

HIỂU SIÊU DỮ LIỆU VÀ MỤC ĐÍCH CỦA NÓ

KAREN COYLE

“Siêu dữ liệu là biên mục được thực hiện bởi con người”¹

Thế giới công nghệ thông tin đang bàn luận ngày càng nhiều về siêu dữ liệu. Tất cả mọi người ngày hôm nay dường như đang tạo ra một định dạng siêu dữ liệu nào đó. Thường có một thẻ <meta> trong một tài liệu HTML để chứa đựng siêu dữ liệu cho các nguồn tài nguyên Internet; các nhà khoa học đã phát triển một siêu dữ liệu để mô tả những bộ gen; các nhà xuất bản có một định dạng siêu dữ liệu để tạo điều kiện cho việc tiếp thị sản phẩm, dịch vụ thông tin và dữ liệu giá tới người bán lẻ. Vậy, điều gì sẽ xảy ra trong một thế giới công nghệ hiện đang dẫn dắt tất cả mọi người tin rằng siêu dữ liệu là câu trả lời? Nếu một sự lựa chọn cho rằng siêu dữ liệu sẽ là câu trả lời, thì câu hỏi đó sẽ là gì và nó có ý nghĩa như thế nào đối với thư viện và những mục lục thư viện (library catalogs)?

ĐỊNH NGHĨA SIÊU DỮ LIỆU (METADATA)

Trước tiên chúng ta hãy định nghĩa siêu dữ liệu là gì? Một định nghĩa chung nhất cho siêu dữ liệu đó là “*dữ liệu về dữ liệu*”. Định nghĩa này thoạt nghe sẽ cảm thấy cuốn hút, nhưng nó không giúp chúng ta hiểu toàn bộ siêu dữ liệu là gì. Những gì giải thích sau đây ít cuốn hút hơn, nhưng nó đưa đến một cách hiểu tốt

hơn. Với cách này, thì siêu dữ liệu là thông tin được cấu trúc, điều này có nghĩa là nó thuộc phạm trù sáng tạo của con người, và chúng ta không tìm thấy nó trong tự nhiên. Một minh họa tốt cho thông tin được cấu trúc là việc sử dụng kinh độ và vĩ độ để mô tả trái đất và các điểm trên đó. Trái đất trong thực tế hiển nhiên không có những đường kinh tuyến vòng quanh nó, mặc dù hiện nay chúng ta đã quen thuộc nhìn thấy bản đồ và quả địa cầu với những đường kinh tuyến đó, song sự sáng tạo ra kinh độ và vĩ độ cho phép chúng ta nói về các vị trí trên hành tinh này và di chuyển chính xác qua những khoảng cách rộng lớn mà không có một cột mốc nào chỉ dẫn cho chúng ta cả. Minh họa đó dẫn chúng ta đến một đặc tính cần thiết thứ hai của siêu dữ liệu: siêu dữ liệu được con người phát triển cho một mục đích hoặc một chức năng nào đó. Bởi vậy, một bản đồ hệ thống tàu điện ngầm dùng cho người lái tàu với các tuyến đường và điểm chuyển tiếp được mã hóa màu sắc. Bản đồ này thường chỉ vừa đủ đại diện cho tỷ lệ và địa lý thực tế của một thành phố mà tuyến xe điện ngầm đó phục vụ, song nó thực sự hữu ích bởi vì nó đưa ra một bức tranh tập trung của hệ thống tàu điện ngầm, mặc dù nó không quan tâm đến sự chính xác về mặt địa lý. Một bản đồ đường bộ của một khu vực tương tự sẽ thực tế hơn về mặt địa lý, song nếu bản đồ đó do một ban quản lý về

du lịch thiết kế thì nó thường làm nổi bật lên vị trí các khách sạn, bảo tàng, điểm di tích và bãi đỗ xe. Một bản đồ của một khu vực do một câu lạc bộ đi bộ sử dụng thường tập trung vào xây dựng những hình học và mốc chỉ ra thắng cảnh tự nhiên. Rõ ràng rằng không có một bản đồ duy nhất nào phục vụ mọi nhu cầu, bởi vậy không có một loại siêu dữ liệu nào cho các tài liệu hoặc các đối tượng thông tin khác nhau. Tại sao lại như vậy, chính bởi vì không phải chỉ chính đối tượng thông tin quyết định siêu dữ liệu mà còn cả nhu cầu và mục đích của người tạo ra nó cũng như đối tượng mà nó hướng đến. Nếu không quá siêu hình thì siêu dữ liệu không phải là một thế giới, nó là cách chúng ta nhìn thế giới như thế nào vào một thời điểm vì một mục đích nào đó.

Siêu dữ liệu cũng thường được sử dụng như là dữ liệu đại diện cho một vật thực tế. Trong một mục lục thư viện, các tiêu đề (headings) là thông tin đại diện cho những quyển sách trên giá sách. Trong khi người sử dụng thư viện thường khó khăn xem mỗi sách để quyết định họ cần quyển sách nào thì ít ra quyển sách cụ thể vẫn ở trên giá để họ có thể xem qua. Trong môi trường số, vai trò đại diện của siêu dữ liệu là rất quan trọng bởi vì nhiều nguồn tài nguyên không dễ dàng có thể xem toàn văn và nhiều nguồn không chứa đựng dữ liệu rõ ràng mô tả về chính chúng. Sự quan tâm ngày càng tăng đến siêu dữ liệu là một phần trong nỗ lực tổ chức lại thế giới hỗn loạn của các nguồn tài nguyên số và cung cấp sự truy cập và dịch vụ hoàn hảo hơn mà nó chưa từng có trước kia. Nó cũng là một cách trao đổi dữ liệu giữa nơi lưu trữ tài nguyên khác hẳn

nhau, đồng thời cho phép tìm kiếm rộng khắp trên nhiều kho dữ liệu số.

XML và RDF

Hai từ viết tắt mà người ta thường nghe và được sử dụng đồng thời khi họ bàn luận về siêu dữ liệu là **XML** và **RDF**. **XML** là từ viết tắt của *eXtensible Markup Language*² và **RDF** là viết tắt là *Resource Description Framework*³. Một số người nói về XML và RDF dường như chúng chính là những định dạng siêu dữ liệu, nhưng điều này là một sự nhầm lẫn giữa khổ mẫu (form) và nội dung. Cả XML và RDF thực tế là những định dạng dữ liệu chung mà người ta có thể dùng cho nhiều ứng dụng. Cụ thể, XML thường được sử dụng như là định dạng tài liệu và là một định dạng mở rộng của HTML. Nếu một nhân viên thư viện chưa quen thuộc với cấu trúc biểu ghi XML thì nó trông có vẻ khá phức tạp và bí ẩn. Thực tế thì định dạng cơ bản của nó khá đơn giản mặc dù người ta có thể dùng nó để tạo ra những biểu ghi dữ liệu phức tạp. Nếu một thủ thư biết về một biểu ghi MARC như là có các trường dữ liệu được mô tả bởi các thẻ trường (tags), ví dụ như sử dụng thẻ “245” ám chỉ cho thẻ “nhân đề” - “title”:

245\$a Hamlet, Prince of Denmark

sau đó XML chỉ là một cách khác tạo ra thẻ nhằm gắn phân tử dữ liệu vào trong đó, mặc dù với tài liệu XML nhất thiết phải đưa ra thẻ bắt đầu và thẻ kết thúc (với một dấu “/” trước tên thẻ đó) bao lấy mỗi phân tử dữ liệu:

```
<title>Hamlet, Prince of Denmark</title>
```

Những thẻ này có thể là bất kỳ cái gì mà chúng ta muốn chúng như vậy, miễn là chúng cần được định nghĩa trước trong một cấu trúc định nghĩa định dạng dữ liệu. Bởi vậy, nếu muốn, người ta có thể định nghĩa bằng bất kỳ kí tự nào cho một thẻ nhan đề:

```
<245>Hamlet, Prince of Denmark</245>
```

```
<ti>Hamlet, Prince of Denmark</ti>
```

XML, cũng như thẻ MARC và trường con, cần thiết phải là một cấu trúc có thứ bậc. Lợi thế đối với MARC21 đó là nó có nhiều mức phân cấp tới mức cần thiết chứ không phải chỉ hai cấp độ của thẻ và trường con. Trong tài liệu XML, sự phân cấp này được lồng vào nhau như “búp bê Nga” tới mức cần thiết.

Định dạng RDF là một hoặc hai bước tiến vượt xa hơn XML. RDF nhấn mạnh vào *mối quan hệ* giữa các phần tử dữ liệu. Một mối quan hệ chính trong RDF đó là “về” (“about”). Ở đâu một nguồn tài nguyên Web là đối tượng của tài liệu RDF, và các trường khác trong tài liệu đó là sự mô tả về nguồn tài nguyên đó. Đó là một cách hiểu đơn giản nhất. RDF cũng có thể tận dụng các mối quan hệ như:

```
subClassOf
subPropertyOf
member
isDefinedby
```

và những mối quan hệ khác. RDF là một cấu thành cần thiết của một nỗ lực mà người ta gọi là “Web ngữ nghĩa”⁴ (“Semantic Web”), một cố gắng của tổ chức 3WC (World Wide Web

Consortium) nhằm bổ sung một cấu thành có ngữ nghĩa vào việc chia sẻ dữ liệu qua Internet. RDF phức tạp hơn và ít được sử dụng hơn XML, và người ta vẫn chưa rõ ràng liệu nó có kế thừa như là một ngôn ngữ chung để mô tả thế giới Web không. Dường như rõ ràng rằng nó đòi hỏi chúng ta phải hiểu sâu những khái niệm có tính triết học nhất định hơn là đối với XML và số người thấy nó vốn đã khó hiểu (và tôi ở trong nhóm đó) lớn hơn nhiều số người cho nó là một giải pháp. (Ví dụ dưới đây của một biểu ghi sử dụng một định dạng đơn giản của RDF).

SIÊU DỮ LIỆU CHO ĐỐI TƯỢNG GIỐNG TÀI LIỆU

Như là các thủ thư, chúng ta sẽ làm việc với siêu dữ liệu mô tả tài liệu và các đối tượng giống tài liệu (Document-like objects) mặc dù chúng ta thấy công việc của mình chính là lưu trữ, tổ chức, và cung cấp dịch vụ liên quan đến nhiều loại siêu dữ liệu khác nhau, ví dụ như siêu dữ liệu khoa học. Song, đối với bài báo này, tôi sẽ tập trung vào siêu dữ liệu mô tả tài liệu, và xoay quanh một vấn đề chính là siêu dữ liệu này sẽ khác với mục lục của thư viện như thế nào? Chúng ta có thể ghi nhận rằng các định dạng siêu dữ liệu giới thiệu trong bài báo này (Dublin Core, MODS và METS) chỉ là ba trong số nhiều định dạng đang được sử dụng hiện nay, song ba định dạng này là ba tiêu chuẩn được sử dụng thường xuyên nhất trong các thư viện số.

Biên mục thư viện rõ ràng là điều kiện tuyệt đối cần thiết của mô tả siêu dữ liệu tài liệu. Nó có thể truy tìm nguồn gốc

về tận giữa những năm 1800 với sự thống trị của Jewett và Panizzi. Nó đã quen thuộc, thậm chí với một người có chút kiến thức trong thế giới nói tiếng Anh (Anglo-American). Sổ mục lục thư viện hiện có vừa đủ để áp đảo bất cứ lược đồ siêu dữ liệu nào khác sẽ được sử dụng cho sách (mặc dù không thể dùng cho bài báo tạp chí). Bởi vậy, khi người phát triển trong các ứng dụng Internet cần siêu dữ liệu cho các tài liệu truy cập trực tuyến, họ đã không chấp nhận một tiêu chuẩn sẵn có của thư viện. Trong thực tế, một tiêu chuẩn siêu dữ liệu tài liệu thường thấy nhất trong những ứng dụng không phải thư viện đó là Dublin Core. Để hiểu được nguyên nhân tại sao lại như vậy, chúng ta cần xem xét đến mục đích của nó.

Dublin Core

Bởi vì từ “Dublin” trong “Dublin Core” ám chỉ cho thành phố ở bang Ohio, trụ sở của OCLC và bởi vì OCLC là tổ chức hỗ trợ tiêu chuẩn Dublin Core (DC), cho nên người ta thường tưởng lầm rằng DC là một sự kế thừa nào đó từ truyền thống của thư viện. Trên thực tế, đã có nhiều nỗ lực để chia tách chuẩn Dublin Core khỏi những tiêu chuẩn truyền thống của thư viện, và nỗ lực đó đã phần lớn thành công. Mục đích của Dublin Core là cung cấp một bộ phần tử dữ liệu đơn giản cho việc mô tả tài liệu và các đối tượng khác trên Internet. Nó sẽ đơn giản tới mức bất kỳ ai cũng có thể tạo ra một biểu ghi cho tài liệu riêng của mình. Dublin Core có 15 phần tử dữ liệu chính (“core”)⁵, mà những phần tử này sẽ được chi tiết hóa hơn bằng việc sử dụng các bộ chuẩn hóa dữ liệu (qualifiers). Những phần tử chính

này có ý nghĩa rất mở rộng, bởi vậy thay thế thế Tác giả (Author), nó sẽ là “Người tạo lập” (Creator), song thế “Người tạo lập” có thể được gán cho tên tác giả, người tạo lập hoặc người sáng tác v...v. Tôi có thể dễ dàng làm một biểu ghi Dublin Core cho bất thứ cái gì, kể cả cho chính bài báo mà tôi vẫn chưa viết xong này:

creator = Karen Coyle
title = Understanding Metadata and its Purpose
date = December, 2004
description = The first draft of an article for Journal of Academic Librarianship
subject = metadata
type = text

Một hy vọng của tiêu chuẩn Dublin Core đó là mọi tài liệu trên Internet sẽ có được sự mô tả thư tịch riêng của chúng, và chúng sẽ có những phần tử dữ liệu được mã hóa cho các phần thông tin như tác giả, nhan đề, ngày tháng. Theo một ý nghĩa khác, tiêu chuẩn này đại diện cho một quan điểm rất giống như một thủ thư, điều này có nghĩa là một tài liệu có thể được tìm theo tên tác giả hoặc nhan đề. Trên Internet ngày hôm nay, Dublin Core quả thật đã được sử dụng ngày càng nhiều mặc dù nó vẫn chưa tạo ra một mục lục của nguồn tài nguyên Internet. Thay vào đó, Dublin Core đã trở thành một siêu dữ liệu mô tả tài liệu cho nhiều ứng dụng dựa trên Web. Một ví dụ của ứng dụng này đó là “Creative Commons License”.

Creative Commons⁶ là cả một dịch vụ Web và một phong trào có tính xã hội. Nó được phát triển bởi Larry Lessig, một

giáo sư luật trường đại học Standford nổi tiếng về việc bị chỉ trích đối với sự tăng hiệu lực của luật bản quyền không có lợi cho quyền của công chúng đối với việc sử dụng và sử dụng lại những ý tưởng của người đi trước⁷. Với sự quan tâm tạo điều kiện cho người tạo lập có thể cho phép sử dụng những tác phẩm của mình, một bộ những cấp phép đã được phát triển để dễ dàng gắn vào các tập tin trên Internet. Những bộ cấp phép này xác lập rằng người tạo lập lên một tác phẩm cho phép sử dụng và sử dụng lại cái gì. Ngoài sự cấp phép này ra, phần mềm Creative Commons cho phép người tạo lập gắn thêm một phần nhỏ những gì mà các thủ thư gọi là siêu dữ liệu “mô tả”: Người tạo lập (creator), nhan đề (title), và ngày tháng (date), và mô tả ngắn về một tiêu đề. Họ sử dụng những phần tử dữ liệu Dublin Core là creator, title, date, description (được mã hóa trong một biểu ghi như là “dc:creator”, “dc:title”, ...). (Hình 1).

Để sử dụng phần mềm Creative Commons License không cần hiểu về luật bản quyền hoặc các thoả thuận hợp đồng, vì gần như bất kỳ ai đều có thể dễ dàng

hiểu các phần tử mô tả. Theo nghĩa này, Dublin Core đã đạt được mục đích cung cấp một bộ phần tử chính có thể dễ dàng nhúng vào nhiều ứng dụng Web khác nhau.

Một trong những điều gì làm cho Dublin Core dễ dàng sử dụng đối với bất kỳ ai, đó là nó không gắn liền với một quy tắc biên mục. Đây là cái gì đi ngược lại khuynh hướng của nguyên tắc biên mục thư viện và nó nhất định làm giảm đi khả năng có thể sử dụng lại nội dung của các biểu ghi Dublin Core. Có nhiều sự mô tả mỗi phần tử dữ liệu theo tiêu chuẩn Dublin Core, bởi vậy ý nghĩa của một phần tử dữ liệu được định nghĩa chung

chung, song biên mục vẫn hợp lệ khi biên mục một phần tử dữ liệu “Creator=Karen Coyle cũng như “Creator = Coyle, Karen”. Lợi thế của điều này đó là Dublin Core có thể được sử dụng bởi nhiều cộng đồng và văn hóa khác nhau; và điều bất lợi hiển nhiên đó là nội

dung của các trường dữ liệu không nhất quán hay thống nhất trên mọi ứng dụng, điều đó làm cho sự vận hành liên kết trở lên khó khăn.

```

<!-- /Creative Commons License -->
<!--
<rdf:RDF xmlns="http://web.resource.org/cc/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  >
  <Work rdf:about="">
    <dc:title>Metadata: Data with a Purpose</dc:title>
    <dc:date>2004</dc:date>
    <dc:description>A general discussion of document/resource metadata and
  some related uses.</dc:description>
    <dc:creator><Agent>
      <dc:title>Karen Coyle</dc:title>
    </Agent></dc:creator>
    <dc:rights><Agent>
      <dc:title>Karen Coyle</dc:title>
    </Agent></dc:rights>
    <dc:type rdf:resource="http://purl.org/dc/dcmitype/Text" />
    <dc:source rdf:resource="http://www.kcoyle.net/meta_purpose.html"/>
    <license rdf:resource="http://creativecommons.org/licenses/by-nc-nd/2.0/" />
  </Work>
  <License rdf:about="http://creativecommons.org/licenses/by-nc-nd/2.0/">
    <permits rdf:resource="http://web.resource.org/cc/Reproduction" />
    <permits rdf:resource="http://web.resource.org/cc/Distribution" />
    <requires rdf:resource="http://web.resource.org/cc/Notice" />
    <requires rdf:resource="http://web.resource.org/cc/Attribution" />
    <prohibits rdf:resource="http://web.resource.org/cc/CommercialUse" />
  </License>
</rdf:RDF>

```

Hình 1
Creative Commons License với Phần tử Dữ liệu Dublin Core in đậm

MODS: Một khổ mẫu MARC nhẹ nhàng và đơn giản hơn

Khổ mẫu MARC một biểu ghi được cấu trúc chuyên biệt ở mức cao cho việc mã hóa thông tin thư tịch. Nó nổi tiếng trong thế giới thư viện và được hậu thuẫn bởi nhiều hệ thống thư viện tại Hoa Kỳ, Canada và nhiều nước khác, đặc biệt trong thế giới các quốc gia nói tiếng Anh. Trong một môi trường nối kết mạng mà ở đó siêu dữ liệu mô tả được chuyển tiếp qua nhiều hệ thống và có thể gắn liền vào trong hoặc với nhiều loại siêu dữ liệu khác, thì quả thật lý tưởng để sử dụng biểu ghi MARC cho mục đích này. Tuy nhiên, vấn đề của MARC đó là việc nhúng dữ liệu này nói chung đòi hỏi sử dụng một cấu trúc dữ liệu XML, và MARC không phải một biểu ghi XML. Thư viện Quốc hội Mỹ đã tạo ra một cách nhằm biên dịch biểu ghi MARC sang XML, nhưng đã không nhận được sự hưởng ứng, vì một lý do có thể nhận thấy, đó là: biểu ghi MARC lớn và chi tiết hơn là mọi hệ thống có thể cần, và việc sử dụng các thẻ số và mã trường con làm cho mọi người khó hiểu nếu không được đào tạo đầy đủ. Bởi vậy, người ta đã cần một phiên bản MARC nhẹ nhàng và đơn giản hơn mà có thể chấp nhận những phần tử dữ liệu chính từ một biểu ghi MARC và chuyển chúng vào một khổ mẫu XML dễ dàng hiểu hơn. Vậy, tiêu chuẩn MODS đã ra đời (Tiêu chuẩn Mô tả Đối tượng Siêu Dữ liệu – Metadata Object Description Standard).

MODS sử dụng những thẻ mà người ta dễ dàng hiểu để thay thế cho những thẻ bằng 3 chữ số và mã trường con của MARC (vd., “Nhan đề” (“title”)

thay thế cho “245”). Nó bỏ qua phần lớn các phần tử dữ liệu của trường cố định, ngoại trừ những mã khổ mẫu có tính vật lý (của mã 007) và nhiều mã dành cho thể loại (của mã 008). Nó cũng giới thiệu một số cách thực hành hiệu quả và cách tân hơn. MODS định nghĩa một cấu trúc gọi là “Tên” (“Name”) đại diện nhiều trường bao gồm cả trường con để chứa đựng thông tin tên cá nhân, công ty và cho tên một hội nghị. Cấu trúc này có thể được sử dụng bất kỳ khi nào một tên xuất hiện, nó có thể là mục dữ liệu chính, phụ hoặc chủ đề. Bởi vậy, với một trường tên như:

```
<byname type="personal">N>
  <namePart>Shakespeare,
    William</namePart>
  <namePart type="date">1564–
    1616</namePart>
</name>
```

Có thể được sử dụng như một trường tác giả, hoặc có thể trở thành một phần của “tiêu đề đề mục” (Subject Heading):

```
<subject authority="lcsch">
  <name type="personal">
    <namePart>Shakespeare,
      William</namePart>
    <namePart type="date">1564–
      1616</namePart>
  </name>
  <topic>Bibliography</topic>
  <topic>Periodicals</topic>
</subject>
```

Mặc dù được sinh ra từ MARC21 và chi tiết hơn nhiều tiêu chuẩn Dublin Core, những MODS lại ít quy tắc biên mục hơn MARC21. Tương tự như Dublin Core, không có nhiều trường phụ thuộc

và tất cả các trường đều có thể lặp lại. MODS chứa đựng nhiều giá trị từ MARC, song nó cũng có những nhiều điểm khác với MARC21: Không có những khái niệm về “dẫn mục chính” (“main entry”) hoặc “dẫn mục phụ” (“added entry”), tất cả tác giả đơn giản chỉ là “tác giả”; và một biểu ghi có thể có nhiều nhan đề mà không có một “nhan đề chính” (“main title”). Khi những biểu ghi MARC21 được biên dịch tới MODS, thì người ta có một biểu ghi trong XML mang dấu ấn của MARC. Biểu ghi MODS có thể được tạo ra từ siêu dữ liệu thư tịch mà không khởi đầu từ mục lục thư viện, ví dụ như trích dẫn của bài báo, và nó thường được sử dụng trong cơ sở dữ liệu có sự trộn lẫn của mục lục thư viện và dữ liệu biên mục khác.

METS – Một siêu dữ liệu cấu trúc

Đó là một định dạng siêu dữ liệu tài liệu và mục đích của nó không phải là “mô tả” theo nghĩa biên mục của thuật ngữ đó. Một định dạng siêu dữ liệu đang được sử dụng phổ biến bởi nhiều thư viện số và lưu trữ số gọi là Tiêu chuẩn Mã hóa và Truyền Siêu dữ liệu (METS – Metadata Encoding and Transmission Standard). METS đóng vai trò như là một tiêu chuẩn “gói siêu dữ liệu” (“wrapper”) và nó giúp tái đóng gói các tập tin cùng tạo lên một đối tượng số. Không giống như một cuốn sách đóng tập, tài liệu số thường được tạo lên từ nhiều tập tin riêng rẽ đại diện cho những trang của cuốn sách hoặc các đơn vị số khác. Và không giống như một cuốn sách in, người ta thường không nhìn thấy trang bìa hoặc trang nhan đề, cũng không có một hình ảnh thu nhỏ

nào thông qua trang sách để tìm thấy một vị trí cụ thể trong cuốn sách. Chúng ta hãy nghĩ về tiêu chuẩn METS như là một tiêu chuẩn đóng tập, đóng gói và di chuyển cho một nhóm tập tin số. Nó cũng bao gồm thông tin kỹ thuật cần thiết để quản lý và hiểu những tập tin đó, như định dạng tập tin, công nghệ được sử dụng để số hóa nếu tiêu đề đó đã bắt đầu đời sống của nó trên giấy, và sự chuyển đổi sang hình thức số và nén số đã được sử dụng đối với những tập tin đó. Điều gì mà tiêu chuẩn METS không định nghĩa đó là siêu dữ liệu mô tả. Thay vào đó, nó cho phép người ta tạo ra biểu ghi METS nhúng bất cứ siêu dữ liệu mô tả nào mà họ mong muốn để sử dụng cho tài liệu đó. Điều này minh họa một đặc tính quan trọng của thế giới siêu dữ liệu, mà chúng ta đã thấy trong ví dụ của phần mềm Creative Commons: siêu dữ liệu có thể được sử dụng lại thay vì sáng tạo ra nó. Những biểu ghi METS thường mang siêu dữ liệu mô tả theo tiêu chuẩn Dublin Core, hoặc MODS.

SIÊU DỮ LIỆU VÀ MỤC LỤC THƯ VIỆN

Vậy tất cả những gì đề cập ở trên có thể làm việc với mục lục thư viện, và quan trọng hơn cả, liệu siêu dữ liệu có thay thế mục lục thư viện không? Như tôi đã đề cập ở trên, một trong những vấn đề chính đối với biểu ghi Dublin Core đó là thiếu qui tắc biên mục, và bởi thế hầu như không có được sự dự báo giữa cộng đồng người dùng hoặc các dự án thực hiện về nội dung của những trường dữ liệu. Cái gì mà việc biên mục và mục

lục thư viện có thể đem đến đó là khả năng tương thích ở mức cao về dữ liệu chứa đựng trong các biểu ghi. Sự tương thích này tạo ra một dịch vụ cho người dùng. Người dùng có thể chuyển từ một mục lục thư viện này đến một mục lục thư viện khác mà vẫn đảm bảo thống nhất về dữ liệu. Song, giá trị chính của sự tương thích này là nó tạo cho chúng ta khả năng biên mục hợp tác và trao đổi biểu ghi thư tịch giữa các thư viện và hệ thống thư viện với nhau. Điều này cũng cho phép các nhà cung cấp hệ thống thư viện tạo ra một sản phẩm có thể sử dụng ở bất kỳ thư viện nào, ví như phiếu mục lục theo kích cỡ tiêu chuẩn có thể đề vừa bất kỳ ngăn kéo mục lục phiếu nào.

Hiệu quả từ sự tương thích này thật là to lớn và cộng đồng thư viện phụ thuộc vào điều này để biên mục các tài liệu gốc của mình. Song, vì thư viện đang chuyển sang việc tổ chức những tài liệu ít có tính truyền thống hơn, cho nên cả qui tắc biên mục và hệ thống thư viện trước kia đều không đem đến các giải pháp khả thi. Hãy tưởng tượng rằng bạn có một bộ lưu trữ ảnh về thành phố của quê hương bạn kể từ đầu thế kỷ 20, và bạn muốn nó sẵn có trên Web. Và chúng ta giả sử bạn có khoảng một nghìn bức ảnh trong số chúng mà hầu như bạn không biết về tác giả chụp chúng là ai, và thường không rõ về ngày tháng chụp. Bạn chỉ có thể thấy rằng ai đó đã ghi bằng bút chì lên phía sau bức ảnh cho biết bức ảnh này chụp cái gì. Vd., “Main Street, circa 1910”. Để biên mục và tạo ra một biểu ghi MARC21 của những bức ảnh này thường rất tốn thời gian và thường kết quả là biểu ghi sẽ có rất ít thông tin. Thay vào đó, bạn có thể tạo ra

một biểu ghi Dublin Core đơn giản như sau:

date = circa 1910

description = Main Street

Biểu ghi này có thể không được nhập vào một mục lục trực tuyến mặc dù biểu ghi như thế này có thể là điểm truy cập cho các công nghệ tìm kiếm siêu dữ liệu cho phép một tìm kiếm duy nhất để có thể tìm kiếm với nhiều cơ sở dữ liệu có định dạng siêu dữ liệu khác nhau. Lợi ích chính đó là những biểu ghi như thế này có thể được nhân viên thư viện tạo ra nhanh và dễ dàng mà không cần phải đào tạo nhiều. Bởi thế, siêu dữ liệu cho một nguồn tài nguyên nào đó dễ dàng có thể được tạo ra, mà nếu theo cách mô tả khác thường thì không có thông tin gì cả.

Tiêu chuẩn siêu dữ liệu như Dublin Core thiếu cấp độ dự báo cho phép sử dụng lại một cách hệ thống trên diện rộng những biểu ghi. Trên thực tế, những khổ mẫu siêu dữ liệu này, và nhiều khổ mẫu dữ liệu khác thường được sử dụng trong các hệ thống riêng biệt và đơn lẻ. Dù rằng thư viện đã bắt đầu trao đổi với nhau từ cuối thế kỷ 19, nhưng chúng ta mới thấy những hệ thống riêng biệt này bắt đầu trao đổi dữ liệu, cho nên các nhà phát triển đi đến một kết luận rằng, không phải cấu trúc dữ liệu, mà là nội dung của những biểu ghi siêu dữ liệu làm nên sự khác biệt giữa một giải pháp trên một hệ thống duy nhất và hệ thống mục lục gắn kết.

Nói một cách khác, chúng ta có thể thấy ***khí siêu dữ liệu tăng lên và trở nên đủ lớn thì nó sẽ trở thành mục lục thư viện.***

CHÚ GIẢI VÀ THAM KHẢO

1. Cách giải thích này lần đầu tiên được đưa ra bởi Tom Delsey của Thư viện Quốc gia Canada nói (“Metadata: Cataloging for men”), và Michael Gorman (“... metadata is cataloging done by men.”).
2. Tiêu chuẩn XML được định nghĩa bởi 3WC (<http://www.w3.org/XML/>), song nhiều tiêu chuẩn ứng dụng XML lại được định nghĩa bởi nhiều nhóm khác, ví dụ như nhóm tiêu chuẩn kinh doanh điện tử (e-business standard group), OASIS (<http://www.oasis-open.org/>).
3. <http://www.w3.org/RDF/>.
4. <http://www.w3.org/2001/sw/>.
5. 15 phần tử Dublin Core là như sau: Contributor, Coverage, Creator, Date, Description, Format, Identifier, Language, Publisher, Relation, Rights, Source, Subject, Title, Type (<http://dublincore.org>)
6. <http://www.creativecommons.org>.
7. Lawrence Lessig là một tác giả của các cuốn sách “Code and Other Laws of Cyberspace (New York: Basic Books, c1999); The Future Of Ideas: The Fate Of The Commons In A Connected World (New York: Random House, 2001); Free Culture: How Big Media Uses Technology And The Law To Lock Down Culture And Control Creativity (New York: Penguin Press, 2004)”.
8. <http://www.loc.gov/standards/mets/>.

“Siêu dữ liệu là Phiếu mục lục ...

GIỐNG NHAU:

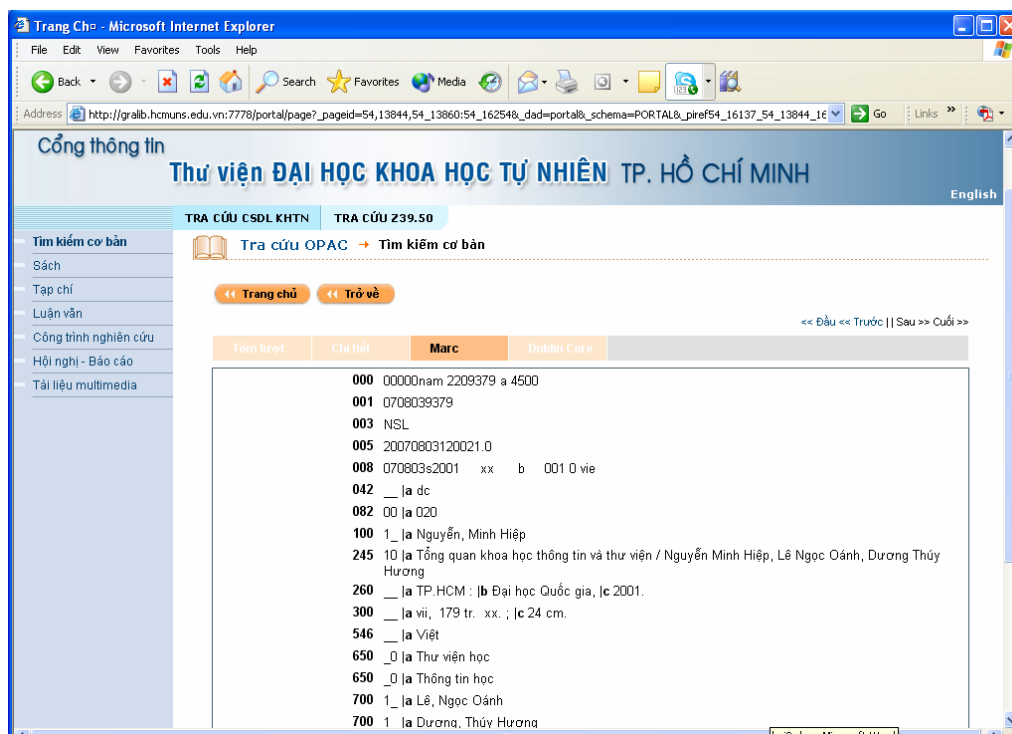
- Được biên mục bởi con người;
- Trình bày lý lịch của tài liệu hay còn được gọi là *thư tịch* (biểu ghi thư tịch – bibliographic record và siêu dữ liệu thư tịch – bibliographic metadata): tác giả, nhan đề, năm xuất bản, vv...; so sánh với *hộ tịch*: họ và tên, ngày sinh, vv...

KHÁC NHAU:

- Phiếu mục lục tách rời kho tài liệu; trong khi siêu dữ liệu thường kèm theo nội dung tài liệu;
- Những thành phần mô tả của phiếu mục lục hay biểu ghi thư tịch thường tách rời theo từng cụm từ hay nhãn trường (MARC) được trao đổi dưới dạng liên biến (analog); trong khi thành phần mô tả siêu dữ liệu được đóng gói bằng công nghệ web và trao đổi dưới dạng kỹ thuật số (digital).

... được thể hiện trong môi trường số”

Minh họa một biểu ghi thư tịch MARC so với siêu dữ liệu thư tịch MARC-XML và Dublin Core trong OPAC của Thư viện ĐH Khoa học Tự nhiên TP. HCM



Biểu ghi thư tịch MARC của nhan đề sách "Tổng quan khoa học thông tin và thư viện"

```

<datafield tag="700" ind1="1" ind2="0">
  <subfield code="a">Lê, Ngọc Oánh</subfield>
</datafield>
<datafield tag="700" ind1="1" ind2="0">
  <subfield code="a">Đương, Thủy Hương</subfield>
</datafield>
<datafield tag="100" ind1="1" ind2="0">
  <subfield code="a">Nguyễn, Minh Hiệp</subfield>
  <subfield code="d" />
</datafield>
<datafield tag="082" ind1="0" ind2="0">
  <subfield code="a">020</subfield>
</datafield>
<datafield tag="300" ind1="1" ind2="0">
  <subfield code="a">vii, 179 tr. xx. ;</subfield>
  <subfield code="c" />
</datafield>
<datafield tag="546" ind1="1" ind2="0">
  <subfield code="a">Việt</subfield>
</datafield>
<datafield tag="260" ind1="1" ind2="0">
  <subfield code="a">TP.HCM</subfield>
  <subfield code="b">Đại học Quốc gia,</subfield>
  <subfield code="c">2001</subfield>
</datafield>
<datafield tag="250" ind1="1" ind2="0">
  <subfield code="a">0 .ed</subfield>
</datafield>
<datafield tag="650" ind1="1" ind2="0">
  <subfield code="a">Thư viện học</subfield>
</datafield>
<datafield tag="650" ind1="1" ind2="0">
  <subfield code="a">Thông tin học</subfield>
</datafield>
<datafield tag="245" ind1="1" ind2="0">
  <subfield code="a">Tổng quan khoa học thông tin và thư viện /</subfield>
  
```

```

<?xml version="1.0" encoding="UTF-8" ?>
<dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/">
  <dc:collection>
    <dc:record>
      <dc:title>Tổng quan khoa học thông tin và thư viện</dc:title>
      <dc:creator>Nguyễn Minh Hiệp</dc:creator>
      <dc:creator>Lê Ngọc Oánh</dc:creator>
      <dc:creator>Đương Thủy Hương</dc:creator>
      <dc:subject>020</dc:subject>
      <dc:subject>Thư viện học</dc:subject>
      <dc:subject>Thông tin học</dc:subject>
      <dc:description />
      <dc:publisherplace>TP.HCM</dc:publisherplace>
      <dc:publisher>Đại học Quốc gia</dc:publisher>
      <dc:date>2001</dc:date>
      <dc:type>Sách</dc:type>
      <dc:format>vii, 179 tr. xx. ;</dc:format>
      <dc:identifier />
      <dc:source />
      <dc:language>Việt</dc:language>
      <dc:relation />
      <dc:coverage />
      <dc:rights />
    </dc:record>
  </dc:collection>
</dc:dc>
  
```

Siêu dữ liệu thư tịch MARC-XML và Siêu dữ liệu thư tịch Dublin Core (được đóng gói bằng ngôn ngữ XML)