

《Metadata 專題》

Metadata 初探

陳亞寧、陳淑君

「數位典藏」、「數位圖書館」或是「數位博物館」，不論名詞如何翻譯、引用或解釋，資訊數位化及服務數位化已是大勢所趨。爲了一探並落實數位圖書館/博物館計畫（digital library and museum）內涵，國科會已成立推動數位圖書館與博物研究計畫，而中央研究院（以下簡稱本院）亦是其中之一。爲了支援本院數位圖書館與博物館研究計畫（以下簡稱 DLM）的物件描述與詮釋，而成立了 Metadata 工作小組，負責規劃 DLM 有關 Metadata 的事宜。

本院 Metadata 工作小組係自民國八十七年六月二十八日成立以來，已針對現行 Metadata 相關標準及研究計畫進行瞭解，同時工作進度已邁入第二階段。Metadata 工作小組將陸續把工作相關的成果依順整理、發表，而本文即爲小組第一次正式對外發表有關 Metadata 的研究成果與心得。

Metadata 定義及範疇與實質意涵

有關 Metadata 的定義、範圍及實質內容方面，已有許多相關文獻提出探討，本文依序分爲定義、Metadata 與資料的異同、範疇與實質意涵分述如下：

1、定義

有關 Metadata 的定義，一般泛稱爲「資料的資料」（data about data）【註一】，此一定義係源自在一九九五年三月由 OCLC（Online Computer Library Center）、NCSA（National Center for Supercomputing Applications）兩單位共同主辦名爲「Metadata Workshop」研討會，廣邀圖書館學、電腦科學、文獻編碼、以及相關領域學者專家等參加。在此會議中，首先提出了「資料的資料」作爲 Metadata 的定義【註二】。自此之後，有關 Metadata 的各種定義亦紛紛出現，試列舉如下：

• 描述資源屬性的資料 (Data which describes attributes of resources.) 【註三】

• 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.) 【註四】

• Metadata 是一種有關全球資訊資源或其他的機讀資訊 (Metadata is machine understandable information about web resources or other things.) 【註五】

• Metadata 就是資料 (Metadata is data.) 【註六】

• Metadata 是包括有關資料特性的資訊，但是 Metadata 必須對資料加以描述、詮釋，層次上應該涵蓋：人、物、時、地、原因及途徑等方向 (Metadata consist of information that characterizes data. Metadata are used to provide documentation for data products. In essence, Metadata answer who, what, when, where, why, and how about every facet of the data that are being documented.) 【註七】

2、Metadata 與資料的異同

究竟 Metadata 與「資料」是否有所不同呢？若從定義而言，Metadata 似乎應包含了資料，其實不然。若從 W3C (World Wide Web Consortium) 提出的「資源描述架構」(resource description framework, RDF) 中的解釋，即可印證發現兩者間並無不同：

資料與 Metadata 間並無明顯絕對性的差異，因為兩者常同時被用來詮釋同一資源。若有分別，也只是在特定的應用層面而已。(The distinction between “data” and “Metadata” is not an absolute one; it is a distinction created primarily by a particular application, and many times the same resource will be interpreted in both ways simultaneously.) 【註八】

因而，也可以推論 Metadata 與資料間似乎存有著某種關聯性或互動關係，而且彼此間的關係並不十分明確。所以針對此點，關於 Metadata 方面也有文獻提出不同的定義與解釋，如：

- Metadata 與資料並無差異 (There is no essential distinction between data and Metadata.)【註九】

- Metadata 可以用來描述 Metadata (Metadata can describe Metadata.)【註十】

- Metadata 與資料在行為及特性方面具有某種程度的雷同性 (Metadata and data have similar behaviors and characteristics.)【註十一】

- Metadata 可以用以描述某一文獻，或是在文獻中的文獻，或是不同文獻間等關係 (Metadata about one document can occur within the document, or within a separate document, or it may be transferred accompanying the document.)【註十二】

- 有關 Metadata 間的關係並不單純 (There is no single “about” relationship.)【註十三】

- Metadata 可以指引至某一具有統一辨識碼的資源 (Metadata may refer to any resource which has a URI.)【註十四】

- Metadata 可以存儲在任一種資源 (Metadata may be stored in any resource no matter to which resource it refers.)

【註十五】

從以上定義的探討，可以歸納出兩種推論：一為，Metadata 就是資料，二則 Metadata 與資料間存在某種互動關係，而且屬性不是單一、簡單化的。

3、範疇與實質意涵

究竟 Metadata 實際範圍與涵義包括了何種特性與項目呢？若從一九九七年九月十六及十七兩日在美國馬里蘭舉辦的第二屆 IEEE

Metadata 會議議題中可窺見一二，如下：【註十六】

- Metadata 的管理

- a. Metadata 的模式及呈現：在 Metadata 延展性的模式選擇方面之衝擊、Metadata 在儲存、檢索及資料使用上的衝擊、Metadata 在 WWW 中的架構邏輯/內容語義之呈現。

- b. Metadata 的採掘/粹取

- c. Metadata 的大量儲存及倉儲 (warehousing)

- d. 整合多層次的 Metadata

- e. Metadata 貯藏之處所

- f. 管理 Metadata 的系統：記錄/提供/檢索/使用、WWW 為架構基礎的 Metadata 管理。

- Metadata 在特定領域的應用

- a. 地球科學：如氣象、生態系統等。

- b. 資源管理：如採礦、農業等。

- c. 資料密集性高的其他應用領域：人口統計、影片檔案等。

- 全球資訊基礎建設中之相關應用

- a. 數位圖書館 Metadata 之議題：多媒體文獻的檢索、多媒體及異質性數位資料之 Metadata、Metadata 之大量儲存。

- b. 以 WWW 為架構管理的資訊資源之 Metadata

- c. 資訊/電子商業

- 標準

a. “政府文獻”：如電子郵件、聲音郵件。

b.磁帶格式及檔案的 Metadata

c.自行描述資料的格式：如 netcdf

d.現有標準的調查

e.進行不同 Metadata 間標準的轉換

• 實際運作中與 Metadata 相關之議題

即使將於今（一九九九）年四月份舉行的第三屆會議，亦可發現內容大同小異，如下【註十七】：

• Metadata 的

a.管理與模式

b.在 WWW 上的語意表達

c.大量儲存及倉儲

d.WWW 模式資訊資源的管理

• Metadata 發掘與萃取

• Metadata 目錄

• Metadata 管理系統

• Metadata 特殊應用

• Metadata 在數位圖書館的議題

• Metadata 有效發展的方式

• 標準

至於真正的 Metadata 內容究竟為何呢？其實圖書館長久以來建立、維護與提供的線上公共目錄的書目資訊即是其中一個典型範例，至於 Metadata 的應用層面，也可以從本文列舉的範例窺知梗概（如表一所示）。但是 Metadata 範疇只限於此？其實依據本院 Metadata 工作小組研究與實作心得證實，Metadata 除了在於描述、詮釋資料與書目資訊相同外，但在層次與深度方面，與傳統書目資訊是有所不同的。主要差異有四個方面：一為，任何物件(object)皆是 Metadata 涵蓋的範圍，諸如圖書館界所熟悉的書、期刊、文章等，乃至器物、人等皆是。二則為，範圍從實體典藏的物件擴大至虛擬典藏的各類物件。三是，詮釋的深度遠比以往更為深入，並不僅限於內容主題的分析標引而已，尚包括了物件彼此互動關係，包括人、時、地、物、主題/事件（events）等五大主軸間的互動、牽引；換言之，從資料、資訊的整理提昇至知識內涵的建構。四為，物件的辨識、保存、展示、取用、篩選與評估、服務及系統管理等方面皆是環環相扣，密不可分，並不可從單一觀點視之或處理。

Metadata 的重要性與功能

究竟 Metadata 對於數位圖書館有何重要性呢？難道沒有了 Metadata，數位圖書館就無法順利建置完成？堅信前述兩個問題應該是許多人好奇，或是想一探真實的熱門話題。如果追溯 Metadata 一詞的起源，其實是來自網路資源的失控，亦即網路用者無法適當地找出所需的資訊。即使 Metadata 起源於此，卻可以明確發現 Metadata 的應用早已超越此一範疇，已延伸至圖書館等不同資料單位既有館藏資源、紙本或其他傳統形式媒體轉為數位形式、以及全然煥新的電子資訊等三大領域。因而，數位圖書館範圍也隨之擴展，而 Metadata 也成為數位圖書館的必備基礎，因為不可能將所有物件內涵全文予以查詢、檢索，成本效益畢竟不高，而且資源有限。

依據 Heery Rachel 觀察分析目前網路資源的無序特性，提出了下列六大發現：資料分散放置各處（locations）、相同資料擁有不同版本（document versions）、資料存放位置易於改變（lack of stability）、重覆過時的資料充斥於網路上（redundant data）、資料索引的層次不夠深入與重覆索引的問題（granularity），以及缺乏相關資源的檢索說明（nature of location data）等。因而，在荷蘭的 DESIRE（Development of a European Service for Information on Research and Education）研究計畫中，就提出

Metadata 應具備下列五大功能：位置標示指引 (location)、尋找或發掘 (discovery)、歷史回溯追蹤 (documentation)、評鑑 (evaluation) 及篩選過濾 (selection) 等【註十八】。若依本院 Metadata 工作小組研究實證結果，其實 Metadata 應具備下列八項功能：

- 1.資料架構與模式 (structure & model)：設計一個共通性組織結構，以容納不同類型與學科領域的 Metadata。
- 2.資料輸入與描述整理 (input & descriptive organization)：為典藏品資料建立一套詮釋性的記錄。
- 3.檢索與索引 (retrieval & indexing)：讓使用者很有效率地進行查詢這些記錄。
- 4.展現與辨識 (representation & identification)：從查得的記錄中，使用者可以清楚的獲得所需的訊息及制訂呈現方式。
- 5.串聯與互動關係 (linkage and interactive relationship management)：建立不同文獻間的串聯架構、方向 (雙向與多向)、模式與管理等。
- 6.取用與認證 (access & authentication)：作為系統安全控制的機置功能之一，以區別不同身份的使用者，包含智財權 (intellectual property rights) 的管理與控制。
- 7.交換與儲存 (interchange, mapping & exchange and storage)：這些詮釋性記錄可以因各種不同需求 (包括書目與全文兩部份) 而進行交換及儲存。
- 8.整合 XML (extensible markup language)、RDF 與 Z39.50 不同協定的應用：除了致力於 Metadata 的制訂，因應文獻結構的制訂、交換、檢索與展現的需求，另結合 XML、RDF 與 Z39.50 等協定的應用，以發揮 Metadata 的功能。【註十九】

全球 Metadata 發展現況與研究計畫

目前至少已有 20 種以上屬國際標準或漸形成標準的 Metadata 格式

存在於各學科領域，Lorcan Dempsey 等人曾依 Metadata 格式的繁簡度及特質概分為三大組：【註二十】

1.第一組：簡單格式（simple formats）

此類多是由某公司獨家製造與專屬。係全文式索引網路資源（例如：Lycos、Altavista、Yahoo 等），通常由機器自動擷取資源後直接索引並提供檢索。

2.第二組：結構化格式（structured formats）

此類係正在標準化中的格式，以欄位為架構來索引相關資源。例如，Dublin Core、IAFA templates、RFC1807、SOIF、LDIF 等。它們所包括的描述資料基本上足以讓使用者找出有興趣之相關資源。這類型格式記錄以簡單的設計為原則，未受專門訓練的資料提供者，也能容易上手建立 Metadata 記錄。但是相對地，格式的設計上較缺乏物件或資源間不同層次的關聯性功能。

3.第三組：複雜格式（rich formats）

此類為已符合國際標準的格式，以詳細的標誌（elaborate tagging）來呈現資源的內涵。例如：ICPSR（Inter-university Consortium for Political and Social Research）、CIMI（Computer Interchange of Museum Information）、EAD（Encoding Archival Description）、TEI（Text Encoding Initiative）、MARC 等。這類型的格式通常需由研究學者或具專門知識者來描述資源的內涵。

而在國內，台大陳雪華教授將 Metadata 依主題歸納為八大類【註廿一】如下：

1.早已普遍使用的格式

- MARC：如 USMARC, UKMARC, UNIMARC
- PICA+（the Dutch Centre for Library Automation）

2.描述科技文獻

- BibTeX
- EELS (The Edinburgh Engineering Virtual Library)
- RFC1807

3.描述人文及社會科學資源

- ICPSR SGML Codebook Initiative
- TEI Headers

4.描述政府資訊

- GILS (Government Information Locator Service)

5.描述地理空間性資源

- FGDC (Federal Geographic Data Committee)

6.描述博物館藏品與檔案特藏

- CDWA (Categories for the Description of Works of art)
- CIMI
- EAD

7.描述大量網路資源

- Dublin Core (Dublin Metadata Core Element Set)
- IAFA/WHOIS++ Templates (Internet Anonymous FTP Archive)
- LDIF (LDAP Data Interchange format)
- SOIF (Summary Object Interchange Format)

- URCs (Uniform Resource Characteristics /Citations)

8.其他

- Warwick Framework

Metadata 相關研究計畫及參考資源，可由下述 UKOLN (The UK Office for Library and Information Networking) 及 IFLA (International Federation of Library Associations) 兩個機構所提供的 Metadata 網站獲得相當完整而集中的訊息：

1. UKOLN Metadata page

(<http://www.ukoln.ac.uk/Metadata/>)

此網站係由『英國圖書館與資訊網路辦公室』(即 UKOLN) 所維護，網站資源包括該辦公室參與的七大資源描述相關之計畫介紹，如：BIBLINK (Linking Publishers and National Bibliographic Services)、DESIRE (Development of a European Service for Information on Research and Education)、ROADS (Resource Organisation and Discovery in Subject-based Services) 等。此外，尚提供豐富的 UKOLN 相關研究報告、各式 Metadata 的網站、相關配合標準 (如：CORBA、MIME、Unicode、SGML、XML、Unique Identifiers)、不同 Metadata 格式間的對照表、處理 Metadata 的相關軟體工具、相關詞彙清單等等。

2. The IFLA Metadata Resources page

(<http://ifla.inist.fr/II/Metadata.htm>)

此由國際圖書館學會聯盟所維護的 Metadata 資源網頁，也提供非常齊全而完整的相關資訊，包括各式 Metadata 格式的相關探討文獻，以及相關會議和研討會的訊息。

Metadata 評估準則要項

現存已有許多 Metadata 標準，如何選擇與評鑑一個適用的 Metadata 格式就顯得十分重要。本節試從探討有關 Metadata 評估要點，作為圖書館事前選用適合的 Metadata，以及使用後的評鑑。茲

分為下列要點說明：【註廿二】

1. 使用環境 (environment of use)

Metadata 格式的使用環境可從下述三方面著手評析：

- 使用群 (constituency of use)

誰是此 Metadata 格式的主要使用群？他們是否屬於某一特別專業領域或學科範圍？例如：USMARC 被普遍使用於美、加及澳大利亞等國家的圖書館界；而 EAD 則是為了檔案及手稿方面的文件利用而發展。瞭解使用群的特性，有助於初步判斷該 Metadata 格式的適用性，並可探索這些使用單位的應用情況及發展趨勢。

- 建立資料的容易度 (ease of creation)

資料提供者（如：學科專家）是否容易自行建檔？抑或需賴資訊服務提供者（如：圖書館員）代為協助建檔？例如：FGDC 標準為一擁有超過 300 項 (Dublin Core 只有 15 項) 資料欄位的複雜型 Metadata，因此必須再另外開發許多工具以協助建檔，同時該標準的工作小組也需要支援相關的教育訓練課程。

- 國際標準化的程度 (progress towards international standardization)

雖然有許多的標準制訂機構近年來積極地發展 Metadata 相關標準與機制，但目前為止尚未誕生任何一種已完全符合國際標準的 Metadata，即便是 USMARC 中有關網路資源的著錄欄位也仍屬討論、研議階段。瞭解不同 Metadata 格式的國際標準之進展，有助研判該 Metadata 的成熟度，以及未來繼續發展的可行性。

2. 格式相關議題 (format issue)

評估 Metadata 格式可從內容要項、描述項目的建構規則、多語言的處理能力、顯示物件間關聯性的能力、完整性等五方面加以探討如下：

- 內容要項 (content)

為讓 Metadata 的功能充分發揮，原則上 Metadata 記錄的內容要項應該包含：

a.基本描述項目

意指對物件本身書目性的描述，例如：物件的題名、作者、日期、出版者、提供者、地理位置、形式、實體描述等。

b.主題描述

對物件的本質作更深入的記載，例如：事件（event）、主題、關鍵詞、材質等。

c.統一化資源識別碼

物件編號包括 ISBN/ISSN、URL、URN、或任何該物件獨一的編號，以資識別於其他文件。

d.資源格式及技術性的特質

常見的電子檔包括 html、pdf、postscript、powerpoint、rtf、latex、gif、jpeg、plain text、sgml 等不同格式，這些不同的檔案格式會有相對應的解讀軟體。

e.系統及使用權管理細節

包括控制權限的管理（如：login in、password、IP、domain name）、使用說明（如：是否有任何時間、地域以及身份限制），URL 及 URN 等考量細節。

f.行政管理之記錄

含括記錄類型、建檔者/日期、審核者/日期、最新審核日期、最新更新者/日期、修改歷史檔等相關訊息。

g.來源出處

物件來源出處的記載。

- 描述項目的建構規則 (rules for the construction of these elements)

建立 Metadata 時，是否提供任何可以遵循的規則？例如：MARC 的描述欄位規定須嚴格地依照英美編目規則 (AACR2) 來著錄，相對地此套規則也深深影響到該格式描述項目結構。

- 多語言的處理能力 (multi-lingual issues)

描述或使用其他語言的相關標準及機制是否已被考量？例如：GILS 的資料著錄項包含資料的語言 (Language of resource) 以及記錄本身的語言 (Language of record) 等指示多語言的相關欄位。

- 顯示物件間關聯性的能力 (ability to represent relationships between objects)

不同內容的物件之間容許呈現多種層次的關聯性描述，例如：參照 (references)、發源 (based on)、版本 (version of)、型式 (format of)、母子 (part of) 等等多種可能的關聯性。

- 完整性 (fullness)

Metadata 設計的原旨會影響其內容的完整性，例如：原來的 Dublin Core 完整性並不高 (12 個資料項)，經過六次的研討會議以及相關單位的實作後，決定朝向較詳盡及完整記錄的方向修正 (增為 15 個資料項，並且開始研議如何藉由修飾詞等機制進一步地詮釋及精緻化記載、標引的內容)。

3. 網際網路相關之協定 (protocol issues)

目前一些資源發掘 (discovery) 系統可經由網路上的分散式檢索及目錄 (如：whois++、LDAP、Dienst) 等方式達成探索功能；另一方面，愈來愈多完整性高的系統 (如：MARC、CIMI、TEI、EAD 等) 開始採用 Z39.50 來進行檢索。

4.實施應用的狀況（implementation）

從各 Metadata 相關的應用計劃可觀察到此格式的優勢與不足之處，同時可得悉此格式發展趨勢，甚至可與相關單位互相觀摩。

結語

從網路資源失控為發端的 Metadata，發展、應用範圍早已跳脫既有形式。因而，Metadata 應用層面也從書目層次擴展至全文，乃至知識內涵與建構。在應用 Metadata 前，除了必須對其定義及範圍加以釐清外，如何選擇與評估亦十分重要，三者缺一不可。未來 Metadata 如何應用發展，且讓我們拭目以待。

附註：

註一：Stuart Weibel, Jean Godby, and Eric Miller, “OCLC/NCSA Metadata Workshop Report,”[Access Date : 25 December, 1997].

http://www.oclc.org/oclc/research/conferences/Metadata/dublin_core_report.html

註二：同註一。

註三：Lorcan Dempsey, et al., Specification for Resource Description Methods.

Part 1. A Review of Metadata : a survey of current resource description formats. Version 1.0. 19 March, 1997. [Access Date : 24 May, 1997].

[http://www.ukoln.ac.uk/Metadata/DESIRE/overview directory.](http://www.ukoln.ac.uk/Metadata/DESIRE/overview_directory)

註四：Alison Chilvers, and John Feather, “The Management of Digital Data : a Metadata approach,” The Electronic Library 16(6)(1998) : 365.

註五：Tim Bernes-Lee, “Metadata Architecture,” January 1997. [Access Date : 7Decemember, 1997].

<http://www.w3.org/DesignIssues/Metadata>

註六：同註五。

註七：“Tools for Creation of Formal Metadata : frequently-asked questions on FGDC Metadata,” 20 July,1998. [Access Date : 29 August, 1998].

<http://geology.usgs.gov/tools/Metadata/tools/doc/faq.html>

註八：“Resource Description Framework (RDF) Model and Syntax Specification,”W3C Working Draft, 8 October, 1998. [Access Date : 25 November, 1998].

<http://www.w3.org/TR/WD-rdf-syntax>

註九：Ron Daniel, and Carl Loagoze, “From Metadata containers to active digitalobjects,” D-Lib Magazine Novemember 1997. [Access Date : 18 November, 1997].

<http://www.dlib.org/dlib/november97/daniel/11daniel.html>

註十：同註五。

註十一：Carl Loagze, Clifford A. Lynch, and Ron Daniel, “The Warwick Framework : a container architecture for aggregating sets of Metadata,” [Access Date : 20 August, 1998].

<http://www.ifla.org/documents/libraries/cataloging>

/Metadata/tr961593.pdf

註十二：同註五。

註十三：同註九。

註十四：同註五。

註十五：同註五。

註十六：” Second IEEE Metadata Conference : call for papers,” [Access Date : 24 Februray, 1999].

http://www.llnl.org/liv_comp/Metadata/metad97.CFP.html

註十七：” Three IEEE Metadata Conferfence,” [Access Date : 24 Februray, 1999].

http://www.llnl.org/liv_comp/Metadata/md99/md99.html

註十八：同註三。

註十九：中央研究院 Metadata 工作小組，「Metadata 工作小組第一階段報告：以平埔族計畫實作為例」，民國八十八年一月七日。

註二十：同註三。

註廿一：陳雪華，「網路資源組織與Metadata之發展」，圖書館學刊，第十二期(民86年12月)，頁23-28。

註廿二：本節係綜合三份文獻而成，請參見下列資料：

Rachel Heery, “Review of Metadata Fromats,”
Program 30(4)(October 1996) : pp. 345-373.
[Access Date : 25 February, 1999].

<http://www.ukoln.ac.uk/Metadata/review.html>

Lorcan Dempsey, et al., Specification for Resource Description Methods.

Part 1. A Review of Metadata : a survey of current resource description formats. Version 1.0. 19 March, 1997. [Access Date : 24 May, 1997].

<http://www.ukoln.ac.uk/Metadata/DESIRE/overview> directory.

中央研究院 Metadata 工作小組，「Metadata 工作小組第二階段報告：以平埔族書目、古文契書及研究人員為例」，民國八十八年二月二十三日。