

Metadata專論

陳淑君

中央研究院計算中心

數位典藏國家型計畫・後設資料工作組

數位典藏專業培訓課程
2005年7月23日・30日

課程內容

- 2005.7.23
 - Metadata概論
 - Dublin Core專論
 - 博物館社群Metadata標準
 - 概論
 - 標準專論：VRA, CDWA, CCO, Getty, AAT
- 2005.7.30
 - 檔案館社群Metadata標準
 - 概論
 - 標準專論：各國檔案描述規則，權威控制
 - 數位學習社群Metadata標準
 - 學習物件的觀念
 - 標準專論：LOM
 - 標準在地化的方法學

Metadata概論

界定Metadata

- 現存很多Metadata的各種定義，端視特定社群或使用情境而不同。
- 字義與抽象的定義：
 - 有關資料的資料（data about data）
- 更明確的定義：
 - 有關資訊物件之結構的資訊（structured information about an information object）

界定Metadata

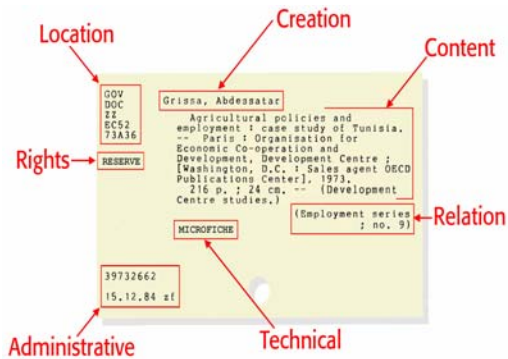
- 其他重要的定義：
 - 圍繞資料物件的一大群附屬（間接）的資訊（A cloud of collateral information around a data object）
 - 描述資源屬性的資料 (Data describes attributes of resources)
 - 有關資料背景與關聯性、資料內涵以及資料控制等相關資訊 (Information about the context of data and the content of data and the control of or over data).
 - 任何協助辨識、描述與放置網路化電子資源的資料。目前存在許多不同的metadata格式，有些在描述上很簡單，有些則相當複雜而豐富。(IFLA, 國際圖書館協會)
 - 記載資料元素或屬性（如：名稱、大小、資料類型等）的資料、有關記錄或資料結構（如：長度、欄位、行列等），以及有關資料的資料（如：位置、關聯、擁有者等）。
 - 當metadata是以結構的層級方式編排時，可以稱為「知識本體」（Ontology）或綱要（schema）
 - metadata就是資料 (Metadata is data)

Metadata的面貌 不同社群的習慣稱呼

- 電腦界：資料字典・綱要（Data Dictionary, Schema）
- 圖書館界：圖書館目錄(Library Catalogue)
- 博物館界：館藏記錄、登錄、資料標準（Documentation, Register, Data Standard）
- 檔案館界：檔案描述、檢索工具（Archival Description, Finding Aids）
- ...

Metadata的實例

圖書館卡片目錄



Metadata的實例

資料字典(data dictionary)

6.2.2 FileSize

Definition	<i>extent of image in number of bytes</i>
Type	positive integer
Required	M
Repeatable	N
Values (Examples)	618 72839 116126
Usage Notes	The file size must record the number of bytes as provided by the system. Do not attempt to record file sizes in terms of KB, MB, or other notation.
Use	System

6.2.3 Checksum

6.2.3.1 ChecksumMethod

Definition	<i>type of error detection technique used, i.e. a checksum (or equivalent)</i>
Type	enumerated type
Required	R
Repeatable	N
Values	
Usage Note	Local repository policies regarding file integrity metadata should govern implementation of this field. The enumerated type values should be defined locally, as should the rule

Metadata的實例 博物館登錄資料庫

University of Queensland - Antiquities Museum Register: Enquiry - Microsoft Internet Explorer

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

網址: http://www.staff.uq.edu.au/antdig/museumsearch.html

Inventory numbers for all objects except coins or certain 'loan' objects are in the form: 03.015 (i.e. two digits followed by a stop followed by three digits). The first two figures represent the year of acquisition. If you wish to see all of the objects acquired in any one year, simply enter the two digits for the year. Inventory numbers for coins are in the form: C037 (e.g. letter C followed by three digits). Inventory numbers for 'objects on loan' may consist of a combination of letters and digits (e.g. 03.11). Important! If you enter an inventory number for your search, do not enter characters in any of the other boxes below.

Use the arrow at the right of the box to make a selection of Civilisation/Culture, Class of Object, or Sub-class of Object. If in your search you do not wish to use any or all of these three criteria, leave the relevant box(es) set at "No selection".

In the Type of Object box, enter the specific type of object for which you are searching (e.g. Kylix, Amphora, Coin, etc.).

Enter the provenance for which you are searching (e.g. Attica, Crete, Metaponto, etc.) and/or a date or period. Important! When entering criteria in the Date/Period box, please use care, as the method of indicating dates/periods/eras/etc. can vary from object to object.

When searching using any of the boxes in this group, enter a word or short phrase which you might expect to find in a description of the object or of its function, or a description of its method of manufacture or of its decoration, or a reference to a painter.

To search by Bibliography or Comparanda, in each case enter a

Inventory numbers:

Civilisation/Culture: No selection

Class of Object: No selection

Sub-class of Object: No selection

Type of Object:

Provenance:

Date/Period:

Description:

Function:

Manufacture:

Decoration:

Painter:

Bibliography:

Metadata的實例 檔案館登記表

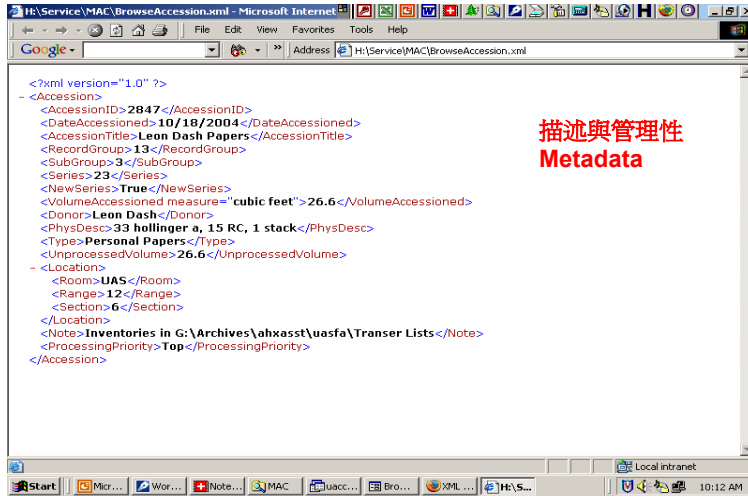
Metadata欄位

University of Illinois Archives and Research Collections: ACCESSION REGISTER

Date/ Acen no.	Title	Record Series #	Donor/Source	Vol. Acqd.	Phys. Descp.	Notes	Loc.
6/5/99	"Museum of the Past"	24/6/99	Archaeological Museum	2	1 book		
6/5/99	US National Bank	24/6/99	From 5th - 6th century	1	1 envelope		
6/11/99	Stamps from Africa	24/6/99	Sh. L. Caraceni	3.0	3 boxes	reboxed into 1.0	
6/11/99	1911 - 1913	24/6/99	Thomas Brown	.6	3 boxes		
6/11/99	Brian Voth Papers	7/15/99	Voth	4	416 boxes	was	44233-9
6/11/99	Mark Naxosidis Papers	15/6/99	Classics Dept	15	21 boxes		ARC 1044
6/10/99	Shao Lee So Papers	11/6/99	MIE Dept	23	23 RC boxes		ARC 1044
6/11/99	R. Oliver Papers	15/6/99	Classics	2	2 boxes		ARC 1044
6/11/99	Ad. Garmel Campaign Notes	11/6/99	Ad. Garmel	5	4 boxes		11.0
6/11/99	YMC A 63	4/11/99	John London	42	38 boxes		ARC 1044
6/11/99	Tamara E. Robertson Papers	11/10/99	Judy Tallner	7	6 boxes		PFE
6/28/99	Rudolf Finkbeiner	15/12/99	Rudolf Finkbeiner	14.0	14 RC boxes	was 346	44233-9
6/28/99	Allen Wicks Papers	12/1/00	L. Wicks	1.0	1 box	Provenance	PFE
6/28/99	Frank Leavitt Papers	4/1/00	Frank Leavitt Papers	.5	1 box		PFE
6/28/99	Isaac, Shalom (L)	11/1/00	Shalom	1.3	3.5 boxes		PFE
6/28/99	Isaac, Shalom (L)	11/1/00	Shalom	3	5 boxes		PFE
7.8.99	Isaac, Shalom (L)	4/1/00	Shalom	1.0	1 box		PFE
7.9.99	Isaac, Shalom (L)	15/1/00	Shalom	2	2 boxes		PFE
7.10.99	Isaac, Shalom (L)	15/1/00	Shalom	1.0	1 box		PFE
7.11.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes	Provenance	PFE
7.12.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.13.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.14.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.15.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.16.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.17.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.18.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.19.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.20.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.21.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.22.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.23.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.24.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.25.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.26.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.27.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.28.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.29.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE
7.30.99	Isaac, Shalom (L)	15/1/00	Shalom	1.6	15 boxes		PFE

Metadata的實例

XML Metadata



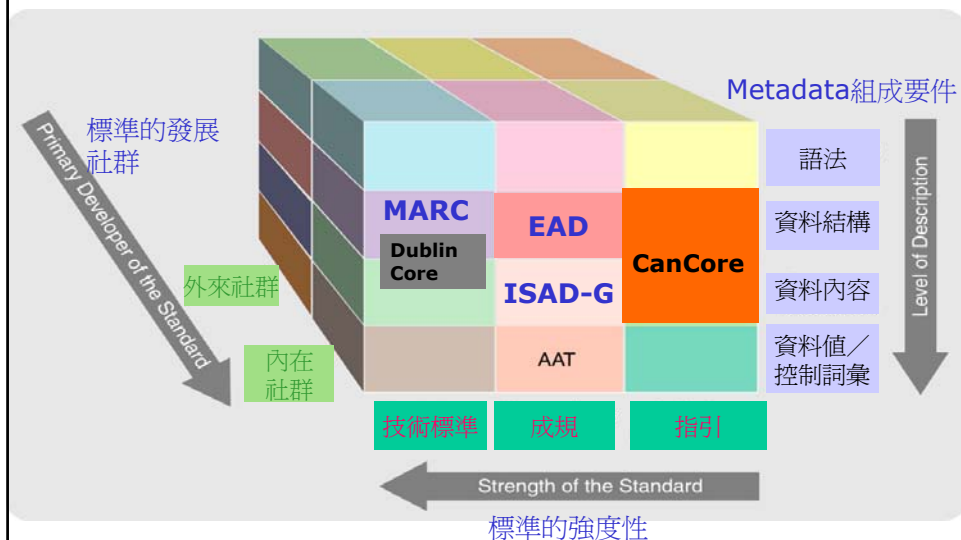
Metadata的類型

- 描述性：描述或識別資訊資源(如：編目記錄)
- 管理性：管理與控制資訊資源(如：採購資訊，版權與重複追蹤，合法取用需求的記載說明，數位化典藏品選擇準則，相似資訊物件的版本控制，文書系統產生的稽核痕跡)
- 技術性：有系統功能或metadata的運轉(如：軟硬體記載文件，數位化資訊，如：格式與壓縮率，系統反應時間追蹤)
- 保存性：資訊資源的保存管理(如：實體與數位版本的保存方式記載，包括資料的更新與轉移)
- 使用性：用者或使用追蹤，展覽記錄，內容再使用
- 分析性
- 結構性
- 智財權性

Metadata的功能

- 資源探索
- 資源描述：包括相關資源的識別、相似資源的組合、不同資源的區別
- 資訊物件的檢索、保存與管理
- 多個部件的连接，提供資源內部的結構
- 追蹤智財權資訊，並指引或限制資源的使用方式
- 促進互通
- 數位識別
- 長久保存

Metadata標準的構面



Borrowed from:
David Bearman's WGSAD Matrix (1989)

完整的Metadata結構 之組成要件

- 語意
- 內容規則
- 控制詞彙
- 語法

Metadata的特質與屬性

- Metadata的來源
 - 內部metadata，來自資訊物件的建立者在產生或數位化的當刻
 - 外部metadata，來自資訊物件產生之後，通常並非物件創造者所建立
- Metadata的產生方法
 - 自動：關鍵字索引，用者使用檔
 - 人工
- Metadata的本質
 - 一般性
 - 專門性
- Metadata的狀態
 - 靜態・動態
 - 長期・短期

Metadata的互通與交換

- Metadata對照
 - 元素，語意，語法
 - 成功的對照與互通
 - 不同metadata格式間的相似性
 - 不同metadata格式的元素間之粒度性
 - 每一metadata格式的元素之內容規則的相容性

Metadata的互通與交換

http://www.getty.edu/research/conducting_research/standards/instrumetadata/pdf/crosswalk1.pdf - Microsoft Internet Explorer

Metadata Standards Crosswalk

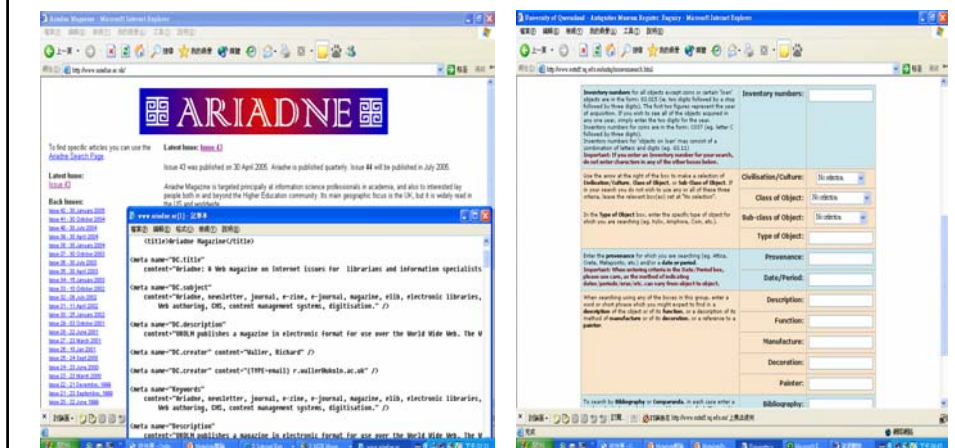
NOTE: To see the full list of categories for any one of the standards below, click on the name of the standard. To print from your Web browser, use the landscape print format.

CDFIA	CCO	YRA 3.0 Core Categories	USMARC	Dublin Core	Object ID	FLA Guide	CIMI	EAD*
OBJECT WORK (core)								
Object Work - Catalog Level						Document Classification-Catalogue Level (core)		LEVEL attribute
ObjectWork- Type (core)	Work Type	Type	ISS Genre-Form	Type	Type of Object	Document Classification-Document Type (core) Purpose-Purpose (Broad) (core) Purpose-Purpose (Narrow)	object NAME	<controlaccess=>genreform< (in <archdesc>)
ObjectWork-Components			ISS Physical Description-Extent	Format-Extent		Document Classification-Extent	quantity	<physdesc=>extent< (in <archdesc>)
CLASSIFICATION (core)	Class		ISS	Subject Classification (scheme)				
ORIENTATION/ARRANGEMENT				Description				
TITLES OR NAMES (core)	Title	Title Title Variant Title Translation Title Series Title Larger Entity	24XX Title and Title-Related information	Title	Title	Group/Item Identification-Repository Title Group/Item Identification-Descriptive Title (core) Group/Item Identification-Inscribed Title	object title bibliographic title	<descripter=>(in <eadheader>) <unititle=>(in <archdesc>)
STATE	Date		54XX Copy and Versions					FRDAOCTATUS attribute (in <eadheader>)

11 x 8.5 英寸 1 / 8

Metadata記錄的儲存方式

- 埋置在數位物件中
- 儲存在資料庫中

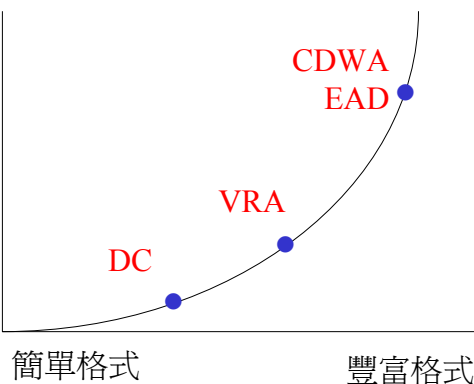


Metadata標準綜覽

標準的詳簡類型

特定性

一般性



Metadata標準綜覽

一般性後設資料

- 特色：簡單，具延展性，跨學科，跨資料類型。
- 典型代表：都柏林核心集 (Dublin Core)，強調支援數位典藏品的資源探索 (Resource Discovery)。

Metadata標準綜覽

特定性metadata¹

- 特色：著重特定領域資訊的共同需求與著錄標準。
- 典型代表
 - 圖書館社群：MARC, MODS
 - 博物館社群：CDWA, VRA, CIDOC
 - 檔案館社群：EAD
 - 政府社群：GILS
 - 視覺資源社群：VRA Core Categories
 - 植物標本館社群：HISPID, DarwinCore
 - 教育社群：IEEE LOM
 - 音樂社群：SMDL

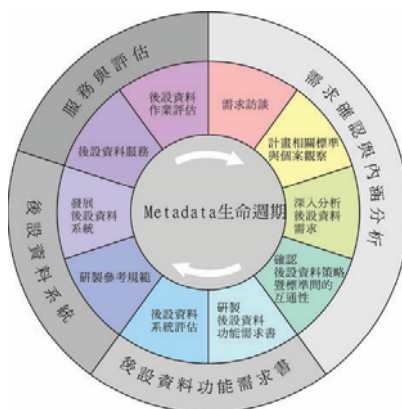
Metadata標準綜覽

特定性metadata²

- 典型代表
 - 出版業：ONIX
 - 電子商務：INDECS, ONIX
 - 數據型資料：ICPSR Data Documentation Initiative
 - 文字型資料：TEI
 - 靜態影像：NISO Dictionary for Technical Metadata for Digital Still Images
 - 地理資料：CSDGM (FGDC)
 - 視聽媒體資料：MPEG-4, MPEG-7 for Audio and Video

Metadata的方法學

- 如何規劃，設計，選擇，建置，評鑑，服務？
- Metadata生命週期作業模型



Dublin Core標準專論

Dublin Core基本概念

- 都柏林核心集(Dublin Core, DC)
 - DC metadata元素集是源自1995年，由OCLC與NCSA在美國俄亥俄州Dublin所召開的研討會，形成共識而成，因此取次會議的地名作為metadata元素集的名稱
 - 目前已成為相關國際與國家標準
 - ISO Standard 15836-2003(2003年2月)
 - NISO Standard Z39.85-2001 (2001年9月)
- 都柏林核心集metadata計劃(Dublin Core Metadata Initiative, DCMI)
 - DC的後續發展及相關規格書，目前由DCMI組織管理，主要活動包括：
 - 標準發展與維護(包括：舉辦國際會議，工作小組會議等)
 - 支援上述所需的相關工具，服務與基礎建設(包括DCMI metadata註冊中心)
 - 教育活動與支援

Dublin Core基本概念

- DC原來發展的旨趣，是以簡要的15個元素描述網路為基礎的文件，推出後廣受各領域及各資料類型的應用，也衍生更複雜的需求。為此，DC製訂修飾語(qualifier)，以及成立相關工作小組製訂應用檔案(application profile)
- 元素使用的特徵：
 - 每個元素皆為選擇性(可以採用或不採用)，及皆為可重複著錄
 - 每個元素可以任何排序呈現
 - 建議一些元素使用控制的值，但並非必備
 - 每個元素的內容著錄規則，由個別建置單位訂定
 - 鼓勵採用所屬特定領域的應用檔案，並遵守這些應用檔案規範的資料內容與資料值之規範或標準

Dublin Core簡介

- 發展背景
 - 1994, Chicago, 2nd Intl World Wide Web Conference, *semantics & the Web*
 - DC-1：1995年由，由OCLC與NCSA於美國Dublin, Ohio召開，提出*metadata semantics*想法（為網路資源發展一套核心組，將有助於網路分類，並進而讓其檢索更加容易）“Dublin Core metadata”由此定名(OCLC/NCSA Metadata Workshop)
 - DC-2：1996年4月由UKOLN (UK Office for Library and Information Networking) 及OCLC於英國Warwick共同舉辦(OCLC/UKOLN Metadata Workshop)
 - DC-3：1996年9月由CNI (Coalition for Networked Information)與OCLC於美國Dublin, Ohio共同舉辦(CNI/OCLC Workshop on Metadata for Networked Images)

Dublin Core簡介

- 發展背景
 - DC-4：1997年3月由OCLC, DSTC (the Distributed Systems Technology Centre) 與澳洲國家圖書館於澳洲坎培拉 (Canberra) 共同舉辦(NLA/DSTC/OCLC Dublin Core Down Under)
 - DC-5：1997年10月由OCLC與芬蘭國家圖書館於芬蘭赫爾辛基 (Helsinki) 共同舉辦(The 5th Dublin Core Metadata Workshop)
 - DC-6：1998年11月由美國國會圖書館與OCLC於美國華盛頓特區 (Washington, D.C.) 共同舉辦(DC-6: The 6th Dublin Core Metadata Workshop)
 - DC-7：1999年10月由德國國家圖書館與OCLC於德國的法蘭克福(Frankfurt am Main)共同舉辦(The 7th Dublin Core Metadata Workshop)

Dublin Core簡介

- 發展背景
 - DC-8：2000年10月由加拿大國家圖書館與OCLC於加拿大渥太華(Ottawa)共同舉辦(The 8th International Dublin Core Metadata Workshop)
 - DC-2001：2001年10月由日本NII (National Institute of Informatics)，DCMI及日本科技公司(JST)於日本東京共同舉辦，同時為第一屆的Dublin Core與metadata應用國際會議(International Conference on Dublin Core and Metadata Applications)
 - DC-2002：2002年10月由DCMI、義大利圖書館學會、義大利國家圖書館、科學史博物館、歐洲大學研究所, 非倫沙大學等於義大利佛羅倫斯共同舉辦(主題：Metadata for e-Communities: Supporting Diversity and Convergence)；同時也召開第10次工作會議

Dublin Core簡介

- 發展背景
 - DC-2003：2003年9月由美國華盛頓大學資訊學院、DCMI、華盛頓大學圖書館、雪城大學(Syracuse University)資訊研究所等共同於美國西雅圖主辦(主題：Supporting Communities of Discourse and Practice-Metadata Research & Applications)
 - DC-2004：2004年10月由中國上海圖書館、中國科學院圖書館、中國國家科技圖書館及DCMI共同於中國上海舉行(主題：Metadata Across Languages and Cultures)
 - DC-2005：2005年9月由西班牙馬德里的卡羅斯三世大學(Universidad Carlos III de Madrid)與DCMI於西班牙馬德里舉行(主題：Metadata Vocabularies in Practice)

Dublin Core的目標與定位

- 更容易找到資源
 - 發展出能夠探索(discovery)跨領域(across domains)資源的metadata標準。
 - 訂定出能夠讓不同套metadata間互通的架構(framework)
 - 發展出與特定領域/社群(discipline-specific) metadata標準間，能夠跨領域探索資源與 metadata 互通(interoperability)的架構。

DC基本結構 元素

Title (題名)	Creator (主要作者)
Subject (主題)	Description (描述)
Type (類型)	Format (型式)
Publisher (出版者)	Source (來源)
Language (語言)	Contributor (其他作者)
Date (日期)	Relation (關聯性)
Identifier (識別碼)	Rights (著作權)
Coverage (時空範圍)	

DC基本結構 元素

Title (題名)	Creator (主要作者)
Subject (主題)	Description (描述)
Type (類型)	Format (型式)
Publisher (出版者)	Source (來源)
Language (語言)	Contributor (其他作者)
Date (日期)	Relation (關聯性)
Identifier (識別碼)	Rights (著作權)
Coverage (時空範圍)	

DC基本結構

修飾語

- 二大類型：
 - 元素精確化(Element Refinement).
 - 這類的修飾語可使一個元素的意義更精細或更特定。
 - 元素編碼表(Encoding Scheme)
 - 此類的修飾語確認出能夠協助一個元素值的解釋之系統，包括：
 - 控制詞彙(controlled vocabularies)
 - 正式的標記方式(formal notations)
 - 或解析規則(parsing rules)
 - 如：來自分類系統或主題標目的詞彙，或與正式標記相符的字串(如：2005-07-23)

DC基本結構

應用原則

- 重要的應用原則：A grammar of DC
 - 簡單化
 - 二大層次的語言：elements, qualifiers(名，形)
 - 為數位遊客設計的洋涇濱語(A Pidgin for Digital Tourists)
 - 元素與修飾語
 - 15個元素：有助於跨資料庫與跨領域的檢索
 - Encoding schemes: 指標到相關資訊，解釋資料值
 - Element refinements: 使元素更特定與明確

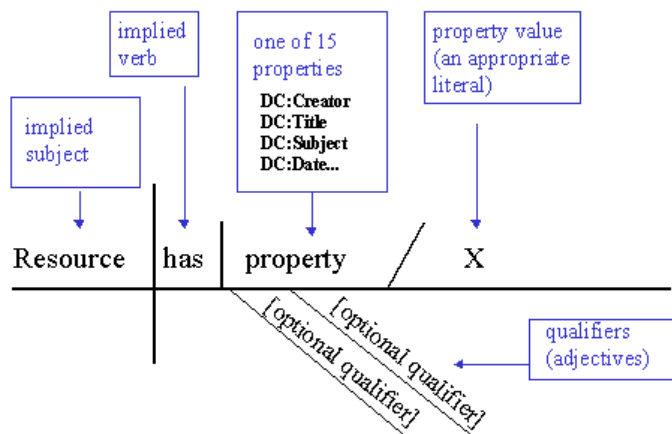
DC基本結構

應用原則

- 修飾語的應用原則：Dumb-Down Principle, refine vs. extend
- 元素值的應用原則：
 - 是否有助於資源探索? 是正確應用的核心觀念。通常，元素值會有預期出現的字串型式。
- 評估應用的良莠

DC基本結構

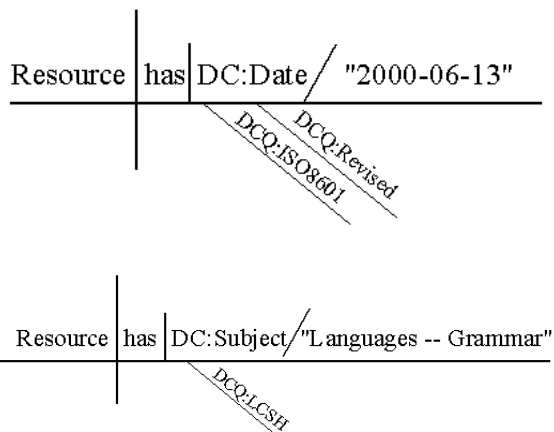
應用原則



資料來源: Baker (2000)

DC基本結構

應用原則



DC工具

- 工具
 - 詳細的工具清單 <http://dublincore.org/tools/>
 - 練習與試用 <http://www.lub.lu.se/cgi-bin/nmdc.pl>

DCMI工作小組

- Accessibility Working Group
- Agents Working Group
- Architecture Working Group
- Citation Working Group
- Collection Description Working Group
- Date Working Group
- DC-Kernel/ERC Working Group
- Education Working Group
- Environment Working Group
- Global Corporate Circle
- Government Working Group
- Libraries Working Group
- Localization and Internationalization Working Group
- Persistent Identifier Working Group
- Preservation Working Group
- Registry Working Group
- Standards Working Group
- Tools Working Group
- User Documentation Working Group