

元資料及其在中國大陸的應用 Metadata and Its Applications in the Mainland China

劉 嘉

Jessica Liu

北京大學資訊管理系

Department of Information Management

Peking University

E-mail : jialiulj@hotmail.com

【摘要 Abstract】

隨著網路的迅速發展，對網路資訊資源的組織日顯必要和迫切，作為描述促進資訊資源並促進資訊資源發現的元資料因而受到益發普遍的重視。本文首先簡述元資料的定義，並對其內涵和重要性做進一步的闡釋；隨後，介紹元資料的類型和作用；最後，列舉元資料在全世界範圍內的典型應用實例之外，著重簡要說明中國大陸目前的元資料應用專案。

With the rapid development of the network, it becomes more and more necessary and emergent to organize the networked information resource. As a means to describe the information resource and provide access to it, metadata has been taken seriously during these years. First, the article gives a brief introduction to the definitions of metadata and, further, explains its connotation and significance. Then, the types and roles of metadata are dealt with in the following discussion. Finally in addition to showing some typical examples of the project in applying metadata all over the world, the writer focuses on introducing several such projects in the mainland China.

關鍵詞 Keyword

元資料 數位圖書館

Metadata ; Digital library



壹、前言

在因特網發展的最初幾年裏，它緊緊依賴於工具和方法，只需要很少或幾乎不需要人類的干涉或系統化。但是，隨著因特網的不斷發展，有一點日趨明顯，即基於因特網站點內容的全文標引的方法，並不是為這些資源提供存取的一種完整或完全適當的解決方法。人們需要擴大和豐富資料的「自我描述」的方法，並鼓勵製作者和其他人參與到這方面的工作中來。增加關於資源的額外資訊或元資料是更好地組織資源的基礎，它能夠提高相關資源被檢索和存取的可能性，提供對主題領域更清晰的全面認識，並提高用戶區別相似資源的能力。正因為如此，元資料逐漸受到普遍的重視。

貳、元資料概述

元資料的英文名稱為 *Metadata*。該詞最早出現於美國航空與宇宙航行局(National Aeronautics and Space Administration, NASA)的《目錄交換格式》(Directory of Interchange Format, DIF)手冊中。(註 1)

根據現有的文獻，在大陸，基本上都將 *Metadata* 翻譯為「元數據」；而在臺灣，它的中文譯名則比較多樣化。臺灣最早對 *Metadata* 提出中文譯名並已有理論和實踐成果的吳政叡教授，把 *Metadata* 譯為元資料，因為在臺灣，“data”一詞被譯作為資料。此外，臺灣還有一些研究者將其譯作詮釋資料、超資料等。在此，為了更便於與臺灣同行交流，本人即採用在臺灣較為通用的名稱「元資料」作為 *Metadata* 的譯名。

一、何為元資料？

元資料是描述和限定其他資料的資料。元資料的典型實例有資料的重要特性（如創作者的名

稱、出版年）、用於資料定位的資料（如圖書館圖書的杜威分類號、電視節目的時間和頻道）以及有助於資料檢索的資料（如數據的自由文本描述或資料的摘要，或者一系列適用於資料的可檢索的主題關鍵字）。

關於元資料，迄今為止，還沒有完全統一的定義，最常規的定義就是：元資料是關於資料的資料(Data about data)。(註 2)

但這個定義過於簡潔，無法清晰地反映出元資料的內涵。於是，一些專家和學者就把這個解釋加以擴展和深化。比較具有代表性的幾種定義有：

(一)元資料是關於資料的結構化的資料(Structured data about data)。這個概念突出了元資料的結構化特徵，從而使採用元資料作為資訊組織的方式同全文索引有所區分。(註 3)

(二)元資料是與物件相關的資料，此資料使其潛在的用戶不必預先具備對這些物件的存在或特徵的完整認識。它支援各種操作。用戶可能是程式，也可能是人。(註 4)

(三)元資料是對資訊包(Information package)的編碼描述(如，用 MARC 編碼的 AACR2 記錄、都柏林核心記錄、GILS 記錄等)，元資料之目的在於提供一個中間級別的描述，使得人們據此就可以做出選擇，確定孰為其想要瀏覽或檢索的資訊包，而無需檢索大量不相關的全文文本。(註 5)

(四)元資料，即代表性的資料，通常被定義為資料之資料。它包含用於描述資訊物件的內容和位置的資料元素集，促進了網路環境中資訊物件的發現和檢索。(註 6)

此外，還有兩種關於元資料的重疊但又有些矛盾的定義值得考慮。

第一個來自 Priscilla Caplan，芝加哥大學圖書館的系統負責人助理，他參與了都柏林核心元資料標準形成的那次討論。他提出：「元資料的確



不過就是關於資料的資料；書目記錄是元資料；TEI 標題也是，或其他形式的描述」。(註 7)

第二個出自國際圖聯(International Federation of Library Associations & Institutions, IFLA)有關「元資料資源」的主頁：「元資料是關於資料的資料。此術語指任何用於幫助網路電子資源的識別、描述和定位的資料」。(註 8)

Caplan 的定義似更可取，因為它兼顧了印刷和聯機環境。它還向人們再度證實元資料並不是什麼新奇的東西，而是出版商以及圖書館員在幾個世紀以來一直在製作和使用著的。同時，更加嚴格的 IFLA 的定義提醒人們元資料源於數位資訊的世界。它意味著元資料能夠採用多種形式，以不同的級別存在，為了滿足不同的目的，可以通過多種方式生成。

搜索引擎利用機器來自動擷取非結構化元資料的方式，較容易且經濟。主題指南經過人工增值，比前者費時、費力、費財。但此二者均無法提供較詳細的資源描述給使用者，無法完全滿足用戶複雜的資訊需求。而到目前為止，惟有元資料才能為網路資源提供用戶所需要的較詳細、清晰和準確的資源描述。在因特網和萬維網飛速發展、網路資訊資源迅猛增長的情形下，為有效率地進行資源檢索，元資料的建立愈顯必要。

二、對元資料的進一步認識

由於到目前為止，人們對元資料的認識還時常會有所混淆甚至產生謬誤，在此，筆者認為有必要就元資料做進一步的說明和解釋。

(一)元資料不一定是數位形式的

元資料的形式是多樣化的，並不一定是數位形式的。在管理人類文化遺產的過程中，有關專家一直在編制元資料。只是隨著電腦技術的發展，人們越來越頻繁地生成為電腦所能識別和採

用的數位式元資料，並將原有的元資料集成到數位資訊系統之中。

(二)元資料不只同對資訊物件的描述相關

儘管許多博物館、檔案館和圖書館的專家們對元資料的描述或編目功能最為熟悉，但是，他們同樣認識到，元資料還能夠說明被描述資源的使用環境、管理、加工、保存和使用等方面的情況。

(三)元資料可以來自各種不同的資源

元資料可以由人類（編制者、資訊專家或使用者）提供，還可以由電腦自動生成，或者通過一項資源與另一項資源的關係來推斷，如超鏈結。

(四)在資訊物件或系統的生命過程中自然增加元資料

在資訊物件或資訊系統的整個生命過程中，元資料首先被生成，繼而被修改，甚或在某個階段被廢棄。

(五)元資料常規定義中的「資料」

在上文所提及的元資料的常規定義中，對前後出現的兩個「資料」，應有較為寬泛的理解。一般而言，資料(Data)表示事物性質的符號，是進行各種統計、計算、科學研究、技術設計所依據的數值，或者說是數位化、公式化、代碼化、圖表化的資訊。(註 9)不過，在上述定義中，更準確地說，第一個定義所代表的含義應是「資源」。且兩個「資料」都不一定是數位形式的，而可以以各種不同的形式出現。

三、元資料是全新的概念嗎？

誠如在介紹元資料的概念部分中已經提及的，事實上，元資料並非什麼新生事物，而是由來已久。從關於元資料的常規定義，即可發現這一點。圖書館的書目記錄就是一種元資料。目錄



記錄通常都包含諸如題名、作者、出版商、主題和載體描述之類的標識資訊，這是印刷環境中最明顯的元資料應用。

然而，儘管元資料的形式和功能於很多方面在聯機環境中仍基本保持一致，但是，就聯機資訊物件而言，與其印刷品相比，技術已經帶來了結構上的革新，從而使網路環境下的元資料發生革新，這些革新可以被分為新的單元、新的功能和新的模式。(註 10)

(一)新的單元

新的單元源於生成資料物件的新技術，而非資訊管理系統中所採用的新技術。優秀的資訊管理者使人們能夠得到各種類型的元資料，而這些元資料對於回答有關某一資源的普通問題而言是必要的。無庸置疑，描述各種類型資源的元資料是不同的，隨著新的資訊資源的產生，對之進行描述的元資料中必將出現新的單元。電子資源具有許多印刷資料所沒有的特徵。而且，存在多種電子資源格式(如數位資料文檔、網站、CD-ROM等)，每種格式都有其特殊的屬性集，而不出幾個月，就會出現一種新的格式。像英美編目條例第二版(AACR2)和機讀目錄格式(MARC Format)這樣的圖書館標準，已經在努力追隨這種發展趨勢，然而，出於對普適性的要求，它們無法根據某些用戶群的需求提供詳細級別的特殊化格式。因此，派生出針對此類資源類型的元資料方案，如地理資訊系統(Geographic Information System, GIS)和人文文本文檔(Humanities text file)。這些新的元資料格式中所包含的許多單元，同適用於以往的印刷型資訊資源的元資料格式中的單元相比，都必然是嶄新的。

(二)新的功能

以往，元資料的功能基本上就是描述，在描述的基礎上提供檢索。元資料和資源本身是分離

的兩個主體。而伴隨著 HTML 的使用，目錄記錄、題名清單或其他地方中元資料單元的熱鏈結為檢索者生成了從外部製作的元資料到資源本身的直接的鏈結。(註 11)通過使用相關性排序和概念映射之類技術的運算法則，搜索引擎可以模仿人工標引者的行為生成元資料。目錄和資料庫檢索軟體還能夠被設定來隱藏元資料，不管是用於簡要記錄的顯示還是自然語言的檢索幫助。

(三)新的模式

電腦儲存和網路資訊資源檢索的發展，已經引發了至少兩種有關元資料的新模式的形成，這使得元資料看起來有了新意，而不僅僅是「新環境中編目那樣的古老藝術」。一種模式為「資料庫模式」(Database Paradigm)。在這種模式中，與資料物件相關的元資料被記錄為以值來表示的屬性；這些屬性及其定義都被登記在一個資料詞典裏。元資料被存儲在記錄中，根據編程所做出的選擇，可以以完整或任何可能的子集顯示(如目錄中簡短的款目或檢索結果的清單)。幾個資料物件共用的元資料可被保存在一個與這些物件的元資料記錄鏈結的單獨的記錄中，尤其是在元資料間有互相參照的情況下(例如，一個作者的名稱有幾種形式)。每個元資料款目都是自治的，只取決於在資料詞典中對它的定義和根據資料物件的含義所給定的識別字，而不依賴於其他的元資料。如此，就可在不損失意義的情況下以各種序列和標識方案顯示該元資料。使用 MARC 格式的圖書館聯機目錄模仿了資料庫模式。但嚴格地講，缺少一個有效的資料詞典，因為記錄仍然是依據 ISBD 那種結構化的規則生成的，然後，被目錄軟體「轉換」為一系列屬性和值。由此而導致的冗餘、不靈活、引起混淆的趨勢在文獻中常被論及。(註 12)另一種模式為「標記模式」(Markup paradigm)，涉及資料物件中元資料的嵌入，通常



使用 SGML 文獻類型定義來標識資料。SGML 非常典型地被用於控制資料的顯示，而非描述資料物件，但通過為對於識別甚為關鍵的資料單元做標記，它能夠標引資料物件並促進對它的直接檢索。此外，這種標識方法還可被用於自動生成目錄記錄。或許，標記模式最大的潛力在於為資料本體內部的關係編碼，用於基於人工智慧的資料挖掘應用和精確的檢索系統。

綜上所述，就其內涵而言，元資料早已存在。但當前的元資料所針對的物件主要是網路資訊資源，其典型的作業環境為電腦網路環境，在功能和模式方面已有所革新。

四、元資料為什麼是重要的？

上面的描述已經反映出，元資料乃可能需要花費良多於生成和維護的複雜架構。那麼，如何證明為其所付出的代價和努力都是值得的呢？對此，萬維網的發展以及其他網路數位資訊系統已經為資訊專家們提供了許多機會，與此同時，又使得他們不得不去面對以前所不曾或偶然才會遇到的問題。通過精心設計而形成的元資料，都盡可能地與國內或國際的標準相吻合，它們業已成為資訊專家們充分利用現有的機遇並解決新問題的工具。

(一)提高可檢索性

通過採用豐富而相互相容的元資料，檢索的有效程度得到了大大的提高。只要在各個地點的描述型元資料是相同的或是可以被映射的，元資料就能夠使在多個館藏間的檢索或為分散在數個收藏地的資料生成數位館藏成為可能。不同的專業群體開發出不同的數位資訊系統和元資料標準，但它們都選用一些通用的資料單元，如編碼檔案描述(Encoded Archival Description, EAD)、文本編碼倡議(Text Encoding Initiative, TEI)和都

柏林核心(Dublin Core)皆如此。這些專業群體正努力使用戶能夠在資訊物件的各種描述替代物與資訊物件本身的不同數位版本之間更容易地穿梭，並且能夠在資訊系統內部和各系統之間以文獻級別和館藏級別對資訊物件進行檢索。

(二)保持情境

博物館、檔案館和圖書館並非簡單地擁有資訊物件，它們還維護著存在複雜的內部關聯的館藏物件，這些物件皆與一定的人物、地點、活動和事件相關。在數位世界裏，並不難做到將一個館藏中的某個單一的物件數位化，繼而將它與它自己的編目資訊及其與同一館藏中其他物件的關係中分離出來。元資料在記錄、維護這些關係以及顯示資訊物件的真實性、結構完整性和綜合性的過程中發揮著重要的作用。例如，當記載一條檔案記錄的內容、情境(Context)和結構時，有助於將此記錄與無關的資訊區分開來的，正是以檔案查找幫助形式出現的元資料。

(三)拓展使用

針對博物館和檔案館館藏的數位資訊系統使得唯一物件的數位版本更易於在全球傳播，否則，由於地理上的原因、經濟上的困難或其他障礙，這些機構的館藏物件不可能為全球的用戶所用。然而，新的用戶群的出現又帶來了新的挑戰，要求專家們解決如何使用戶最便利地檢索到這些資料的問題。這些新用戶群的需求可能與一些傳統用戶的需求大相徑庭，而現有的資訊服務却是為這些傳統用戶設計的。例如，教師和孩子們所要檢索和使用的資訊物件與學術研究者所要檢索和使用的就相差甚遠。元資料能夠記錄系統使用和內容方面的變化，並且將此類資訊反饋給系統開發的決策部門。結構良好的元資料還能夠為檢索資訊、顯示結果甚至在不影響完整性的情況下對資訊物件進行特殊處理提供幾乎無限多樣的途



徑。

(四)版本多樣化

如果資訊和文化物件以數位形式存在，為其製作多個不同的版本便不費吹灰之力，因而提高了人們在這方面的興趣。這種加工簡單得就像既可以出於保存或學術研究的目的製作高解析度的副本，同時還可以製作為快速參考目的在網上迅捷傳遞的低解析度的極小的圖像。或者，可以針對不同的用途，諸如為出版、展覽或學校教室使用，生成變化的或派生的形式。在這兩種情況下，是元資料將多個版本鏈結起來並捕捉到各個版本之間的異同。元資料還能夠識別數位版本和原件或母件之間質量上的差異。

(五)法律事宜

元資料使得擁有者能夠跟踪到許多資訊物件及其多個版本的許可權層面和複製方面的資訊。元資料還能夠記錄有關物件的其他法律或捐贈人方面的要求，如隱私或正當權益。

(六)保存

如果通過電腦硬體和軟體不斷的升級換代，或轉移到全新的傳遞系統，有為正在製作中的數位資訊物件連續移植的機會，這些資訊物件就需

要能夠使其獨立於系統存在的元資料，而此系統是正被用於存儲和檢索這些資訊物件的。技術、描述和保存方面的元資料，記錄數位資訊物件如何被製作和維護、如何運作以及它如何同其他資訊物件發生關聯。這些元資料都是非常基本的。要保持資訊物件始終能夠被檢索和可以被理解，就得注重元資料。此外，對元資料的保存和移植，也是十分重要的。

(七)系統改進和經濟考慮

電腦將自動收集一些作為基準點的技術資料，這對於評估和改進系統、從而使系統效率更高和從技術和經濟學角度考慮更經濟至關重要。此類資料還將被用於新系統的設計。(註 13)

參、元資料的類型

Paul Miller 將元資料粗略分為兩類：專家層次 (Expert approach) 與搜索引擎層次 (Search engine approach)。前者包括 MARC、TEI 標題等較複雜的資源描述架構；後者則在 HTML 文件中隱藏 META 語法，使之可以被搜索引擎用於檢索。(註 14)顯然，這種最簡單的分類方式無法透徹地反映出各種元資料之間的區分。

表一：同類型的元資料及其功能

類型	定義	例子
管理型元資料	在管理資訊資源中利用的元資料	<ul style="list-style-type: none"> ● 採購資訊 ● 權利和複製品追蹤 ● 法定檢索所要求的文獻 ● 位置資訊 ● 用於數位化的挑選標準 ● 版本控制
描述型元資料	用來描述或識別資訊資源的元資料	<ul style="list-style-type: none"> ● 編目記錄 ● 查找幫助 ● 特殊化的索引 ● 資源之間超鏈結的關係



保存型元資料	與資訊資源的保存管理相關的資訊	<ul style="list-style-type: none"> ● 用戶的注解 ● 資源實體條件方面的文獻 <p>保存資源的物理和數位版本中所採取行動，如數據更新和移植方面的文獻</p>
技術型元資料	與系統如何行使職責或元資料如何發揮作用相關的元資料	<ul style="list-style-type: none"> ● 硬體和軟體文獻 ● 數位化資訊，如格式、壓縮比例、縮放比例常規 ● 系統反應次數的追蹤 ● 真實性和安全性資料，如密碼、口令
使用型元資料	與資訊資源利用的等級和類型相關的元資料	<ul style="list-style-type: none"> ● 展覽記錄 ● 使用和用戶追蹤 ● 內容再利用和多個版本的資訊

資料來源：Anne J. Gilliland-Swetland, "Defining Metadata," In *Introduction to Metadata: Pathways to Digital Information*, ed. by Murtha Baca (U.S.A. : Getty Information Institute, 1998), p.3.

1998 年，美國 Getty 資訊研究所 (Getty Information Institute) 曾就元資料進行過一次專項研究。(註 15) 在有關於此的專著中，Anne J. Gilliland-Swetland 根據功能將元資料劃分為管理型元資料、描述型元資料、保存型元資料、技術

型元資料和使用型元資料 5 種類型，其各自的定義及相應的例子如表一所示。

而 Lorcan Dempsey 與 Rachel Heery 則根據結構和語意的遞增，將元資料分為以下三組，如表二所示。

表二：元資料格式的類型

第一組	第二組	第三組
全文索引	簡單結構化的普通格式	結構更加複雜，特殊領域內的格式
專屬格式	專屬格式	TEI 標題
	都柏林核心	ICPSR
	IAFA/WHOIS++ Template	EAD
	RFC 1807	CIMI
...

資料來源：Lorcan Dempsey and Rachel Herry, "Metadata: A Current View of Practice and Issues," *The Journal of Documentation* 54 : 2(Mar. 1998), p.156

在表二中，第一組是取自資源本身的全文索引的資料，通常根本都不使用「元資料」這個術語。隨著摘要和抽取技術的提高，其質量也會有

所提高。因特網標引服務中使用的系統所生成的資料，即被包括在第一組內。第二組是簡單的結構化的格式，使用範圍相當普遍。它包括幾個在



電腦科學領域已經出現的支援檢索和目錄服務的格式。都柏林核心是此類格式中的一個例外，它是一個得到普遍認同的簡單的元資料元素集。第三組是特殊領域內所使用的格式。在此，人們發現，更完整、更加結構化的格式通常都是為了滿足特殊的功能需求和在特定的領域或醫學傳統內發展起來的。第三組中較新的格式通常都是基於 SGML 的，或正在向這個方向發展。第三組內某些格式的一個特徵是，它們很明顯的是包含了「內容」標識（如 TEI）的更寬泛的框架中的部分。當然，還有許多在表中未提及的格式。（註 16）

肆、元資料的作用

顯而易見，元資料具有傳統目錄的「著錄」功能，目的在於使資源的管理維護者及使用者可通過元資料瞭解並辨別資源，進而利用和管理資源，為由形式管理轉向內容管理奠定必要的基礎。

Renato Iannella 及 Andrew Waugh 詳細地指出元資料主要用來：

1. 概述資源的內涵；
2. 讓用戶查找到該資源；
3. 讓用戶確定該資源是否是他所需要的；
4. 避免用戶存取該資源（如禁止兒童使用某類資源）；
5. 讓用戶檢索、複製資源；
6. 指示應如何解釋該資源（如說明資源的格式、編碼、加密等情況）；
7. 用來決定可檢索哪一項資源（若資源可以多種格式存在）；
8. 說明資源使用的合法情況；
9. 說明資源的歷史，如說明其原始資源如何及其它變化；
10. 說明資源的聯繫人，如擁有者；
11. 指示該資源與其他資源的關係；
12. 控制資源的管理。（註 17）

本人以為，對於元資料在網路資訊資源組織方面的作用，基本上可概括為以下幾個方面：描述、定位、搜尋、評估和選擇。

一、描述 (Description)

根據元資料的定義，它最基本的功能就在於對資訊物件的內容和位置進行描述，從而為資訊物件的存取與利用奠定必要的基礎。

對資訊物件描述的詳簡和深淺，隨具體採用的元資料格式之不同而各不相同。例如，都柏林核心元素集所提供的，是就資訊物件的識別而言最為基本的描述資訊；而機讀目錄格式，則為資訊物件提供詳細級別的描述。

二、定位 (Location)

由於網路資訊資源沒有具體的實體存在，因此，明確它的定位至關重要。

元資料包含有關網路資訊資源位置方面的資訊，因而由此便可確定資源的位置之所在，促進了網路環境中資訊物件的發現和檢索。

此外，在資訊物件的元資料確定以後，資訊物件在資料庫或其他集合體中的位置也就確定了。這是定位的另一層含義。

三、搜尋 (Discovery)

元資料提供搜尋的基礎。在著錄的過程中，將資訊物件中的重要資訊抽出並加以組織，賦予語意，並建立關係，將使檢索結果更加準確，從而有利於用戶識別資源的價值，發現其真正需要的資源。

四、評估 (Evaluation)

元資料提供有關資訊物件的名稱、內容、年代、格式、製作者等基本屬性，使用戶在無需瀏覽資訊物件本身的情況下，就能夠對資訊物件具備基本瞭解和認識，參照有關標準，即可對其價值進行必要的評估，作為存取與利用的參考。



五、選擇(Selection)

根據元資料所提供的描述資訊，參照相應的評估標準，結合使用環境，用戶便能夠做出對資訊物件取捨的決定，選擇適合用戶使用的資源。

此外，從系統的角度審視元資料，元資料的功能還包括提供瀏覽及檢索的功能、管理功能以及組合各個物件以及藏品的再呈現等。(註 18)

伍、元資料的應用

元資料的使用範圍非常廣，Susanne Boll 等認為元資料可被應用於：

1. 圖像檢索、導航和圖像集中的瀏覽
2. 視頻
3. 聲頻和演講
4. 結構化的文獻管理
5. 地理和環境資訊系統
6. 數位圖書館
7. 支援資訊存取的混合多種媒體(註 19)

一、元資料在整個世界範圍內的應用

在整個世界範圍內，元資料正在被用於越來越多的領域中。比較突出的實例即有：

- (一)教育資料閘道(Gateway to Educational Materials, GEM)(註 20)
- (二)歐洲數學圖書館和電子資源(The EULER Project, European Libraries and Electronic Resources in Mathematical Sciences)(註 21)
- (三)北歐萬維網索引(Nordic Web Index, NWI)(註 22)
- (四)亞歷山大數位圖書館(Alexandria Digital Library)(註 23)
- (五)DESIRE(Development of a European Service for Information on Research and Education, 歐洲研究與教育資訊服務之發

展)專案(註 24)

- (六)北歐元資料項目(The Nordic Metadata Project)(註 25)

數位圖書館是元資料應用非常突出的領域。數位圖書館提供範圍很廣的數位文獻方面的服務和收集，為元資料框架的發展和實施提供了頗具挑戰意義的應用領域。目前，所有正在開發中的數位圖書館專案都必須解決元資料方面的問題。這些專案中的大部分已經使它們的元資料框架能夠適用於數位資料館藏的描述，這些資料不僅包括全文資料，還包括參考資料集以及視頻和聲頻那樣的多媒體資料等。在數位圖書館中，元資料的主要作用是為分散式資料發現和檢索奠定基礎，元資料體系具備描述、整合、控制和代理四個基本功能。(註 26)目前，在全球範圍已有近二十個國家和地區在創建或已經創建了數位圖書館。它的第一批規模型成果已開始在因特網上陸續出現。

此外，元資料在知識管理中同樣能夠發揮重要的作用。當時間的脚步進入二十一世紀，人類也由資訊時代邁進知識時代。科學技術的發展及其相互影響造成了人類知識的快速增長，而因特網及萬維網的快速發展更促成了對於網路空間中數位資訊和知識的傳遞和獲取。有鑒於此，知識管理成為當今的重要課題。元資料能夠通過完整描述揭示物件的內涵，自然能夠在知識管理中發揮作用。元資料貫穿於資訊物件的整個生命過程，滲透於知識管理的各個環節。知識從資料到資訊的增值過程中所採取的方式，主要與資訊物件的形式特徵相關，最根本的依據是關於資訊物件的元資料；從資訊到知識的增值過程，則不僅限於外部的資訊加工與組織，而需要對資訊物件內容的加工和分析，其根本依據仍然是元資料。在知識管理的前期階段資訊增值（資料→資訊→知識）的階段中，元資料發揮著重要的作用；而



在知識管理中的其他階段中，元資料同樣不可或缺。充分利用元資料，是知識管理成功的關鍵之所在。

二、元資料在中國大陸的應用

無疑，在中國大陸，迄今為止，最成熟的元資料標準是中國機讀目錄格式。除此之外，其他元資料在中國大陸的應用與發展尚處於起步階段，應用的領域還不够廣泛，其中，比較突出和初顯成效的是數位圖書館中元資料的應用。數位圖書館為元資料框架的發展和實施提供了頗具挑戰性的應用領域，所有數位圖書館專案都必須解決元資料方面的問題，中國大陸正在開發中的數位圖書館專案自然亦無例外。

早在 1995 年，北京圖書館（現名為中國國家圖書館）就開始對發達國家研究數位圖書館的跟蹤，並先後在文化部和國家計委申報了若干有關專案進行研究，在 1999 年又建立了數位圖書館的實驗環境，在國家有關部門的支援下，進一步驗證了一些關鍵技術。

目前，在中國大陸，影響最大的數位圖書館有：

（一）中國試驗型數位式圖書館

中國試驗型數位式圖書館專案(China Pilot Digital Library Project, CPDLP)，是 1997 年度由國家計委批准立項的國家重點科技專案。專案實施期限為 1997 年 7 月至 2000 年 12 月。專案實施包括兩大部分：首先是研製一套初步成形的數位式圖書館的實現技術，它應具有與國際主流技術接軌的特點，同時，要適合在中國推廣；其次，是逐步建設一個規範化的分散式數位式資源庫，在資源組織和描述上，強調符合本專案總體技術組的統一要求。(註 27)該項計劃由北京圖書館、上海圖書館、深圳圖書館、中山圖書館、南京圖

書館、遼寧圖書館參與，模仿美國數位圖書館倡議計劃(Digital Library Initiative)，側重技術方案的實現，兼顧資源的數位化，並設立了一項軟課題研究。計劃完成後將在網上建立多館合作的數位圖書館網頁。(註 28)1998 年 4 月 12 日至 13 日，「中國國家試驗型數位式圖書館專案組」在北京召開技術組會議。根據會議對元資料方案所做出的結論，該數位圖書館專案最小元資料集合採用都柏林核心；最小元資料集合都柏林核心的實現採用 RDF 方式；元資料分成兩層，第一層為都柏林核心，第二層為 MARC 或 TEI 標題，在資源建設中要求按這兩層實現。技術組據此提出專案所需採用的規範的元資料元素集及其定義、元資料元素的限定規則以及元資料元素的具體表達方式。(註 29)

（二）中國數位圖書館工程

考慮到因經費不足，「中國試驗型數位式圖書館」專案進展得不盡如人意，1998 年 8 月 25 日，「中國數位圖書館工程」籌備領導小組成立，有關研究、論證工作正式展開。(註 30)2000 年 4 月 5 日，由文化部牽頭召集的中國數位圖書館工程聯席會議第一次會議在國家圖書館隆重召開。這次會議標誌著中國數位圖書館工程歷經幾年的籌備，已經正式啓動，即將由試驗階段進入有規劃、有組織的、科學有序的實質性操作階段。據悉，中國數位圖書館工程建設已經啓動，確立工程組織架構、制定「中國數位圖書館工程實施方案」、數位圖書館相關標準規範的研究與應用、規模型數位圖書館資源庫建設、探討解決數位圖書館建設中知識產權問題的方案、數位圖書館關鍵技術的研發、建設數位圖書館樣板工程等項工作。(註 31)在這個龐大的專案中，元資料方案的制訂是最基本和至關重要的部分，需要建立數位圖書館國家中心的元資料共用、檢索系統，建立一個元資



料資源中心。該中心使用並行資料庫技術和分散式電腦系統來支撐海量的元資料系統。

(三)清華大學建築數位圖書館

清華大學建築數位圖書館(THADL, Tsinghua University Architecture Digital Library)是由清華大學圖書館、電腦科學與技術系、清華大學建築學院三方精誠合作共同研製開發的。1999年10月,清華大學建築學院開始策劃構建清華大學建築數位圖書館。在經過一系列的可行性調研之後,2000年3月5日,清華大學圖書館、電腦科學與技術系、建築學院三方領導決定成立清華大學建築數位圖書館課題組,並將此專案立項,開始正式投入運作。日前,在清華大學數位圖書館站點上,已能夠提供中國營造社史資訊導航、學社資料成果資訊導航以及新營造學社資料方面的服務。清華大學數位圖書館的元資料基本上採取的是都柏林核心。(註 32)

此外,中國大陸有關方面正在著手於基於中文資源的元資料標準的制訂。廣東省中山圖書館制訂的《數位式中文全文文獻通用格式》(註 33)是這方面的嘗試和努力結果之一,而北京大學數位圖書館研究所在元資料標準的研究方面也展開了積極的工作。北京大學數位圖書館研究所、北京大學圖書館數位圖書館工程將「中文 metadata 標準研究」作為重點研究專案。該專案試圖以各類「中文文獻」(如具有鮮明的中文特徵的金石拓片和敦煌古卷、善本、輿圖等)為資源實體物件,研究中文文獻元資料標準,並在其基礎上制訂中文文獻元資料標準框架,建立若干中文文獻元資料示範資料庫。(註 34)

目前,中國大陸數位圖書館的針對物件大都偏重於古籍、珍貴史料。這自然與版權問題有一定的關聯。同時,更是因為一方面為了更好地保存此類文獻且其閱讀條件較為苛刻,使得此類文

獻的利用大受限制;而另一方面,此類文獻數量極少,可獲得性較低,需求者較難獲取。數位圖書館却能夠較好地解決這兩方面的問題。隨著數位圖書館在中國大陸的發展,數位圖書館的物件將逐漸豐富、擴大。

陸、結語

由上述中國大陸中文元資料的開發與利用之現況與發展,本人提出以下幾點思考與建議:

一、與國際相關標準相容

在制定中文元資料標準方面,於認清中文元資料處理方面的特殊性之同時,必須考慮到與國際相關標準相容的問題,從全球的資訊大環境著眼,以免形成的系統無法國際化,成為閉關自守的系統。

二、積極參與國際性的相關專案

通過一系列試驗專案和總結,國際上對於網路資訊資源有無編目的必要性以及如何進行編目業已形成定論,中國大陸的圖書館應儘早加入國際上的此類專案,同時,積極開展對中文網路資訊資源的組織,並進一步推動中文網路資訊資源的對外輸出。

三、採用都柏林核心作為資源描述格式是發展趨勢

從元資料在整個世界範圍的發展及其利用情況來看,將都柏林核心作為資源描述格式已是大的趨勢,都柏林核心的彈性、互操作性等日漸得到廣泛的認同,以至於逐漸取代其他元資料格式成為元資料格式中的主流。而圖書館界是否直接開發以都柏林核心為資源描述架構的系統,還需做長期的打算。至少在目前的情況下,首先仍要考慮到與現有系統的相容以及使用者的認識和習慣。



四、合作開發，避免重複

在以往的資料庫開發等許多工作方面，兩岸都有重複的現象發生。良性競爭固然有利於質量與效率的提高，但畢竟需要開發的專案太多，合作開發，互補長短，應更爲可取。只是臺灣已有「數位博物館專案計劃」等多項成功案例；而大陸在元資料開發與利用方面還剛剛起步，要與對

岸共用這方面的成果，更需儘快行動，爭取早日取得實際成效，以求齊頭並進。

元資料在中國大陸的開發和利用還剛剛開始，所謂「路漫漫兮，其修遠矣」，相關的從業者尙任重而道遠。

(收稿日期：2001年4月24日)

註釋：

- 註 1：NASA, Directory of Interchange Format Manual Version 1.0 · (Jul. 13 1988), NSSDC/WDC-A-R&S. pp.88-89 ·
- 註 2：Gisle Hannemyr, "The Internet Is Not Your Friend: Using the Dublin Core on the Internet,"
<<http://www.ddb.de/partner/dc7conference/results/dc7-hannemyr/sld001.htm>> (22 May 2000).
- 註 3：Renato Iannella, "Mostly Metadata: a Bit Smarter Technology,"
<<http://www.dstc.edu.au/RDU/reports/VALA1998>> ·
- 註 4：Lorcan Dempsey and Rachel Herry, "Metadata: a Current View of Practice and Issues," The Journal of Documentation 54:2 (Mar. 1998), p.149.
- 註 5：Arlene G. Taylor, The Organization of Information (Englewood, Colorado: Libraries Unlimited, Inc., 1999), p.246.
- 註 6：Amanda Xu, "Metadata Conversion and the Library OPAC," The Serials Librarian 33:1/2 (1998), p.180.
- 註 7：Priscilla Caplan, "You Call It Corn, We Call It Syntax-independent Metadata for Document-like Objects," The Public-Access Computer Systems Review 6:1 (1995), <<http://info.lib.uh.edu/pr/v6/n4/capl6n4.html>>.
- 註 8：International Federation of Library Associations, "Digital Libraries: Metadata Resources," 29 Sep. 1999,
<<http://www.ifla.org/II/metadata.htm>> (12 Mar. 2000) ·
- 註 9：周文駿主編，圖書館學情報學詞典（北京：書目文獻出版社，1991年），頁403。
- 註 10：Gregory Wool, "A Meditation on Metadata," The Serials Librarian 33:1/2(1998), pp.170-172.
- 註 11：W3C, "Hypertext Markup Language (HTML),"
<<http://www.w3.org/pub/WWW/MarkUp/MarkUp.html>> (13 Mar. 2000).
- 註 12：Gregory H. Leazer, "An Examination of Data Elements for Bibliographic Description: Toward a Conceptual Schema for the USMARC Formats," Library Resources & Technical Services 3(Apr. 1992), pp.189-208.
- 註 13：Anne J. Gilliland-Swetland, "Defining Metadata," Introduction to Metadata: Pathways to Digital Information, Ed. by Murtha Baca (U. S. A.: Getty Information Institute, 1998), pp.6-8.
- 註 14：Paul Miller, "Metadata for the Masses," <<http://www.ukoln.ac.uk/ariadne/issue5/metadata-masses>> ·
- 註 15：同註 13，p.2。
- 註 16：同註 4，p.157。
- 註 17：同註 3。
- 註 18：陳雪華、陳昭珍、與陳光華，「數位圖書館/博物館中詮釋資料之理論與實作」，圖書館學刊，13（民國 87 年 12 月），頁 41-42。



- 註 19 : Amit Sheth, and Wolfgang Klas, ed. Multimedia Data Management: Using Metadata to Integrate and Apply Digital Media (New York: McGraw-Hill Co., Inc.,1998), pp.5-10.
- 註 20 : S. A. Sutton and S. G. Oh., "GEM: Using Metadata to Enhance Internet Retrieval by K~12 Teachers," Bulletin of the American Society for Information Science 24:1 (Oct./Nov. 97), pp.21-24.
- 註 21 : M. Jost, "EULER-Ein EU-Projekt zur Integration Heterogener Informationsquellen," Bibliotheksdienst 32:3 (1998), pp.513-517.
- 註 22 : A. Ardo, and S. Lundberg, "A Regional Distributed WWW Search and Indexing Service --the DESIRE Way," Computer Networks and ISDN Systems 30:1/7 (Apr. 1998), pp.173-183.
- 註 23 : F. von Essen, Metadata Initiative of German Libraries (META-LIB) =Metadaten-Projekt Deutscher Bibliotheken : [a project]. 1997-1999.
- 註 24 : Michael Day, "Metadata: Mapping Between Metadata Formats,"
<<http://www.ukoln.ac.uk/metadata/interoperability/>> (2 Feb. 1999).
- 註 25 : "The Nordic Metadata Project," <<http://linnea.helsinki.fi/meta/index.htm>> .
- 註 26 : 林海青,「數位化圖書館的元資料體系」, 中國圖書館學報 2000 年第 4 期 (2000 年), 頁 59-69。
- 註 27 : Zhuang Jing,「中國試驗型數位式圖書館專案」, 1 Oct. 1998,
<<http://www.nlc.gov.cn/dlib/dle1.htm>> (15 May 2000).
- 註 28 : 劉焯,「數位圖書館概要」<<http://www.istis.sh.cn/istis/dlib/report/dlibintro.html>>(16 May 2000).
- 註 29 : 劉焯, 趙亮,「元資料方案實施意見」, 15 Jun. 1998,
<<http://www.istis.sh.cn/istis/dlib/report/metadata.html>> (25 Jan. 2000).
- 註 30 : 中國數位圖書館工程籌備領導小組辦公室,「中國數位圖書館工程籌備領導小組成立,有關研究、論證工作正式展開」,簡報第 1 期 (1998 年 8 月 26 日),
<<http://www.ccnt.gov.cn/digilib/dilijb1.htm>>(16 May 2000) .
- 註 31 : 國家圖書館,「中國數位圖書館工程建設聯席會議第一次會議召開,中國數位圖書館工程建設正式啓動」,
<<http://www.nlc.gov.cn/shuzi/index.htm>> (14 Apr. 2000).
- 註 32 : 清華大學建築數位圖書館專案組,「清華大學建築數位圖書館」,
<<http://166.111.68.208/digital/default.asp>> (7 Jan. 2001).
- 註 33 : 廣東省中山圖書館,「數位式中文全文文獻通用格式:徵求意見稿」,(2000 年 1 月 22 日)。
- 註 34 : 北京大學圖書館,「中文 Metadata 標準研究專案」,<<http://162.105.138.98/metadata/>>(4 Jan. 2001).

