

# 數位典藏計畫異質系統互通機制： 以OAI建立聯合目錄的理論與實作

陳 昭 珍

國立臺灣師範大學圖書資訊學研究所教授

## 【摘要】

2001 年國科會開啟數位典藏國家型計畫，為使各機構及一般大眾能分享及整合檢索各單位所建置的數位資源，聯合目錄的建置是當務之急。聯合目錄的建置模式可分為兩種：一種是集中式的聯合目錄，另一種為分散式查尋的虛擬聯合目錄，集中式聯合目錄的優點是檢索效能佳，但其缺點是建置成本昂貴。虛擬聯合目錄的優點是建置成本低廉，但是檢索效能差。為保留上述集中式聯合目錄及虛擬聯合目錄的優點，且避免其缺點，在數位時代，一個新的分散擷取metadata 的協定——OAI (Open Anehives Initative) 就應運而生。本文主要在分析幾種不同的聯合目錄建立模式，並以數位典藏國家型計畫為例，說明選擇OAI 做為國家數位典藏聯合目錄建置機制的的原因，及此系統的架構、功能、實作時所遇到的問題。

關 鍵 詞：數位典藏國家型科技計畫、互通性、異質系統、開放檔案詮釋資料擷取協定、聯合目錄

## 一、聯合目錄的必要性

網際網路的普及，使得我們可以便利的查尋遠端資源，全球蔚為風潮的數位化計畫，更使我們得以從網路上取得更多有價值的數位內容。然而，一個個異質且分散的資訊系統，對使用者而言，是福音也是負擔，如果沒有很好的機制，將

分散各地的異質系統加以整合，則使用者將需要個別去連結並使用其不同的檢索功能；更有甚者，除非，有一個好的聯合目錄，否則可能各機構辛苦建立的資訊系統，將不易被人得知。一本好期刊若未被期刊論文索引收錄，將不易被人發覺，一本好書若未印在出版目錄上，不易被遠方的買者得知，一個網站若未被搜尋引擎抓取，將難以被人拜訪，同樣的，資訊系統若未被聯合目錄收錄，也難以被人檢索。

聯合目錄除了提供整合檢索的便利外，對數位典藏國家型計畫而言，它更是展現計畫整體成效的最佳工具。由聯合目錄，可以得知數位化的現況，而若各種主題、類型、地區、時期及典藏單位的資料可以被排比利用，則無論對學習者、研究者、加值者而言，都是便利的工具，更使得數位內容的價值有加乘的效果。

## 二、聯合目錄的建置模式

聯合目錄的建置模式主要有兩種，一種為實體聯合目錄，又稱為集中式聯合目錄；另一種為虛擬聯合目錄，又稱為分散式聯合目錄。所謂實體聯合目錄是指在聯合目錄伺服器上已建好索引檔，使用者查尋時，系統告知使用者找到多少符合查尋條件的資料，再由此索引檔連到實際的網站，如搜尋引擎，或叫出書目紀錄，如圖書館界的書目中心。虛擬聯合目錄則指未在伺服器建立實際的索引檔，只是透過檢索介面將使用者的查尋問句送到各資料庫，並將查尋結果彙整後顯示在螢幕上。例如以 Z39.50 協定所建立的查尋介面，就是這種做法。這兩種聯合目錄各有優缺點，以下即分別說明其特色。<sup>[1]</sup>

### (一)集中式聯合目錄

#### 1.集中式聯合目錄的類型

集中式聯合目錄的建立方式主要有兩種：一種是大家共用一套資訊系統，各個共用的機構以使用者端程式將資料鍵入共用的資料庫管理系統，這種資訊系統大部分的功能採參數化設計，使之能適用於各個共用此系統的機構；第二種方式是參與合作的單位先在本身的系統上建立資料，再將資料以批次方式上傳到聯合目錄，如國家圖書館的全國書目資訊網，即為此種模式；或先將資料建到聯合目錄，再即時下載到自己的系統，如 OCLC (Online Computer Library Center) 書目中心即採用這種方式。

[1] 陳昭珍，《電子圖書館整合檢索之理論與實作》(臺北：文華，2000)，頁 32-41。

## 2. 集中式聯合目錄的特性

集中式的聯合目錄，其特性可由一致性、整合性、系統效能與管理（performance and Management）三個角度分析之。

### （1）查尋與索引的一致性

這類聯合目錄的資料，雖由不同的單位上傳而來，但由於已透過標準軟體處理，集中儲存在聯合目錄中，所以實際在查尋時只針對一個資料庫查尋，因此查尋功能及索引方式都是一致的。

不過由於各館的編目原則並不完全一致，因此送到聯合目錄的資料也無法完全一致，所以各聯合目錄系統都會處理品質控制問題。

這類聯合目錄的查尋與檢索技術乃以發展了四十年的資訊檢索技術為基礎，也可以根據資料的屬性或統計屬性來排序檢索結果，換言之，任何適用於單一資訊系統的技術都適用於集中式的聯合目錄。

### （2）紀錄的整合

各聯合目錄系統對資料的整合處理，詳盡情況不盡相同，如 OCLC 有很多計畫在做紀錄的整合，不過他們的整合，主要在刪除重複，並保留一筆正確的紀錄，而不記錄各館編目的差異。

加州州立大學的 MELVYL 系統則花很多時間在整合上，他們對於每一筆新輸入的資料都逐欄的與既存的紀錄做比較，當發現有不一致的情形時，系統會一欄欄的記錄及保存不一致之處，以致載入速度很慢，每小時只能載入1,000—2,000 筆新紀錄。

雖然整合要花時間，對集中式的聯合目錄而言，高品質的整合目標是可以達到的，換言之，可以做到在書目資料庫中沒有重複的紀錄，但分散式的聯合目錄則不可能。

### （3）系統效能與管理

集中式聯合目錄已有數十年的發展歷史，它在管理大量資料的技術已相當成熟，但因過去的使用者主要是從業人員或檢索專家，因此相對的，與一般終端使用者之間的互動，經驗尚淺。

集中式的聯合目錄也已有很好的系統效益評估工具，如對於系統反應時間及系統的使用情況的掌握相當完整。

若要加入一個新的合作單位，所需成本不高。

## (二)分散式的聯合目錄

### 1.分散式聯合目錄的類型

分散式的聯合目錄主要可分為以標準協定來建立主從架構的分散式查尋系統，最有名的就是 Z39.50 及 OpenURL 系統；及非依標準設計的分散式查尋系統。其中 Z39.50 廣被圖書館自動化系統所支持，而 OpenURL 則是目前整合商業性資料庫及圖書館資訊系統非常著名的標準。非依據標準設計的分散式查尋系統，其查尋介面需要能轉換查尋問句到各個不同的系統，當異質系統不多或只檢索同一種資訊組織模式（如都是 MARC 系統）時，還可以應付，但若要跨不同資訊組織模式或不同國界做檢索，困難度較大。

### 2.分散式聯合目錄的特性

分散式聯合目錄最大的優點，是不需花任何的軟硬體成本及人力資源，去建立及維護實體的聯合目錄。不過就查尋與索引的一致性、紀錄的整合、系統的管理與效能而言，都較集中式的聯合目錄差。茲說明如下：

#### (1) 就查尋與索引的一致性而言

理論上，就功能而言，分散式應可做到和集中式一樣的效果，但事實上有兩個問題：

所有的系統，必需支援最小共通的查尋功能，如果其中一館無法做到（如某館無法支援 truncate，或索引欄位不同），就會產生不一致的查尋結果。介面設計的愈複雜，各館也更需正確的支援這些功能，才能執行複雜的查尋動作。

不同的系統必需用共通的語意設計 Z39.50 的查尋屬性，以及以一致的方法處理這些屬性。由於 Z39.50 並非資料庫的索引標準，因此屬性集也不是依資料庫的架構來定義。例如很多系統會接受並回覆 Z39.50 問句的作者與題名查尋，但在資料庫中，這些系統並不見得用相同的欄位做作者與題名索引，如有的系統，題名索引包括正題名、副題名、並列題名、其他題名。有的未包括這麼多。

#### (2) 從資料整合的角度而言

有些 Z39.50 client 只用來以相同的介面檢索遠端的系統，而未做多資料庫同時查尋（broadcasting，或稱廣域查尋），因此沒有整合的問題。即使有做整合，也是根據單一鍵如 ISBN 等來刪除重複，但大部

分的系統都無此功能。

如果要做到像 MELVYL 的整合，必需將檢索出來的紀錄反覆查尋各參與的系統；即使不做如 MELVYL 的整合，但要做任何的整合必需由客戶端將送回的資料加以整合 (merging)，或在伺服器端將資料以相同的方式排序，而這並不是一件簡單的事。

### (3) 從系統效能與管理的角度而言

分散式系統的效能往往視 client / server 之間網路速度而定。此外，也常受速度慢的伺服器的影響，因為它必須等所有的 server 都回覆後，才能顯示查尋結果。所以有的系統會考慮設定等待時間，因此是否要設定等待時間 (time-out)，以及等待時間要設多久，是這類系統必需考慮的問題。

分散式查尋會增加各系統 (local system) 的查尋負擔，因為每一個查詢都會送到各 local system 處理。

查尋反應時間比聯合目錄慢很多。

各系統專屬的查尋系統功能，往往比虛擬聯合目錄還多，這是無可避免的，但是如果各系統的資訊組織模式一致，則共通介面所能提供的功能也會和專屬的查尋功能越接近。

小系統參與虛擬聯合目錄做分散式查尋，會使得查尋速度更慢。

虛擬聯合目錄的可靠性也是一個問題，因為隨時會有某一臺主機當機，或停止提供服務的情況發生。

## 三、OAI-based的數位典藏聯合目錄

由上節有關集中式聯合目錄及分散式聯合目錄的優缺點的分析中可知，若以類似 Z39.50 協定及 OpenURL 的模式，即時分散查尋各資訊系統，可能會浪費使用者的查尋時間；而若以人工方式，由合作單位批次傳送資料到聯合目錄，則不但浪費各合作單位的人力，若合作館未定時將資料傳送，則聯合目錄將無法適時反應各合作單位的最新資料狀況。所以一個簡單、容易開發、可自動擷取各合作館的 metadata 的協定——OAI 正好可以彌補二者的缺憾，因此我們以 OAI 技術開發數位典藏聯合目錄的原型系統 (Prototype)。

OAI (Open Archives Initiative) 最初是由 Paul Ginsparg、Rick Luce、

Herbert Van de Sompel 等人，在 1999 年 10 月於 Santa Fe 的 Universal Preprint Service 會議中所促成的。有鑒於各資料庫系統，彼此互不隸屬，相關資料分散而難以統整，使得資料的流通有所限制，該會議的與會代表認為有必要對於學術性電子期刊的預刊本及相關數位典藏，發展出一套可以互通 (interoperability) 的標準架構，因此成立開放典藏計畫 (Open Archives Initiative, 簡稱 OAI)。<sup>[2]</sup> 並於 2001 年 1 月，發表了名為 Open Archives Initiative Protocol for Metadata Harvesting (簡稱為 OAI-PMH) 的網路通訊協定，提供異質性資料庫間互通搜尋的解決方案。OAI-PMH 運用網際網路 (Internet) 及詮釋資料 (metadata) 兩種技術，在增強整合檢索功能及簡化系統開發難度上，達成了極佳的平衡。發佈到網路上的學術資料，藉助 OAI 詮釋資料擷取協定，將可不受限於系統平臺、應用程式、領域、國界及語言，達到廣泛流通的目的。也因此，透過 OAI 詮釋資料擷取協定，使用者能利用最經濟的時間，取得更豐富且精確的資料。<sup>[3]</sup>

#### 四、OAI 協訂內容

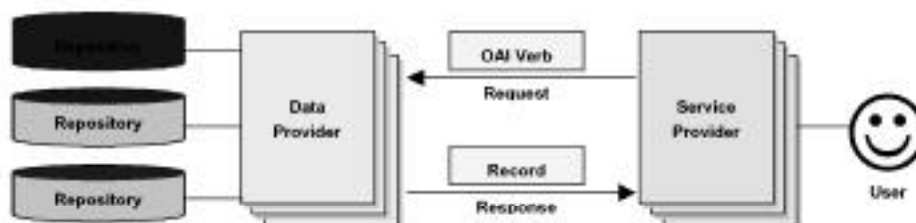
OAI 是一個簡單、容易設計程式的協定，主要乃在透過指定的命令集，提供前端向後端儲存器提取所需資訊的協定，OAI 元件主要分為 OAI Service Provider 與 Data Provider。以 OAI 協定為基礎的聯合目錄架構，主要是由 OAI 的 service provider (服務系統) 定期向 data provider (各典藏單位的資料庫系統) 抓取 metadata 資料，建立集中式的聯合目錄。由於 OAI 是架構在 HTTP 上的應用協定，因此其命令集即是透過 HTTP 所使用前端與後端傳輸的變數名稱與其內容，觸發後端對應的伺服器程式，依據變數內容處理後傳回的結果，並須遵照 OAI 協定 XML Schema 所規範的 XML 格式傳送資料。Service Provider 主要在集中維護從各系統擷取來的 metadata，並進行加值性服務的查詢檢索服務；Data Provider 主要在維護資料倉儲 (repositories)，並且支援 OAI 協議來取得其倉儲

---

[2] Carl Lagoze and Herbert Van de Sompel (2001), "The Open Archives Initiative: Building a low-barrier interoperability framework," *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries, Roanoke VA* (June 24-28, 2001), pp. 54-62, from <http://www.cs.cornell.edu/lagoze/papers/oai-jcdl.pdf>.

[3] Herbert Van de Sompel and Carl Lagoze (2000), "The Santa Fe Convention of the Open Archives Initiative," *D-Lib Magazine* 6:2, from <http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>.

內的資料內容。其關係以圖一表示：<sup>[4]</sup>



圖一：OAI 基本架構

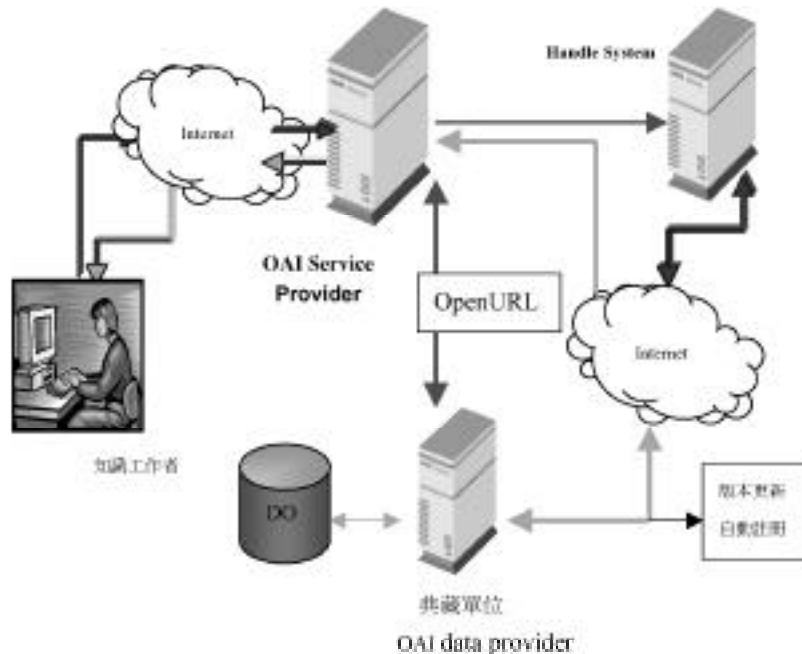
## 五、數位典藏聯合目錄實作

### (一) 系統功能

為建立數位典藏國家型計畫的聯合目錄，數位典藏計畫邀請國內參與數位典藏機構的代表成立OAI test-bed 小組<sup>[5]</sup>，以 OAI 技術建構國家數位典藏的聯合目錄；OAI 是一個簡單、容易設計程式的協定，但是在實際的聯合目錄系統設計上，尚有一些是 OAI 未考慮到的問題，如與各單位的資料庫應如何連結、如何透過詮釋資料擷取數位物件、資料服務端的介面應如何設計等，都是 OAI 未訂定，也是無法訂定的部分，但在實際環境中，則是一定要處理的問題。以下即介紹我們為國家聯合目錄所定義的系統功能及架構。整體系統架構如圖二所示：

[4] 國家圖書館，《研擬圖書資訊相關技術規範座談會會議資料》（臺北：國家圖書館，2001），頁 3-1。

[5] 本系統的開發由數位典藏國家型計畫辦公室補助經費；感謝一起參與討論及設計的夥伴：陳雪華教授、陳亞寧組長、陳淑君小姐、蘇國盛先生、歐陽慧小姐、張陳基先生、顏學勇先生、洪筱盈小姐、張懷文小姐、徐代昕先生、沈漢聰先生等的參與。更感謝參與第一期系統測試的典藏單位：中央研究院、故宮博物院、國立自然科學博物館、國立臺灣大學圖書館。



圖二：整體系統架構

如上圖所示，此聯合目錄的主要功能包括：

#### 1. Service Provider 的功能

##### (1) Metadata 擷取及維護功能

定期向各 data provider 抓取 metadata 資料，以更新聯合目錄中的 metadata，抓取條件包括：set、date、identifier、metadata 類型等。更新頻率可以參數設定。

具備讓 data provider 登錄功能 (register)，並以親和的介面維護基本資料，及上傳 metadata、DTD 及 schema。

Service provider 為 Web services 的服務系統，提供使用者查尋、瀏覽功能。

Service provider 可儲存維護各典藏單位較詳細的 metadata，並具有 metadata cross mapping 的功能。

##### (2) 瀏覽功能

可設定將 metadata 中，典藏單位、作品類型、創作者、主題、時代、地點、典藏單位等欄位資料另外擷取出來，做為瀏覽點選功能的



主題目錄。

目錄編製功能應模組化，並可分辨已處理及未處理的資料。

可將各單位的影像圖檔，以 thumbnail 存檔，作成圖示索引，讓使用者以作品類型、創作者、主題、時代、地點等點後，將點選結果以小圖示呈現，再連到各典藏機構，呼叫大圖。

### (3) 查詢功能

系統能提供欄位檢索、模糊檢索（含同音、同義、簡繁體等）及不限欄位的全文檢索功能。

共通的查尋模式包括，簡單查詢及進階查詢。進階查詢欄位包括：作品名稱、作品類型、創作者、主題、時代、地點、典藏單位等。

系統具布林邏輯查詢、切截查詢（truncate）完整查詢的功能，並以年代、資料類型、地點、檔案格式、館藏單位等條件縮小查詢範圍，限制條件儘量以點選方式處理，不需使用者鍵入資料。

除整合檢索的共通介面外，使用者可選擇只要查尋某幾個單位的資料。查尋後的資料以共通的欄位呈現，並可設定是否要連結更豐富的原始 metadata。

系統可讓使用者選擇呈現資料的排序方式，並可前後翻頁顯示。可選擇排序方式，含相關度、主題、創作者、時代、典藏單位等。此外可選擇以文字清單呈現或圖示清單呈現。

### (4) 全文、多媒體連結管理功能

系統應提供正確的全文或多媒體連結、呈現或播放。連結機制包括透過 handle system 轉為 URL 的連結。

### (5) 權限及系統管理功能

Service provider 具權限設定功能，可依據 service provider 系統管理者、目錄維護者、一般使用者，及 data provider 登錄編輯等，設定權限。

### (6) 擴展功能

此 Service provider 也可擔任 data provider 角色，提供 Dublin Core 格式紀錄，與國內外其他機構進行互通。

具有 CCCII、big-5、UTF-8 轉換功能，可將各單位送來的資料轉為

UTF-8。

首頁除了查詢與瀏覽外，尚可整合查詢 Internet 上的網站，具入口網站功能。

## 2.Data Provider 的功能

### (1) 支援 metadata 擷取功能

提供 Service Providers 可以根據 set、identifier、datestamp、metadata prefix 或是其他查詢指令個別及批次取得資料，且可設定自動擷取的時間與條件，也可以手動擷取。

建立 Identify 資料：建立儲存器相關資訊，包括行政管理、系統標示等資訊。

根據 OAI 協定規範，至少必須包含以下資訊：

- \* repositoryName
- \* baseURL
- \* protocolVersion
- \* adminEmail

#### Metadata 資料格式需求

OAI Data Provider 提供各種文件與資料的 Metadata，Metadata type 可為各單位本身使用的詳細格式，但也需提供 Dublin Core 格式資料：

#### a. Metadata 結構包含三部分

- \* Header：標記有關此筆資料的資訊，主要可分為二部分，包括 unique identifier、datestamp。
- \* Metadata：資料錄實際的 Metadata 內容。
- \* about：非必要部分，其他相關說明，如版權宣告。

b. DateStamp：有關資料增修的日期，提供 Service Provider 程式判斷處理。

c. OAI Data Provider 必須根據所典藏的資料種類，提供不同的 Metadata，並且根據不同 Metadata 設計各種 XML schema。但是最起碼必須提供 Dublin Core 格式 Metadata 並符合 oai\_dc OAI-defined XML schema 格式，oai\_dc 規格內容規定見 <http://www>。

openarchives.org/OAI/openarchivesprotocol.htm#dublincore。

- d. 根據不同 Metadata 所設計各種 XML schema 需符合 XML schema defined for the response ( 必須通過 XSV 驗證 XML schema validator )
- e. OAI Data Provider 提供的 Metadata 格式必須是 valid XML 格式 ( valid XML 指的是裡面的標籤用法符合某一套 schema 或 DTD 所規定的規則 )
- f. 提供具有唯一性的 OAI identifier：每一筆 metadata record 都必須有唯一的 unique identifier，identifier 的格式需為 URL，並且在 respective repository 中是唯一的。採用 OAI identifier scheme，確保在 OAI namespace 具有唯一性。OAI identifiers 格式正式規定在 <http://www.openarchives.org/OAI/oai-identifier.xsd>。

OAI identifier 基本上分成三部分，並以「：」為分隔符號。

「 schema：Repository ID：local ID」

A scheme that must be equal to oai

A repository identifier, unique within the oai scheme. The repository identifier is case sensitive and may contain any alphabetic or numeric character (no punctuation characters are permitted).

A local identifier that is the unique identifier of a record within a repository.

OAI identifier 範例：oai:Metalogy:abcdef/0002/01

### 3. 全文及多媒體連結功能

- (1) data provider 需能支援 service provider 端的 Handle System 送來的連結請求。
- (2) data provider 端應可管理在 handle system 上的數位物件的儲存位置，一旦位置改變時可以更新 Metadata 資料中的連結。

### 4. 字碼轉換功能

- (1) 以 UTF-8 將 metadata 送給 service provider。
- (2) 並具有 CCCII、big-5、UTF-8 轉換功能。

### 5. Handle system

- (1) 各單位的檔案應以 handle system 自動檢查檔案名稱的正確性、唯一性及與 metadata 連結的正確性，以及一個 file name 與多個 location 對照表維護的功能，此對照表由各單位同時維護，所以需有嚴格的權限控制。
- (2) 可整批或單筆增刪、修改多媒體檔案的唯一名稱及其對應的URL。
- (3) Handle system 需具定時、自動核對連結正確性的功能。
- (4) Handle system 需具有排序及查尋的功能。
- (5) Handle system 需具統計功能，可按照單位、資料類別、日期等統計資料量及使用數量。
- (6) 若一數位物件同時存放多處，系統應可設定提供連結那一處的物件優先順序。

## 6. 系統註冊

- (1) 系統需向 OAI registry 註冊。
- (2) 註冊規定

進行符合測試 Your repository will be listed in the registry only after you reply to this email.

定期檢測。

儲存相關資料

- \* repositoryName: a human readable name for the repository;
- \* baseURL: the BASE-URL for making protocol requests to the repository;
- \* protocolVersion: the version of the OAI protocol supported by the repository;
- \* adminEmail: the e-mail address of the administrator of the repository;
- \* additional repository-specific description packages;
- \* Metadata format, DTD, XML schema.

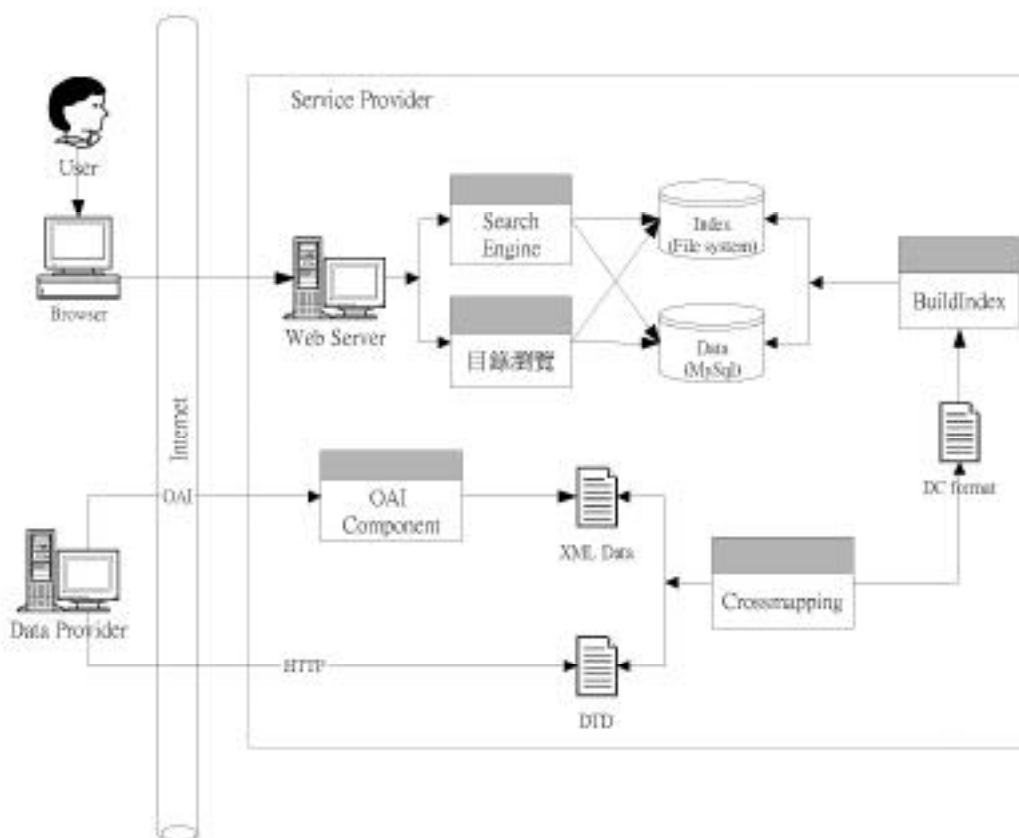
資料庫開放檢索。

## (二) 系統實作

### 1. Service Provider

目前以 OAI 為基礎的聯合目錄原型系統 (Prototype) 已大致完成，Service

Provider 的功能主要分成三個部分，即資料擷取、資料查詢、瀏覽目錄，系統架構如圖三所示：



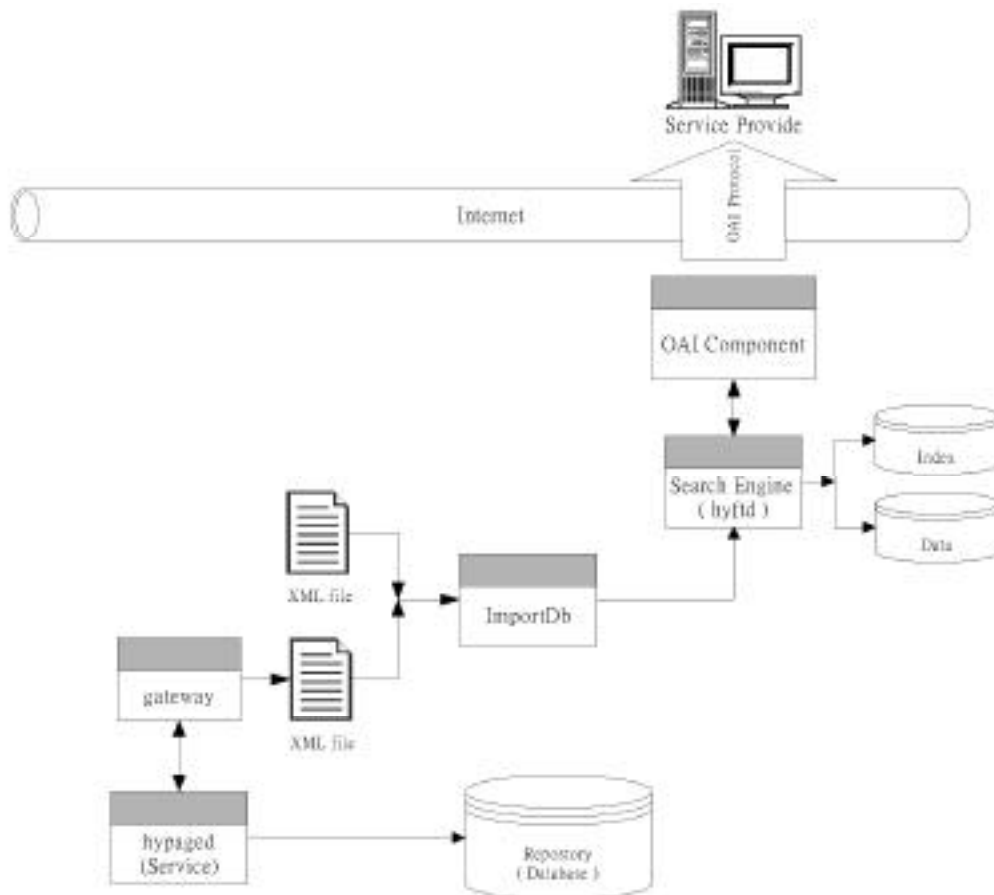
圖三：Service Provider 系統架構圖

此外，Service Provider 端的系統環境主要如下：

- (1) 作業系統：Linux RedHat 6.1, kernel 2.2.13 以上。
- (2) 硬體需求
  - CPU：Intel Pentium4 1.4GHz 以上
  - RAM：512 MB 以上
  - HD：系統程式約 2.5 MB，資料容量視情況而定，建議預留 20G MB 以上空間。
- (3) 軟體需求
  - JDK 1.3.1 版以上。

## 2.Data Provider

Data Provider 包含資料閘門子系統 (DB Gateway)、資料匯入子系統 (Data Import) 與 OAI 子系統 (OAI Component), 系統架構如圖四所示:



圖四：Data Provider 系統架構圖

### (1) 資料閘門子系統 (DB Gateway)

主要功能在將 repository 的資料從資料庫中轉出，並產生 XML 格式的文字檔。

資料閘門包含「資料處理程式」( gateway ) 與「資料庫連結程式」( hypaged )

- a. 資料庫連結程式：為一個服務程式 (service), 負責接收「資料處理程式」所提出的需求並連結資料庫，最後將相關資料傳回給「資

料處理程式」。

b. 資料處理程式：負責將資料庫的資料轉成xml 格式檔案。

目前資料閘門支援的資料庫包含：

Windows NT/2000 + ODBC, Linux + MySql, Linux + oracle8,

Sun Solaris + Informix, Sun Solaris + oracle8, Sun Solaris + Sybase。

實際應用時，採用不同的資料庫時只要使用相對應的「資料庫連結程式」( hypaged ) 即可。

#### (2) 資料匯入子系統 (Data Import)

資料匯入子系統處理程式為 ImportDb，主要功能包含將 xml 檔案透過「檢索引擎」儲存成索引檔與資料檔。

資料匯入子系統所讀取的xml 檔案來源有兩種：

a. 可由 DB Gateway 產生。

b. 由 Repository 系統程式或是由相關系統人員產生。

#### (3) 檢索引擎

檢索引擎為一個服務程式 ( service ) 主要負責索引資料的建立與查詢。

資料匯入時 ImportDb 會將 xml 資料傳送給 hyftd，hyftd 會將 xml 資料建立到 data 與 index 兩個子目錄中。

資料查詢時 OAI 子系統會將 OAI Protocol 的查詢參數傳給 hyftd，hyftd 則將查詢結果傳回給OAI 子系統。

#### (4) OAI 子系統 (OAI Component)

可接受 Service Provider 提出的 6 個 OAI protocol 並回應相關的資訊。

OAI 子系統係以 Java Servlets 所撰寫而成，因此必須具備 JDK 1.3.1 版以上，及 Tomcat 4.0.3 版以上。

#### (5) Data Provider 系統環境

作業系統：Microsoft Windows 2000 or Windows NT 4.0。

硬體需求

CPU：Intel PentiumIII 800 MHz 以上

RAM：256 MB 以上

HD：系統程式約 3.5 MB，資料容量視情況而定，建議預留 2G MB 以上空間。

軟體需求

JDK 1.3.1 版以上

Tomcat 4.0.3 版以上。

## 六、問題檢討

建置聯合目錄最大的困難不在系統的開發，而是在合作單位的共識。換言之，各合作單位是否願意加入聯合目錄的建置，而將 metadata 放在 Data Provider 上，讓 Service Provider 擷取；以及各合作館是否願意遵循 OAI 所要求的 metadata 格式，將 metadata 格式轉換為 Dublin Core，並以 XML 語法傳送資料。前者問題技術無法解決，但後者可以。為了解決有些合作單位可能一時無法提供 Dublin Core 格式的問題，我們在 Data Provider 及 Service Provider 端都有一支 Metadata mapping 程式，供合作館或 Service Provider 轉換 metadata 格式。而在系統環境方面，由於 Service Provider 是負擔使用者檢索的部分，所以採用 Linux 作業系統，而 Data Provider 則採用較普及的 Microsoft Windows 2000 or Windows NT 4.0 系統平臺，以便於各單位的安裝與維護。

為求最大的共同點，整合檢索所能提供的檢索點以及所顯示的 metadata，一定沒有各合作館本身的系統來得完整及專指 (specific)，此外，各館由於編目規則不一致，轉換為 Dublin Core 的原則也不一致，因此聯合目錄的品質也不可能一致，這都會使得聯合目錄的檢索效能不佳。

以 OAI 所建立的聯合目錄，其數位物件不會儲存在 Service Provider 端，而仍儲存在各典藏單位，但可透過 Service Provider 所查到的 metadata 連回原典藏系統並展現之，各典藏單位可以自行掌控要以那一種解析度的數位檔提供使用者觀看，以及數位檔是否要加浮水印，是否允許使用者下載，甚至是否要收費。所以，我們可以將聯合目錄視為一個行銷櫥窗，它不會取代原系統，不會構成原系統的威脅，所有的著作權都保留在原典藏單位，聯合目錄只是一個查尋入口而已，這對所有的典藏單位而言，應該是一個可以安心參與的合作模式。



# Interoperability of Heterogeneous Systems of the National Digital Archives Program : The Theory and Implementation of the OAI-based Union Catalog

**Chao-chen Chen**

Professor, Graduate Institute of Library and Information Science

National Taiwan Normal University

## 【 Abstract 】

On January 1, 2002 the National Science Council of Taiwan launched the National Digital Archives Program (NDAP). In order to share and conserve all valuable collections, retrieve the digital collections of the content holders via a union interface, and allow the general public to access the digital collections in Taiwan, the need to build a union catalog of National Digital Archives has been urgent. In this article, we define the functions and the system architecture of the OAI-based union catalog system and explain the problems we encounter in developing it.

**Keywords:** National Digital Archives Program, Interoperability, Heterogeneous System, OAI, Union Catalog