

電子資訊的組織模式

Organizational Models of Electronic Information

陳昭珍

Chao-chen Chen

輔仁大學圖書資訊學系副教授

Associate Professor

Department of Library and Information Science

Catholic Fu-Jen University

【摘要】

隨著網際網路的普及，電子資訊已漸漸成為重要的出版形式，而有效的電子資訊組織模式，也成為電子資訊檢索與利用的重要課題。相關電子出版或處理單位，紛就此一課題提出解決之道，本文之目的即在探討目前已發展出來的電子資訊組織模式，說明其發展背景、應用環境及相關問題。

【ABSTRACT】

Along with the popularity of Internet, electronic information is getting more and more important. To find efficient organization methods of electronic information is the main issue for information retrieval. Many solutions have been proposed.

The aim of this article is to discuss the existing organizational models of electronic information. We explain their backgrounds of developments, their environments of applications and some related issues.

關鍵詞 Keywords:

資訊組織；電子資訊

Information organization ; Electronic information

壹、前言

由於網路資源無限制的增加，使得網路資源的找尋成為資訊界與使用者共同關心的課題。有效的網路資源查尋，端賴於有效的資源管理，而有效的網路資源管理，則有賴於好的組織與索引模式。面對網路電子世界的衝擊，傳統圖書館也在急劇的轉型中，過去以自有館藏為主的服務，目前已擴充為三種層次的資訊的服務：（註1）

第一層：自有館藏及合作館之資訊；

第二層：除了第一層外，再加上商業資訊提供者所提供的資訊，如以專線或透過Internet來檢索的線上資料庫；

第三層：除了第一、二層外，再加上網路上廣大的免費資訊。

由上可知，轉型後的電子圖書館將擁有大量的電子資訊。而如何有效的管理與存取電子圖書館資料庫？吾人認為除了提昇CPU的性能、擴充儲存容量、及改善網路頻寬等方法外，有效的組織模式是電子資訊查尋、檢索、交換、管理等的基礎。因此本文將探討目前已發展出來的電子資訊組織模式，說明其發展背景、應用環境、優缺點及相關問題。

貳、為何電子資訊的組織問題值得探討

圖書館界對於資訊的組織已有很長的經驗及很好的基礎，為何目前有關電子資訊的組織又會成為廣被探討的問題，我們認為主要有下列幾點因素：

一、網路資源的種類與特色

網路資源的種類相當多，幾乎各種主題、各種類型都有。在MARBI 第49及54號報告書中，將線上資訊資源分成兩大類：（註2）

1. 電子資料資源：如軟體、文件、資料檔、多媒體、書目資料庫等。
2. 系統或服務系統：如校園資訊系統、圖書館目錄系統、電子佈告欄等。

這二種資源存在著一對多或多對多的關係，即一個資訊資源可能會存在很多個系統上；一個系統上也會存很多的資訊資源。

此外，網路電子資源還具有下列特色：

1. 易變性 -- 電子資源極易被複製及更改，版本及修訂部份之追蹤不易；
2. 易動性 -- 資源的存放位置極易改變；
3. 分散性 -- 一份資源可能會被分解成數個檔，分放在距離很遠的不同機器上。
4. 多元性 --
 - a. 有很多不同的metadata已存在，如圖書館目錄，記載有關資料屬性、格式、位置等資訊。
 - b. 有很多不同的使用者，他們對資訊有不同的需求。
 - c. 有很多不同的資源類型，有的即興存在、有的具學術價值，有的單純、有的複雜。
 - d. 有很多不同的資訊提供者；有的是商業性質、有學術研究為主、有的 metadata可能和資源在一起、也可能分開。

上述特性是傳統的資訊媒體未具備的，因此當網路資源越來越多時，圖書館界對於如何有效的掌握電子資訊，一時之間變得有點兒束手無策。

二、傳統圖書館界的資訊組織模式

圖書館界經過幾百年的努力，已累積了相當多有關資訊描述、主題分析、權威控制之理論與實務經驗，且也建立了實際可行的系統及館際間的資訊交換、分享與系統互通模式。不過，傳統圖書館的編目主要以實體資料（如書、期刊、錄音帶、光碟、磁片）的描述為主，目錄和所描述的資訊是在同一地點，並強調以一致性的方法來組織及排列各種資料。如圖書館有關電子資源的編目，主要是針對自己擁有或租用的磁片、光碟等資料。編目依據是 AACR2 Computer File，並根據內容給主題語詞、索書號，再將資料排架，讓讀者透過目錄來查尋。圖書館所有的資料類型都以此模式處理，相當強調『實體管理』的特質。這種特質可由 AACR對“目錄”一詞的定義看出，AACR認為目錄乃是：(1)一份根據一定的方法，針對圖書館的館藏，整理而來的資料清單，或(2)一份為特殊目的準備的資料清單（如展覽目錄、銷售目錄）。（註3）

綜合言之，傳統圖書館的資訊組織模式主要如下：

1. 記述編目：以編目規則為依據
2. 主題分析：以分類表、標題表做控制詞彙式的分析
3. 權威控制：對於檢索點皆控制其一致性與參照關係，以提升檢索品質

4. 機讀化：將書目記錄機讀化，作為館際間資訊交換與分享的基礎
5. 資料排列上架
6. 國家書目控制
7. 全球書目控制

此外，圖書館界所描述的資料層次是整份出版品（即所謂的macrolevel），而將文章或目次的分析留給索引、摘要界來做。對於出版品則就其關係分為：單行本、合訂合刊、選集全集、叢書等幾種層次；在描述的詳簡層次上，又分為簡單編目、完整編目、詳細編目三種；而就所花費的人力、時間而言，可分為抄錄編目、原始編目。這些原則對網路資源而言是否適用呢？

顯然地，傳統圖書館的編目原則並不全然適用於網路資源或電子圖書館，因為在網路上有些資源並未正式出版或以不固定格式出版，這並不符合AACR2對所謂出版品的定義；網路上有些資源內容並未改變，但當位置改變時，其外在型式可能也變了(如由HTML檔變成Postscript檔)，這到底應算另一新版或視為同一版本？其次，網路資源到底以一份文獻為描述對象？以一個檔為描述對象？以一個網頁為描述對象？或以一個伺服器為描述對象呢？編目規則也未預見；另外，網路資源分散各處，要取用它必需知道其位置與存取方法等資訊，這些都是傳統的電腦檔所沒有的問題。所以，在網路資源未出現前，只要各國將其國家書目收集整理好了，則國際書目控制的理想也就快達成了。但現在出版品以另一種全新的模式出版，且增加速度驚人，使得資源的找尋又重新回到大海撈針的時代。復次，傳統圖書館的資訊組織花了很多時間在款目、分類號、標題的的正確、一致上，但事實上使用者從自動化系統所獲得的功能只有：

1. 簡單的作者檢索；
2. 簡單的書名檢索；
3. 簡單的標題檢索；
4. 簡單的關鍵字檢索；
5. 布林邏輯檢索。

甚多的描述資訊在檢索或統計上皆未派上用場，這種費時的描述方式是否適用於成長快速，但生命又有點短暫的『浮游資訊』？

三、分散式資料庫互通(Interoperability)之需求

資訊的查尋已不可能僅限於單機單館之資源，藉著網際網路的建立，所有網

路上的資訊已連結成一整域性資訊，雖然這些資訊分別存在不同的系統平台、不同的資料庫管理系統上，但藉著網際網路，使得分散式資源的查尋成為可能，互通也成為現代系統設計時最重要的考慮。

所謂的互通，有好幾個層次，在作業平台層次，互通的問題包括：電腦硬體、作業系統、網路協定、網路架構等，而在資料庫層次，異質的問題包括：資料模式的差異、限制(constraints)、查詢語言、交易控制機制(transaction control mechanisms)等，這些都屬於比較低階的問題。即使這些問題都解決了，資料庫的互通還是有問題，因為還有語意異質(semantic heterogeneity)的問題未解決，語意異質包括：屬性命名、格式、結構、流失及衝突值的確認等，而這些問題就得有賴使用共通的資訊描述格式不可了。

參、電子資訊組織模式

資訊組織是有效地使用資訊最基礎的工作，尤其對龐大的網路資源而言，不組織簡直無法使用。在網路世界，一般將描述資訊的資訊稱為元資料 (metadata) 或元資訊 (meta-information)，類似於圖書館的書目記錄。由於描述資料的方法不同，所產生的元資料也就不同，所以，一般也將 metadata 泛指描述資料的方法或格式。而當我們將 metadata 泛指為資訊組織的方法及格式時，通常在此方法及格式裡面所包含的意義還包括：編碼方式、標記方式、資訊描述項目及描述依據、主題分析、權威控制、字碼等問題，此外，也一定會牽涉到其使用範圍、系統開發、通訊協定等相關問題。

目前在網路世界中，已開發或還在開發的 metadata 模式相當多，在此我們仿照 戴柏斯(Lorcan Dempsey)及希瑞(Rachel Heery)的分法，依metadata的結構性、完整性、專業性，將網路資源的組織模式可分為下列三類：(註4)

第一類：一般性的網路查尋工具（即檢索引擎及主題目錄）

第二類：以蒐尋為目的之超資料（metadata for discovery）

第三類：以詳細記錄資源為目的之超資料（metadata for documentation）

茲分別說明如下：

一、一般性的網路查尋工具

網路查尋工具主要可分為二種，一為完全自動化作業的search engine，另一種

則為由程式自動抓取，再經人為分類或做簡單的、非結構化描述及分類的主題目錄。這一類資訊的特色是：

1. 資料少有明確的語意，也不能支援欄位檢索。
2. 目前這些資料主要是以web crawlers來建立，已有很多系統存在，其目的在做整域索引之服務，且廣被使用。
3. 優點是檢索回收率大，缺點是缺乏精確度。以網路資源呈指數成長的驚人速度，將越來越難分辨這些查尋結果，同時以 search engine建立的資料庫，有一天可能會比實際的網路資源還要來得大。
4. 由於索引範圍太廣，且描述性資料有限，所以蒐尋(discovery)的效果很有限。使用者可能會找到很多資源，但因無合適的索引詞，也無足夠的描述可以判斷是否和要找的主題相關。
5. 這些 search engine並未索引非文件式的檔案，如執行檔、影像檔、壓縮檔、聲音、動畫等。

網路上的檢索引擎相當多，如 Alta Vista、Lycos Intlomi、InfoSeek、WebCrawler 等等，它們都以數量取勝，同時也在檢索工具上不斷推陳出新，如：自然語言檢索、布林檢索、限制地區、切截檢索、整合蒐尋、混合蒐尋等。主題目錄也相當多，且各有特色，有按字母、年代、地區排列者，也有按主題及系統類型及其他方式排列者。其中依主題排列者如 Yahoo，其類目、主題層級的安排和傳統圖書館分類理念相同。

二、以蒐尋為目的之元資料(Metadata for discovery)

這類資訊描述格式主要有：Dublin Core、IAFA/Whois++ Templates、RFC 1807、Summary Object Interchange Format (SOIF)、LDAP Data Interchange Format (LDIF)、MARC、Uniform Resource Characteristics (URC)。這些metadata主要的特色是：

1. 包含完整的描述，可讓使用者找到感興趣的的資源；
2. 屬於結構化的資料，可提供欄位查尋；
3. 結構簡單，可由非專家建立，不必具備深厚的學科知識；
4. 傾向於個別物件的描述，不做多元的物件關係描述；
5. 其資料可能由人工建立，或由程式自動抓取再由人工加值處理；
6. 大都以屬性/值組模式建立資料；

7. 資料建立的成本昂貴；
8. 各種metadata之間，資料項目容易對映。

在這類metadata中，目前最受矚目的是 Dublin Core，有很多導航計畫在進行；而被普遍且實際使用的則是MARC模式，以下即就這二種分別敘述之，加以比較、並說明其在圖書館界被採用的潛力。

(一) Dublin Core

1. 簡介

Dublin Core 是最近提出來的metadata，它的格式及項目很簡單，只有十三項data elements。此十三項內容是：（註5）

Subject : The topic addressed by the work

Title : The name of the object

Author: The person primarily responsible for the intellectual content of the object

Publisher: The agent or agency responsible for making the object available

OtherAgent: The person(s), such as editors and transcribers, who have made other significant intellectual contributions to the work

Date: The date of publication

ObjectType: The genre of the object, such as novel, poem, or dictionary

Form : The data representation of the object, such as Postscript file

Identifier: String or number used to uniquely identify the object

Relation: Relationship to other objects

Source : Objects, either print or electronic, from which this object is derived, if applicable

Language: Language of the intellectual content

Coverage: The spatial locations and temporal durations characteristic of the object

這十三項主要可分成三類：（註6）

檢索點：即 title, subject, author, other agent, identifier 等項

協助辨識之資訊：即publisher, date, object type, language, form, coverage 等項

與其他物件之關係的資訊：即 relation, source 等項。

Dublin Core 設計目的主要在：a.鼓勵作者與出版者提供描述資料；b.鼓勵廠商發展可讓作者直接填寫metadata的authoring tools；c.再由其他團體在此基本的metadata上加上詳細的書目資料，如控制性詞彙等。Dublin Core 的設計相當具彈性，可與其他 metadata 整合在一起，且便於修訂與擴充。

2. Dublin Core 與 MARC 的比較

Dublin Core的設計乃希望由程式直接在網路環境上抓取由製作者提供的描述資料並索引之，另一個目標，則在協調整合 Dublin Core 及 MARC、TEI header等標準。

此外，它主要是為找尋、檢索等存取(access)之用，並未考慮到資訊的安全、確認、購買等需求，從它所設計的項目來看，我們發現它所描述的也是文獻類型的物件，並未解決系統與服務(system and service)描述之需求。另外，由於它並未規定資料描述的依據，所以各機構所描述的內容勢必不一致，如在 AACR2中，有如何記述正題名、並列題名、及如何維持作者等名稱的一致性等規則，而Dublin Core並不規定這些，所以即使兩個單位都以相同的項目描述資訊，所描述出來的內容卻未必一致，所以檢索的可靠性也成問題。

Dublin Core的結構很簡單，描述項目也很精簡。它比一般性的網路資源檢索工具精確，但和專業所提供的書目品質仍有差距。

3. 在圖書館使用的潛力

網路上的資訊物件若能伴隨Dublin Core的metadata，將可改善search engine的品質，但 Dublin Core 是否可提供有效的檢索，仍需視資訊提供者如何發展系統而定。以此標準來著錄的資訊將不如MARC嚴謹，但因較簡單，極適用於網路資源的描述，也適合由作者自己描述資訊，所以我們認為未來圖書館的系統可能需具備Dublin Core的資料轉入，再由圖書館人員擴充為完整的 MARC records。

(二) MARC

MARC是圖書館界為描述、儲存、交換、處理及檢索資訊，而於1960年代設計的標準，經過修訂後，目前也被用來描述及存取網路資源，透過 856欄，使MARC記錄可與電子文件作超連結，實驗證明MARC已是一個找尋與檢索網路資源的可行標準。

1. MARC 和新的 metadata 之比較

MARC是一控制得很嚴謹的標準，任何的規則修訂都需經過長時間的討論，因此解決新問題的機制相當遲緩。MARC可對資訊物件作詳細的描述分析，且用編目規則、分類表、標題表等工具要求描述的一致性，並使用權威檔支援完整的檢索，雖然這些工具亦可應用在其他的metadata上，但這些metadata發展單位似乎無意採用，因為太花時間了。

專業的建立資料、嚴謹的編碼規則及複雜的系統，將使資訊的存取品質較有保證，圖書館自動化系統使用MARC來交換資料，並透過欄位856所記載的URL（實際上URL並非記錄於此，但本文對此問題暫且不表），Webpac等設計使得書目得以和文件連結，再加上Z39.50的介面，使用者不但可用傳統的查尋方法如關鍵字、布林邏輯檢索等，也可同時作多資料庫的查尋。

此外，MARC也可用來描述複雜的資訊物件關係。如它可記載連續性關係及參照關係，使讀者可透過書目連結，找到相關的資訊物件，這麼詳細及完整的關係追蹤也是其他metadata所沒有的。

2. 在圖書館使用的潛力

在InterCAT Project中，經過圖書館界的測試，發現MARC可用於網路資源的整理。但在增加新項目上，USMARC的彈性最差；完整性方面，排名第二，緊次於TEI header；與現有的書目資料庫的相容、檢索的可靠性、資料結構的複雜度上都排第一。（註7）雖然search engine及一般的Web page都可提供檢索，但圖書館的功能遠超過此，他們是資訊加值中心，各種資訊都被謹慎的選擇、組織、儲存及傳佈。若某份網路資源品質好、符合讀者的興趣、且和館藏相關，則它會被選入圖書館的目錄中，以此方式，提供使用者一個在主題內較正確且較完整的主題資訊。

雖然它也是一個費時、費力且昂貴的機制，有時也會被錯誤的使用，但都不能否認它可用於網路資訊存取的事實、它是一個成熟的工具，可讓我們將重要、相關、高品質的網路資訊整合到目錄中。所以Amanda Xu認為以MARC及AACR2來組織網路資源有下列優點：（註8）

- 資源經過主題專家的選擇以符合使用者的需求；
- 他們有書目描述、權威控制、主題分析等嚴謹的控制；
- 圖書館自動化系統早已在處理大量資訊的檢索；
- 更重要的是網路資源可與上千萬存在於機讀格式中的書目資訊結合。

但MARC並非沒有缺點，它的優點正好就是它的短處；此外它還有下列缺點：

- 有些描述是屬於程序性(procedure)而非描述性(descriptive)；
- 欄位重複(雖然有其使用背景)；
- 只能由第三者(編目員)來建立資料，無法由作者或使用者建資料，互動性差。

三、以詳細記錄資源為目的之元資料(Metadata for documentation)

這類metadata主要有：ICPSR SGML Codebook Initiative、FGDC-Content Standards for Digital Geospatial Metadata、Consortium for the Interchange of Museum Information (CIMI)、Text Encoding Initiative (TEI)、Encoding Archival Description (EAD)、Government Information Locator Service (GILS)等。其共通的色是：

1. 資訊描述的完整性最高；
2. 資料最結構化；
3. 除用來記錄位置及找尋資訊外，也扮演詳細記錄物件或物件集的角色；
4. 這類metadata的使用對象，通常和學術及研究團體有關，需要專家知識來建立及維護，且迎合專家的資訊需求；
5. 這些metadata都是為特定領域或特殊資料而訂定；
6. 它們詳細表達物件間複雜的關係；
7. 這些特定領域的metadata 有很多都以SGML為編碼標準；以Z39.50作為跨資料庫之資訊查尋與檢索協定。

以下即就幾個已被採用或可能被採用的metadata說明之：

(一) TEI Header

文獻編碼計畫 (Text Encoding Initiative, TEI) 是一項自1987年就開始研究的計畫，並於1994年出版了1400 頁的Text Encoding Initiative Guidelines。TEI計畫之目的主要是在定義一套表達電子資料的指南，使各領域的研究者可交換及再利用資源，而不受軟硬體及應用環境的限制。

TEI 計畫是由：The Association for Computers and the Humanities、The Association for Computational Linguistics、The Association for Literary and

Linguistic Computing三單位一起合作的計畫，經費由US Endowment for the Humanities及European Union 3rd Framework Programme for Linguistic Research and Engineering 聯合贊助。歐、美學術界也紛紛成立委員會研究相關問題。

TEI Guidelines 使用SGML的欄位及屬性，分析各種文獻特性，定義成幾套適用於各種文獻類型之標誌符號(tag)。TEI 的標誌符號可分成四類：1. Core set，這是所有文獻都需要的項目；2. Base sets適用於特定類型文獻之項目，如詩、散文、戲劇；3. Additional sets: 適用於不同文獻作特殊處理之項目；4. Auxiliary sets: 賦予特別角色之項目，如 TEI header即屬此類。TEI Guideline 規定每一份 TEI文件前面都要有 TEI header，以描述該文件，這份TEI header是由歐洲、北美圖書館界、檔案界所組成的委員會 Committee on Text Documentation 所訂定。

TEI header有幾種作法：1.可和TEI文件放在一起，成為文件的一部份。若此，則該header希望由作者或出版者來建立、或在文件編碼時建立；2.TEI header可由圖書館、研究單位、檔案單位等來蒐集建立成一書目資料庫，以指引存放在遠端的TEI文件，這也就是 Guideline中所說的independent header；3.TEI header也可用來描述非TEI文件，如網路上的資源。

TEI header包括：1. file description：說明文獻之書目特性及出處；2. encoding description：編輯過程之決定的記錄；3. profile description：描述語言、參與者、主題等非書目性資料；4. revision description：更新、修訂等記錄。（註9）

1. TEI Header and MARC 的比較

TEI header的編碼架構很有彈性，其必備項只有title statement，publication statement，source description statement，這些都包括在file description這組資料項中。此 header 的描述項目可以就需要，以 SGML DTD來擴充，如可就地域性需求決定擴增項目。

TEI header可附隨全文或單獨存在，使用者不只可查尋全文也可查書目；描述項目由作者等直接填寫，不需接受訓練。而其缺點是未規定描述標準，所以資料可靠性不高，各機構所建立的也會不一致，這將影響合作編目時的資料交換。

TEI header的結構比USMARC鬆散且不精簡(compact)，如Publication Statement，分散成三個項目，而USMARC則以260的三個分欄含括，所以MARC在資料載送上較精簡、有效，一個TEI header書目記錄可能比MARC來得長。

TEI header中記錄書目關係的項目是source description statement，此處所描述的是原始文件，並未說明此電子文件和其他作品的關係。但MARC可記載書目作品間複雜的關係。

TEI Guidelines中有說明TEI header與USMARC的對映關係，但要達到MARC的品質還要人工的介入，如TEI header不要求作者等名稱形式的一致性，且有些非書目性資料在MARC中也無對映欄位，只能轉到Notes欄中。

2. 在圖書館使用的潛力

根據1989年及1991年兩次調查顯示，圖書館的電子文件館藏增加了兩倍，且這些電子文件大都以SGML標誌，此外，網路上以SGML標誌的電子文件相當多，大多數電子圖書館的文獻也都以SGML標誌；大型出版社及大企業、跨國公司，其出版品也紛紛建為SGML文獻。這是採用SGML方式編碼的TEI-DTD及TEI header不容忽視的實力。事實上，我們也發現除了TEI header外，很多的metadata也都以SGML作為編碼架構，它們可以說就是一種SGML-DTD，如Dubline Core、CIMI、CDWA、EAD、FGDC、ICPSR、URC等。

TEI及MARC都可用於電子文件的描述，兩種格式之編碼都很耗時間及人力，TEI可較生動的描述資訊，MARC則長於資訊分享及傳佈，若TEI之編碼格式及規定和MARC一樣嚴謹，且圖書館的自動化系統可自動轉換TEI header與MARC資料，則我們認為它在圖書館界被採用的可能性也很高。

(二) GILS

政府資訊形式有公報、法令、連續性出版品、統計、會議紀錄等，內容多樣化，包含各種學科領域，且多為集體創作、經常修訂、內容深入，故相當具權威性及參考價值。為使大眾了解美國政府資訊，協助政府做資訊傳佈的工作，故建置了美國政府資訊系統（Government Information Locator Services，簡稱GILS）。GILS系統最早是在1992年由Syracuse University資訊研究所的馬克庫爾(Charles R. McClure)所設計，現在已成為NII計畫的一部份，由資訊基礎建設任務小組（Information Infrastructure Task Force，簡稱為IITF）負責。GILS是一分散式系統，由各政府機關之資料庫組成，各自負責本單位資料的製作與維護，而由IITF負責統籌。（註10）該系統最主要的核心部份稱為GILS Core，也就是政府資訊的metadata，其欄位依規定順序排列共有下列22項，茲列舉如下：（註11）

1. Title
2. Originator
3. Controlled Vocabulary
 - index Term-Controlled
 - Thesarurs
4. Local Subject Index
5. Abstract
6. Purpose
7. Agency Program
8. Spatial Reference
 - Bounding Rectangle
 - Western-most
 - Eastern-most
 - Northern-most
 - Southern-most
 - Geographic Name
 - Geographic Keyword Name
 - Geographic Keyword Type
9. Time Period of Content
 - Time Period-Structured
 - Time Period-Textual
10. Availability
 - Distributor
 - Distributor Name
 - Distributor Organization
 - Distributor Street Address
 - Distributor City
 - Distributor State
 - Distributor Zip Code
 - Distributor Country
 - Distributor Network Address
 - Distributor Hours of Service
 - Distributor Telephone

- Distributor Fax
- Resource Description
- Order Process
- Technical Prerequisites
- Available Time Period
 - Time Period-Structured
 - Time Period-Textual
- Available Linkage
- Available Linkage Type
- 11. Source of Data
- 12. Methodology
- 13. Access Constraints
- 14. Use Constraints
- 15. Point of Contact
 - Contact Name
 - Contact Organization
 - Contact Street Address
 - Contact City
 - Contact State
 - Contact Zip Code
 - Contact Country
 - Contact Network Address
 - Contact Hours of Service
 - Contact Telephone
 - Contact Fax
- 16. Supplemental Information
- 17. Cross Reference
 - Cross Reference Title
 - Cross Reference Linkage
 - Cross Reference Type
- 18. Schedule Number
- 19. Control Identifier
- 20. Record Source
- 21. Original Control Identifier
- 22. Date of Last Modification

GILS系統是一分散式架構，所以跨系統的查尋乃採用Z39.50標準，同時提出一份Z39.50 Profile作為系統設計的標準，GILS Profile第二版是於1997年2月8日提出來的，此份文件主要在說明GILS系統的應用範圍、目標、採用的屬性集(attribute set)、記錄語法(record syntax)、查尋方式、GILS架構等。

(三) FGDC (註12)

1992年美國聯邦地理資料委員會(Federal Geographic Data Committee, FGDC)首次於1992年召開，會中討論描述各種地理資訊的專業用語及其定義，並於1994年制訂通過一項應用於電子地理資訊的metadata標準，名為Content Standards for Digital Geospatial Metadata (CSDGM)。CSDGM雖為該標準的正式名稱，但一般仍慣以FGDC稱之，該標準也於1995年，成為美國地理資訊metadata的國家標準，在此之前，地理資訊只有最基本的識別性資料及聯絡資料而已，經由此 metadata的訂定，將metadata的意義、資料類型及其從屬物(inherent dependencies)都做了明確的定義，以維護地理資料的品質，避免因為時間變遷或人員變動等因素，導致資料內容與品質無法掌握。

1995年，FGDC首次在ISO的Technical Committee中被提出討論，與ISO達成修訂的共識，並於1997年7月提出修訂版。目前FGDC可以對應的標準有DIF、GILS、USMARC、Dublin Core等。FGDC的欄位有300個之多，相當複雜，這些欄位間有從屬關係，以階層式結構呈現及儲存。FGDC只規定metadata的內容，而未規定其編碼方式，所以負責推動FGDC標準的單位National Geospatial Data Clearing-house建議以SGML來支援資料的載入、交換、展示，並發展出一參考的DTD。

此標準也未規定通訊協定，該Clearinghouse曾使用了不同的Z39.50系統，並計畫提出一份Z39.50 GEO Profile，此外，由CNIDR所設計的Isite ISearch軟體也修訂為可支援地理資訊的查尋及SGML款目的檢索。

FGDC的資料欄位包含7個主要資料段(main sections of metadata)及3個輔助資料段(support sections)，內容如下：

Main sections

Section 1: Identification information (辨識資訊)

如：title、geographic area covered、currentness、citation等

Section 2: Data quality information (資料品質)

如：positional and attribute accuracy、completeness、consistency、the source of information、methods used to produce the data等

Section 3: Spatial data organization information (資料組織方式)

如：the methods used to represent spatial positions directly (such as raster or vector) and indirectly (such as street addresses or county code)、the number of spatial objects in the data set

Section 4: Spatial reference organization information (地理參考資訊)

如：the name of and parameters for map projections or grid coordinates systems、horizontal and vertical datums、the coordinate system resolution

Section 5: Entity and attribute information (類型及屬性資訊)

如：the names and definitions of features、attributes、and attribute values

Section 6: Distribution information (資料獲得方式)

如：a contact for distributor、available format、information about how to obtain data set online or physical media、fees for the data

Section 7: Metadata reference information (metadata參考資訊)

如：currentness and information about the organization that provided the metadata

Support sections

Section 8: Citation information (引文資料)

如：originator、publication date、title、edition、series information、larger work citation等

Section 9: Time period information (時間資訊)

如：single date/time或多個日期/時間或日期/時間範圍

Section 10: Contact information (聯絡方式)

如：contact person or organization primary、contact position、address、hours of service等

(四) CIMI

博物館資訊交換聯盟 (Consortium for the Interchange of Museum Information, 簡稱CIMI) 是由十六個會員單位所組成的組織。該組織的成立主要是1988年由於

Museum Computer Network (MCN) 呼籲探討博物館資訊交換標準，而獲得的成果。CIMI 委員會於1990-1992年間由David Bearman、John Perkins提出“A Standard Framework for the Computer Interchange of Museum Information”，此標準中規定了資訊交換協定、格式、低階網路與通訊標準、資料內容標準等，採用很多現有的標準。其中有關資訊查尋與檢索協定採用Z39.50、藏品資料描述標準則採用SGML。該聯盟除制定標準外，並進行一項展示系統設計計畫，即「文化遺產資訊線上系統」(Cultural Heritage Information Online，簡稱CHIO)，此系統建置主要包括兩部份：1. CHIO Structure：乃在探討SGML的使用，設計出博物館資訊所需的CIMI DTD，此DTD也被TEI納入其TEI Tags中；2. CHIO Access：即在探討Z39.50的使用，並建立一份CIMI Z39.50 Profile，作為系統設計的標準。(註13)

CIMI將博物館資訊分為八種類型：(註14)

exhibition catalogues 展覽目錄

labels for objects displayed on the walls in exhibitions (wall labels) 牆上標籤

object records 物件記錄

images 影像

bibliographic records 書目記錄

exhibit brochures 展覽指南

authorities (Art & Architecture Thesaurus) 權威資料

同時該計畫也考慮一般大眾及藝術研究人員之資訊檢索需求，訂定出博物館資訊應具備的檢索點，這些檢索點和TEI header都會被標誌在文件中以便檢索，其中必備的檢索點有：

award

bibliography

collection

concept

copyright.restriction

credit-line

date-range

even

identity-number

inscription.mark

material

nationality.culture.race
object
occupation
organization
person
place
process.technique
quote
styles.movements
subject.title.name
type.classification.

肆、與Metadata相關之問題分析

由上述分析我們可以看出一些電子資訊組織的發展趨勢及相關問題，茲分為下列幾項討論之：

一、全域索引或區域索引（Global Indexing vs. Local Indexing）

在上述三類metadata中，對資訊的索引範圍恰好也可分成三種層次：

1. Search engine 及 Subject catalog 傾向於做全域的索引，廣泛的蒐集網路資訊，希望使用者只要透過該服務即可掌握全部的資源；
2. 第二類以蒐尋為主的metadata，並未規定伺服器蒐錄的範圍，但在應用上，由於透過人工處理，所以會有取捨，而在實際環境裡，其取捨的原則通常是資訊的品質與使用者的需求為主。
3. 第三類 metadata 則傾向於專門主題或資料類型之分散式資料庫的建立，希望同質機構利用相同的格式整理的自己的收藏，再透過相關通訊協定互通資訊。

二、語意模糊或語意互通

由於第一類metadata並未對資源作描述，所以索引語意模糊，造成檢索不精確的問題，第二類metadata除了MARC外，只規定資料描述格式，未規定資料描述標準及也未做詞彙控制等處理，所以在應用時，若未配合其他的標準使用，如編目規則、標題表、索引典等，雖有欄位檢索，但語意的模糊問題仍在；第三類metadata

主要希望在同一領域中語意互通 (semantic interoperability) ，所以在語意處理上則相當嚴謹，並採用相關標準來建資料。

三、Metadata的格式

網路資源的描述格式將多元並存，是不可避免的事實，所以任何一種 metadata 若相當具排他性，將無法存在。以 Dublin Core 為例，它相當具有包容力，任何 metadata 架構都可套用到 Dublin Core 中。這也造成 Dublin Core 內容的不一致。為使 Dublin Core 之各種使用者能交換不同類型的 metadata，又提出 Warwick Framework。這是一個組合及交換各種 metadata objects 之架構，不過此架構尚未被詳細定義。(註15)

此外，描述格式所包含的內涵不是只有資料描述項目而已，還必需規定其編碼標準、描述標準、主題分析、權威控制等相關問題；我們分析現存的幾種 metadata 後，發現大多數的 metadata 都以 SGML 標準編碼，少數直接以 ASCII 格式填寫，只有 MARC 是採 ISO 2709 的編碼架構。

四、Metadata的建立問題

Metadata 到底由誰來建立，以及建立的速度是否夠快，是各種 metadata 決勝的重要關鍵。由於各類 metadata 之目的不同，本來就不可能有相同的建立方式；不過，第一類 metadata 為彌補其資料不精確的缺失，以及第二類 metadata 為彌補其人工建資料時效性差、速度慢的短處，都在尋求由作者或出版者提供基本資料的途徑。而第三類 metadata 乃在對自己的收藏做詳盡、精緻的描述，所以需要主題專家的投入。

五、通訊協定問題

第二類和第三類 metadata 都是屬於分散式查尋架構，所以需有相關的通訊協定配合，如 TCP/IP 協定、Z39.50 及做目錄式查尋的 Whois++，都有被討論或指定使用，其中，第三類 metadata 使用 Z39.50 的不少，而在圖書館界，Z39.50 已是一個實際運作的系統。

六、系統開發

好的架構必需有可用的好系統配合，否則也只是空中樓閣。上述三類

metadata中，普遍存在的是第一類的search engine、subject catalog；第二類的MARC系統，全球的圖書館界都在用，且已建立了上千萬筆的資料。此外 GILS、TEI header、CIMI、FGDC都有系統採用，而Dublin Core則有一些導航計畫在進行。RFC 1807被用在NCSTRL計畫中。SOIF被廣為使用作為 Harvest 的內部格式，但對內容無一致的定義。LDIF也處在同樣的情況，對資源描述缺乏一組架構，LDIF and SOIF已被作為Netscape's directory server and catalog server products on LDAP and Harvest的基礎。此外，在無國界的網路時代，系統必需考慮多國文字處理的問題，如此才有可能應付網路上多元的資訊處理與資訊需求。

陸、結語

戴柏斯認為，metadata未來之發展趨勢如下：（註16）

1. 當web成為更成熟的出版環境時，網路資源之品質將被改進；
2. 蒐尋系統奠基於較完整的描述資料，在資訊不同的處理階段將會由不同的人加上不同的描述資訊；
3. metadata將包含更多更複雜的資訊，供不同目的使用；
4. 系統將以不同的途徑、方法收集處理資料；
5. 未來將是一個使用目錄協定(Whois+,...)及查尋與檢索協定(Z39.50)的分散式環境。

戴柏斯的看法，和筆者二年前在「圖書館員如何在資訊時代扮演資訊組織與分析者的角色」一文中對於未來的目錄特色看法不謀而合，筆者認為，未來的目錄特色為：（註17）

1. 它是分散但卻能整合性查尋的目錄，這些目錄將會有不同的描述格式；
2. 它是利用知識樹相連的目錄；
3. 它會利用索引典、權威檔、語料庫、語法規則等工具；
4. 它有自動索引與自動檢索機制；
5. 它和全文相連結；
6. 它是整個資訊界合作的成果。

時隔二年，再回過來看今日索引目錄的發展，我們發現有很多當時預測的現象已開始實現，而我們也可以預測，未來實現的速度會超乎我們的想像。所以，我們不該只停留在研究階段，而應開始有具體的行動，圖書館學會或國家圖書館應結合相關力量，探討各種可行方案，帶領整個圖書資訊界在未來二十一世紀真正扮演資訊領航員的角色。

註 釋

- 註1： 陳昭珍，「電子學術圖書館的館藏發展與維護」，*大學圖書館*一卷一期（民86年1月）：頁30。
- 註2： Caplan, Priscilla, "Providing access to online information resources: a paper for discussion", PACS-L.
- 註3： Gorman, Michael and Paul Winkler, *Anglo-American Cataloguing Rules*, 2d ed. rev. (Chicago: American Library Association, 1988), 616.
- 註4： Dempsey, Lorcan and Rachel Heery, "Metadata: an Overview of Current Resource Description Practice",
<http://www.ukoln.ac.uk/metadata/DESIRE/overview/doc0000.htm>
- 註 5： Weibel, Stuart, Jean Godby and Eric Miller, "OCLC/NCSA Metadata Workshop Report",
http://www.oclc.org:5046/oclc/research/publications/weibel/metadata/dublin-core_report.htm
- 註 6： *ibid.*
- 註 7： Palowitch, Casey and Lisa Horowitz, "Meta-information Structures for Networked Information Resources", *Cataloging & Classification Quarterly* 21 (3/4) (1996): 115.
- 註 8： Xu, Amanda, "Accessing Information on the Internet: Feasibility Study of USMARC Format and AACR2",
<http://www.oclc.org/oclc/man/colloq/xu.htm>
- 註 9： 同註4。
- 註10： Government Information Locator Service (GILS), <http://www.usgs.gov/public/gils>
- 註11： Johnson, Julie, "GILS: Government Information Locator Service: A Brief Description", *Australian Library Journal*, v44 (Nov 1995): 217。Christian, Eliot J, "Application Profile for GILS-Version 2",
http://www.usgs.gov/gils/prof_v2.html
- 註12： 同註4。
- 註13： 同上註。
- 註14： Sander, Margaretta B, " Project CHIO: Cultural Heritage Information Online", *Final Report of Consortium for the Computer Interchange of Museum*

Information 12。

- 註15：Dempsey, Lorcan and Stuart L. Weibel, "The Warwick Metadata Workshop",
<http://hosted.ukoln.ac.uk/mirrored/lis-journals/dlib/july96/07weibel.html>.
- 註16：Dempsey, Lorcan, "Meta Detectors", available at:
<http://www.ariadne.ac.uk/issue3/metadata/>
- 註17：陳昭珍，「圖書館員如何在資訊時代扮演資訊組織與分析者的角色」，*中國圖書館學會會報* 55 期（1995），頁47。