

HIỂU VÀ SỬ DỤNG DUBLIN CORE

PHẠM MINH QUÂN

Phòng Tài nguyên thông tin
Thư viện ĐH Khoa học Tự nhiên, ĐHQG TP. HCM



I. TỔNG QUAN VỀ METADATA VÀ DUBLIN CORE

1.1 METADATA

Metadata là dữ liệu về các dữ liệu hay còn gọi là siêu dữ liệu, là những thông tin chuyển tải ý nghĩa của các thông tin khác. Metadata bao gồm một tập hợp các phần tử thiết yếu để mô tả nguồn thông tin.

Một biểu ghi Metadata bao gồm một hệ thống các thành tố hay còn gọi là các yếu tố cần thiết để mô tả nguồn thông tin. Siêu dữ liệu Metadata trong thư viện - hệ thống mục lục thư viện - bao gồm một tập hợp các biểu ghi với các yếu tố mô tả của một cuốn sách hay một tài liệu thư viện như: tác giả, nhan đề, xuất bản, đề mục, ký hiệu xếp giá.

Mối liên hệ giữa nguồn thông tin và biểu ghi Metadata có thể ở dạng độc lập: tương tự như phiếu mô tả mục lục truyền thống hoặc ở dạng kết hợp tương tự như phần biên mục tiền xuất bản CIP (Cataloguing In Publication), các phần tử Metadata được thể hiện ngay trong bản thân tài liệu hoặc nguồn thông tin.

Siêu dữ liệu Metadata có hai chức năng chính:

- Cung cấp một phương tiện khai thác những cơ sở dữ liệu đã có sẵn,
- Thể hiện nội dung, chất lượng và đặc điểm của một CSDL cũng như sự thuận tiện trong sử dụng chính CSDL đó.

1.2 DUBLIN CORE

Chuẩn Dublin Core là chuẩn dùng để mô tả dữ liệu trong các Metadata nhằm khai thác các tài liệu trong thư viện và trên các web site thông qua mạng Internet. Chuẩn Dublin Core bao gồm 15 yếu tố được thiết lập từ các cuộc hội thảo mang tầm cỡ quốc tế và mang ý nghĩa kết hợp của các ngành khoa học: thư viện, tin học, bảo tàng, mã hoá văn bản và các lĩnh vực khác có liên quan.

➤ **Đặc điểm của Dublin Core:**

- Đơn giản trong tạo lập và bảo trì: được thiết kế nhằm phục vụ những người không chuyên; dễ sử dụng và rẻ nhưng mang lại hiệu quả lớn.

- Ngữ nghĩa thông dụng: khắc phục những khó khăn trong việc hiển thị các thuật ngữ. Vd.: yếu tố <tác giả> (Creator) được gán cho người tạo lập, nhà soạn nhạc, đạo diễn, trong vai trò là tác giả chính.
- Phạm vi quốc tế: tháng 11 - 1999, đã có phiên bản của hơn 20 thứ tiếng: Phần Lan, Na Uy, Thái Lan, Nhật, Pháp, Đức, Hy Lạp, Indonesia, Tây Ban Nha. Tổ chức WWW phát triển Chuẩn Dublin Core trên nền tảng kết hợp đa ngôn ngữ, phục vụ cho môi trường tài nguyên thông tin điện tử mang tính chất đa văn hoá và đa ngôn ngữ.
- Khả năng mở rộng: với cơ chế mở, Chuẩn Dublin Core có thể được mở rộng bởi các chuyên gia bằng việc gắn kết thêm các yếu tố mở rộng. Khả năng này còn được thực hiện một cách đơn giản thông qua việc kết nối nhiều CSDL khác nhau thông qua mạng Internet.

➤ **Ý nghĩa của Dublin Core trong Thư viện số**

- Là một phương thức mô tả nguồn thông tin, đặc biệt là nguồn thông tin điện tử một cách có hiệu quả. Dublin Core càng đặc biệt phát huy tác dụng khi được sử dụng để mô tả tư liệu điện tử vốn khó xác định được loại hình và nội dung các yếu tố cần thể hiện.
- Thay thế cho các dạng thức trình bày thông tin trước đây như MARC do sự đơn giản trong cấu trúc mà người sử dụng có thể tự thiết kế theo yêu cầu của riêng mình.
- Cung cấp cho người sử dụng một phương án tiếp cận thông dụng thông qua các giao diện quen thuộc như Web.
- Tạo cho người cán bộ thư viện sự thuận tiện trong công tác khi không còn phải gõ bó trong các trường, các yếu tố vốn dĩ đã rất đa dạng và phức tạp.

1.3 CÁC YẾU TỐ CỦA DUBLIN CORE:

➤ **Phân loại các yếu tố:**

NỘI DUNG	SỞ HỮU TRÍ TUỆ	THUYẾT MINH
Nhan đề	Tác giả	Ngày tháng
Đề mục	Tác giả phụ	Mô tả vật lý
Mô tả	Xuất bản	Định danh
Loại hình	Bản quyền	Ngôn ngữ
Nguồn gốc		
Liên kết		
Nơi chứa		

- **Các yếu tố cơ bản:** Các yếu tố cơ bản của Dublin Core đều mang thuộc tính lựa chọn và có thể lập lại. Mỗi yếu tố cũng có một giới hạn những hạn định, thuộc tính nhằm diễn giải chính xác ý nghĩa của các yếu tố.
1. **Nhan đề (Title):** Tên của nguồn thông tin thường do tác giả hoặc nhà xuất bản đặt cho tài liệu.
 2. **Tác giả (Creator):** Người hoặc cơ quan chịu trách nhiệm chính về nội dung trí tuệ của nguồn thông tin
 3. **Đề mục (Subject):** Chủ đề của nguồn thông tin và được thể hiện bằng từ vựng có kiểm soát gồm tiêu đề đề mục, số phân loại,...
 4. **Mô tả (Description):** Phần thể hiện nội dung của nguồn thông tin bao gồm cả phần tóm tắt của tư liệu văn bản hoặc nội dung của tư liệu nghe nhìn
 5. **Xuất bản (Publisher):** Cơ quan, tổ chức chịu trách nhiệm tạo lập, xuất bản nguồn thông tin trong định dạng thực.
 6. **Tác giả phụ (Contributor):** Cá nhân hay tổ chức có những đóng góp về mặt trí tuệ cho tư liệu nhưng không phải là tác giả chính.
 7. **Ngày tháng (Date):** ngày tháng có liên quan đến việc tạo lập, xuất bản hay công bố tư liệu.
 8. **Loại hình (Type):** hình thức vật chứa nội dung tư liệu
 9. **Mô tả vật lý (Format):** Định dạng vật lý và kích thước của tư liệu như kích cỡ, thời lượng,.. Định dạng cũng còn được dùng để chỉ rõ phần mềm và phần cứng cần thiết để sử dụng tư liệu.
 10. **Định danh tư liệu (Identifier):** Là một dãy ký tự hoặc số nhằm thể hiện tính đơn nhất của tư liệu như: URLs và URNs, ISBN, ISSN,...
 11. **Nguồn gốc (Source):** Nguồn gốc mà tư liệu được tạo thành, yếu tố này có thể bao gồm siêu dữ liệu về nguồn thông tin thứ hai nhằm khai thác tư liệu hiện hành.
 12. **Ngôn ngữ (Language):** Ngôn ngữ của nội dung tư liệu, được thành lập theo quy tắc RFC 1766.
 13. **Liên kết (Relation):** Một định danh cho nguồn thứ hai và những mối quan hệ của nó với tư liệu hiện hành. Yếu tố này thể hiện những kết nối giữa những nguồn tư liệu có liên quan.
 14. **Nơi chứa (Coverage):** Những đặc tính về không gian và/hoặc thời gian của tư liệu. Không gian nơi chứa chỉ ra một vùng sử dụng địa danh hoặc toạ độ. Đặc tính thời gian trong yếu tố này chỉ ra khoảng thời gian mà tư liệu đề cập tới và thường sử dụng tên thời kỳ như thời kỳ Đồ đá.
 15. **Bản quyền (Rights):** Thông tin về tình trạng bản quyền, kết nối tới thông tin về tình trạng bản quyền hoặc dịch vụ cung cấp thông tin bản quyền cho tư liệu.
- **Các yếu tố mở rộng:** Thực tế sử dụng Dublin Core cho thấy mỗi yếu tố cơ bản còn gộp chứa trong nó một vài thành tố phụ nhằm diễn đạt chi tiết hơn nội dung chính yếu tố đó. Các thành tố phụ được coi là các yếu tố mở rộng và được thể hiện thông qua những khung mã hoá cụ thể. Ví dụ khi thể hiện nội dung của một tài liệu, người ta

cung cấp một vài cách tiếp cận khác nhau như qua ký hiệu phân loại, tiêu đề đề mục, từ khoá.

YẾU TỐ	YẾU TỐ MỞ RỘNG	KHUNG MÃ HOÁ
Nhan đề	Nhan đề thay thế	
Tác giả		
Đề mục		LCSH, MeSH; DDC, LCC, UDC
Mô tả	Mục lục (Table of Contents) Tóm tắt (Abstract)	
Xuất bản		
Tác giả phụ		
Ngày tháng	Tạo lập (Created) Có giá trị (Valid) Có hiệu lực (Available) Xuất bản (Issued) Hiệu đính (Modified)	Bảng thời kỳ của DC Định dạng ngày tháng của W3C
Loại tài liệu		Thuật ngữ về loại hình của Dublin Core
Mô tả vật lý	Kích thước và thời lượng (Extent) Vật mang tin (Medium)	IMT loại tư liệu
Định danh		URI Uniform Resource Identifier
Nguồn gốc		URI Uniform Resource Identifier
Ngôn ngữ		ISO 639-2 RFC 1766
Liên kết		
Nơi chứa		
Bản quyền		

➤ So sánh đối chiếu với các yếu tố mô tả AACR2 và MARC

DC	AACR2	MARC
Nhan đề	Nhan đề chính	245\$a
Tác giả	Tác giả chính	100, 245\$c
Đề mục	Điểm truy cập khác	050, 082, 650
Mô tả	Phụ chú nội dung, yếu tố bổ sung nhan đề	245\$b
Xuất bản	Nơi và nhà xuất bản	260\$a, 260\$b
Tác giả phụ	Tác giả liên quan	
Ngày	Năm xuất bản	260\$c
Loại tài liệu	Phụ chú hình thức	
Mô tả vật lý	Mô tả vật lý	300
Định danh		
Nguồn gốc		
Ngôn ngữ		
Liên kết	Phụ chú	
Nơi chứa		
Bản quyền		

II. SỬ DỤNG DUBLIN CORE

2.1 CÁC QUY TẮC SỬ DỤNG

- **Kiểm soát từ vựng:** Từ vựng được sử dụng khi mô tả biểu ghi Dublin Core phải là từ vựng có kiểm soát, được lựa chọn từ trong các hệ thống tiêu đề đề mục như LCSH, MeSH nhằm mục đích kiểm soát tính nhất quán trong các hệ CSDL và hỗ trợ việc tìm kiếm, biên mục tự động.

- **Sử dụng Khung mô tả nguồn (RDF - Resource Description Framework):** RDF là một khuôn mẫu trao đổi và thể hiện thông tin trong môi trường Web. Ngoài ra, RDF còn được coi là khung chuyển đổi giúp nhận biết nội dung các yếu tố cho dù chúng ở trong nhiều loại CSDL khác nhau.
- **Hồ sơ áp dụng:** Trên lý thuyết, tất cả 15 trường đều mang thuộc tính lựa chọn và lặp lại. Tuy nhiên, mức độ tối thiểu theo khuyến cáo của các tổ chức có liên quan bao gồm các yếu tố như: Nhan đề, Tác giả, Ngày tháng, Mô tả, Ngôn ngữ. Tùy theo mức độ chi tiết được đòi hỏi trong việc mô tả dữ liệu, người ta cũng có thể nhập thêm một số yếu tố bổ trợ từ các Metadata khác. Trong trường hợp đó, thuật ngữ sử dụng để mô tả cần được định nghĩa một cách chặt chẽ.

2.2 SỬ DỤNG VỚI NGÔN NGỮ HTML

HTML có 2 thẻ (tag) để trình bày dữ liệu là <META> và <LINK>. Do đó, nếu siêu dữ liệu được nhúng trong văn bản thì chúng phải được đặt trong vùng giữa 2 thẻ <HEAD> của văn bản đó nhằm giúp cho các hệ thống tìm kiếm và chỉ mục tự động có thể nhận biết và tìm ra giá trị của phần mô tả của tài liệu.

* Ví dụ:

```
?<HTML>
<HEAD>
<TITLE>Mating Habits of the Northern Hairy Nosed Wombat</TITLE>
<META NAME= "DC.Creator" CONTENT="Smythe, Pearl">
</HEAD>
<BODY>
<H1>Northern Hairy Nosed Wombats</H1>
<P>The Northern Hairy Nosed Wombat is an animal native to
Australia....</P>
</BODY>
</HTML>?
```

Giống như một văn bản Web thông thường, HTML Metadata có những đặc điểm:

- Thuận tiện cho việc tìm kiếm và xuất dữ liệu một cách tự động với mọi trình duyệt.
- Mỗi yếu tố được bắt đầu bằng "<META" và kết thúc bằng ">"
- Ngữ pháp được bao gồm 2 thành tố: TÊN (NAME) và NỘI DUNG (CONTENT)

Sử dụng cú pháp HTML:

- Mỗi yếu tố đều bao gồm 2 phần: NAME và CONTENT,

* Ví dụ:

```
<META NAME="DC.Creator" CONTENT="Hồ Chí Minh">
<META NAME="DC.Title" CONTENT="Nhật Ký Trong Tủ">
```

- Bất kỳ yếu tố nào cũng có thể được lặp lại hoặc bỏ qua. Trong trường hợp lặp lại, các giá trị lặp có thể được thể hiện thành từng dòng riêng biệt hoặc bao gồm chung nhưng được phân cách bằng dấu chấm phẩy (;),

* Ví dụ 1:

```
<META NAME="DC.Creator" CONTENT="Viện Ngôn ngữ học">
<META NAME="DC.Creator" CONTENT=" Cao Xuân Phổ">
<META NAME="DC.Creator" CONTENT="Chu Khắc Thuật">
```

* Ví dụ 2:

```
<META NAME="DC.Creator" CONTENT="Viện Ngôn ngữ học; Cao Xuân
Phổ; Chu Khắc Thuật">
```

Tuy nhiên trong quá trình thực hiện, các yếu tố lập được khuyến cáo liệt kê thành từng dòng cho mỗi yếu tố.

- Quy ước cho việc nhúng Dublin Core Metadata trong HTML nhằm xác định và tập trung các yếu tố quy định sử dụng tiền tố "DC." để khai báo và được thể hiện như sau:

```
META NAME="DC.Creator"
```

```
META NAME="DC.Title"
```

Không sử dụng:

```
"DC.CREATOR" hoặc "DC.creator" hoặc "dc.Title"
```

Trong trường hợp không sử dụng bộ mã ASCII (các chương trình nhập liệu, các file nguồn sử dụng các ngôn ngữ không thuộc hệ Latin như Trung Quốc, Nhật, Nga, Ả Rập,..) thì cũng vẫn sử dụng cùng những quy ước này như trong phần thân của văn bản.

2.3 SỬ DỤNG VỚI NGÔN NGỮ XML

Việc sử dụng Dublin Core trong XML là kết quả của sự phát triển ứng dụng CNTT trong công tác thư viện. XML cho phép người sử dụng tự thiết kế khuôn dạng trình bày theo yêu cầu của mình. Tất cả những gì cần thiết chính là việc khai báo các yếu tố trong file định nghĩa loại hình dữ liệu (.DTD - Data Type Definition) và file thể hiện dữ liệu.

File định nghĩa loại hình dữ liệu DTD chứa các giá trị khai báo hay nói cách khác, là nơi khai báo và định nghĩa loại hình dữ liệu của các yếu tố Dublin Core cần thể hiện. Định nghĩa các yếu tố được bắt đầu bằng dấu ! đặt trong dấu ngoặc nhọn < và kết thúc bằng dấu ngoặc nhọn >

* Ví dụ:

```
<?xml version="1.0" ?>
<Book (Title, Creator, Date, Publisher, Format, Subject, Description)>
  <!Element title (# PC Data)>
  <!Element creator (# PC Data)>
  <!Element date (# PC Data)>
  <!Element publisher (# PC Data)>
  <!Element format (# PC Data)>
  <!Element subject (# PC Data)>
  <!Element description (# PC Data)>
```

File thể hiện dữ liệu trình bày nội dung các yếu tố trong phần mô tả của file XML nằm trong vùng <rdf:Description> </rdf:Description>. Trong trường hợp có giá trị lặp, giá trị của yếu tố có thể được trình bày lặp lại cũng giống như trong phần trình bày với HTML. Các yếu tố Dublin Core được trình bày đúng như đã khai báo trong file DTD và được thể hiện như sau:

```
<dc:tên yếu tố>giá trị</ dc:tên yếu tố>
```

* Ví dụ:

```
<?xml version="1.0" ?>
<rdf:RDF      xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.0/"
  xmlns:dcq="http://purl.org/dc/qualifiers/1.0/">
  <rdf:Description>
    <dc:type>text</dc:type>
    <dc:language>En</dc:language/>
    <dc:creator>Richard, John E.</dc:creator />
    <dc:title>Resource and environmental
    economics</dc:title>
    <dc:publisher>London : </dc:publisher>
    <dc:date>1995</dc:date/>
    <dc:format> <dc:format/>
    <dc:coverage> <dc:coverage/>
    <dc:relation> <dc:relation/>
    <dc:identifier> <dc:identifier/>
    <dc:rights <dc:rights/>
    <dc:description> <dc:description/>
    <dc:subject>Environmental
    economics</dc:subject/>
  </rdf:Description>
</rdf:Description>
```

2.4 DUBLIN CORE TRONG SIÊU DỮ LIỆU METADATA ĐỘC LẬP

Metadata độc lập có thể tồn tại trong bất kỳ một CSDL nào. Dưới đây là một ví dụ minh họa mô tả một tấm ảnh được chứa trong một file khác trên mạng Internet. Toàn bộ biểu ghi được trình bày như sau:

```
<META NAME= "DC.Title" CONTENT="Kita Yam (Japan)">
<META NAME= "DC.Creator" CONTENT="Kertesz, Andre">
<META NAME= "DC.Date" CONTENT="1968">
<META NAME= "DC.Type" CONTENT="image">
<META NAME= "DC.Format" CONTENT="image/gif">
<META NAME= "DC.Indentifier"
CONTENT="http://foo.bar.zaf/kertesz/kyama">
```

2.5 DUBLIN CORE TRONG SIÊU DỮ LIỆU METADATA BAO GỒM

Dưới đây là một ví dụ về Metadata bao gồm trong một văn bản được viết bằng ngôn ngữ HTML. Ngoài những phần chính của văn bản, người ta đã nhúng các yếu tố Dublin Core trong phần HEAD của văn bản đó. Khi hiển thị trên các trình duyệt Web, các yếu tố Dublin Core không được thể hiện ra trong giao diện nhưng khi sử dụng các bộ duyệt hoặc các phần mềm tìm kiếm có hỗ trợ việc hiển thị các yếu tố DC sẽ được hiển thị.

```
<HTML>
<HEAD>
<TITLE>Song of the Open Road</TITLE>
<META NAME="DC.Title" CONTENT="Song of the Open Road">
<META NAME="DC.Creator" CONTENT="Nash, Ogden">
<META NAME="DC.Type" CONTENT="text">
<META NAME="DC.Date" CONTENT="1939">
<META NAME="DC.Format" CONTENT="text/html">
<META NAME="DC.Identifier"
CONTENT="http://www.poetry.com/nash/open.html">
</HEAD>
<BODY><PRE>
I think that I shall never see
A billboard lovely as a tree.
Indeed, unless the billboards fall
I'll never see a tree at all.
</PRE></BODY>
</HTML>
```



TÀI LIỆU THAM KHẢO

1. HILLMAN, DIANE I. *Using Dublin Core*. (<http://www.dublincore.org>)
2. DILLON, MARTIN. *Metadata for Web Resources: How Metadata Works on the Web*. (http://www.loc.gov/catdir/bibcontrol/dillon_paper.html)
3. SUGIMOTO, SHIGEO. *Dublin Core: Process and Principles*/ Shigeo Sugimoto, Thomas Baker, Stuart L. Weibel. (<http://www.cais.ntu.edu.sg:8000/icadl2002/invited3.doc>)
4. BAKER, THOMAS. *A Grammar of Dublin Core*. D-Lib Magazine October 2000. Volume 6 Number 10. (<http://www.dlib.org/dlib/october00/baker/10baker.html>)