

從元資料看未來資料著錄的發展趨勢

吳政叡 (Cheng-Juei Wu)
輔仁大學圖書資訊系專任副教授
Associate Prof.
Department of Library & Information Science
Fu-Jen University
E-mail: lins1022@fujens.fju.edu.tw

中文摘要

元資料對電子文件（或檔案）所扮演的角色，正可對比於目錄之於傳統的印刷媒體資料，因此元資料可說是『電子目錄』。元資料除了要負起傳統目錄具有的指引資料和協助檢索的功能外，在格式的設計上，也須能顧及電子檔案所獨有的一些特性，如檔案格式的種類繁多、資料轉換需求頻繁、版本辨識困難等問題。在歸納和分析了FGDC的地理元資料、目錄交換格式、政府資訊指引服務、都柏林核心集等四種元資料後，作者認為未來在資料著錄的發展趨勢上，將有如下的特色：個別資料類型有其支配性的元資料、格式精簡且適合各種專業人士、全球單一資訊網、分散式網路系統、易擴展原則、可修飾原則。本文最後介紹一個資料著錄的實驗系統--元資料實驗系統 (Metadata Experimental System, 簡稱 MES)，透過這個系統，讀者對元資料及其未來的可能運作方式，將有更具體的認知。

關鍵字：元資料，元資料實驗系統，MES，地理元資料，目錄交換格式，政府資訊指引服務，都柏林核心集，電子圖書館。

Abstract

Metadata is a resource which can assist information retrieval of digital documents on the Internet. In the course of designing a metadata, we need to consider the characteristics of digital documents, such as the variety of file formats, frequent format-transformation, and the difficulty of version-distinguishability. After a brief analysis of some existing metadata formats, we introduce several principles for future developments of cataloging on the Internet. Finally, we briefly describe the Metadata Experimental System (MES for short), a metadata-related system currently under development. MES is located on the author's homepage (URL: <http://blue.lins.fju.edu.tw/~wu>).

Keywords: Metadata, MES, Dublin Core, Digital Geospatial Metadata, DIF, GILS, DIF, GILS, Digital Library

一、前言

自古以來人們即不斷尋求更好的材料來儲存知識，以便流傳後世，從以前的泥土、動物骨頭、龜殼、紙張，到今日最新興的電子儲存媒體（如光碟片和磁碟片）。但有了材料來記載知識後，隨著儲存材料的不斷累積，如何快速找到所需要的資料，也成為人們關心的一個課題。於是有目錄的產生，來提綱挈領的整理資料，和對資料加以適當的描述，以協助資料的檢索。因此製作目錄的主要目的之一，是希望透過對資料的著錄和描述，來減少不必要的調閱和取得原件的次數。雖然今日電腦科技突飛猛進，電子媒體儲存資料的能力大增，電腦的運算速度驚人，但是有效率的檢索，仍是一個重要的問題亟待解決。換言之，為了資料檢索和管理的需要，對資料的適當描述仍是必須的，因此某種形

態的電子目錄有其必要性。

全文檢索是近年來興起的一個熱門話題，在某種意義上，似乎有了全文檢索，即不再須要對資料加以著錄和描述，但細究之下，全文檢索雖然有其優點和適用場合，但也非萬靈單，有一些應用上的限制存在，例如非文字資料基本上是不適用全文檢索的，像地圖、聲音、影像等。即使是文字資料，也並非所有檢索需求均可由全文檢索式的處理來滿足，例如現今檢索引擎運作的方式基本上即是屬於全文檢索，主要是透過自動抓取程式在網際網路上抓取網頁，然後以自動拆字(或詞)作索引的方式來建立其資料庫，做為檢索的基礎。此種運作方式固然可滿足部分檢索需求，但很明顯有其他問題產生，低效率和無法有效過濾資料最為人所詬病，正因為如此，元資料近來日益受到重視。

自電腦問世後，圖書館也注意到這種資料處理上的利器，並加以引進到圖書館界，於是有機讀編目格式(MARC)[註 1]的產生，但是機讀編目格式基本上是目錄的電子化，其所描述的資料，大多數是以紙張媒體的形式存在的。因此機讀編目格式雖是目前圖書館自動化系統的核心和相關資料處理的要角，但就其本質而言，可說是卡片目錄的電子化。由於科技的進步，以電子(數位)形態出現的資料，將逐漸成為圖書館資料處理的主流，全球資訊網(World-Wide Web, 簡稱 WWW)[註 2]的盛行，和國家資訊基礎建設(NII)[註 3]的大力推行，可說是此一趨勢的最佳見證。由於電子檔案有很多異於紙張媒體的特性，如電子檔案格式的複雜和多變性，使傳統的機讀格式面臨嚴重挑戰，因此另一種針對電子檔案和全球網路作業環境的資料著錄格式--元資料(metadata)[註 4]正在逐漸成形中，本文擬從電子檔案的特性和元資料的相關發展，來探討資料著錄的發展趨勢。

二、電子檔案的特性

紙張媒體有很多的優點，如質輕、易使用等，其中一個優點是電子媒體所沒有的，那就是用肉眼即可閱讀。紙張媒體的資料，祇要不腐爛，數千年後的子孫，依然可以用他們的肉眼來閱讀；但是用電子媒體儲存的資料，可就沒這麼幸運了，相對於人眼祇有一種，解讀電子檔案的電子眼則是千變萬化的。首先，資料是以數位的0和1存在於電子媒體(如磁碟片)上，須有適當的設備(如磁碟機)來讀取，而此週邊設備又須有對等的中央處理器(CPU)來指揮。當一串的0和1載入主記憶體後，須有適當的軟體來詮釋這當一串的0和1，因此又牽涉到所使用的字元集(如ASCII、Big 5等)和檔案格式，以電子檔案格式而言，現已有成千上萬種格式存在。另一方面，軟體也須有對等的作業系統(OS)和中央處理器的支持才能工作。因此電子檔案的儲存，不僅牽涉到資料檔本身而已，還涉及眾多的軟體和硬體設備。

舉例而言，作者個人的博士論文，是大約三年前用微軟(Microsoft)的文書軟體--Word 2.0版製作的，結果在短時間內，此軟體已演進到第七版，當作者嘗試用新版Word來讀取以前博士論文的電子檔案時，已發現有些數學公式會出現亂碼，雖然不用重打一遍即可加以克服，但也須耗費時間加以適當處理。更糟的是圖表部分，無法在新版中做任何修改，為了這個問題和其他因素，有一陣子作者被迫同時維持二個版本(第二版和第七版)的存在，由這個例子可知電子檔案在儲存和管理上的複雜性。

由以上的例子可知，隨著電子圖書館的普及，和電子媒體資料的大量存在，眾多檔案格式的處理，將成為資料儲存和管理上的一大問題；同樣的，同

一系列檔案格式間新舊版本的差異，也將是圖書館的一大棘手問題。以新舊版本的差異問題來說，解決方案有二：一是保存每一主要版本的相對應使用軟體；一是隨著新版檔案的產生，即將所有舊版檔案格式的電子檔案轉換成新的版本格式。以第一個方案而言，保存軟體並不如想像中的單純，如前面所分析的，每個軟體有其工作所須的獨特平台（platform）[註 5]，即作業系統和中央處理器，因此有可能會演變成須要保留整組設備（如整部個人電腦加上許多軟體），就長期而言，這幾乎是不可能的，也是不切實際的。另一個解決方案是做轉換的工作，雖然可行但是工程浩大。尤其以目前科技進展的速度，一個技術、產品、軟體的生命週期，有越來越短的趨勢。因此每隔一段時間，將儲存的電子檔案翻新或檢查一遍，恐將成為資料管理者的例行工作和頭痛的問題之一。同時其龐大的數量，唯有利用電腦來自動執行才可能負荷。因此如何利用對檔案的適當描述，來協助電腦進行有效率的更新或轉換格式，是一個重要的課題。

版本的問題雖非電子檔案所獨有，[註 6] 像紙張媒體的資料也有版本的差異存在，如很多教科書，都有由同一作者用同一書名不斷出新版的情形，但畢竟其更版是以年為單位，同時紙張媒體的資料也沒有格式的問題。但由於電子檔案可輕易加以修改或轉換格式，使得其版本問題變得格外複雜。以判別二個不同電子檔案是否為同一作品為例，目前一般是以檔名、作者、檔案創造或修改時間為主，但這些資訊是不太可靠的。例如以下的情況可能發生：

假設在站台B複製了站台A的一個檔案X後

- (一) 站台A開啟檔案X，刪除其中一個標點符號後，隨即又添加回去，並加以存檔，則站台A上檔案X的修改時間會被更動。
- (二) 站台A將檔案X轉換成另一種格式儲存，如副檔名從 .txt 換成 .doc。

則當檢索者同時查到站台A和B的檔案X時，該如何來辨別這二個檔案的異同或關係呢？

由以上例子可窺知電子檔案版本辨識的複雜性。

三、元資料的發展現況和特性

在資訊的傳播方式上，網際網路和WWW盛行前，是主要以下面的方式進行：

資料提供者-->圖書館-->資料使用者

因此圖書館可以說是主要的媒介者，來溝通資料提供者（如出版社）和資料使用者（如個人），所以圖書館扮演了資料儲存和傳播者的主要角色。為了有效達成其做為媒介和橋樑的角色，使圖書館能夠有效率的來管理擁有的資料，以便使用者可以很快找到所需的資料，圖書館須要有一套很好的方法，來描述所收藏的資料，於是有目錄的興起。其後隨著資料處理科技的進步，從卡片目錄演進到目前的機讀編目格式（MARC），其編製手法和處理方式，或有人工與電腦操作的差別，但它們的基本功能和扮演的角色卻是相同的。

近幾年來，網際網路（Internet）和 WWW 的迅速興起，對資訊傳播的方式產生了重大的衝擊。同時網際網路和WWW的相互結合，大幅降低了資訊傳播的障礙。資訊傳播障礙的移除，引發了二個看似迥異卻又相關的問題，一是如何

來有效率的過濾資料，一是如何來有效率的描述資料。就前者而言，目前在使用 WWW 上的搜尋引擎（如Locys [註 7] 等）來收集資料時，大家經常會面臨到的問題之一，是所得到的資料回覆量太多，經常可有上萬條款目，實無法一一來加以過濾，更糟的是，排在前面的款目，又往往不是你所真正需要的，頗使人進退維谷，祇有瞎猜亂挑。很明顯的，我們需要更多的資訊，來從回覆的款目當中，挑選我們真正需要的資料，而這些資訊必須由資料提供者來提供，因此如何制定一套資料描述格式，來有效率的描述收藏的資料，成為一個重要的課題，這正是元資料（Metadata）日漸受到重視的原因。如經由微軟一些與網際網路相關的軟體所製作的 HTML 文件中，已在文件開頭處，加入許多元資料項目，來紀錄此文件是由那個軟體製作的。

元資料（Metadata）最常見的英文定義是 "data about data"，[註 8] 可直譯為描述資料的資料，就其本義和功能而言，可說是電子目錄（Electronic Catalogue）。編製目錄的目的，即在描述收藏資料的內容或特色，進而達成協助資料檢索的目的。而元資料即是用來揭示各類型電子文件或檔案的內容和其他特性，其典型的作業環境是電腦網路作業環境。

地理資料由於無法直接檢索其內容，故必須經由適當的著錄，來揭示其內容，以利資料的整理、交換、檢索，所以地理方面的元資料是發展較早和完備的。美國聯邦地理資料委員會的地理元資料標準，和美國太空總署所制定的目錄交換格式，即是二個代表產品。

就目前元資料的發展現況來說，最成熟和完整的元資料格式，首推美國聯邦地理資料委員會（Federal Geographic Data Committee，簡稱FGDC）的地理元資料（Digital Geospatial Metadata）標準。FGDC 的地理元資料 [註 9] 是根據 1994 年美國總統行政命令（Executive Order 12906）[註 10] 而來，命令 FGDC 負責制定一個關於地理資料收藏、著錄和交換的標準，用來統合政府和民間地理資料的散布和傳播。此命令也強制所有聯邦機構，自此命令公布的 9 個月後（即 1995 年 1 月起），必須遵循此一標準來描述所收藏的地理資料，包含所有聯邦機構參與的活動（如經費贊助）。在此行政命令的規範下，FGDC 的地理元資料可說是成了美國地理資料著錄的標準。

FGDC 的地理元資料，就其設計理念而言，是基於類似 NII 的國家空間資料基礎建設（National Spatial Data Infrastructure，NSDI）架構，在資料的交換、查詢、散布上，以電腦網路為其假想作業環境，因此在 US Executive Order 12906 中，明令要設置全國性地理資料的交換所（national geospatial data clearinghouse）-- 即一個連結地理資料生產者、管理者和使用者的分散式電子網路，換言之，交換所（clearinghouse）是機構或組織用來協助地理資料的搜尋、評價和傳布的一套軟體。[註 11] 交換所雖是由 FDGC 贊助，但它基本上是一個分散式的網路系統，由散布在網際網路上的眾多伺服器（server）組成，利用 ANSI Z39.50 協定作為主從式架構下，主/從站間資料查詢和結果回傳的協定。此外 FDGC 也製作了一個元資料的格式規範（profile）-- GEO [註 12] 作為在 Z39.50 架構下，操縱其資料項的依據。

為了確著作錄資料的語法（syntax）正確，在網際網路上已有此元資料的語法查核軟體 [註 13]，來協助維持著錄的品質。此外也有一個教導初學者如何寫作的指引文章。[註 14] 另外也有多種協助寫作的文書軟體。[註 15] 同時也可在 WWW 上運作。整體來看，FDGC 的地理元資料已自成格局，有著錄內容的標準，有資料交換的架構，有語法查核器，整個架構已大致完備，事實上在美國

已在網際網路上實際運作了。對 FGDC 有興趣的讀者，可參考以下二個站址：

1. <http://www.dwr.ehnr.state.nc.us/metadata/notes.htm>
2. <http://www.blm.gov/gis/nsdi.html>

另外一個與地理資料相關的元資料，是由美國太空總署（NASA）所制定的資料著錄格式 -- 目錄交換格式（DIF）[註 16]，NASA 因為擁有大量的地理方面衛星照片和遙測資料，因此 DIF 也成為全世界地球科學界通行的一個標準。為了與 FDGC 的地理元資料標準相容，新版的 DIF 做了不小程度的翻修，總共含三十三個欄位，同時也有一個 DIF 和地理元資料的欄位對照表 [註 17]。有關 DIF 的詳細介紹，請參考作者的另一篇文章--目錄交換格式 [註 18]。

政府資訊指引服務（GILS）[註 19] 是一個美國聯邦政府所推動的資源指引器系統，管理暨預算局（Office of Management and Budget，簡稱 OMB）認為須要一種能協合各機構資源和資料傳布的工具，於是推動一個以機構為單位的政府資訊指引服務，來協助大眾對聯邦機構各項資源的檢索。[註 20] 同時美國國家標準暨技術局（National Institute of Standards and Technology，簡稱 NIST）已規定所有聯邦機構必須採用 GILS 來指引其文件出處。[註 21]

GILS 是一個利用網路，和 ANSI Z39.50 標準來執行檢索的系統，因此又被暱稱為虛擬卡片目錄（virtual card catalog）。透過 GILS，大眾可以在網路上輕易掌握所有政府提供的資源、服務、和文件，並且知道如何去取得所須的資料。GILS 不但已在聯邦政府間廣為設立，美國許多州政府和其他機構也逐漸倣效設立。目前已有所謂的政府資訊指引服務核心項目標準（GILS Core Element Standard）和著錄指導文件，來協助人們建立 GILS 記錄。在美國 GILS 也被列為 NII 的一個組成部份。

做為一個資源指引者，GILS 是指引使用者到資源或檔案的儲存處，因此資源或檔案的詳細內容並非其描述重點，讀者須再進一步到被指引處，才能獲得與內容相關的資訊。在系統的設計上，主要是透過 ANSI Z39.50 [註 22] 做為資訊搜尋和檢索的工具，[註 23] 所以使用者可以直接利用網際網路來使用 GILS，或者透過中介機構。而中介機構可有多樣化的方式，來提供 GILS 的服務，除了直接上網查尋外，也可以使用電子郵件、電子布告欄、印刷文件等方式來提供 GILS 的服務。[註 24]

系統設計上的另一個特色是採用分散式架構，即由各機構自行建構和維護其本身的 GILS 系統，但允許使用者利用網路來一次搜尋許多不同的 GILS 站，換言之，使用者並不需要一次次簽入個別 GILS 站來查尋，因此某些聯邦機構，如國家檔案暨紀錄管理處（National Archives and Record Administration，簡稱 NARA）和政府出版局（Government Printing Office，簡稱 GPO）會維持一份完整的聯邦機構 GILS 站的清冊，即所謂的政府資訊指引服務核心（GILS Core）[註 25]，它是由所有聯邦機構的 GILS 系統所組成，同時在紀錄的著錄上，都遵循政府資訊指引服務核心項目標準（GILS Core Element Standard）[註 26]，此核心項目祇有 14 個必須著錄項，同時根據估計，每個 GILS 紀錄的平均長度會少於 1000 個字，此外其著錄規範文件，包括說明、實例、和附錄，全文長度祇有 43 頁，因此在著錄上，GILS 可以說是符合電腦和網路時代，易學好用的原則。有關 GILS 的其他細節，請參考作者的另一篇文章--政府資訊指引服務 [註 27]。

都柏林核心集（Dublin Core）這個元資料格式，是 1995 年 3 月由 Online Computer Library Center（OCLC）和 National Center for Supercomputing

Applications (NCSA) 所聯合贊助的研討會，經過五十二位來自圖書館、電腦和網路方面的學者和專家，共同研討下的產物。目的是希望建立一套描述網路上電子文件特色的方法，來協助資訊檢索。因此在研討會的報告中，將元資料定義為資源描述 (resource description)，而研討會的中心問題是--如何用一個簡單的元資料記錄來描述種類繁多的電子物件？[註 28]

因為研討會的目標，是發展一個簡單有彈性，且非專業人員也可輕易了解和使用的資料描述格式，所以都柏林核心集祇規範那些在大多數情況下，必須提及的資料特性，最初規範有 13 個資料項 [註 29]，在 1996 年 OCLC 等所舉辦的一場研討會上，根據與會影像處理專家的建議，都柏林核心集新增了二個資料項--簡述 (Description) 和版權規範 (Rights Management)，並修改了部分資料項名稱 [註 30]，在此以扼要的方式，將 1996 年 12 月公布的資料著錄項目列表如下：[註 31]

- (一) 主題和關鍵詞 (Subject and Keywords)：作品所屬的學術領域。
例子：Subject = Digital Geospatial Metadata
- (二) 題名 (Title)：作品名稱。
例子：Title = Geospatial Support Staff Metadata Tutorial
- (三) 著者 (Author or Creator)：作品的創作者或組織。
例子：Creator = Abeyta, Carolyn
- (四) 簡述 (Description)：文件的摘要或影像資源的內容敘述。
- (五) 出版者 (Publisher)：負責發行作品的組織。
- (六) 其他參與者 (Other Contributors)：對作品創作有貢獻的相關人或組織。
- (七) 出版日期 (Date)：作品公開發表的日期。
例子：Date = 1995/05
- (八) 資源類型 (Resource Type)：作品的類型或所屬抽象範疇。
例子：Type = Dictionary
- (九) 資料格式 (Format)：告知檢索者在使用此作品時，所須的電腦軟體和硬體設備。
例子：Format = text/html
- (十) 資源識別代號 (Resource Identifier)：字串或號碼可用來唯一標示此作品。
例子：Identifier (scheme = URL) = http://www.blm.gov/gis/meta/barney/tut_met1.html
- (十一) 關連 (Relation)：與其他作品 (不同內容範疇) 的關連，或所屬的系列和檔案庫。
例子：Relation (type = ContainedIn) (identifier=URL) = <http://www.blm.gov/>

(十二) 來源 (Source)：作品從何處衍生而來 (同內容範疇)。

(十三) 語言 (Language)：作品所使用的語言。

例子：Language = English

(十四) 涵蓋時空 (Coverage)：作品所涵蓋的時期和地理區域。

(十五) 版權規範 (Rights Management)：作品版權聲明和使用規範。

其中某些資料項，是針對電腦作業環境而設計的，如資料格式 (Format)，其他如資料類型 (Resource Type)、關連 (Relation) 和來源 (Source)，也和網路或電子作業環境有密切的關係。[註 32] 同時此資料描述格式可說是非常簡單和易使用，幾乎所有資料項都有自我解釋的功能，大部份人在短時間內就知如何使用。有關都柏林核心集的其他細節，請參考作者的另一篇文章--三個元資料格式的比較分析 [註 33]。

四、未來資料著錄和檢索的發展趨勢

以下我們就上述的元資料標準，歸納出下面的一些共同特色和發展趨勢：

- (一) 個別資料類型有其支配性的元資料：地理元資料和目錄交換格式，都是針對地理資料特性而設計的元資料，此二種元資料，與前面所介紹的政府資訊指引服務和都柏林核心集，就資料項來加以比較，可以發現它們之間有很大的差異。因此正如柏林核心集研討會與會者指出的，沒有任何單一的元資料格式，足以適用於任何的作業環境，所以作者認為在元資料的未來發展趨勢上，會是針對每種主要資料類型有一個支配性的資料描述格式，如地理資料中的地理元資料。
- (二) 格式精簡且適合各種專業人士：元資料基本目的之一，是提供眾多非圖書館專業人士，一套簡單好用的資料描述格式，並且盡量降低紀錄的製作成本，來應付快速增加的資料量，因此都柏林核心集祇規範了十五個項目且無必須著錄項。DIF 總共規範了三十三個項目，但其中祇有六個是必須著錄項。GILS 的核心項目有二十二項，但祇有十四個是必須著錄項。由於可知，著錄項目的精簡和簡單好用，是元資料未來的發展趨勢。
- (三) 全球單一資訊網：未來的資料描述格式，必須是要以類似網際網路的全球資訊網為其設計標的，前面所介紹的四種元資料都是如此。另外一個正在發展中的元資料--資源特性統一描述格式 (Uniform Resource Characteristics, URC) [註 34]，更清楚的揭示了這個原則，URC 是網際網路工程任務小組 (Internet Engineering Task Force, IETF) 設計用來銜接現在 WWW 上的主角 -- 資源位置統一描述器 (Uniform Resource Locator, URL) [註 35] 和未來的主角 -- 資源名稱統一描述格式 (Uniform Resource Name, URN) [註 36]，URN 設計的基本目的，在確保世界上沒有二個不同文件 (或檔案) 會有相同的 URN 碼，因此如圖書館的登錄 (流水) 號一樣，具有唯一性的特質。有關 URC 的詳細介紹，請參考作者的另一篇文章--三個元資

料格式的比較分析 [註 37]。

- (四) 分散式處理系統：地理元資料和GILS在設計上的基本原則之一，即是採用分散式的處理方式。因為在全球單一資訊網的作業環境下，沒有任何一台主機，可以儲存所有的資料和負荷所有的檢索需求，故而分散式處理為時勢所趨，元資料的設計須能顧及此原則。
- (五) 易擴展原則：此原則是為了適應全球網路的作業環境，因眾多的站台各有自己獨特的資料種類和需求，因此必須有適當的彈性。能隨時加入新資料項，而不會影響到使用舊版本的應用程式，即能以逐漸增加的方式來擴大其功能和應用範圍；同時應用程式碰到不認識的資料項，能加以忽略，繼續處理其他資料項。此外元資料也須能隱含版本資訊，使應用程式能以適當方式，來同時處理不同版本的元資料。
- (六) 可修飾原則：都柏林核心集即採用這原則來同時滿足圖書館專業和非專業人員的需求。對於非專業人員來說，他們基本上不須要去查閱專業書籍來進行著錄的工作，這將大大減輕項目的著錄成本和時間。另一方面，對欲維持一定品質的專業人員而言，透過在（）內加修飾語，可明確指出所使用的資訊來自何處，如：Subject (=LCSH) =UNIX (Computer System) [註 38]。作者非常贊同這個可同時兼顧專業和非專業人員的設計理念。

五、元資料實驗系統 (MES) 簡介

最後簡介一個資料著錄和檢索的實驗系統--元資料實驗系統 (Metadata Experimental System, 簡稱 MES, 網址: <http://blue.lins.fju.edu.tw/~wu>)，作者建立 MES 目的，除了是讓讀者透過這個系統，對元資料及其未來的可能運作方式，有更具體的認知外；也希望利用此一實驗系統來測試和驗證元資料的功能和效用，例如都柏林核心集這種簡易的資料描述格式，是否如制定者們所預期的，足以滿足大部分網路文件著錄和檢索的需求。MES是一開放性的實驗系統，歡迎任何人上站著錄自己的網頁或文件，以供他人查詢和檢索。以下就簡介 MES 目前的狀況、特色、功能、以及未來的發展方向。

- (一) 目前狀況: MES 尚處於 Alpha 版的測試狀態，然而這並不代表系統運作不穩定，事實上初步測試顯示運作良好，而是意味內容版面在未來半年內會經常調整，但是不論如何調整，作者會盡力確保已著錄資料的完整。截至1997年6月止，祇包含兩種元資料類型--都柏林核心集 (Dublin Core) 和 IETF 正在規劃中的 URI 架構 (包含 URN, URL, URC)。
- (二) 系統特色和功能:
 - (1) 同時提供著錄和檢索兩種功能：為何強調著錄和檢索的合併呢？因為目前的主要檢索引擎都是採用自動拆字(或詞)作索引的方式來建立其資料庫，做為檢索的基礎，這固然快速和滿足了部分的檢索需求，但是隨著網際網路資料的不斷快速膨脹，很多使用者已抱怨檢索引擎無法有效替他們來過濾查詢所得資料，這明白顯

示此種方式有極大的缺失。而圖書館界很早以前就體認到資料描述的必要性，這正是元資料所扮演的角色之一，無怪乎元資料越來越受到重視。MES建立的目的之一，即在測試元資料在檢索上的效用。值得一提的，MES的處理對象並不祇限於網路文件，傳統印刷媒體資料亦歡迎著錄，因此MES也可提供傳統書目資料查詢的功能。事實上，MARC已被視為元資料的一種。

- (2) 開放性設計：著錄和檢索部份均開放給任何人使用。由於目前在人工智慧(Artificial Intelligence)和類神經元網路(Neural Networks)上的發展，尚無法創造具有類似圖書館員素質的自動著錄系統，事實上，連模仿三歲兒童說聽故事的能力都還辦不到，因此在可預見的將來，以人工著錄仍為無法避免的事實。至於由誰來擔負主要的資料描述工作呢？由WWW的運作方式和網際網路上資料的快速膨脹來看，圖書館員是無法負荷此龐大的著錄工作份量，所以由文件創作者自行著錄實為惟一的解決之道，這正是MES開放著錄功能的主要原因，雖然可能產生著錄品質不劃一的問題，但兩害相權取其輕。
- (3) 使用URN作為資源(或文件)的唯一識別碼：由於未來的趨勢是以URN來取代URL作為資源(或文件)的識別名稱，因此MES採用URN作為文件的唯一識別名稱，同時未來MES中的所有元資料格式，均可藉由URN連結在一起。但由於IETF尚未對URN作出最後規範，因此MES目前採取以下過渡措施：

資源(或文件)的URN為
MES:{您的IP位址}:{您的MES識別名稱}:{您自訂的文件編號}
例子: URN: MES:140.136.85.1:wu:yk00001

為了確保唯一性，文件擁有者請維護您文件編號的完整和唯一性。非文件擁有者請先使用MES提供的查詢功能進行URN確認。

- (4) 提供模糊檢索功能：適當欄位如主題(subject)，加入關連值(weight)來提供模糊檢索功能。傳統關鍵字和主題的處理方法，以及布林邏輯的運作方式，是屬於二元邏輯(即祇有真和假兩種可能值)。這已無法滿足檢索需求，為能更精確來過濾資料，某個關鍵字或主題在文件的重要性，可更有彈性的以(0,1]--即0到1(不含0,但含1)來表示，關連值為1代表此文件和這個關鍵字或主題完全相關。(注意:由於篇幅有限，無法詳細介紹模糊邏輯(Fuzzy Logic)所牽涉的種種概念和專有名詞，為避免造成讀者可能的閱讀負擔，此處使用名詞非為正統，但求能使讀者有一直覺式的了解。)

(三) 未來的發展方向

- (1) 加入其他的元資料格式如GILS等，或提供評價認證(Seals of Approval, SOAP)等功能。
- (2) 進一步加強模糊檢索功能，由於模糊檢索目前尚有若干理論瓶

頸，因此未來的努力重點是研究如何突破理論瓶頸，來建立一個完整和高效率的模糊檢索系統。

六、結論

網際網路和WWW的結合，大幅降低了資訊傳播的障礙，於是全球單一資訊網的架構已在逐漸形成中，但這引發了資訊量過多的問題。而如何有效率來過濾和處理大量資料，乃成為亟待解決的課題，作者以為這個問題的解決方案，將主要依賴二個方法，一是資料提供者運用元資料來提供與文件相關的充分資訊給檢索者，一是檢索系統採用模糊邏輯原理，來加強對資料的過濾和處理能力。以前者而言，綜觀目前大多數的搜尋引擎，在資料的回覆畫面上，都祇有顯示標題、密合百分比、簡短的數行文字、URL（路徑+檔名）、有些系統有附上檔名大小和製作時間。如此簡略的設計，無怪乎檢索者無法判斷某筆資料到底是否為其所需，而惟有將整個檔案下載，直接閱讀後才能得知。這種操作是很沒有效率的，因為網路的傳輸部分，往往是系統最慢的一個環節。解決之道應是透過元資料來對資料加以適當的描述，提供給檢索者更多的資訊來做判斷，而達到減少不必要傳輸的目的，事實上，這正是目錄的基本功用。

元資料對電子文件（或檔案）所扮演的角色，正可對比於目錄之於傳統的印刷媒體資料，因此元資料可說是『電子目錄』，正如目錄過去所扮演的角色一樣，元資料將可大幅減少不必要的檔案傳輸次數，提高資料檢索的效率。

元資料中也可放入其他資訊來協助檢索，其中一個可能的應用是評價認證（Seals of Approval, SOAP），[註 39] 這是將資料所屬領域專家對此資料的評價資訊放入 URC 中，而此種專業的評價資訊，在檢索時可作為過濾資料的一個標準。在設計上應允許多個電子簽名和評價認證，同時存在一個元資料紀錄中，因此一個文件可有多個評價。

在配合系統的安全防護和防止不當的資料存取上，元資料中也能放入一些項目來配合系統的安全體系，如將電子簽名（digital signature）的資訊，加入與收藏資料相對映的元資料中，則系統可提供安全認證（authentication）的功能，因為資料接收者可利用電子簽名，來查證所接收的資料，是否的確來自指定機構或指定人。

總結來說，元資料是因為全球資訊網的作業環境，和電子檔案逐漸成為資料主流等趨勢而興起的資料描述格式。元資料除了負起傳統目錄指引資料和協助檢索的功能外，在格式的設計上，也須能顧及電子檔案所獨有的一些特性，如檔案格式的種類繁多、資料轉換需求頻繁、版本辨識困難等問題。在未來的發展趨勢上，作者認為將有如下的特色：個別資料類型有其支配性的元資料、格式精簡且適合各種專業人士、全球單一資訊網、分散式網路系統、易擴展原則、可修飾原則。

最後簡介一個資料著錄的實驗系統--元資料實驗系統 (Metadata Experimental System, 簡稱 MES, 網址: <http://blue.lins.fju.edu.tw/~wu>)，透過這個系統，讀者對元資料及其未來的可能運作方式，可有更具體的認知；也可利用此一實驗系統來測試和驗證元資料的功能和效用。MES 系統特色和功能如下：同時提供著錄和檢索兩種功能、開放性設計、使用 URN 作為資源(或文件)的唯一識別碼、提供模糊檢索功能。

註釋：

- 註 1：IFLA, "UNIMARC: An Introduction," <www.nlc-bnc.ca/ifla/VI/3/P1996-1/unimarc.htm> (26 Sept. 1996).
- 註 2：Tim Berners-Lee and Robert Cailliau, "WorldWideWeb: Proposal for a HyperText Project," 1990, <<http://info.cern.ch/hypertext/WWW/Proposal.html>>.
- 註 3：NIST, "National Information Infrastructure (NII): General Information," 1996, <<http://nii.nist.gov/nii/niiinfo.html>>.
- 註 4：吳政勸，「三個元資料格式的比較分析」，中國圖書館學會會報 57 期（民 85 年 12 月），頁 35-45。
- 註 5：黃中生等，電腦大辭典（普及本），（台北市：松崗，民 85 年），頁 p-24。
- 註 6：Stuart Weibel, Jean Godby, Eric Miller, and Ron Daniel, "OCLC/NCSA Metadata Workshop Report," 1995, <http://www.oclc.org:5047/oclc/research/publications/weibel/metadata/dublin_core_report.html>, p. 12.
- 註 7：Infoseek Corp., "Infoseek Home Page," <<http://www.infoseek.com/>> (18 Feb. 1996).
- 註 8：E.P. Shelley and B.D. Johnson, "Metadata: Concepts and Models," in Proceedings of the Third National Conference on the Management of Geoscience Information and Data (Adelaide, Australia: Australian Mineral Foundation, 1995), pp. 1-5.
- 註 9：FGDC, "Content standards for digital geospatial metadata -- FGDC," 1994, <<http://fgdc.er.usgs.gov/fgdc.html>>.
- 註 10：William Clinton, Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure (Washington D.C.: The White House, 1994).
- 註 11：FGDC, "Making your Spatial Data Discoverable: Building an NSDI Clearinghouse Node," <<http://www.fgdc.gov/clearinghouse/itraining/toc.html>> (26 Sept. 1996).
- 註 12：Brandon Plewe, "GeoWEB Site Home Page," 1995, <<http://wings.buffalo.edu/geoweb/>>.
- 註 13：Peter Schweitzer, "Metadata Validation Service," <<http://www-mel.nrlmry.navy.mil/meta-val.html>> (26 Sept. 1996).
- 註 14：Carolyn Abeyta, "Geospatial Support Staff Metadata Tutorial," 1995, <http://www.blm.gov/gis/meta/barney/tut_met1.html>.
- 註 15：BLM, "Metadata and WWW Mapping Home Page," 1996 <<http://www.blm.gov/gis/nsdi.html>>.
- 註 16："Directory Interchange Format (DIF) Writer's Guide, Version 5.0a," Oct. 1996, <<http://gcmd.gsfc.nasa.gov/difguide/difman.html>>.
- 註 17："FDGC Metadata Standard to GCMD DIF," <http://gcmd.gsfc.nasa.gov/dif_mapping/fgdc_dif_map5_0.html>, (9 Dec. 1996).
- 註 18：吳政勸，「目錄交換格式」，臺北市立圖書館館訊 3 卷 4 期（民 86 年 6 月），頁 16-31。
- 註 19："Guidelines for the Preparation of GILS Entries," March 1995, <<http://gopher.nara.gov:70/0/managers/glis/guidance/gilsdoc.txt>>.
- 註 20："GILS 5/2/94 Document Executive Summary," May 1994, <<http://www.usgs.gov/gils/gilsexec.html>>, p. 1.
- 註 21："GILS 5/2/94 Document Table of Contents," May 1994, <<http://www.usgs.gov/gils/gilstoc.html>>, p. 3.
- 註 22：National Information Standards Organization, ANSI/NISO Z39.50-1992, Information Retrieval Application Service Definition and Protocol Specification

for Open Systems Interconnection (Gaithersburg, MD: National Information Standards Organization Press, 1992).

- 註 23：同註 21，頁 8。
- 註 24：同註 21，頁 5。
- 註 25：同註 21，頁 9。
- 註 26："GILS 5/2/94 Document Appendix A GILS Core Elements," May 1994, <<http://www.usgs.gov/gils/gilsappa.html>>.
- 註 27：吳政勳，「政府資訊指引服務」，國立中央圖書館臺灣分館館刊 3 卷 4 期（民 86 年 6 月），頁 16-31。
- 註 28：同註 6，頁 2。
- 註 29：同註 6，頁 7-11。
- 註 30：S. Weibel and E. Miller, "Image Description on the Internet: A Summary of the CNI/OCLC Image Metadata Workshop," D-Lib Magazine (Jan. 1997), <<http://www.dlib.org/dlib/january97/oclc/01weibel.html>>.
- 註 31："Dublin Core Metadata Element Set: Reference Description," 15 Jan. 1997, <http://purl.org/metadata/dublin_core_elements>.
- 註 32：同註 4，頁 40。
- 註 33：同註 4，頁 39-40。
- 註 34：Daniel LaLiberte and Michael Shapiro, "Universal Resource Characteristics (URCs)," 1996, <<http://union.ncsa.uiuc.edu/HyperNews/get/www/URCs.html>>.
- 註 35：T. Berners-Lee, L. Masinter, and M. McCahill, "Uniform Resource Locators (URL)," 1994, <<ftp://ds.internic.net/rfc/rfc1738.txt>>.
- 註 36：K. Sollins and L. Masinter, "Requirements for Uniform Resource Names," 1994, <<ftp://cnri.reston.va.us/internet-drafts/draft-ietf-uri-urn-req-00.txt>>.
- 註 37：同註 4，頁 41-42。
- 註 38：同註 6，頁 7。
- 註 39：J. Rhine, "Interpedia Homepage," 1994, <<http://www.hmc.edu/interpedia/index.html>>.