

圖書自動化

從書目控制到網路資源探索談 Metadata—由 Dublin Core 談起（上）

陳亞寧

編按：網路資源的無序化與混沌化等現象，已使得網路電子資源的應用與搜尋皆發生了不同層次的問題。面對這個問題，圖書館試圖以書目控制的方式來達成，而電腦界則趨向採取網路資源探索的方式來完成，但是兩者皆有優缺點。

因而，為了因應解決這類問題，Metadata 遂應運而生。但是 Metadata 與圖書館目錄有何不同？本文從討論網路資源組織整理方式的現況，進而以 Dublin Core 為例進行探求 Metadata 的應用與發展，次而從 Metadata 的定義、目的及索引層次與深度，反思 Metadata 的定位。最後，作者提出一連串問題作為應用、發展 Metadata 的參考依據。

前言

昔日圖書館是以所謂的書目控制 (bibliographic control) 方式進行資料的整理，以作為提供各項服務的基礎機制。近年來，隨著網際網路與電子出版品的盛行，已有許多電子資源經由網路及其相關設施的應用，將各項資訊提供給一般使用者使用。基於此種環境需求下，強調的重點是資源探索 (resource discovery)，而 Metadata 也應運而生。原先書目控制的理論與應用是否仍然適用在網路資源？本文試從圖書館學的書目控制為起點進行探討，進而探索 Dublin Core 的發展現況與應用，最後回歸研究 Metadata 的起源、目的與應用。

整理網路資源的現況分析

目前網路資源已隨著網路的普遍應用而散佈在各處，一般而言，整理方式約可分為微觀 (micro approach) 與宏觀 (marco approach) 兩種型態，分述如下：

1. 微觀：依據美國圖書館學會圖書館與資訊科學詞典 (The ALA Glossary of Library and Information Science) 解釋，所謂

「書目控制」(bibliographic control) 包括了出版品完整的書目記錄，書目描述的標準化，經由圖書館聯盟、網路與其他合作方式提供實體的取用，以及經由聯合目錄、主題目錄與書目服務中心提供書目取用等活動，皆稱為書目控制【Young, 1983, p.21】。基於網路資源的興起與應用，圖書館深刻體認網路資源的重要性與必要性，因而也開始著手整理。就圖書館而言，網路資源或電子資源只是眾多資訊載體形式之一，書目控制並不因媒體形式不同而有截然不同的處理方式。因而，圖書館基於擴大館藏可利用資源的前提下，亦著手進行蒐集整理。但是圖書館是以現有館藏重點與使用者需求為主要導向，將現有館藏與網路資源合而為一，且經過篩選。通常以圖書館現有的 OPAC 為基礎，並結合 WEB Technology 形成所謂的 WebPAC。此法的優點是資訊精準度高，但是相對可提供的資訊數量較少、成本高、速度慢、新穎性低，同時整理人員需要具備圖書館專業的技能與素養。若從書目控制理論與實務觀點而言，現存網路資源存有下列各項缺失，致使網路資源的整理、檢索與查詢產生許多始料未及的問題，主要項目條舉如下：

- 無法明確標示文獻內容；
- 各文獻間關係模糊不清；
- 參照關係不存在；
- 實體與虛擬館藏無法順利融合為一；
- 內容不確定性；
- 網路資源資訊數量多，但品質方面未達應有的要求與控制。

2.宏觀：所謂的網路資源探索，乃是利用資訊檢索的技術來查詢、找尋存放在網路上的資源，如 Archie、Gopher、WAIS，乃至近年來最為盛行的全球資訊網（World-Wide Web，WWW）。一般而言，網路資源探索的方式是以某一學科主題或區域（實體與虛擬兩種，前者有如現有的地理或政治實體，後者如網路的空間領域）為主要訴求範圍，通常是以結合搜尋引擎與主題指引兩大類型為主，並以 Web Homepage 方式呈現。前者如 AltaVista、

MetaCrawler 等，後者如 Yahoo。一般而言，所謂的搜尋引擎是以電腦程式形式(即 robot)，自動在網際網路空間蒐尋各式 Web Page HTML (HyperText Markup Language) 內的 Meta Tag 為主要依據，再做成索引提供給網路用者使用。而主題指引則是除了依 Robot 自動蒐集的資訊外，並經由人工的介入與整理分析而成。這兩種方式各有優缺，前者是省時省力，但資料精準度有待商榷；後者雖然資料較為精確（類似以圖書館整理組織的方式），但所需人力與成本相對提高，同時資訊新穎性亦不如前者。圖書館除了應用現有 OPAC 或 WebPAC 結合現有網路資源外，亦有採取此種方式，同時結合搜尋引擎與主題指引，其中最負盛名者如英國的 BUBL Link (<http://www.bubl.ac.uk>)、美國的 Internet Public Library (IPL, <http://www.ipl.org>) 等皆是。

由於網路資源數量眾多且龐雜，因而圖書館與電腦界採取了不同方式試圖有效組織現有的各式網路電子資訊資源，然而優缺互見，遂有人提出 Metadata 一詞與觀念，嘗試以不同的途徑與概念來突破現有的窠臼模式。因而，各學科領域也針對本身需求提出了不同的 Metadata 格式 (format)。本文將以 Dublin Core 作為 Metadata 的起始探討，藉以一窺 Metadata 的真實面貌。

Dublin Core 緣起

為了因應網路上各式電子資訊資源，圖書館也嘗試以既有的分類編目理論與實務為基礎，進行各式網路化電子資訊資源的書目控制。對一般使用者而言，既不具分類與編目基礎，更不熟悉分編等相關工具，因為這些原本專業工作是需要相關專業技能的。為能加速網路電子資訊資源的整理與組織，而且加強對現有網路資源的找尋與檢索的精準度，及篩選不必要的資訊，以減緩資訊超載與爆炸 (information overload and explosion) 的問題，Metadata 應運而生，而 Dublin Core 即是其中之一。

追溯 Dublin Core 的起源發端，可從一九九五年三月 OCLC 與 NCSA 聯合召開第一屆 Metadata Workshop 談起。在此研討會中，主要宗旨有二：

1. 對於現有資料擁有者 (stakeholders) 的需求，以及現有整理資源方式的優缺點等方面進行通盤性瞭解。

- 2.對於描述網路資源的 Metadata 核心集（core set）達成共識。【Weibel, Godby, Miller, and Daniel, 1995】

基於上述兩大宗旨前提下，此次研討會歸納 Dublin Core 具備下列六大目的導向：

- 1.以 Dublin Core Elements 來描述網路資源；
- 2.制訂小而全球通用的 Metadata Elements；
- 3.達成不同網路資源探索工具（resource discovery tools）間的透通性（interoperability）與應用；
- 4.由作者、出版者自行提供 Metadata，以利網路蒐尋引擎自動地蒐集與擷取。
- 5.鼓勵網路出版工具（network publishing tools）能夠內含簡化的制式資料表格，以提供作者、出版者輸入 Metadata 相關資訊；
6. Dublin Core 可以作為圖書館書目控制記錄的基礎（basis），以利圖書館作為進一步再利用。一旦類似 Dublin Core 的 Metadata 變成標準時，各領域使用者皆可理解【Weibel, Godby, Miller, and Daniel, 1995】。

在此次研討會中，Dublin Core 已完成 13 個元素 (elements) 的制訂【Weibel, Godby, Miller, and Daniel, 1995】，隨著 Dublin Core 的應用、科技發展與使用者需求等情形下，截至一九九九年十一月為止，已召開了七次會議，約略可歸納出下列現象：

1. Dublin Core 的研發制訂一開始是經由年會的舉辦，召集相關應用單位共同討論 Dublin Core 的制訂。截至第六次會議，Dublin Core 年會決議將各元素的制訂委交各工作小組（working group）來執行制訂後，再依序提交給政策委員會（Policy Advisory Committee, PAC）與技術委員會（Technical Advisory Committee, TAC）審查准可。
- 2.由單一 Metadata 的發展，逐漸朝向整合其他 Metadata 格式的事實；如 Warwick Framework、RDF（Resource Description Framework）、XML

(eXtensible Markup Language) 等。

- 3.原本 Dublin Core 是朝向簡單為主要導向，但現況應用發展上而言，已隨著附元素 (subelement) 或限定詞 (qualifier) 的應用，而趨向複雜。
- 4.由原有以電子資源為主，逐漸擴展應用至紙本式或由紙本轉為電子媒體的資訊資源。

Dublin Core 發展現況

自一九九五年以來，Dublin Core 的發展已引起全球的矚目，而 Dublin Core 研發至今究竟有何重要性，本節試從 Dublin Core 的制訂原則、元素說明與重要成果等三方面進行探求。

- 1.制訂原則：在 Dublin Core 設計時，即已鎖定使用容易、全球通用、彈性高等三項為主要訴求，因而制訂下列六大原則，說明如下：【Weibel, Godby, Miller, and Daniel, 1995】
 - a.內在固有性 (intrinsicity)：應用 Dublin Core 時有一重要前提，即是以物件(object)本身為範圍，並不涉及其他相關物件。因而在應用時，詮釋範圍僅以整理的物件對象為限，並不包含其他；至於其他相關物件則利用 Relation 元素予以標示、指引。
 - b.延展、擴充性 (extensibility)：為能包含各類電子資源的特性，同時能廣泛應用在不同學科領域，因而 Dublin Core 本身制訂的各項元素是以全面通盤性為考量基礎，並不為任何學科或特定某一資料類型而設立。如果使用者或單位必須進一步描述、分析、詮釋時，各使用者與單位可以視實際情形，就現有元素再進一步層次地制訂更詳盡的附元素，或是利用限定詞來達成。
 - c.語法獨立性 (syntax-independence)：由於 Dublin Core 的定位是在全球通用的基本考量前提下發展的，目前著錄語法是按 SGML (Standard Generalized Markup Language)、DTD (Document Type Definition) 的要求來記載描述的，並不因語文、文化與學科的不同，而有不同的著錄語法；亦即語法跨越學科領域。
 - d.選擇性 (optional)：在 Dublin Core 集中的任一個元素皆是可以視資

料內涵特性及單位需求，進而選擇性使用，而非強迫性全盤使用。但在使用前必須詳閱每個元素的定義、範圍與應用層面，方能正確使用。

- e.可重複性 (repeatability)：此一原則與上述選擇性雷同，每個元素皆可重複使用，並非只能使用一次。例如，DATE 元素可以標引出版日、複製日、發行日、出土日、發現日等不同類型的日期。
- f.可修改性 (modifiability)：有時資料內涵特性或單位使用需求無法與現有 Dublin Core 十五個元素完全符合應用時，使用者可以找出同質或雷同性較高的元素，予以解釋、重新定義後再使用，以符合實際作業需求。例如前述的附元素與限定詞的制訂，皆是針對現有 Dublin Core 十五個元素進行實作後提出的建議。

在中央研究院實作經驗中，亦可發現 Dublin Core 尚有一大特點—連鎖 (concatenation)；亦即不同元素間可以聯合串接使用，並達成類似附元素或限定詞的功能，類似英國的普列式索引 (Preserved Context Indexing System, PRECIS)。例如 DATE 元素，如前述範例所舉，有關出版日期其實可以和現有的 PUBLISHER 元素予以連鎖使用，達成資料的整體化與結構化的要求，對於資料庫設計、資料建檔與取用辨識等方面皆有很大的助益。

- 2.元素說明：一九九五年起，Dublin Core 只有十三個元素，至一九九六年第三次會議時，又擴增了 Description 與 Rights-Management 等兩個元素。大體而言，Dublin Core 元素可分為三種形式：內涵 (content)、智財權 (intellectual property) 與指示性 (instantiation) 等。所謂內涵乃是標引資訊內容與外在形式的特徵；而智財權則兼具使用與系統管理等兩種層次，期以對資訊的創作、發行、使用與管理能符合智財權相關規定的要求。至於指示性，則是將資料具有明白告知的屬性予以標示，而使用者無需特別指引或說明即可瞭解。目前國內吳政叡教授已將 Dublin Core 譯成中文 (<http://dimes.lins.fju.edu.tw/dublin/>)，至於十五個元素分佈、歸屬情形如下：

- a.內涵：Coverage、Description、Resource Type、Relation、Source、Subject/Keyword、Title。
- b.智財權：Contributor、Creator、Publisher、Rights-Management。

c.指示性：Date、Format、Identifier、Language【Weibel, and Haka la, 1998】。

至於限定詞與使用手冊（ user guide ）等方面， 目前仍處於發展與討論階段， 尙未定案， 而有些 Dublin Core 元素亦有相同狀況； 如 Resource Type、Relation 等元素。

3.重要成果：截至目前為止， Dublin Core 已對現有網路資源探索與 Metadata 具有相當大的影響力。就網路資源探索方面而言， Dublin Core 已可放在標準 HTML 語法與格式中， 因而網路上的各式搜尋引擎可以利用 Dublin Core 擷取資料， 而使用者亦可找到較為精準的資訊。就 Metadata 觀點而言， 由於 Dublin Core 具有核心（ core）、通用（ universal）的屬性特色， 因而現今各式 Metadata 格式皆以 Dublin Core 作為基礎研發， 同時 Dublin Core 亦作為各式 Metadata 轉換的對象； 亦即 Metadata 中的交換格式， 一如 UNIMARC 之於一般 MARC（ Machine-Readable Cataloging）。至於 Dublin Core 重要成果列舉如下：

a.結合使用 HTML 標準：即應用 HTML 4.0 的 META TAG 來標示 Dublin Core。

b.獨立自主的（ stand-alone） Metadata。

c.可以放置在電子資源內（結合利用 HTML 之故）。

d. DATE 元素遵循 ISO 8601。

e.創立 Warwick Framework 一詞與觀念，促成 RDF 的研究發展。

f.多語文：目前已有英、阿拉伯、捷克、丹麥、荷蘭、法、芬蘭、德、希臘、印尼、義大利、日、韓、挪威、葡萄牙、西班牙、泰、土耳其與中文等不同版本。

g.完成附元素報告初稿（ sub-elements draft report）。

h.結合應用 Z39·50。

i. Dublin Core 已納入在 Z39 · 50 的 List of Bib-1 Use Attributes 內。

j. 已成為 IETF' s RFC 2413 及 2713 文獻。【Weibel, 1999】(下期續)

參考書目

- Weibel, S. (1999). The state of Dublin Core metadata initiative. D-Lib Magazine. [Online, Access Date : 19 April, 1999]. Available : <http://www.dlib.org/dlib/april99.04weible.html>
- Weibel, S., Godby, J., Miller, E., and Daniel, R. (1995). OC LC/NASA/ Metadata workshop report. [Online, Access Date : 24 August, 1998]. Available : http://www.oclc.org:5046/oclc/research/conferences/metadata/dublin_core_report.html
- Weibel, S., and Hakala, J. (1998). DC-5: the Helsinki metadata workshop: a report on the workshop and subsequent developments. D-Lib Magazine. [Online, Access Date : 20 August, 1998]. Available : <http://www.dlib.org/dlib/february98/02weibel.html>
- Young, H., ed. (1983). The ALA glossary of library and information science. Chicago: American Library Association.