH bằng cách nói rõ cái gì không phải thuộc về nó.

Đầu tiên, mặc dù nó là một sản phẩm của một nỗ lực cộng đồng nhằm nâng cao lợi ích của công việc xuất bản học thuật tự lưu trữ, OAI-PMH vốn không phải là một ứng dụng truy cập mở. OAI-PMH cũng cung cấp ngữ cảnh bổ sung cho OAI-PMH bằng cách mô tả mối quan hệ giao thức với những kho lưu trữ ePrint, các kho cơ sở, và các tạp chí truy cập mở. Người đọc bây giờ nên lưu ý rằng giao thức chỉ hữu ích cho việc chia sẻ siêu dữ liệu mô tả một loạt nội dung trong nhiều loại định dạng, OAI-PMH có thể được và đã được sử dụng để chia sẻ, tổng hợp siêu dữ liệu mô tả truy cập mở và hạn chế truy cập nội dung có phí (chẳng hạn như tác phẩm văn chương

được xuất bản vì mục đích thương mại trên các bài báo của tạp chí). Nhận thức được khả năng sử dụng rộng rãi của OAI-PMH, vào cuối năm 2003 Ủy ban Bản Quyền OAI (OAI Rights Committee) được thành lập nhằm đưa ra những nguyên tắc tư vấn cho những người thực hiện OAI-PMH về cách tốt nhất để liên kết những quyền sở hữu trí tuệ với các bản ghi siêu dữ liệu được những nhà cung cấp dữ liệu OAI-PMH phổ biến. Đến năm 2006, đây vẫn là một lĩnh vực hấp dẫn cho những người thực hiện OAI-PMH.

Dù theo tên gọi của nó, OAI-PMH không phải là một tiêu chuẩn hoặc một đặc tả kỹ thuật cho lưu trữ thực tế theo nghĩa học thuật. "Lưu trữ" được sử dụng trong tên OAI-PMH cần được hiểu trong ngữ cảnh khái quát của nó, không phải có nghĩa hạn chế hơn, được giả định khi thảo luận những chức năng lưu trữ của các tổ chức trường đại học, chính phủ, và di sản văn hóa. Mặc dù đã có và vẫn còn một số nhầm lẫn trong cộng đồng thư viện kỹ thuật số (Flirtle 2001), OAI-PMH là hoàn toàn khác và tách biệt với mô hình tham chiếu (Reference Model) cho một hệ thống thông tin lưu trữ mở (Open Archival Information System, OMS), nó thực sự là một mô hình thực tiễn tốt nhất về việc xây dựng các phiên bản kỹ thuật số cho những kho lưu trữ truyền thống. OAI-PMH không cụ thể về hình thức những hồ sơ lưu trữ ở tất cả các trường hợp. Thay vào đó, việc sử dụng những tài liệu lưu trữ dạng lệnh trong OAI-PMH xuất phát từ hoạt động tham chiếu của cộng đồng ePrint đến các kho của ePrint như là những kho lưu trữ trực tuyến.

OAI-PMH cũng không đồng nghĩa với định dạng siêu dữ liệu đơn giản DC hay Dublin Core Metadata. Mặc dù OAI-PMH tham khảo tới định dạng siêu dữ liệu DC đơn giản và đòi hỏi những chấp nhận giao thức với việc sử dụng định dạng mẫu, DC không phải là một phần của giao thức, cũng không phải là một phần OAI-PMH của DCMI. Những định dạng siêu dữ liệu khác vẫn có thể được sử dụng với OAI-PMH. Định dạng văn bản của giao thức thực sự khuyến khích những người thực hiện sử dụng những định dạng siêu dữ liệu khác. Giao thức được viết sao cho thay đổi đặc điểm kỹ thuật định dạng DC đơn giản được thực hiện bởi DCMI thường không yêu cầu thay đổi OAI-PMH (và ngược lại). Về phần OAI-PMH, các hồ sơ metadata xuất và thu bằng cách sử dụng giao thức được xác nhận bằng cách sử dụng các đặc điểm kỹ thuật toàn bộ bên ngoài OAI-PMH. Phần lớn OAI-PMH chỉ đòi hỏi các bản ghi metadata được định nghĩa như có một định dạng metadata cụ thể có thể được xác nhận bằng cách sử dụng các tài liệu ngôn ngữ XML Schema được tham khảo rõ ràng và từ bên ngoài.

Cuối cùng, OAI-PMH thường được sử dụng để giúp kích hoạt chéo nhằm tìm kiếm các nguồn tài nguyên thông tin, bản thân nó không phải là một giao thức để tìm kiếm. Không giống như Z39.50 và giao thức SRU liên quan (Tìm kiếm / Lấy thông qua URL), OAI-PMH không được thiết kế để hỗ trợ công việc tìm kiếm thời gian thực tùy ý và linh động. OAI-PMH cũng không đồng nghĩa với công nghệ "mạng nhện" được sử dụng bởi những công cụ tìm kiếm web như Google. Thay vì dựa vào việc khai thác

siêu dữ liệu mô tả, công nghệ “mạng nhện” dựa vào các liên kết nhúng để di chuyển từ trang web này đến trang web khác, tập hợp các đối tượng có nội dung toàn văn để trích lọc sau khi chúng lướt qua. Cách tiếp cận này không tận dụng lợi thế của các thông tin bên ngoài mà thường ở bên trong trong siêu dữ liệu mô tả và có thể bỏ sót nội dung không truy cập tự do hay không rõ ràng và riêng biệt được liên kết từ các trang web. Ví dụ, các tài nguyên được bảo quản trong một cơ sở dữ liệu và thường chỉ được phát hiện ra thông qua sự tương tác với cơ sở dữ liệu đó. Đây thường được gọi là Web "ẩn". (Google gần đây đã bắt đầu chấp nhận các URL cung cấp dữ liệu OAI thay cho những bản đồ trang Web thông thường.)

Sự khác biệt giữa Z39.50/SRU và OAI-PMH làm nổi bật lên một số khác biệt cơ bản cho công cụ tìm kiếm được xây dựng bằng cách sử dụng hai phương pháp tiếp cận.

Trong khi cả hai mô hình Z39.50 và OAI-PMH liên quan chủ yếu với siêu dữ liệu mô tả định rõ tính chất DLOs, và cả hai giả định rằng những nguồn tài nguyên thông tin kỹ thuật số chủ yếu vẫn thuộc quyền kiểm soát và quản lý bởi các nhà cung cấp phân phối dữ liệu rộng, những hệ lụy kéo theo là làm thế nào một người dùng ở đầu cuối phát hiện ra sự khác nhau giữa các nguồn tài nguyên theo các phương pháp xử lý quan trọng. Trong Z39.50 hoặc SRU, truy vấn tìm kiếm của người dùng đầu cuối được phát tán rộng rãi đồng thời bằng công cụ tìm kiếm đến những nhà cung cấp đa dữ liệu liên quan. Mỗi lần tìm kiếm được thực

hiện song song dựa vào siêu dữ liệu mới nhất và cập nhật mà mỗi nhà cung cấp dữ liệu đã có sẵn. Điều này có nghĩa rằng công việc tính toán liên kết chức năng tự tìm kiếm của nó được phân phối trong tất cả các nhà cung cấp dữ liệu hơn là trách nhiệm duy nhất của một dịch vụ trung tâm. Các công cụ thông tin Z39.50 hoặc SRU chỉ có trách nhiệm cho việc chuyển đổi và phát tán lệnh tìm kiếm của người dùng theo một ngôn ngữ truy vấn chuẩn và để tổng hợp các kết quả tìm kiếm trả về bởi các nhà cung cấp dữ liệu tham gia.

Mặt khác, bởi vì tìm kiếm phải được thực hiện bởi mỗi nhà cung cấp dữ liệu tham gia trong thời gian thực, bất kỳ sự gián đoạn hoặc chậm trễ ở bất kỳ nhà cung cấp dữ liệu nào, hoặc trong mạng lưới thông tin liên lạc với bất kỳ trang web cung cấp dữ liệu, dẫn đến kết quả thu được không đầy đủ hoặc chậm trễ cho người dùng cuối. Mỗi nhà cung cấp dữ liệu phải thực hiện giao thức Z39.50 hoặc SRU khó khăn hơn và đòi hỏi khắt khe hơn về mặt kỹ thuật. Mỗi nhà cung cấp dữ liệu cũng được đưa ra tính linh hoạt đáng kể trong việc làm thế nào để thực hiện các tính năng tìm kiếm được mô tả trong Z39.50 hoặc tiêu chuẩn SRU. Tiêu chuẩn hóa tập trung trước khi phối hợp hoặc mở rộng của siêu dữ liệu cho sự khám phá tăng cường không thể được thực hiện. Tất cả những yếu tố này mô tả sự thỏa hiệp rằng phải chú ý đến các nhà cung cấp dữ liệu và cả cung cấp dịch vụ. Tuy nhiên, người đọc cần lưu ý rằng các phương pháp phát tán (Z39.50, SRU) và thu nhận (OAI-PMH) hướng đến sự khám phá tài nguyên chéo giữa các kho, mặc dù các phương pháp khác

nhau, không phải là không thể dung hoà. Hoàn toàn khả thi và hợp lệ để tạo ra một dịch vụ tìm kiếm chéo giữa các kho, để tìm kiếm một tập cục bộ của siêu dữ liệu được tổng hợp thu nhặt từ một tập các nhà cung cấp dữ liệu và ở cùng thời điểm phát tán những truy vấn tìm kiếm đến một tập thứ hai của các nhà cung cấp dữ liệu. Thực hiện đúng, phương pháp này có thể tận dụng hoàn toàn những lợi ích của cả hai phương pháp (Sanderson và cộng sự năm 2005).

### **Kết luận**

Giao thức OAI-PMH từ khi ra đời đến nay chứng tỏ được những ưu thế, tiện lợi của mình so với các giao thức khác, mang đến sự tiện dụng cho người sử dụng cuối cũng như các nhà cung cấp dịch vụ và dữ liệu. OAI-PMH hứa hẹn đem đến một sự phát triển vượt bậc cho

ngành thông tin thư viện trong thời đại web 2.0 hiện nay. Đặc biệt OAI-PMH sẽ hỗ trợ rất nhiều cho việc phát triển các thư viện số, kết nối các kho dữ liệu của các thư viện lại với nhau, tạo nên nguồn thông tin dồi dào, hữu ích cho độc giả sử dụng.

Qua bài viết (được tóm lược từ phần mở đầu của “*Using the Open Archives Initiative Protocol for Metadata Harvesting*” của tác giả Timothy W.Cole và Muriel Foulonneau), chúng tôi hi vọng đã cung cấp cho các bạn những thông tin, hiểu biết cơ bản nhất về OAI-PMH. Để chúng ta có thể dễ dàng hình dung được cách ứng dụng OAI-PMH vào hệ thống của chúng ta, để cùng nhau đưa ngành thông tin thư viện nước nhà lên một tầm cao mới, vươn ra hội nhập với thế giới.

## **THƯ MỤC**

1. COLE, TIMOTHY W. and FOULONNEAU MURIEL. – *Using the Open Archives Initiative Protocol for Metadata Harvesting*. – Westport, Connecticut : Libraries Unlimited, 2007.
2. LESK, MICHAEL. – *Practical Digital Libraries: Books, Bytes, and Bucks*. – San Francisco, California : Morgan Kaufmann, 2005.
3. RHYNO, ART. – *Using Open Source Systems for Digital Libraries*. – Westport, Connecticut : Libraries Unlimited, 2004.
4. WITTEN, IAN H. and BAINBRIDGE, DAVID. – *How to Build a Digital Library*. – New York : Morgan Kaufmann, 2003.