

元資料實驗系統和都柏林核心集的發展趨勢

吳政叡（輔仁大學圖書資訊系專任副教授）

摘要

本文主要在介紹都柏林核心集從創始(1995年3月第一次研討會)到(1997年3月)第四次研討會的發展沿革，以及作者所發展的元資料實驗系統（MES，網址：<http://blue.lins.fju.edu.tw/~wu>或<http://140.136.85.194/>）。都柏林核心集的創立，是為了改正目前WWW上檢索引擎無法有效過濾資料的缺失。透過對每一次研討會成果的詳細介紹和分析，勾勒出整個都柏林核心集的發展歷程，和WWW上資源描述的發展趨勢。最後本文詳細介紹了元資料實驗系統的設計和功能，使讀者能輕易來使用MES，而透過實地的操作，讀者對元資料和都柏林核心集將有更深刻的體認。

關鍵詞：元資料 元資料實驗系統 MES 都柏林核心集 電子圖書館

一、前言

資訊的傳播方式在網際網路和WWW盛行前，主要是以下面的方式進行：資料提供者→圖書館和其他中介機構→資料使用者，其主要特色是間接傳播，也就是資料提供者（如出版社）和資料使用者（如個人）間，由於空間和距離等的限制，並無有效率的直接溝通管道，因此知識的傳播和銷售，往往需要透過一些中介機構，如圖書館和書店的幫助，其中圖書館是社會公共機構的一環，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者和橋樑的角色，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。在電腦尚未興起前，資訊檢索的效率幾乎全依賴人工製作的卡片目錄的品質。在電腦興起後，圖書館自動化逐漸盛行，於是有機讀目錄（Machine Readable Cataloguing, MARC）的產生（註1），利用電腦來提昇編目效率，和結合資料庫來改善檢索效率（註2）。不過MARC的整體架構可說是承襲卡片目錄，換言之



，是卡片目錄的數位化版本。

因為網際網路和 WWW 的緊密結合，資訊傳播的障礙已大大的降低，兩者的結合提供一條非常方便和快速的傳播管道，使資料得以日夜不息的在全球流動。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。這是因為目前在 WWW 上使用的檢索引擎（如 Locys（註3）等），基本的運作方式是屬於全文檢索，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是最為人詬病的。

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，來有效率的描述資料，成爲一個重要的課題，這正是元資料（Metadata）日漸受到重視的原因。

元資料（Metadata）最常見的英文定義是“data about data”（註4），可直譯爲描述資料的資料，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。都柏林核心集是近來頗受重視的元資料之一，本文透過對每一次研討會成果的詳細介紹和分析，勾勒出整個都柏林核心集的發展歷程，和 WWW 上資源描述的發展趨勢。最後本文詳細介紹了元資料實驗系統（MES，網址：<http://blue.lins.fju.edu.tw/~wu>或<http://140.136.85.194/>）的設計和功能，使讀者能輕易來使用MES。

二、都柏林核心集的發展現況和特色

如今網際網路和 WWW 提供了一條直接的管道，使資料提供者和資料使用者可以直接接觸，毋須透過圖書館來作爲媒介者。這固然降低了資訊傳播的障礙（少了一個中介機構），但另一方面，資料提供者如今必須自己擔負起圖書館所提供的一些功能，其中之一是對所擁有的資料加以描述（著錄）。但圖書館所發展出來的資料描述格式，雖然完整和嚴謹，但卻較適合圖書館專業人員使用，對大多數的非專業人員而言，是過於繁瑣和不易學習的。都柏林核心集（Dublin Core）（註5）即是在這一背景下興起的產物，試圖提供一套簡易的資料描述格式，來滿足大多數非圖書館專業人員的需求。



(一)第一次研討會

都柏林核心集這個元資料格式，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，在邀請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 (註6)：

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？

根據研討會的報告，都柏林核心集處理的對象，將祇限於「類文件物件」(document-like objects, 簡稱DLO) (註7)，那何謂DLO呢？簡言之，是可用類似描述傳統印刷文字媒體方式，加以描述的電子檔案。同時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 and 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。

就項目的基本設計原則而言，基於與會者認為沒有任何單一的元資料格式，足以適用於任何作業環境的認知，他們主張先建立一套描述資料的最小核心資料項。因此都柏林核心集的設計原理，是使此元資料的資料項，同時擁有意義明確、彈性和最小規模三種特色。在設計上所秉持的原則是：內在本質原則、易擴展原則、語法獨立原則、無必須項原則、可重覆原則、和可修飾原則。以下是它們的簡要敘述 (註8)：

(1) 內在本質原則 (Intrinsicality)：祇描述跟作品內容和實體相關的特質，例如主題 (subject) 屬於作品的內在本質。但是收費和存取規定，則屬於作品的外在特質，原則上不屬於核心資料項，將透過其他機制來加以處理。

(2) 易擴展原則 (Extensibility)：應允許地區性資料以特定規範的方式出現，也應保持元資料日後易擴充的特性，以及保有向後相容的能力。

(3) 語法獨立原則 (Syntax-Independence)：在此元資料成熟前，將盡量避免制定特定語法。

(4) 無必須項原則 (Optionality)：所有資料項都是可有可無，以保持彈性和鼓勵非專業人士參與製作。

(5) 可重覆原則 (Repeatability)：所有資料項均可重覆。

(6) 可修飾原則 (Modifiability)：資料項可用修飾詞 (qualifier) 來進一步修飾其意義。

以上各原則的詳細分析，及其對資料著錄所造成的影響，請參見「從都柏林核心集看未來資料描述格式的發展趨勢」一文 (註9)，根據以上的原則，研討會的與會者制定了13個資料項 (註10)。



(二)第二次研討會

第一次研討會制定出核心著錄項後，在美國、歐洲、和澳洲等地引起廣泛的研究興趣，於是一年後在英國的 Warwick (1996年 4月) 由 OCLC和 UKOLNC (United Kingdom Office for Library and Information Networking) 聯合舉辦了第二次研討會。鑒於未來各種不同元資料會共榮共存的必然趨勢，以及須為都柏林核心集制定更明確的實作機制，此次研討會有二大目標 (註11)

(1) 協助跨越不同語言和增加語意互通性 (semantic interoperability) 。

(2) 制定一套機制來增加DC的擴充性和跟其他元資料的連結能力。

此次研討會的成果即是一套初步的架構，能用來達成上述二個目標，此架構以開會的地點命名為「瓦立克架構」(「Warwick Framework」)。

以下根據『The Warwick Framework: A Container Architecture for Aggregating Sets of Metadata』一文簡介其基本架構和特性 (註12)。「瓦立克架構」主要包含二種元件，一個是「封裝物」(Package)，一個是用來容納各式「封裝物」的「容器」(Container)。這二種元件的關係，可用水果罐頭禮盒來類比：禮盒即是「容器」，用來將各式各樣的水果罐頭(「封裝物」)組合成一個易於攜帶(傳輸)的單元，同樣的，「容器」的主要功能也在

(1) 組裝各式不同的「封裝物」為一體。

(2) 易於在網路上傳輸，通常以IETF的URI (註13) 做為存取的標籤。

何謂「封裝物」(Package)呢？它正如水果罐頭，基本上是自成一體的小單元，所謂的自成一體，是指「封裝物」本身含有足夠的資訊，讓接收方在拆開「容器」後，可以被單獨來加以處理。(換言之，個別「封裝物」可以獨立於「容器」和「容器」內其他「封裝物」來使用。)「封裝物」基本上有三種類型：

(1) 元資料封裝物：用來裝載元資料本身。

(2) 間接指引封裝物：其功用是用來指引到其他物件(或資源)，所扮演的角色如同URL和URN。

(3) 容器封裝物：如同「容器」(Container)，用來容納許多「封裝物」(Package)，因此可形成巢狀結構，好比大禮盒內再有小禮盒。「瓦立克架構」(「Warwick Framework」) 架構的特色如下：

(1) 允許個別設計者專注於其特殊的元資料設計，因為「封裝物」是自成一體的獨立小單元。

(2) 個別元資料的一切操作由包含它的「封裝物」負責，網路上的資料傳輸者祇須處理「容器」本身。

(3) 提昇相互操作性和擴充性，接收者可自由取其所需的「封裝物」而忽略其餘。



(4) 描述同一文件的不同元資料，可個別分開控制和處理，如USMARC和都柏林核心集。

(5) 可自由加入新版本元資料，而不妨礙舊版本的繼續流通和使用，祇要將新版本另外放入一個「封裝物」即可。

關於「瓦立克架構」的實作方式，研討會中初步建議三種方式—HTML、MIME、SGML。本文僅舉HTML的實作方式以供參考，HTML的實作規格，主要是藉由HTML 2.0版中提供的META和LINK二種標籤，其規格如下（註14）：

(1)<META NAME="<schema_name >.<element.name>" CONTENT="string data">

(2)<LINK REL=META.<schema_name> HREF="<URL>">

以作者開發的元資料實驗系統（MES，下一節中有詳盡的介紹）首頁為例，其HTML式的規格如下：

<HTML>

<HEAD>

<TITLE>元資料實驗系統 (Metadata Experimental System, MES) </TITLE>

<META NAME = "DC.title" CONTENT = "元資料實驗系統 (MES)">

<META NAME = "DC.author" CONTENT = "吳政勸">

<META NAME = "DC.identifier" CONTENT =

"http://blue.lins.fju.edu.tw/~wu/metadata/">

<LINK REL="SCHEMA.DC" REF="http://meta.org/meta-reg/Dublin- Core.html">

</HEAD>

<BODY>

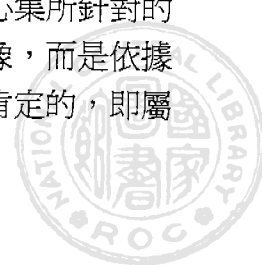
</BODY>

</HTML>

至於MIME和SGML的實作規格，請參見「The Warwick Framework: A Container Architecture for Aggregating Sets of Metadata」一文（註15）。

(三)第三次研討會

1996年9月24-25日CNI (Coalition for Networked Information) 和OCLC舉辦了一場研討會，地點與第一次研討會相同，都是在美國俄亥俄州的都柏林，此次研討會邀請了70位網路圖像 (image)資源描述專家與會，討論都柏林核心集在圖像資源描述上可扮演的角色。出乎意料的，與會的專家經過討論後，認為柏林核心集祇要稍加修改和擴充，即可用來描述大多數的圖像資源，主要原因是與會專家認為，都柏林核心集所針對的「類文件物件」(DLO)，其劃分的依據，並非物件的呈現形式是文字或圖像，而是依據物件的內容，是否對所有使用者來說，其意義是大致相同的，如果答案是肯定的，即屬



於 DLO。相反的如抽象畫，每個人對畫的解讀均不同，則為非DLO物件，其他的非DLO物件有虛擬情境 (Virtual Experience)、資料庫 (Database)、互動式應用 (Interactive Application) 等 (註16)。

雖然圖像和文字資源可用大致相同的一組資料項來加以描述，但是圖像資源有其特殊性，例如：

- (1) 圖像型態：位元對映 (Bit-Mapped) 或向量 (Vector)。
- (2) 圖像檔案格式：GIF、TIFF等格式繁多。
- (3) 壓縮方法和壓縮比率：如JPEG等。
- (4) 解析度。

換言之，圖像資源的使用，所須資訊甚多。

因此根據與會專家的建議，以及會後的討論，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，使其名稱較不具文字導向色彩，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17)：

- (1) 主題和關鍵詞 (Subject and Keywords)：作品所屬的學術領域。
例子：Subject = Digital Geospatial Metadata。
- (2) 題名 (Title)：作品名稱。
例子：Title = Geospatial Support Staff Metadata Tutorial。
- (3) 著者 (Author or Creator)：作品的創作者或組織。
例子：Creator = Abeyta, Carolyn。
- (4) 簡述 (Description)：文件的摘要或影像資源的內容敘述。
- (5) 出版者 (Publisher)：負責發行作品的組織。
- (6) 其他參與者 (Other Contributors)：對作品創作有貢獻的相關人或組織。
〔註：如書中插圖的製作者。〕
- (7) 出版日期 (Date)：作品公開發表的日期。
例子：Date = 1995/05。
- (8) 資源類型 (Resource Type)：作品的類型或所屬抽象範疇。
例子：Type = Dictionary。
- (9) 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。
- (10) 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。

(11)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/。

(12)來源 (Source)：作品從何處衍生而來 (同內容範疇)。

(13)語言 (Language)：作品所使用的語言。

例子：Language = English。

(14)涵蓋時空 (Coverage)：作品所涵蓋的時期和地理區域。

(15)版權規範 (Rights Management)：作品版權聲明和使用規範。可能值如下 (註18)：

(a) 空白 (Null)：無特別聲明，使用者須自行參考其他來源。

(b) 無限制 (No Restriction on Reuse)：可複製再傳播。

(c) 參考處 (URI or Other Pointer)：使用的相關說明，在所指定的出處。

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。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。

(四)第四次研討會

1997年 3月在澳洲坎培拉 (Canberra) 的國家圖書館，舉辦了都柏林核心集的第四次研討會 (簡稱DC-4)，與會者是65位包括電子圖書館研究者、網際網路專家、圖書館員等人士。以下根據『The 4th Dublin Core Metadata Workshop Report』一文 (註19)，簡述DC-4的主要目標和成果，由於在第三次研討會後，都柏林核心集15個資料項的架構已大致確立，DC-4乃在前三次研討會的基礎上進一步發展，所以此次研討會的主要目標為

(1) 確立項目結構：將資料項結構正式化，規範可能的修飾詞使用方式。

(2) 擴充連結機制：都柏林核心集和其他元資料的連結方式。

(3) 項目再精純化：少部分項目其語意的定義須再加以明確化，如版權規範 (Rights Management)。

會議的主要成果是「坎培拉修飾詞」(Canberra Qualifier)，正如文法中的修飾詞功用一樣，都柏林核心集中的修飾詞，是對項目的內容和語意，做進一步的界定或說明，使其意義更明確，目前包括三種修飾詞：

(1) 語言 (language) 修飾詞：指出資料項所使用的語言。



例如：Lang = EN指出此資料項是使用英文。

[註：此修飾詞不是用來指示原始文件（或資源）所使用的語言。此外原始文件和元資料的關係，正如同書籍和目錄間的關係。]

(2) 架構 (Scheme) 修飾詞：指出項目內容的詮釋方法。

例如：SCHEME = LCSH指出這個主題來自LCSH。

(3) 類別 (Type) 修飾詞：其功用同於次項目名稱 (sub-element name)，主要在指示資料項內容涵蓋的範圍。

例如：Creator.PersonalName = "C. J. Wu"，更明確的界定此資料的屬性是（著者）姓名。

至於「坎培拉修飾詞」的實作方式目前有二種建議（註20）：

(1) 完全遵循HTML 2.0規格，其寫法如下：

```
<META NAME="DC.subject"
```

```
CONTENT=" (SCHEME=LCSH) (LANG=EN) Computer Cataloging of Network Resources">
```

此法的優點是與目前的系統和瀏覽器完全相容，缺點是修飾詞和項目的內容，混合在同一字串內，須有特殊程式將其從字串中抽離出來。

(2) 完全遵循HTML 2.0規格，其寫法如下：

```
<META NAME="DC.subject"
```

```
SCHEME="LCSH"
```

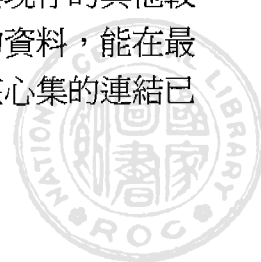
```
LANG="EN"
```

```
CONTENT="Computer Cataloging of Network Resources">
```

此法的優點和缺點大致與上面的方法相反。

綜合來說，此次研討會的成果，是折衷了二派人士的觀點，而產生了「坎培拉修飾詞」，一派人士希望保持都柏林核心集單純化，如此才能使其項目意義明確，且適合眾多非專業人士使用（都柏林核心集創立的初衷）。另一派人士則希望能廣泛的使用修飾詞，來豐富都柏林核心集的內涵和擴大其應用範圍，雖然修飾詞的加入，會使都柏林核心集複雜化和格式不易明確化。最後折衷的成果是「坎培拉修飾詞」，都柏林核心集正式收納了三種修飾詞，不過每種修飾詞的內容並未定案下來，例如那些項目可有那些種修飾詞，每個修飾詞又有那些名詞可使用，這些問題相信在以後的研討會中會被詳加討論。

由於都柏林核心集祇是一套最小核心資料項，須要有一個機制，來與現存的其他較完整描述格式，如USMARC等，來做資料的對照和轉換。使都柏林核心集的資料，能在最小成本下，轉換成更完備的描述格式。在這方面，有關USMARC和都柏林核心集的連結已



在發展中（註21）。

三、元資料實驗系統 (MES) 簡介

有鑒於元資料對資料著錄和檢索的重要性，作者建立了一個相關的實驗系統—元資料實驗系統 (Metadata Experimental System, 簡稱MES, 網址：<http://140.136.85.194/>或<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) (註22)。其中尤以都柏林核心集的功能最為完整，可達成直到第四次研討會的要求，採用與HTML 2.0完全相容的格式。

(二)系統特色簡介:

(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建立的目的是，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已被吸納入元資料的架構中，成為元資料的一種。

(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(註23)，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)系統功能簡介：

(1) 註冊子系統：為了確保您所著錄的資料不會遭到別人的任意刪改，系統提供著錄者註冊的功能。註冊的識別名稱可自訂，不須用自己的姓名，最長可有 255個字元，中文、英文、數字皆可使用。由於識別名稱如圖書館的登錄號須唯一，因此若您欲註冊的識別名稱與他人重覆，則系統會給您錯誤訊息，此時請您試用其他的識別名稱。請您在註冊時也留下其他的個人資料，如姓名、地址、電話、E-mail等，以便於日後的聯絡，您的個人資料是用密碼保護，他人無法取得，敬請放心。若您沒有註冊，或者在著錄時未使用您的識別名稱，則您將無法來修改資料。註冊子系統有提供以下的功能：

(a) (新增)註冊：識別名稱和密碼是必須提供的資料，其餘資料可免填，惟便於日後的聯絡，仍請盡量提供，因為您的個人資料是用密碼保護，並不會外洩，請放心。

(b) 修改註冊資料：可用來更正您先前登錄的個人資料，包括姓名、職稱、地址、電話、傳真、電子郵件地址。識別名稱無法自行修改，若須更動請聯絡系統管理者。



(c) 修改密碼：用來更改個人密碼。

(d) 查詢：輸入您個人的識別名稱和密碼，即可查詢您先前登錄的個人資料，此查詢功能有密碼保護，所以他人無法得知您的個人資料。

(2) 都柏林核心集子系統：都柏林核心集目前有15個資料項，若要使用前面所提的「坎培拉修飾詞」，請使用第一種方式著錄，即將修飾詞內含在內容的字串中，如“(SCHEME=LCSH) (LANG=EN) Computer Cataloging of Network Resources”。此外由於MES將會提供模糊邏輯檢索的功能，所以都柏林核心集中的主題(subject)資料項，有加入關連值，以評斷此主題與文件的關連程度。都柏林核心集子系統主要包含兩個次系統—著錄次系統和查詢次系統。

(a) 著錄次系統：有四個項目如下

[1] 全部項目著錄：可一次輸入15個資料項，也可祇著錄部分的資料項，(不要著錄的資料項是內定為"none")。著錄者內定為"Meta"，請改成您個人的識別名稱，以免失去對資料的修改權。

[2] 單一項目著錄(一般項目)：因為都柏林核心集的特色是無必需著錄項，同時所有資料項都是可重覆項，因此都柏林核心集基本上是允許一次著錄一個資料項，所以MES也提供此種功能以供選擇，此功能包含14個資料項，主題(subject)資料項因為可使用關連值而單獨置入下一個功能。

[3] 單一項目著錄(關連值項目)：即都柏林核心集中的主題(subject)資料項，以1.0代表此文件(或資源)完全契合這個主題，值越小表示相關的程度越少，值須在(0,1)之間(可有小數點)。

[4] 更新：您可隨時更新您先前著錄過的資料，更新時系統會先核對識別名稱和密碼，因此可確保資料的安全。為了擁有資料的修改權，請您務必先在MES註冊，並且在著錄資料時使用您自己的識別名稱。

(b) 查詢次系統：有五個項目如下

[1] 關鍵字查詢：可找出所有資料項中含指定關鍵字的資料。

[2] 單一欄位查詢(一般項目)：查詢都柏林核心集中，某個特定資料項的資料，可用萬用字元(即關鍵字)來查詢。(資料項不包含主題(subject)資料項，因為它可使用關連值，而單獨置入下一個功能。)

[3] 單一欄位查詢(關連值項目)：查詢主題(subject)項中符合指定關鍵字的資料，同時可併用關連值來查詢，祇列出大於指定關連值的資料。

[4] URN對映URL查詢：參見下面URL子系統中的敘述。(此即是URC的主要功能，對映URN到URL(註24))。



[5] 單一文件資料：取得某一文件的所有都柏林核心集資料，若選擇HTML格式，則可得到符合第四次研討會所訂的輸出規格。

(3) 查詢子系統：此部分的查詢，基本上是針對 MES整體，跟個別元資料直接相關的資料查詢，在個別元資料子系統內，如跟都柏林核心集直接相關的查詢，在都柏林核心集的查詢子系統內。有四個項目如下：

(a) 關鍵字查詢：可找出所有資料項中含指定關鍵字的資料。

(b) 單一欄位查詢：查詢選定元資料中，某個特定資料項的資料，可用萬用字元（即關鍵字）來查詢。使用此功能時須自行輸入正確的欄位名稱。

(c) 布林邏輯查詢：熟悉布林邏輯者，可直接使用布林邏輯複合式來進行較複雜的查詢。

(d) URN對映 URL：參見下面 URL子系統中的敘述。（此即是URC的主要功能，對映URN到URL。）

(e) URL查詢：利用關鍵字找出相關的URL資料。

(f) URN查詢：利用關鍵字找出相關的URN資料。

(4) URL子系統：跟 URL 相關的處理，有四個項目如下。

(a) URL著錄：雖然許多元資料都有項目可放入 URL相關的資料，但URL並不必然附屬於特定元資料，在 MES 中可單獨著錄特定URN相關的URL資料。

(b) URL 更新：修改您先前著錄過的資料，更新時系統會先核對識別名稱和密碼。爲了擁有資料的修改權，在著錄資料時，請使用您自己的識別名稱。

(c) URL查詢：利用關鍵字找出相關的URL資料。

(d) URN對映URL：MES中所有的資料都是用URN來識別，爲方便網路資源的取得，在查詢到資料的 URN後，可透過此功能找到URL來取得文件。（此即是URC的主要功能，對映URN到URL。）

(四)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。

四、結論

透過目前在 WWW上的檢索引擎來收集資料時，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，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。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。

元資料正是因應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。

本文介紹都柏林核心集從創始（1995年3月第一次研討會）到（1997年3月）第四次研討會的發展沿革，第一次研討會確立都柏林核心集處理的對象，將祇限於「類文件物件」（document-like objects，簡稱DLO），同時目標是發展一套描述資源的最小核心資料項，使其簡單有彈性，且非專業人員也可輕易來了解和使用，研討會的與會者，根據內在本質原則、易擴展原則、語法獨立原則、無必須項原則、可重覆原則、和可修飾原則等6個原則，制定了最初的13個資料項。

1996年4月的第二次研討會制定了一套機制，來增加都柏林核心集的擴充性和跟其他元資料的連結能力，此架構為「瓦立克架構」（「Warwick Framework」）。它主要包含二種元件，一個是「封裝物」（Package），一個是用來容納各式「封裝物」的「容器」（Container）。這二種元件的關係可用水果罐頭禮盒來類比：禮盒即是「容器」，用來將各式各樣的水果罐頭（「封裝物」）組合成一個易於攜帶（傳輸）的單元，而「封裝物」裏面裝載元資料。

1996年9月第三次研討會，將柏林核心集的處理對象，進一步擴充到圖像（image）資源，同時修改了部分資料項名稱，使其名稱較不具文字導向色彩，以及新增了二個資料項—簡述（Description）和版權規範（Rights Management），總共成為15個資料項。

1997年3月第四次研討會正式在柏林核心集中加入「坎培拉修飾詞」（Canberra Qualifier），修飾詞是對項目的內容和語意，做進一步的界定或說明，使其意義更明確，目前包括三種修飾詞：語言（language）修飾詞、架構（Scheme）修飾詞、類別（Type）修飾詞，並且建議了二種實作方式。



都柏林核心集雖然祇是一個較簡略的料描述格式（或元資料），但它是圖書館界試圖解決電子文件處理難題上的一個新嘗試，其所立下的原則和典範，如無必須項原則、可重覆原則、可修飾原則等，是令人印象深刻和激賞的。以作者個人的觀點，都柏林核心集對未來資料描述格式所揭示的方向和途徑是正確的。

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://140.136.85.194/>或<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。MES系統目前有四大子系統如下：

（一）註冊子系統：提供的功能有（新增）註冊、修改註冊資料、修改密碼、查詢等四種。

（二）都柏林核心集子系統：包含兩個次系統—著錄次系統和查詢次系統。

（1）著錄次系統：提供的功能有全部項目著錄、單一項目著錄（一般項目）、單一項目著錄（關連值項目）、更新等四種。

（2）查詢次系統：提供的功能有關鍵字查詢、單一欄位查詢（一般項目）、單一欄位查詢（關連值項目）、URN對映URL查詢、單一文件資料等五種。

（三）查詢子系統：提供的功能有關鍵字查詢、單一欄位查詢、布林邏輯查詢、URN對映URL、URL查詢、URN查詢等六種。

（四）URL子系統：提供的功能有URL著錄、URL更新、URL查詢、URN對映URL等四種。

註 釋：

註 1 IFLA, "UNIMARC: An Introduction," <www.nlc-bnc.ca/ifla/VI/3/P1996-1/unimarc.htm> (26 Sept. 1996).

註 2 徐小鳳，自動化書目的資訊服務（台北市：學生書局，民73），頁43。

註 3 Infoseek Corp., "Infoseek Home Page," <<http://www.infoseek.com/>> (18 Feb. 1996).

註 4 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 5 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>.

註 6 同註5，頁2。

註 7 同註5，頁3。

註 8 同註5，頁5-6。

註 9 吳政叡，從都柏林核心集看未來資料描述格式的發展趨勢，圖書館學刊，第26期（民86年5月），頁16-17。

註10 吳政叡，三個元資料格式的比較分析，中國圖書館學會會報，第57期（民85年12月），頁40。

註11 C. Lagoze, C. A. Lynch, and R. Daniel, Jr., "The Warwick Framework: A Container Architecture for Aggregating Sets of Metadata," *D-Lib Magazine* (July 1996), <<http://www.dlib.org/dlib/july96/lagoze/07lagoze.html>>, p. 3.

註12 同註11，頁8-9。

註13 Ron Daniel, Jr., "Uniform Resource Identifiers (URIs)," 1996, <<http://www.acl.lanl.gov/URI/uri.html>>。

註14 同註11，頁12-13。

註15 同註11，頁14-17。

註16 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>>, p. 2.

註17 "Dublin Core Metadata Element Set: Reference Description," 15 Jan. 1997, <http://purl.org/metadata/dublin_core_elements>.

註18 同註16，頁5。

註19 S. Weibel, R. Iannella, and W. Cathro, "The 4th Dublin Core Metadata Workshop Report," *D-Lib Magazine* (June. 1997), <<http://www.dlib.org/dlib/june97/metadata/06weibel.html>>.

註20 同註19，頁6-7。

註21 Rebecca Guenther, "Mapping the Dublin Core Metadata Elements to USMARC", 1995, <<gopher://marvel.loc.gov/00/.listarch/usmarc/dp86.doc>>.

註22 同註10，頁41-43。

註23 同註10，頁41。

註24 同註10，頁42。

