

論 述

以資訊處理觀點 論Metadata之本質與意涵

余顯強

副教授兼圖書館館長
世新大學資訊傳播學系
E-mail: ysc@cc.shu.edu.tw

摘要

Metadata最普遍的解釋是「data about data」，但這一個解釋太過於簡單與籠統，對此一概念的傳遞不僅有限且容易被誤解。由於metadata在電子環境已運行多年，以致其內涵在不同資訊處理與應用領域具備不同的意義。因此，本文嘗試從資訊處理角度討論metadata之功能與作用、字根延伸意義、語法及結構模式三個不同應用的角度分析metadata的內涵，藉以歸納適合的定義，並透過現有國內中文譯名的文字字義，探討中譯的適切性，提供使用者對metadata此一字彙與內涵能有更詳細了解與應用。

關鍵詞：定義，資源描述，內容分析，字義分析

前 言

國科會於1998年6月成立「數位博物館專案計畫」，為了支援計畫的物件描述與詮釋，成立了「Metadata工作小組」負責規劃計畫有關metadata的事宜，有系統地發展國內的metadata相關標準(註1)。之後，metadata逐漸普及應用在國內各個資訊應用的領域，包括各典藏計畫、數位內容、政府的共通作業平台(註2)等領域，對知識產業、電子化政府、電子商務等資訊管理與應用造成了非常大的影響。不過metadata一詞不僅目前有許多不同的中文譯名，且許多文獻在解釋時，常有觀點的差異而有不同的定義，因此常常造成應用上的困擾(註3)。

依據字面，metadata是由meta與data兩字複合而成，因此metadata與metadata、meta-data代表相同的意義。Meta源自於希臘文μετα，表示其後(beyond)的意義(註4)。使用metadata這一個字彙，最早可追溯到1960年代，不過直到

2007/12/04投稿; 2007/12/31修訂; 2008/01/03接受

1980年代才大量在資料庫管理系統(Database Management System, DBMS)的文獻上出現(註5),作為資料字典或描述資料庫中資料的目錄(註6)。metadata這一個字彙通常是用來描述在「資料庫中資訊內容特性」的文件資訊(註7)。在DBMS領域,電腦內部包含了被描述的資訊與記述的資料兩個部分,因此,metadata主要便是在此一環境下管理與運作的資訊。

在圖書館尚未自動化之前,編目紀錄型態是非數位格式的書目卡片時,圖書館目錄是使用「書目資料」(bibliographic data)或「目錄資料」(cataloging data)來稱呼此一類資訊的型態。當這些圖書館目錄由紙本轉變到數位化機讀編目(Machine Readable Cataloging, MARC)格式時,圖書館界仍繼續沿用這些稱呼。但是,當編目人員開始要描述網路上的電子資源時,縱使使用相同的書目紀錄型態,但用語卻由MARC紀錄變成了metadata,編目人員所熟悉的英美編目規則(Anglo-American Cataloguing Rules, AACR)和MARC格式所規範的規則也跟著改變。這些轉變起因於廣大資訊機構的聚合效果,組織資源的方法從個別的圖書館界、電腦界,和資訊科學等領域,均交會在網際網路的應用環境之中,也因此「metadata」成為各個學科領域共通能接受的字彙。

圖書館界有許多人並不願全盤接受這一個字彙,認為它只不過是目錄或書目資料的另一個名稱(註8)。事實上在許多的應用,metadata與MARC並無不同,但採用這一個字彙則有一些理由,即metadata這一個字彙在電子化環境早已有其具備的內涵,而目錄這一類的字彙會則沒有。更重要的是,不同資訊機構的編目人員對metadata可能會有各自不同觀點,但metadata卻能提供在不同資訊機構間溝通的一個通用名詞,它允許資訊之間都可以相同且持續不斷地互通有無,也就是藉由metadata達成資訊交換的功能,包含不同metadata記錄之間資訊的對應(mapping)、跨系統之間資訊交換的剖析(parsing),以及紀錄內容的資料擷取(mining),這些對於自動化系統在跨越多種資料庫之間處理metadata資訊之存取、轉換,並提供給使用者確實有用的資訊,具有極大的意義與重要性。

Metadata最普遍的解釋是「data about data」,比較精確的說,這一個解釋太過於簡單與籠統,對此一概念的傳遞不僅有限且容易被誤解,因此並不適合做為定義。由於其在電子環境已運行多年,以致其內涵在不同資訊處理與應用領域具備不同的意義,因此本文嘗試從應用的角度來分析metadata的內涵,藉以歸納適合的定義,並透過現有國內中文譯名的字義,探討中譯的適切性,提供使用者對metadata此一字彙能有更詳細了解與的應用。

二、應用說明

為了提供metadata一個明確的使用功能說明,Sherry L. Vellucci提出了下列的解釋:Metadata是描述資源屬性的資料,以及彼此間關係的特性,使資源

能夠在電子環境中有效地被找尋、管理和利用(註9)。其強調的重點在於 metadata 是電子式的資料，因此傳統紙本的資料，例如圖書館的卡片目錄，縱然其具備描述其他資料的資料屬性，仍不屬於 metadata 的範疇。

但單就電子式的資料角度來解釋 metadata 仍有許多完全不同的面向，有將 metadata 表示為資料格式著錄的規範(註10)，或有將資料的標示語言歸屬於 metadata(註11)。由於此一類型 metadata 的用途在於規範文件的結構，因此這一類型的 metadata 完整名稱應稱之為「 metadata 格式」(format/scheme)。

參考開放檔案計畫組織(Open Archives Initiative, OAI)對其協定所傳輸之資料記錄(record)的解釋為「依據特定 metadata 格式所編製的 metadata」，此處的 metadata 可解釋為針對個別物件的描述資料記錄，其完整的名稱應稱之為「 metadata 紀錄」。因此，OAI 組織所制定之分享與獲取協定(Open Archives Initiative Protocol for Metadata Harvesting, OAI-PMH)，其傳輸的"metadata"，所指的便是 metadata 紀錄。此外，國內數位典藏機構的網站亦多使用 metadata 一詞作為其資料記錄的數量單位(註12)，所指的也是 metadata 紀錄。

針對上述的分析，可獲知 metadata 存在著系統內在與外在的形式，涵蓋了標示語言、資料格式、依據資料格式所著錄的資料記錄三個層面。因此，判斷 metadata 的意義時，必須考量資料的用途、內涵與產生的方式。若以 metadata 常見的「資料描述資料」的解釋，前後兩個「資料」在實務上所代表的意義來探討 metadata，方可明顯區隔出其在資訊處理上不同層面的意義。

以數位典藏為例，將資源在數位化描述的作業區分成：格式定義、著錄資料，及被描述的實體典藏品三個角色的分析。從編目規則之資料著錄格式的角度來看，編目所涵蓋的欄位範圍必須依據資料格式的定義，此時資料格式的定義規範(描述)了著錄資料的內涵，此種情況如圖1所示，格式定義便是 metadata；若從資料著錄的角度，編目所著錄的內容描述了實體典藏品的各種屬性，此種情況如圖2所示，著錄資料便是 metadata。

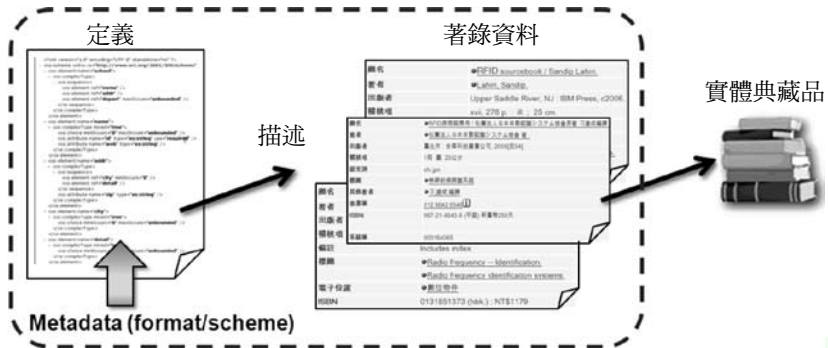


圖1 格式定義規範著錄資料的內涵

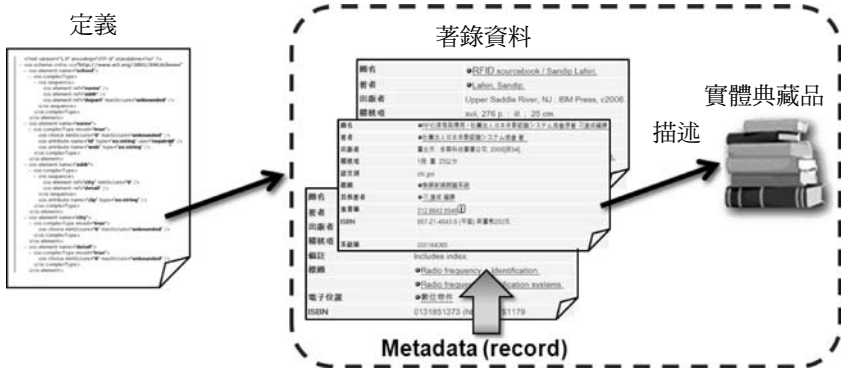


圖2 著錄資料描述實體典藏品

經由上述討論，可歸納出如圖3所示，資訊處理不同層級中metadata意義的差異。依據metadata作為資料結構與資料記錄的兩種類型，並參考metadata在資訊處理的應用，以下分別就metadata之功能與作用、字根延伸意義、語法及結構模式三個角度探討metadata所代表的意義。

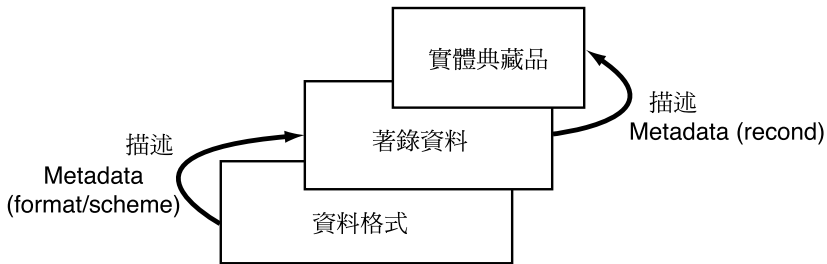


圖3 資訊處理各種層級的metadata

(一)由功能與作用探討metadata 意義

從商業應用角度來看，依據metadata的使用，可分為提供使用者使用所需的業務資料 (business metadata)，提供編目人員和資料庫管理或其它軟體程式處理的技術資料 (technical metadata) (註13)，這些資料形態應用上的差異也就是造成metadata意義混淆的主要原因之一。依據metadata應用於電子資源資料著錄的角度，配合「資料描述資料」的解釋來探討metadata，資訊系統處理應用metadata處理資料的功能與作用包括：結構定義、資料描述、文內詮釋，與屬性界定四種。

1. 結構定義

XML支援語言中立 (language neutral) 的定義和平台中立 (platform neutral)，並能提供定義在Web環境上結構化文件交換的資料格式，使得XML成為多數資訊系統資料處理的主要格式標準 (註14)。XML文件有兩種形式，一種稱為Well-formed XML，另一種稱為Valid XML，兩者最大差別在於Well-formed

XML表示符合文法規範的XML文件，而Valid XML則不僅符合文法規範，還包含文件的結構定義。透過結構定義能夠提供應用系統或剖析器(Parser)驗證XML文件結構的有效性，或提供編輯工具能夠遵循定義的架構來產生XML文件。XML採用結構定義的語法包括文件型別定義(Document Type Definition, DTD)與XML Schema兩種。藉由DTD或XML Schema規範了XML文件的內容型態與結構，即如先前圖1所示定義與著錄資料間的關係，可知DTD與XML Schema即用於描述著錄資料結構的metadata(format/schema)。

2. 資料描述

資料描述的metadata，即資訊系統針對典藏品或特定物件，就是原生資料(raw data)所記錄的相關描述，作為該原生資料在數位資訊的代用性記錄(surrogate record)(註15)，也就是metadata紀錄，例如圖書館書目資料，描述的實際對象為圖書、期刊等典藏品。現今在各類型文獻提及的metadata多指用於資料描述的metadata紀錄，其最常見是透過人工方式編目著錄而產生的資料記錄，並以XML文件格式儲存於電腦環境，資訊系統再依據使用者介面或人機介面設計，而有不同的內容呈現方式。因此，資訊系統便可依據資料描述的metadata內容及系統使用的範圍提供指引(location)、發現(discovery)、文件記錄(documentation)、評鑑(evaluation)、判斷選擇(selection)的功能(註16)。

3. 文內詮釋：

如圖4所示，電腦資訊技術無法很精確地自動區辨文獻中相關人、事、時、地、物的關係，以致無法將許多文獻之間各事件的關聯性串連起來，因此透過標示(markup)方式可將內文各資料的屬性予以標記，以提供系統自動化處理。使用屬性標示描述該段文字的性質，達到詮釋文獻內文之人、事、時、地、物的資訊，因描述的資料是著錄資料，也就是metadata紀錄某一欄位的內容，因此內文的屬性標示應屬於詮釋資料的metadata。此類型的metadata無法歸類於metadata格式或metadata紀錄，僅作為內容詮釋的標示。

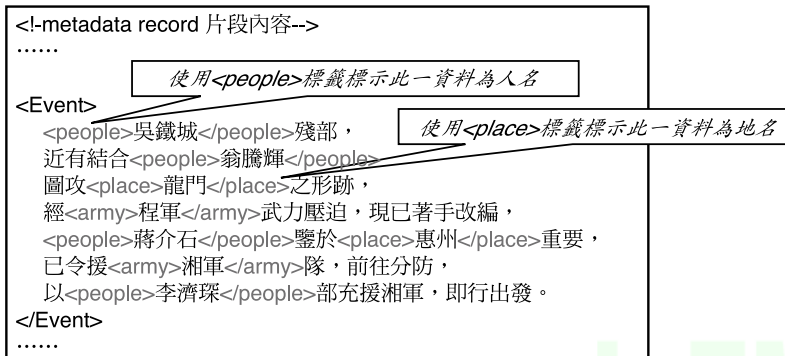


圖4 使用標籤詮釋內容的屬性

4. 元素標示

資料描述的 metadata 欄位無論採用 XML 文件格式或其他任何標示方式，欄位名稱即用來標明內容的型態，例如中國機讀編目格式 (Chinese MARC) 欄號 200 的欄位名稱為「題名及作者敘述項」，描述了該欄位的內容記載了原生資料 (典藏品) 的題名與作者等相關資訊。而 XML 的標籤或屬性亦標示其元素 (element) (註 17) 的內容型態，因此也被視為 metadata。如圖 5 所示 XML 文件內容，<colorimeter_reading> 元素的內容，包含 <device>、<patch>、<RGB> 三個子元素，其中 <RGB> 元素的屬性 <resolution> 內容值 "8"，此一數值對於該屬性本身並沒有直接意義，而是描述處理 <colorimeter_reading> 元素的指示，因此該屬性屬於一個 metadata。若將此一 <colorimeter_reading> 元素解構成一個資訊內容時，<RGB> 元素的內容的資訊既不完整也不明確，因該內容是提供系統處理畫素的色值，因此這些元素亦屬於 metadata (註 18)。此類型的 metadata 亦無法歸類於 metadata 格式或 metadata 紀錄，其特性如同圖書館目錄的款目 (entry)，提供使用者欄位意義的辨識，或作為系統處理與建置索引的依據。

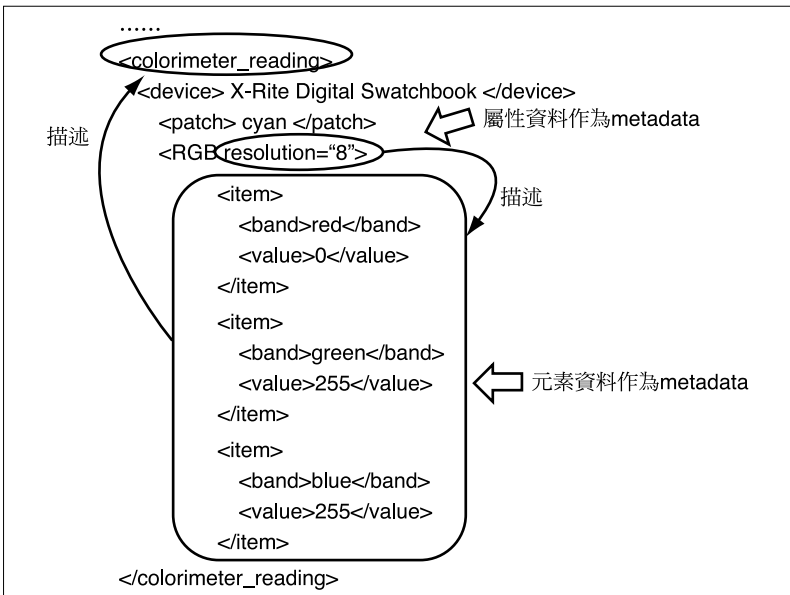
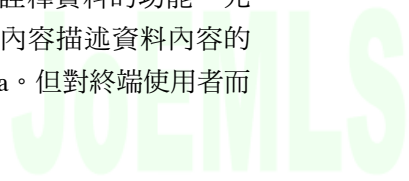


圖 5 元素與屬性作為 metadata

資料整理來源：Akmal B. Chaudhri, Awais Rashid & Roberto Zicari, XML data management: native XML and XML-enabled database systems (Boston: Addison-Wesley, 2003), pp.10-12.

由上述分析可知，文內詮釋的標示方式具備很明確詮釋資料的功能，元素使用標籤名稱描述資料內容的型態或使用標籤之屬性內容描述資料內容的型態，此時標籤名稱與該標籤的屬性均可被視為 metadata。但對終端使用者而



言，通常使用者在檢索或瀏覽資訊時，系統並不會呈現其詮釋標示部分，而會透過標籤名稱描述呈現的資料性質。此外，結構定義與元素標示屬於後端系統運作的層面，終端使用者接觸到的均是資訊系統處理後的人機介面，因此容易被忽略 metadata 此一部分的功能，而造成認知上將 metadata 視為僅用來描述原生資料或呈現資料的誤解。

(二)由字根延伸探討 metadata 意義

依據 Meta 字根意義來探討其各種延伸應用所代表的意義，來反思 metadata 可能具備的意義，及中文譯名內涵是否也能擴及這些 metadata 延伸的應用。Metadata 與 meta 字根在資料著錄的延伸應用包括 Meta-language、Super-metadata、Meta-metadata 三種，分析如下：

1. Meta-language

如圖6所示，方形表示延伸之標示語言，橢圓表示該標示語言是 Meta-language，也就是可再延伸出其他標示語言的語言。許多早先開發的應用程式是使用 SGML 作為資料處理的標示語言，SGML 延伸出 TEI、DocBook、Edgar、HTML、XML 等，而 XML 不僅是現今許多文書處理、資料交換的標準，也延伸出許多標示語言。Meta-language 具備衍生新標示語言的能力，因此 SGML 與 XML 均是 meta-language。若將標示語言視為軟體物件，則 meta-language 便與物件導向軟體工程「物件繼承」的行為完全相同，繼承出的子物件具備父物件的基本型態，亦可再擴充出新的型態，且物件之間可完全獨立運作。

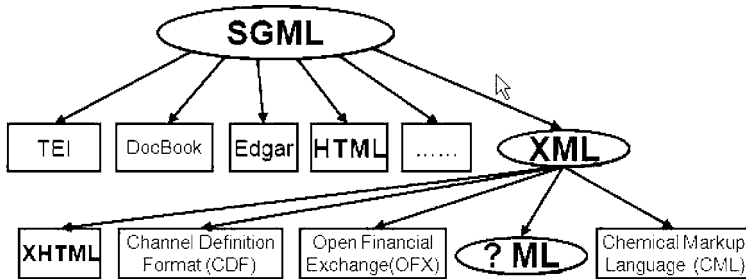


圖6 標示語言關係簡圖

2. Super-metadata

如同軟體繼承的概念一樣，當一個 metadata，繼承出下一代的 metadata，則原始的 metadata 便是下一代 metadata 的 super-metadata (註 19)。無論 metadata 的標式語言、metadata 格式與 metadata 紀錄均可使用這一種應用。例如 meta-language 可再延伸出其他標示語言的語言，也就是說，meta-language 是其衍生標示語言的 super-metadata，例如 SGML 即 HTML 與 XML 的 super-metadata。而

都柏林核心集 (Dublin core element set, DC) 亦是一種典型的 super-metadata，依據 DC 制定出新的 metadata 格式主要是提供相關 metadata 格式之間 metadata 記錄互通與轉換的可行性，減少資料在交換中退化 (degradation) 所造成的資料喪失。此外，不僅應用在 metadata 格式的標示語言、資料著錄所產生的 metadata 紀錄，亦可結合物件導向繼承的概念，應用 super-metadata 的繼承功能，如圖 7 所示，國際圖書館協會聯盟 (International Federation of Library Associations, IFLA) 的「書目記錄功能需求」(Functional Requirements for Bibliographic Records, FRBR) 將目錄區分出 Work, Expression, Manifestation, Item 四個部分，提供子類別繼承上一層類別，使子類別能夠具備上一類別所有的內容，即屬於 super-metadata 的一種形式。

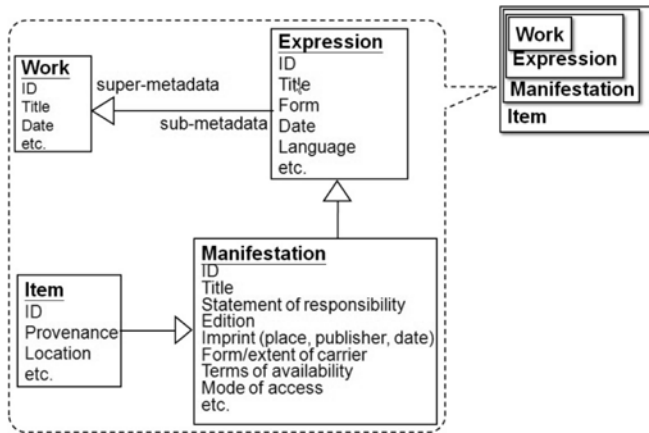


圖7 FRBR的super-metadata架構圖

3. Meta-metadata

Meta-metadata 用於描述 metadata 紀錄之實際個體片段的資料 (註 20)。透過群組方式將相關性質的元素聚合在一起，並以一名稱描述該群組，提供使用的便利性，或提供系統使用 metadata 格式應用上的彈性。例如機讀編目格式的「段」(fields)，不僅可依描述不同資料型態 (圖書、期刊、善本……等) 選擇編目段的種類，還可配合編目規則八大著錄項目的順序提供編目作業的進行。

由上述延伸 meta 字根且應用於 metadata 資訊處理專業術語的作用，可獲知使用 meta 字根的應用大多具備物件導向軟體工程「繼承」與「封裝」的意義，具有中文「元」的本意，且與哲學形上之「後設」一詞相當，表達了根本原理。但就作為「資料描述」的 metadata 紀錄而言，卻缺乏「繼承」與「封裝」的應用。metadata 以「元資料」或「後設資料」雖能夠很有力地表達 metadata 的本質，但 metadata 在實際的應用卻並沒有發揮在眾多學術文獻上所表述的層面，使得在「元」或「後設」字面意義的涵蓋上，似乎仍有過於之處。

(三)由語法及結構模式探討metadata 意義

應用 metadata 作為描述原生資料的語法及結構，包括 HTML 的 <meta> 標籤、XML 文件結構有效性的 DTD 和 XML Schema、關聯式資料庫的系統綱要 (schema) 三種不同應用類型，針對三種結構的應用分述如下：

1. <meta> 標籤

基於 HTML 不具備結構的特性，使得自動化系統無法詮釋各網頁的內容。因此，全球資訊網協會 (World Wide Web consortium, W3C) 於 1999 年 12 月 24 日公佈的 HTML 4.01 版增加了 <meta> 標籤 (註 21)。**<meta>** 標籤屬於非顯示的 HTML 標籤，用於提供網頁的擁有者或作者設定網頁被索引的控制程度 (註 22)。**<meta>** 標籤包含 http-equiv 與 name 兩種類型的屬性。http-equiv 用來描述 HTML 網頁文件的相關性質，提供前端瀏覽器在下載該 HTML 網頁文件前判斷文件的型態，以便控制瀏覽器處理該網頁的行為 (註 23)。Name 屬性對應於 content 屬性的內容，採用 DC 元素描述網頁的內涵，以便於搜尋引擎機器人尋找、判斷與分類。因此，HTML 的 <meta> 標籤亦被視為一 metadata 的簡單型式 (註 24)。

2. DTD 和 XML Schema

XML 最初公佈時採用 DTD 作為定義和規範 XML 文件架構的一組規則，以便能提供應用系統或剖析器驗證 XML 文件結構的有效性，或提供編輯工具 (authoring tool) 能夠遵循定義的架構來產生 XML 文件。不過 DTD 由於是沿用 SGML 的型別定義，本質上有許多限制，因此 W3C 於 2001 年 5 月 2 日正式提出以 XML 作為語法的 XML Schema 標準 (註 25)，使 DTD 與 XML Schema 均可作為描述 XML 文件結構的標準語法。

系統藉由 DTD 或 XML Schema 文件規範或檢驗 XML 文件是否符合 metadata 結構的有效性 (validation)，也就是說資訊系統的運作，是透過 DTD 或 XML Schema 的描述來控制 XML 文件的結構，而 XML 文件又是現今主要 metadata 紀錄的標示語法，因此，DTD 或 XML Schema 即是描述 metadata 紀錄的 metadata 格式語法，提供元素結構定義與文件資料形態呈現的兩種關鍵性資料 (註 26)。由此可知，metadata 紀錄與格式均有一標準標示語法來描述，使得資訊系統可有一致的剖析程序。

3. 關聯式資料庫綱要

資料庫中的 metadata，其主要功能即是規範系統中各物件的結構，這些物件包括使用者、資料庫、表格、索引……等，作為系統內部的資料字典，提供資料庫內容的管理、轉換、運作等各方面的資訊，亦包括存取方法 (註 27)、索引策略、資料分佈、異質資料管理等 (註 28)，也就是現今於關聯式資料庫系統使用的綱要 (Schema)，提供了資料庫內部結構的主要描述。

三、定 義

由上述分析可歸納出 metadata 核心的意義即在於「資料描述資料」，配合其應用的差異，「資料」之間所代表的意義即有不同，這也就是現今 metadata 意義多元的主要因素。由 metadata 應用在電子環境承載各種不同資訊的本質，從它在系統的角色、對資料提供者及使用者的應用三個層面，都各有合乎事實的意義。因此，在資訊處理上可總結出下列三個不同層面的定義：

1. 就系統角度，metadata 是規範著錄內容的格式，包括欄位名稱、順序、資料型態等。因此完整的名稱應稱為 metadata 格式。
2. 就資料提供者角度，metadata 是為達到原生資料（典藏品）的數位資訊應用，包括描述、管理、保存、利用等目的，而著錄的所有描述內容。因此完整的名稱應稱為 metadata 紀錄。
3. 就使用者角度而言，metadata 即是系統呈現任何有關原生資料（典藏品）的資訊內容。

針對這些定義來檢視現有資訊系統的應用，在系統、資料提供者與使用者各層面均能表述 metadata 表示資源與屬性的關聯性，可符合大部分 metadata 的應用方案。但反映至目前大部分實務的應用，可發現現有 metadata 的實務應用比較著重於發現與檢索資源，而不是著重在以詳細的描述行為與管理，這種行為似乎只將 metadata 當成「目錄」的替代品，並未完全發揮「物件」的完整功能。

四、中譯探討

國內對於 metadata 的中譯名稱一直沒有統一，常見的名稱包括元資料、詮釋資料、後設資料等。要恰當地理解和運用 metadata 中譯名詞的意義，必須對其中譯詞彙的來源追溯與脈絡分析，尤其在古籍文獻上的語義描述。因此，以下將就元資料、詮釋資料與後設資料三者之間探討其意義，以供分析判斷其適用性：

(一)元資料

「元」字於古籍中出現的意義如表1所示，歸納其主要為「最大」、「開始」之意。因此，可將元資料視為資料之根本，如此也可呼應易經之「大哉乾元」，表達為事物之根本。

表1 「元」字之古籍釋義

古籍出處	釋	義
爾雅	元，始也	
廣韻	長也，大也	
前漢書	基事之本也	
公羊	元年者何？君之始年也	

資料來源：張玉書，康熙字典，台北市：商務，1968。頁127。

(二) 詮釋資料：

「詮釋」一詞若分開解釋，「詮」字，首出「淮南子」之「詮言訓」一篇。漢朝高誘解釋：詮，就也；萬物之指，以言其微，事之所謂，道之所依也(註29)。康熙字典紀錄古籍出現的意義如表2所示，歸納其主要為「具體」、「解說」之意。而「釋」字為古籍文獻的常見字，本身就有語言解釋的意思。於古籍中出現的意義如表3所示，主要為「取捨」、「釋去」之意。因此，可知詮釋一詞，含有具體的闡述，並分別內容、取捨內容之意。與「詮釋」相近的詞語，還有「詮解」、「詮言」、「詮說」、「詮證」、「詮訓」等，在古籍中往往被交互使用，彼此間區別並不大，可見「詮」字在詞組方面，有著相當的靈活性，可以和字義相近的字彙搭配，把詮釋的意義融合其他學科範疇的意涵，構成新的語彙。

表2 「詮」字之古籍釋義

古籍出處	釋	義
說文	具也	
玉篇	治亂之體也	
廣韻	平也	
類篇	擇言也，又解喻也	
晉書	具說事理也	

資料來源：張玉書，康熙字典，台北市：商務，1968。頁1291。

表3 「釋」字之古籍釋義

古籍出處	釋	義
說文	解也，取其分別物也	
廣韻	捨也	
前漢書	消也，散也	
爾雅	服也	

資料來源：張玉書，康熙字典，台北市：商務，1968。頁1441。

(三) 後設資料

從語言及認知過程，人類使用語言描繪一事物，或敘述一個事件，常會受制於個人主觀的認知。因此，若從文學小說角度來解釋「後設」一詞，其有「一句話的出現，會使讀者受到前後章句的影響」的認知發生，這也就是尼采主張任何一項事實在認定之前，就滲入主觀認知的意思(註30)。若從應用角度來看，把問題推展到窮極究竟的地步，就會遇到該學科本身所不能解答的問題，這時已經由該學科的內部問題轉到哲學問題之上了。所以，在這種構想之下，哲學是一種「殿後」之學，也就「後設」之學，正如哲學 Metaphysics 中譯為形上學，physics 表示物理學，因此 Metaphysics 也就是在物理學之後的學問，也就是根本原理的學問。比如探索科學所不能解答的基礎問題之學，是為「後設科

學」；探索數學所不能解答的基礎問題之學，是為「後設數學」（註31）。

上述三者中譯的名稱，均能達成使用者對metadata意會的目標，但如本文對metadata不同應用層面的定義而言，單一中譯名稱無法涵蓋所有metadata代表的意義，也就是對於metadata這一個名詞而言，並無法使用單一的中譯名稱滿足所有metadata所具備的意義。正如與原文意義完全無關的「原子筆」(ball-point pen)，或直接採用音譯的「雷射」(laser)，縱使中文字面意義與原文有所差異，但透過實用的普及或平易的解釋，現代人都已能了解這些名詞所代表的意義。因此，理解、釐清這些名詞背後所代表的意義應比制定望文生義的中譯名稱更有其必要性。

五、結 論

名詞終究只是方便稱呼某一概念或事物而已，找出貼切適意的中譯名詞固然重要，但如何理解、釐清這些名詞背後所代表的意義才是更為重要的事。本文從系統、資料提供者與使用者三個不同角度解釋metadata的本質，可知Metadata在目前大部分實務的應用似乎只是將metadata當成「目錄」的替代品，並未完全發揮「物件」的完整功能。物件導向的功能，包含了內部的屬性（也就是資料），而屬性內容亦可能又是另一筆metadata紀錄，如同RDF (Resource Description Framework)的結構，資源的屬性可指向另一資源的原理一般；物件亦包含內部的方法（程序、演算邏輯），作為資訊系統剖析某一元素（element）內容時的運作原則，以避免不同資訊系統剖析內容上的差異，而造成資訊失真的狀況。因此，將物件導向的特性融入metadata格式的規範內，應可作為metadata進一階段的設計考量。

註 釋

註1 陳亞寧、陳淑君，「中文Metadata規範指引第一版：Metadata初探」，文獻處理實驗室，http://www.sinica.edu.tw/~cdp/project/04/6_1.htm（檢索於2002年2月19日）。

註2 財團法人資訊工業策進會，電子化政府共通作業平台規範，第一版（台北市：行政院研究發展考核委員會，2002），<http://www.gsp.gov.tw/downloadfile/guidebook/RFP0073.pdf>（檢索於2007年6月4日）。

註3 鑒於Metadata一詞目前有許多不同的中文譯名，造成使用上的困擾。中研院資訊科學研究所、計算中心及數位典藏國家型科技計畫「技術研發計畫」於2007年7月2日召開「界定Metadata中文名詞論壇」（論壇網址：<http://www.sinica.edu.tw/~metadata/bibliography/achievement/conference/96metadata/96metadata%20conference.htm>），目的即是針對Metadata的中文名詞與意涵進行多角度的探討。

註4 David Haynes, *Metadata for Information Management and Retrieval* (London: Facet, 2004), 6.

註5 Sherry L. Vellucci, "Metadata," *Annual Review of Information Science and Tech-*

nology 33 (1998): 187-222.

註6 Chris J. Date, *An Introduction to Database Systems*, 2nd ed. (Reading, MA: Addison-Wesley, 1975), 39.

註7 John T. Phillips, "Metadata: Information about Electronic Records," *ARMA Records Management Quarterly* 29, no. 4 (1995): 52-57.

註8 Mary Lynette Larsgaard, "Cataloging Planetospatial Data in Digital Form: Old Wine, New Bottles- New Wine, Old Bottles," in *Geographic Information Systems and Libraries: Patrons, Maps, and Spatial Information*. Edited by Linda C. Smith and Myke Gluck. (Urbana: University of Illinois, Graduate School of Library and Information Science, 1996), 17-30.

註9 Sherry L. Vellucci, "Metadata and Authority Control," *Library Resources & Technical Services* 44, no. 1 (2000): 33-43.

註10 范紀文、何建明、李德財，「從典藏資料交換角度探討Metadata設計與標準化問題」，於新世紀數位圖書館與數位博物館趨勢研討會（新竹市：交通大學，2001），IX-1~IX-12。

註11 Lynne C. Howarth, "Creating a Metadata-enabled Framework for Resource Discovery in Knowledge Bases," http://www.cais-acsi.ca/proceedings/2000/howarth_2000.pdf (accessed July 4, 2007).

註12 參考數位典藏國家型科技計畫聯合目錄建置的組織與歷程（網址http://catalog.ndap.org.tw/dacs5/System/Introduction/Introduction_1.jsp），內文所提計畫年度成果所編製的資料筆數，均以metadata作為數量的單位。

註13 Thomas Vetterli, Anca Vaduvaz, and Martin Staudty, "Metadata Standards for Data Warehousing: Open Information Model vs. Common Warehouse Metamodel," *ACM SIGMOD Record* 29, no. 3 (2000): 68-75.

註14 余顯強，「機讀編目格式與延伸標示語言互轉之定義研究」，*圖書與資訊學刊* 53期（2005）：18-30。

註15 Arlene G. Taylor, *The Organization of Information* (Englewood: Libraries Unlimited, 1999), 78.

註16 Lorcan Dempsey et al., *Specification for Resource Description Methods*, version 1.0 (May 1997), <http://www.ukoln.ac.uk/metadata/desire/overview/overview.pdf> (accessed May 11, 2001).

註17 在XML標示語言的規範中(<http://www.w3.org/TR/2006/REC-xml-20060816/#dt-element>)，元素(element)是指起始標籤與結束標籤之間所包含的內容，每一元素具備一個可以用來識別的名稱，有時可稱之為通用標識(generic identifier)。

註18 Akmal B. Chaudhri, Awais Rashid, and Roberto Zicari, *XML Data Management: Native XML and XML-enabled Database Systems* (Boston: Addison-Wesley, 2003), 10-12.

註19 Chilvers Alison and Feather John, "The Management of Digital Data: A Metadata Approach," *The Electronic Library* 16, no. 6 (1998): 365-372.

註20 Kal Ahmed et al., *Professional XML Meta Data* (UK: Wrox Press, 2001), 108-110.

註21 W3C Recommendation, "HTML 4.01 Specification," <http://www.w3.org/TR/html401/> (accessed May 29, 2001).

註22 Dariush Alimohammadi, "Meta-tag: A Means to Control the Process of Web Index-

ing,” *Online Information Review* 27, no.4 (2003): 238-242.

註23 Jennifer Niederst, *Web Design in a Nutshell: A Desktop Quick Reference* (Sebastopol: O’Reilly, 1999), 98.

註24 Deborah Christensen, “Golden Retrievers,” *School Library Journal* 45, no.11 (1999), <http://www.schoollibraryjournal.com/article/CA153030.html> (accessed May 29, 2001).

註25 W3C XML Schema, “XML Schema,” <http://www.w3.org/XML/Schema> (accessed May 29, 2001).

註26 Kurt Cagle, *XML Developer’s Handbook* (San Francisco: Sybex Inc, 2000), 272.

註27 物件導向所稱之「方法」(method)，是指某一物件內具備操作(operation)的實作，可以供其他物件呼叫執行的程序或運算邏輯。資料庫系統內的限制(constraint)是定義主鍵(PRIMARY KEY)、非虛值(NOT NULL)、唯一性(UNIQUE)、外來鍵(FOREIGN KEY)、或欄位內容檢查(CHECK)的條件約束，用來強制資料完整性的方法。

註28 Bhavani M. Thuraisingham, *XML Databases and the Semantic Web* (New York: CRC Press, 2002), 110.

註29 高誘，淮南子注釋(台北市：華聯，1968)，235。

註30 蔡源煌，從浪漫主義到後現代主義(台北市：雅典，1992)，189-193。

註31 何秀煌，哲學智慧的尋求(台北市：東大，1983)，149-150。

BRIEF COMMUNICATION

From Information Processing to the Essence of Metadata

Shien-chiang Yu

Associate Professor and Director of the Library
Department of Information and Communications, Shih-Hsin University
Taipei, Taiwan
E-mail: ysc@cc.shu.edu.tw

Abstract

The most common definition for metadata is “data about data”. However, this expression is rather simple and ambiguous. It is also insufficient and easy to be misunderstood. Metadata has already been used widely in the past years in the electronic environment, thus researchers and practitioners have come up with various meanings for metadata on different information applications. This paper analyzes three perspectives of metadata from the domain of information processing, including function, literal, syntax and structure. The paper, therefore, concludes with appropriate definitions for each of the three views of metadata. In addition, this paper discusses the popular Chinese translations of metadata and their appropriateness. It is believed that clear definitions for metadata in different scenarios will eventually help users to understand the term and its applications better.

Keywords: *Metadata; Definition; Resource description; Content analysis; Denotation analysis*

SUMMARY

Metadata is popularized and applied to various information science related fields in recent years, including digital archives, digital content, e-government common platform,¹ and other related domains. It has made significant impact on such areas as knowledge industry, electronic government, and e-commerce. But in numerous documents, there are usually different views with regards to definitions of metadata in the literature, therefore, misuses in applications are found very often.

Based on the literal meaning, the word metadata is composed of two words—“meta” and “data”, therefore, “metadata”, “meta data”, and “meta-data” all represent the same meaning. “Meta” stems from the Greek word, “μετα”, which means “beyond”.² The use of the term “metadata” can be traced back to the 1960s, but it began to appear more frequently in the literature on database management systems (DBMS) in the 1980s.³ The term metadata was used to describe

the information that documents the characteristics of information contained within databases.⁴ In the field of DBMS, there are two kinds of data — descriptive information and schema records. Metadata is mainly the information used for management and operation purpose in the database system environment. Before libraries implemented automation systems, the type of catalogue was bibliographic card of non-digital format. Libraries use the term “bibliographic data” or “cataloging data” to name this type of information. As paper-based library catalogs changed to the Machine Readable Cataloging (MARC) format, libraries continue with the naming. But as cataloguers began to describe the electronic resources on the web, even though the same type of bibliographic record is used, the term librarians used to name bibliographic data become metadata, instead of MARC. This change came from the cooperation of various information institutions in the digital age worldwide. The methods to organize digital resources have been discussed and proposed by fields such as library science, computer science, and information science, which all have been cooperating in the web environment. As a result, “metadata” becomes the term that all communities of professionals can accept and apply to each of their unique disciplines.

The most common definition for metadata is “data about data”. But this definition is rather simple and ambiguous. Such an expression of concept is not only insufficient but easy to be misunderstood. Metadata has already been used for many years in the electronic environment, thus, has led to a number of different meanings in different information applications. From the domain of information processing, this paper aims to analyze the meanings of metadata from three perspectives, which are function, literal, syntax and structure, in order to come up with an appropriate definition for metadata. In addition, this paper examines the popular Chinese translations of metadata and discusses their appropriateness. The author believes a clear definition for metadata can eventually promote users to understand its essence and applications.

Analysis

In order to offer metadata with a clear function-wise description, Vellucci⁵ proposed the following expanded definition, which incorporates statements of functionality and environment: “Metadata are data that describe the attributes of a resource; characterize its relationships; support its discovery, management, and effective use; and exist in an electronic environment.” The main point of this definition emphasizes that metadata are focused solely on electronic resources, i.e., metadata is data of certain electronic resources. Therefore, paper-based card catalogues of the library, even provided with attributes for collection description,

is still not metadata.

There are a number of drastically different ways to explain metadata, examining the term from the electronic resources perspective. One could say metadata represents the standard of records in the form of a catalogue;⁶ one could also argue that metadata is a markup language.⁷ Because the purpose of this type of metadata is to standardize the structure of a document, the proper name for metadata of this kind should be “metadata format” or “metadata scheme”.

Besides, according to the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), which was proposed by Open Archives Initiative, a data record is metadata expressed in a single format. In addition, various institutions of digital archives also use “metadata” as a quantity unit for their data records. In these cases, the appropriate name for metadata of this kind should be “metadata record”.

Consequently, the essence of metadata can be seen with three levels including markup language, data format, and the data which is recorded by the data format. As we attempt to define the meaning of metadata, the use, the intension, and the ways to produce the data record must all be considered. Taking digital archive as an example, metadata has three meanings according to different manipulation of information: definitions of format/schema, record data, and the descriptions of physical archives. From cataloguing perspective, definitions of format/schema must be included in each cataloguing field, which describes the essence of record data. In this case, the definitions of format/schema are metadata. On the other hand, when various attributes of the physical archive are described in a catalogue or finding aid, the record data is metadata.

Definition

According to the above analysis, it is suggested that “data about data” is the core meaning of metadata; however, with different use and applications, meanings of metadata are different, which is the main reason why the two data represent two different meanings. Examining the essence of metadata in the electronic environment from the perspectives of systems, data providers, and users, each definition has its unique meaning based on different applications. This paper thus summarizes the definitions from the three perspectives as follows:

1. From the system perspective, metadata is the format that represents the record standard including field name, orders, data type, etc. So, the appropriate name should be “metadata format”.

2. From the perspective of data provider, metadata is used as a digital information application to describe all kinds of information, including description,

management, preservation, utilization for the raw data (collection). So, the appropriate name should be “metadata record”.

3. From the user’s perspective, metadata represent any information about the raw data (collection) or any labeling of the present information.

Examining current information systems from the aspects of systems, data providers, and users, it is found that the above definitions of metadata can express most of the relationships between resources and attributes clearly.

Conclusion

A name, in a sense, is just a convenient way to call a certain concept or object. How to understand and distinguish the meanings of the name should be more important than finding the proper translations. This paper examines the essence of metadata from three different perspectives, i.e. systems, data providers, and users; and finds that most of the current practices view metadata as a substitute of categories, instead of an object. Object-oriented functions include attributes (the internal data of one object), and attributes could be another forms of metadata record. This concept is illustrated in Resource Description Framework (RDF), with which a structure points to the attributes of a resource to another resource. Besides, object also includes the method (the internal procedure, algorithm of one object) as the operation principle for information system to parse elements of content, in order to prevent different information system from parsing content differently and inaccurately. Therefore, integrating the characteristics of object-oriented functions with the framework of metadata format should be the next stage for application scheme.

Notes

¹ 財團法人資訊工業策進會，電子化政府共通作業平台規範，第一版(台北市：行政院研究發展考核委員會，2002)，<http://www.gsp.gov.tw/downloadfile/guidebook/RFP0073.pdf> (檢索於2007年6月4日)。

² David Haynes, *Metadata for Information Management and Retrieval* (London: Facet, 2004), 6.

³ Sherry L. Vellucci, “Metadata,” *Annual Review of Information Science and Technology* 33 (1998): 187-222.

⁴ John T. Phillips, “Metadata: Information about Electronic Records,” *ARMA Records Management Quarterly* 29, no. 4 (1995): 52-57.

⁵ Sherry L. Vellucci, “Metadata and Authority Control,” *Library Resources & Technical Services* 44, no. 1 (2000): 33-43.

⁶ 范紀文、何建明、李德財，「從典藏資料交換角度探討 Metadata 設計與標準化問題」，於新世紀數位圖書館與數位博物館趨勢研討會(新竹市：交通大學，2001)，IX-1~IX-12。

⁷ Lynne C. Howarth, “Creating a Metadata-enabled Framework for Resource Discovery in Knowledge Bases,” http://www.cais-acsi.ca/proceedings/2000/howarth_2000.pdf (accessed July 4, 2007).