

行政院國家科學委員會專題研究計畫 成果報告

數位典藏互通性理論架構之研究(I)

計畫類別：個別型計畫

計畫編號：NSC91-2413-H-002-023-

執行期間：91年08月01日至92年07月31日

執行單位：國立臺灣大學圖書資訊學系暨研究所

計畫主持人：陳雪華

共同主持人：陳昭珍

計畫參與人員：協同主持人：項潔； 研究生：徐代昕、張懷文

報告類型：精簡報告

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中 華 民 國 92 年 11 月 11 日

目次

摘要.....	2
第一章 數位典藏互通性的內涵	4
第一節 前言	4
第二節 互通性的意義與類型	6
第三節 互通性模式	10
第二章 各類型互通協定介紹	12
第三章 OAI 詮釋資料擷取協定.....	19
第一節 OAI 詮釋資料擷取協定制訂緣起.....	19
第二節 OAI 詮釋資料擷取協定的內涵與規範.....	21
第三節 OAI 詮釋資料擷取協定之技術架構.....	23
第四節 OAI 詮釋資料擷取協定指令說明.....	25
第五節 OAI 詮釋資料擷取協定的優勢.....	32
第四章 以 OAI 互通性架構為基礎的相關應用.....	33
第一節 國外 OAI 相關計畫之發展.....	33
第二節 國內 OAI 相關計畫之發展.....	37
第五章 結論與建議	40
參考書目	42
附錄一： 出席國際學術會議心得報告（2002 亞洲數位圖書館國際會議 (International Conference of Asian Libraries(ICADL)2002)心得報告)	
附錄二： 出席國際學術會議發表之論文（Building an OAI-based Union Catalog for the National Digital Archives Program in Taiwan）	

數位典藏互通性理論架構之研究

The Interoperability Framework Theory of Digital Collections

摘要

目前世界各國均致力於數位典藏的發展，參與的單位與典藏的內容多元且豐富。欲使各典藏單位所建置的系統彼此分享、使用者從一個介面即可檢索到所有典藏機構的資料、以及讓民眾看到數位典藏的全貌，互通性機制的建立與達成為其關鍵。

過去廣泛用來解決數位博物館檢索互通性需求之機制首推 Z39.50，但由於 Z39.50 在互通性檢索的效能不盡理想，且難以建立彈性的瀏覽介面，造成數位圖書館互通性檢索機制難以有效普及。2001 年 1 月，開放性資料庫發展協會(Open Archives Initiative, OAI)發表了名為 OAI 詮釋資料擷取協定 (Open Archives Initiative Protocol for Metadata Harvesting, OAI-PMH)，提供互通性網路搜尋問題一項可行的解決方案。OAI 詮釋資料擷取協定運用目前高度發展的網際網路及詮釋資料兩種技術，在增強功能及簡化實行難度上，達成了極佳的平衡。發布到網路上的學術資料，藉助 OAI 詮釋資料擷取協定，不限於相異的系統平台、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，取得較諸先前更豐富且精確的資料。目前已有許多機構與研究單位著手進行 OAI 等互通性架構之研究與實際系統的建置與使用。本計畫將針對數位典藏之互通性理論架構、協定與發展進行研究與分析，並深入探討 OAI 互通性架構，提供數位典藏相關研究發展之參考。

關鍵詞：數位典藏、數位圖書館、互通性、OAI、OAI-PMH、詮釋資料

Abstract

The development of digital preservation has received much attention. In order to make all the digital information systems be shared, retrieve the digital collections of these content holders just via a union interface, and allow general public to access the digital collection, creating an interoperable framework is a key point.

At previous, the most common method to solve the problem of interoperable retrieval is Z39.50. However, the performance for using Z39.50 is not satisfying and it is difficult to build flexible browse interface. Thus, the mechanism of interoperable retrieval encounters constraints in making it popularity over the community of digital libraries. In January 2001, Open Archives Initiative (OAI) announced OAI Protocol for Metadata Harvesting to provide a practicable solution for interoperable retrieval over digital libraries. OAI Protocol for Metadata Harvesting applies two currently well-developed techniques, the Internet and Metadata, to make an excellent balance between the reinforcement of functionality and the simplification in implementing. Through OAI Protocol for Metadata Harvesting, the academic resources released on World Wide Web can circulate boundlessly over heterogeneous platforms, application programs, academic disciplines, countries and languages. In this way, users will be able to get more abundant and accurate information more efficiently. Currently, many academic organizations and universities have been undertaking the theoretical study and system design of OAI framework. This project intends to conduct deeper research and analysis on OAI theory for digital preservation.

Keywords: Digital Preservation · Digital Library · Interoperability · Open Archives Initiative · Open Archives Initiative Protocol for Metadata Harvesting · Metadata

第一章 數位典藏互通性的內涵

第一節 前言

目前世界各國均致力於數位典藏的發展，參與的單位與典藏的內容多元且豐富。數位圖書館可視為提供使用者存取和找尋數位典藏資源的一個重要工具，但由於網路分散的特質，數位圖書館多半為各自獨立的系統，各有其獨特的發行和檢索介面、系統結構、通訊協定和管理政策等，不能互相透過地溝通以及分享彼此的資源。¹欲使各典藏單位所建置的系統彼此分享、使用者從一個介面即可檢索到所有典藏機構的資料、以及讓民眾看到數位典藏的全貌，互通性(interoperability)機制的建立與達成實為關鍵。

互通性為發展數位圖書館的核心課題。然而，互通性指涉層面廣、觸及許多不同的面向，諸如技術、組織、內容、使用者等層面皆可討論之，與其相關的主題則有詮釋資料、資源檢索與發掘、服務架構等皆包含之。目前已有許多互通性相關的研究與發展，強調跨領域的整合，並期望能提供一般層級的資訊分享架構。

Z39.50 過去被廣泛用來解決圖書館自動化系統的整合檢索需求，然而在數位圖書館環境中，Z39.50 在互通性檢索的效能並不盡理想，且難以建立彈性的瀏覽介面，因此有其他相關之互通性協定正紛紛發展與制訂中，而目前能用於數位圖書館且最受矚目者當屬 OAI(Open Archives Initiative，簡稱 OAI)。²

OAI 的提出，乃為發展最小層級但具高度便利性的互通性架構，其在增強功能與簡化實行難度上，達成了極佳的平衡。在歐、美、及台灣地區，已經有許多機構與研究單位著手進行 OAI 等互通性架構之研究與實際系統的建置、使用。許多的結果亦顯示其具有易於建置與架構簡單的特性，且使用者能利用最經濟的時間，取得比先前更豐富且精確的資料。³

是故，本研究將針對數位典藏互通性作理論探討與內涵分析，介紹各類型相關的互通性協定，進而深入探討 OAI 互通性架構，並概述國內外數位典藏相關

¹余顯強，「淺談數位圖書館典藏資料互通之存取協定」，書藝 39(民國 92 年)。

²陳雪華、項潔、陳昭珍、洪筱盈、徐代昕、張懷文，「數位典藏互通性架構之探討」，中國圖書館學會會報 68(民國 91 年)：頁 1-13。

³同 2。

計畫應用 OAI 的情形，以作為理論探討的具體說明。藉由本研究的執行，期望能提供數位典藏在互通性議題上，相關研究發展的參考與建議。

第二節 互通性的意義與類型

自 1990 年起，隨著網路頻寬的日益增加，出現越來越多的電腦系統、大型資訊貯藏庫、各式應用程式、以及不可勝數的使用者之後，互通性議題開始受到高度的關注。然而，互通性卻缺乏一明確的定義。⁴

互通性普遍被視為系統或技術上的術語，最常被以系統觀點定義之。其認為互通性的問題存在於當系統希望獨立開發與運作各個組成元件，但又希望它們能方便、有效率的互相呼叫與使用彼此的功能來完成系統交付的任務時。⁵

在資訊技術及網路情境中，互通性亦有許多不同的詮釋。如 Bailey 提出所謂的互通係指能在不同的通訊基礎結構上，以個別之技術交換資訊/資料而不會降低整體的運作效能。互通性在網際網路技術層面屬相當重要的概念，且可促進網際網路成為新的資訊媒體平台。Moen 則認為互通性的核心概念在於網路環境中元件如何共同運作，其將互通性簡單地定義為不同元件共同運作以完成預定之過程或任務的能力。Miller 則提出比系統觀點涉及層面更廣的互通性定義，認為必須從複合的層面，包括技術、語義、組織及法律的概念加以闡釋，且強調互通性需要組織環節，特別為對資訊態度上的改變。其認為互通性乃為確保組織之系統、過程、及文化的管理，能夠擴大資訊交換與再使用的機會，而持續主動涉入的過程。歸納而言，互通性在資訊技術及網路情境中，乃用以概括欲達到透通取用資訊的目標所可能遭遇的議題與挑戰。

就數位典藏的觀點而言，由於早期的數位圖書館/博物館多半是各自獨立開發系統並採用不同的規範，但在應用上，卻往往需要跨系統地進行檢索、瀏覽、或其它目的資料交換(如詮釋資料的分享)。因此，當進行這類需要依賴互相合作才能完成的工作時，互通性的問題即浮上檯面。互通性的重要目標之一，即在

⁴ Moen, W. E., "Assessing Interoperability in the Networked Environment : Standards, Evaluation, and Testbeds in the Context of Z39.50" (September 2000)
<<http://www.unt.edu/wmoen/publications/InteropEvalPreprint.pdf>>(24 June 2003)

⁵徐代昕，「數位典藏互通性架構之建置」(國立臺灣大學資訊工程學研究所，碩士論文，民國 92 年)。

整合這些系統與規範，提供使用者一致性的服務。欲達此目標必須包含三個層面的互通需求：⁶

(一)技術層面(Technical agreements)：包含系統彼此之間交換訊息的格式、應用及溝通協定、安全系統等。

(二)內容/呈現層面(Content agreements)：包含內容範圍、語言、詮釋資料、命名方式、語義及使用者介面等。

(三)組織層面(Organizational agreements)：個人與組織的權利、責任，包括系統間有關資料存取、館藏的保存及服務、付費、認證等規則。

基於不同層面的互通性，美國國家科學基金會(National Science Foundation)將現有的開放式典藏資料互通的協定區分成三個型態：⁷

(一)聯盟(Federation)

聯盟模式可說是最傳統的互通性模式，乃指所有參與的系統均使用相同的協定。每一個互通與互動之間都有正式的定義，且每個組織之間均確實地遵循這些標準與規格建立彼此的服務。通常這種模式常會要求所有組織採用相同的電腦系統或軟體套件以便能在預定行程內實現互通性的目標。例如圖書館界使用Z39.50，以分享其線上圖書目錄。組成聯盟最主要的挑戰在於每個成員都必需使用，並且隨時跟上所有的現行規格，參與的成本高。

(二)聚合(Gathering)

聚合是一個最基本的互通模式，如果各種組織之間並未採取任何正式的互通協定，仍可以採用此種模式達到存取彼此間公開的資訊。最常見的就是Web上的搜尋引擎，因為不需花費任何成本即可聚合大量的數位圖書館來提供服務與分

⁶同5。

⁷同5。

享彼此間的資訊。不過，這些服務通常會比較貧乏且缺乏品質，除非有經過額外的加工處理或品質管制。

(三)擷取(Harvesting)

由於創造一個大型的聯盟有其困難性，因此發展出以較為鬆散的方式建立群組的構想。擷取背後的想法是參與者花費少量的努力，來建立某些基本的共通服務，而非像聯盟方式必需採用一整套的標準規格。OAI 的架構，就是環繞著詮釋資料擷取(metadata harvesting)而設計。每個參與單位都以簡單的詮釋資料交換格式--Dublin Core 來描述其典藏品，之後這些詮釋資料可以被服務提供者擷取，應用在它們的資訊服務中。擷取所能提供的服務通常不像聯盟那麼強有力，但是參與者的負擔卻大大地減低，故能吸引更多參與者，這正是 OAI 為何能這麼迅速為大眾所接受的原因。

綜合互通的定義與層面類型，數位典藏互通性概括多個主題：詮釋資料、檢索與發掘(search and discovery)、資源命名(resource naming)及服務架構(service architecture)等。影響互通性因素極其複雜且非單一面向，Moen 認為包括有：

(一)複合且不同的操作與檢索系統(Multiple and disparate operating and IR systems)

(二)複合的協定(Multiple protocols)

(三)複合的詮釋資料集(Multiple metadata schemes)

(四)複合的資料型式(Multiple data formats)

(五)複合的語言及字集(Multiple languages and character sets)

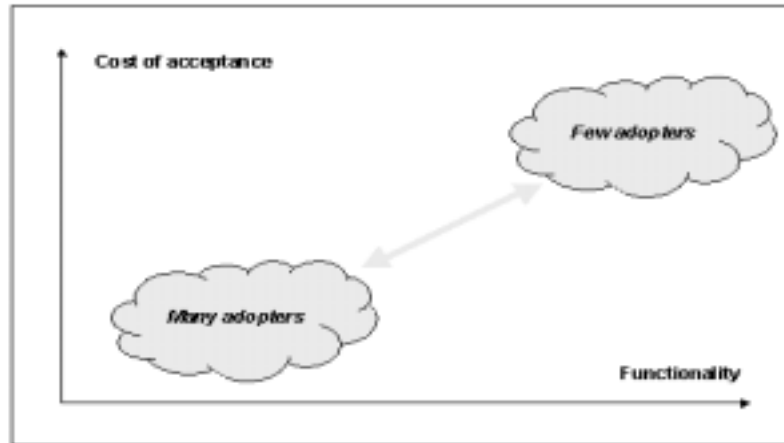
(六)複合的辭彙、學科(Multiple vocabularies, ontologies, and disciplines)

歸納而言，數位典藏互通性的達成可分為技術、內容及組織三個層面，而這三個層面皆與標準有關。數位典藏互通性涉及標準的創造與採用，包含各種類型的詮釋資料標準，如管理、識別，讓詮釋資料於外部系統呈現的方式，如擷取(harvesting)，支持互通性的開放架構，涵括如使用者介面、貯藏庫、索引或管理系統(indexing or handle system)及檢索系統等。而在動態的網路環境中，必

須持續發展互通性的評估策略及方法；達成互通性的實效層面亦需被強調，即解決互通性的成本問題，必須了解不同使用者社群所可接受的互通程度，才能提出促進互通之不同成本的方法。

第三節 互通性模式

互通性涉及多個層面，欲為每個層面制定出通用的協議極為不易，而且如何創造出使各個組織都願意採納的動機，會是更艱鉅的挑戰。Saratos Kapidakis 在 1998 年提出以成本與功能比較的方式，來分析互通性。根據此建議，Arms 於 1999 提出圖一之模式：⁸



圖一：Cost of adoption against functionality

傳統的互通方法為所有的參與者皆遵循同樣的標準；然而，過去的經驗也顯示此法不易達成。雖然採用相同的標準能夠提供數位圖書館較具價值的功能，但也必須付出相當的成本。其中，有些成本直接與經費相關；但絕大部份的成本來自於組織層面的問題。採用一個新標準必須要更改與其相關的既有系統、修正工作流程、改變與供應者間的關係等。上圖一即顯示了組織在接受標準時所做的權衡。縱軸代表組織接受新標準的成本，橫軸表示組織所能獲致的功能。若採納標準的成本極高，將只被高度重視功能的組織所接受；相反地，當成本降低，即使功能有限，多數的組織也願意採用。

而隨著應用領域、設計目標、設計前提的不同，評估一個互通性模式是否成功，或是比較兩個互通性模式的優劣，極其困難。應列入評估的要素可有以下幾點：⁹

⁸ Arms, W. Y., *Digital Libraries*. (Cambridge, Ma.: MIT Press, 2000)

⁹ 柯皓仁、黃夙賢、楊維邦，「詮釋資料與數位圖書館系統互通性之探討」，*大學圖書館* 5(1)(民國 90 年)：頁 49-78。

- (一)元件自主性(Component autonomy)：儘量不需遵循共同的規範來設計數位圖書館的系統或元件。
- (二)建置基礎建設的代價(Cost of infrastructure)：儘量不需花費龐大的成本來建置支援元件互通的基礎建設。
- (三)元件的加入(Contributing components)：要很容易地讓新的元件加入互通的行列。
- (四)元件的使用(Using Components)：元件的使用越簡單越好。需考慮服務元件及元件互動的成本與複雜度。
- (五)元件功能的複雜度和廣泛度(Breadth of task Complexity supported by the solution)：元件所能達成的功能儘量多樣化。
- (六)元件數目的成長空間(Scalability in the number of components)：所能支援參與互通的元件數量越多越好。

第二章 各類型互通協定介紹

為達成互通性的目的，不斷有許多協定標準被開發出來。下面茲簡列一些數位典藏領域常用的協定標準，視不同的功能需求而各有其優缺點。

一、DIENST

DIENST 是由 DARPA(Defense Advanced Research Projects Agency) 與 NSF(National Science Foundation)贊助 DLI(Digital Libraries Initiative)所發展的計畫。DIENST 事實上是 OAI-PMH 的前身，其做法也是利用詮釋資料擷取(metadata harvesting)來分享數位化檔案。但是在求功能強大的同時，也使得整個通訊協定複雜化，因此不得不促使其精簡化的版本--OAI-PMH 的產生。由於兩者承先啟後的關係，事實上可以透過 DOG(Dienst/OAI-PMH Gateway)做為 DIENST 與 OAI-PMH 伺服器間的相互溝通。¹⁰

二、NCSTRL, NCSTRL+

NCSTRL+是由包括美國太空總署 NASA Langley Research Center 專門負責科學與技術資訊(Scientific and Technical Information, STI)計劃所發展的 Networked Computer Science Technical Reference library(NCSTRL)分散式架構所擴充的工具集，最初計劃的目標是希望建立一個簡化出版與存取的使用模式，以使用來整合 NASA STI 館藏中各種不同的學科和資料格式。NCSTRL 結合了超過 100 以上的機構，彼此分享資訊科學相關的技術資料。NCSTRL 達成館際間資料互通的主要協定為 Dienst。Dienst 包含了：儲存體、索引、Meta、使用者介面、圖書館管理等五個服務組件，並採用 HTTP 作為資料傳輸的協定。所以 NCSTRL+ 可以說是應用 Dienst 協定之 NCSTRL 的擴充版本。NCSTRL+包含兩個主要的技術：Clusters 和 Buckets。Clusters 允許依據主題目錄、典藏檔案型態、出版、組織等條件，任意地結合或分割典藏資料。而 Buckets 則是用於聚集數位圖書館線上出版的智慧型代理人(agent)。Buckets 不僅能夠獨立運作且能負責處理典藏資

¹⁰ 同 5。

料的管理、與遠端協調顯示的內容等等。

NCSTRL 和 NCSTRL+可以說是現今美國大學院校內資訊科學報告線上出版最主要的系統之一，而且也提供了數位圖書館間的互通功能。不過由於協定規範的範圍過大，所以如果有一個新的數位圖書館欲加入此一聯盟，必須修改採取 Dienst 協定，並安裝相關的程式庫，以便能與聯盟的服務提供者作業協同一致。不過當軟體的版本修定時，所有加入聯盟的數位圖書館系統必須全部配合修正，否則便無法正常運作。¹¹

三、NESSTAR

就整體而言，Networked Social Science Tools and Resources (NESSTAR)是一組藉由 Internet 傳播資料的基礎建設，其核心結構為應用 Data Documentation Initiative (DDI)的 XML 版本 metadata 系統。NESSTAR 是由 Norwegian Social Science Data Services (NSD)、UK Data Archive (UKDA)和 Danish Data Archive (DDA)聯合發展的軟體系統，並依據此系統的應用目標而設計的協定。其希望能達成的目標包括下列四點：

- (1)能找出在不同組織與國家之間各種資料的原始來源
- (2)瀏覽這些資料的細節，包括記述資料與原件實體。
- (3)能夠因應不同使用者的專業程度，快速且容易的列表與呈現資訊內容
- (4)提供資料和文件的選粹服務，包括完整或部分的資訊，甚至製作成適當的表格以方便使用者應用。

NESSTAR 整體系統架構包括下列三個部分：

- (1)Explorer：提供圖形式的使用者介面。
- (2)Publisher：各資料提供者處理可分享之典藏資料的伺服器。
- (3)Protocol：負責在其間傳遞以 XML 為標示的 metadata 訊息之交換規範。

¹¹ 同 1。

因此其主要實現數位圖書館之間互通性的協定也稱之為 NESSTAR。不過 NESSTAR 處理的資料種類過於簡單，因此有後續 FASTER(Flexible Access to Statistics, Tables and Electronic Resources)的擴充計劃，希望能改善資料處理的深度與範圍。¹²

四、Metadata search engine

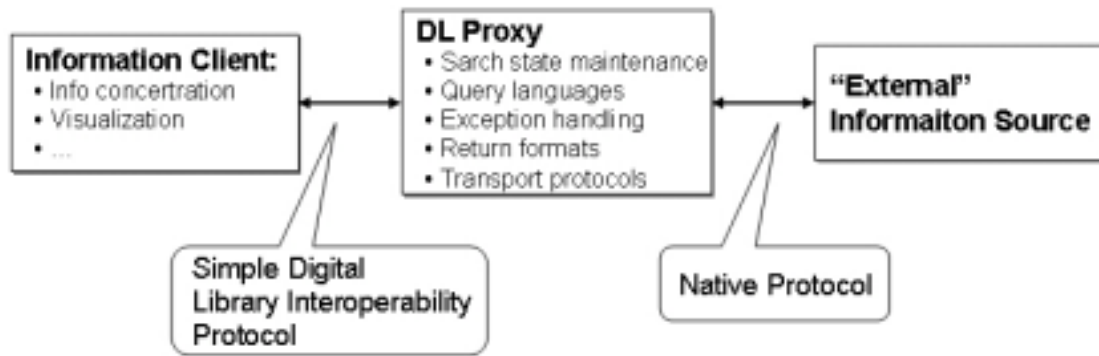
嚴格來說，如 Google, JumpStation II, Northern Light, MetaCrawler, AltaVista Search, EasyASK, Inktomi Enterprise Search, Mercado, Teapot, HotMeta 等 Metadata 搜尋引擎，並不是為了數位圖書館互通性而發展出來的檢索工具與協定，因為數位圖書館主要是強調服務的品質，而搜尋引擎則是著重再搜尋結果的品質。也就是說，搜尋引擎的目標通常是效率高於服務。但是 Meta 搜尋引擎是 Internet 環境裡取得 metadata 最基本的模式，而且 Metadata 搜尋引擎和一般的網頁搜尋引擎所使用的技術與功能類似。一般網頁由於 HTML 屬於程序性標示語言(Procedural Markup Language)，無法提供自動化分析文件的語意內涵，而 metadata 本身屬於描述性標示語言(Descriptive Markup Language)，因此能夠很方便地藉由標籤(tag)自動化分析資料內的結構與內涵。所以對於數位圖書館可以隨時藉由各種 Metadata 搜尋引擎，檢索與獲取其他 DL 的 metadata，達成典藏資料分享的互通需求。¹³

五、SDLIP

簡易數位圖書館互通協定(Simple Digital Library Interoperability Protocol，SDLIP)是由史丹佛大學所發展一個定位為檢索的中介軟體(middleware)。如圖二所示，在 SDLIP 架構中，資訊前端與後端的資料來源之間是透過數位圖書館 proxy 來達成互通檢索的需求。介於前端與 proxy 之間，SDLIP 定義了包括傳輸協定、查詢語言以及存取介面等。

¹²同 1。

¹³同 1。



圖二：SDLIP 系統架構圖

不過 SDLIP 使用 proxy 的方式，允許各數位圖書館採用各自專屬的協定方式，對數位圖書館而言也有下列三項缺點：

(1) proxy 與各資料來源之間的訊息處理，允許採用各自專屬的協定。採取此種方式主要是為了考量各數位圖書館實作方便，但是當要新增或註冊一個數位圖書館時，proxy 必須針對該數位圖書館所使用協定的特性重新修改程式，因此反而導致整體效率不佳。

(2) 正因為各資料來源端允許使用既有的協定，因此加入 SDLIP 時必須在前端額外開發處理 SDLIP 的程式以便和 proxy 溝通，無法提供公用的程式庫給各數位圖書館共同使用。

(3) 雖然 SDLIP 提供了整合性的檢索介面，不過使用著仍需逐一送出檢索的需求，無法透過單一檢索步驟查詢所有聯盟的數位圖書館。¹⁴

¹⁴同 1。

六、GINF

通用互通框架(Generic Interoperability Framework, GINF)主要的目的是希望能達成各數位圖書館之間存取資料的協定、程式語言、資料模式、格式都能彼此獨立，並且能夠動態的發掘元件(component)之間的功能，以及統一各元件之間的介面。因此，GINF 可以說是一種簡化異質系統之間執行元件的整合協定，而其最大貢獻便是各數位圖書館系統能保有各自獨立的協定、程式語言、資料和介面描述。GINF 是基於 RDF 所實際運作的框架，透過此框架來作為語意導向的中介軟體，系統便可藉由此中介軟體開發出高度彈性的主從(client/server)應用模式。¹⁵

七、OAI-PMH(Open Archives Initiative Protocol for Metadata Harvesting)

OAI-PMH 是一個簡易且具延伸性的輕量化(lightweight)通訊協定。它在設計上是利用 HTTP/XML 來當作底層的通訊協定。其優點是不仰賴任何特殊的資訊元件技術或是程式語言。OAI-PMH 的主要目標，是藉由資料提供者分享詮釋資料，而能有效率的傳遞資訊給服務提供者，提供資訊服務。¹⁶

八、OpenURL

OpenURL 為一種分散查詢語法，藉由一組已經定義好的標籤(tag)，以增進 Web 超連結的能力。組成元件有資料提供者(source)與服務提供者(target)，服務提供者只要遵照此機制，就可以輕易解析資料提供者所傳送的要求。而資料提供者也可經由此規範，輕易對服務提供者送出深度連結服務要求。¹⁷

九、SOAP(Simple Object Access Protocol)

SOAP 最早是由微軟、DevelopMentor、UserLand 所共同提交 IETF (Internet

¹⁵同 1。

¹⁶同 5。

¹⁷國家圖書館,「數位圖書館分散檢索協定」(民國 90 年 10 月)<http://www.ncl.edu.tw/bbs/pdf/6-3-3.pdf>

Engineering Task Force) 的標準。與 OAI 類似的是它同樣是一個輕量化的通訊協定，也同樣以 HTTP/XML 的組合做為資料載具。不同的是 SOAP 並沒有像 OAI-PMH 一樣有預先制定的六個命令，而是讓傳遞資料的雙方自行再定義其傳輸的 XML 如何解讀。其優點是能使用的範圍更廣泛，但常需與其他通訊協定合併使用。由於 SOAP 目前是由 W3C 組織所制定發展，並且將會成為 W3C 的標準。因此若有需要，OAI-PMH 也可以開發以 SOAP 為通訊載具的版本。¹⁸

十、Z39.50

Z39.50 已經是圖書館系統使用很久的開放式檢索協定，主要是由「原始系統」(或稱 Z39.50 Client) 依靠線上即時連結一個以上的「目標系統」(或稱 Z39.50 Server)，使用複雜的通訊協定集執行所需的查詢行為，再逐一取得由目標系統獲得的查詢結果、排序、去除重覆，最後將處理結果呈現出來。應用在圖書館界的 Z39.50 協定是最普遍的聯盟式典藏資料互動協定之一。而 Z39.50 以 MARC 和 Dublin Core 為資料處理的依據，著重在點對點(peer to peer)的協定服務模式，透過線上即時查詢目標系統的資訊，然後處理所有遠端回應的資料集之後，再呈現給前端使用者。強調的是即時性的聯盟檢索服務，中間不存在任何中介者的角色，前端也不強調資料加值的處理與分析。¹⁹

¹⁸ 同 5。

¹⁹ 同 1。

十一、各類型互通協定綜合比較

總結前述之各類型開放式典藏資料互通協定，可整理比較如下表所示：²⁰

	NCSTRL, NCSTRL +	OAI-P MH	Meta web search engine	SDLIP	GINF	Search Light	Z39.50
互通型態	聯盟式	獲取式	聚合式	聚合式	聚合式	聚合式	聯盟式
提供多個 數位圖書 館之間同 步互通	可	依服務提 供者功能 而定	可	不可	不可	可	可
提供非同 步檢索	不可	依服務提 供者功能 而定	不可	可	不可	不可	可
資料提供 模式	主動式	被動	被動	被動	被動	被動	被動
包含檢索 協定	無	無	有	有	有	有	有
傳輸協定	HTTP	HTTP	HTTP	TCP, HTTP, CORBA 物件	TCP, HTTP	HTTP	不限定
Metadata 格式	任何 metadata 均 可	XML 格 式的 metadata	無	任何 metadata 均可	RDF	無	MARC, DC
建置成本	資料提供 者：高 服務提供 者：低	資料提供 者：低 服務提供 者：(中 等)，視服 務內容而 定	資料提供 者：無 服務提供 者：高	資料提供 者：無 服務提供 者：高	資料提 供者： 高 服務提 供者： 高	資料提供 者：無 服務提供 者：高	資料提供 者：高 服務提供 者：高

²⁰同 1。

第三章 OAI 詮釋資料擷取協定

第一節 OAI 詮釋資料擷取協定制訂緣起

有鑑於數位圖書館或學術機構的資料庫與系統彼此互不隸屬，相關資料或不同領域的資料不僅分散儲存而且難以整合，使得資料的分享與流通有所限制，Paul Ginsparg, Rick Luce, Herbert Van de Sompel 等人在 1999 年 10 月於 New Mexico 的 Santa Fe 所舉行的 Universal Preprint Service 會議中推動開放性典藏庫發展協會(Open Archives Initiative，簡稱 OAI)的成立，由研究者、圖書館員、出版商、檔案學家等所組成，為一強調互通性議題的組織。組織目標在於提昇學術社群間具體的溝通方式，藉由技術定義及組織的支持兩層面，發展最小層級但具高度便利性的互通性架構。在技術層面將可透過各種終端使用者服務，例如檢索引擎等，以獲取資料；組織層面則定義資料提供者(data providers)與服務提供者(service providers)，規範提供者、資料、服務三者間的溝通機制。OAI 組織的最初目的是針對期刊論文與預刊本之電子資源的互通性檢索而成，但這與各類型數位圖書館所遭遇的互通性問題相當類似，所以在 2000 年上半年，OAI 更將其適用範圍擴展至數位圖書館領域，由 DLF 與 Andrew W. Mellon Foundation 兩個組織在 Harvard University 所舉行的 Cambridge Meeting，討論如何將數位圖書館之館藏資訊散播到網路搜尋引擎上。會中一致認為利用互通的模式在網際網路交換詮釋資料，將是數位圖書館資料散佈至網路上的重要方式。²¹

OAI 認為達成互通性的第一步，需發展詮釋資料擷取協定(metadata harvesting protocol)以利於詮釋資料在資料庫間交換，因此於 2001 年 1 月發表 OAI 詮釋資料擷取協定(Open Archives Initiative Protocol for Metadata Harvesting，簡稱 OAI-PMH)，提供數位圖書館檢索互通機制上的一可行方案。OAI-PMH 運用目前高度發展的網際網路與詮釋資料兩種技術，在增強功能、簡化實行難度上，達成了極佳的平衡。發佈到網路上的學術資料，藉助 OAI-PMH

21 Lagoze, C., Van de Sompel, H., "The Open Archives Initiative : Building a Low-Barrier Interoperability Framework" (2001) <http://www.openarchives.org/documents/oai.pdf> (10 January 2002)

能不限於相異的系統平台、應用程式、領域、國界及語言，達到廣泛流通的目的。也因此，使用者能利用最經濟的時間，取得較諸先前更豐富且精確的資料。OAI-PMH 已向 W3C 組織提出申請，目前 OAI-PMH 的最新版本為 2002 年 6 月所提出的 2.0 版，預計將成為開放性的全球標準。

第二節 OAI 詮釋資料擷取協定的內涵與規範

OAI-PMH 協定的內涵主要包含詮釋資料的發佈(expose)與擷取(harvest)兩個類型，其在協定上分別定義此兩部分：

- (一)定義一個資料提供者能夠透過 HTTP 為基礎的協定，發佈其詮釋資料(metadata)的機制。
- (二)定義一個能夠從儲存器(Repository)獲取含有詮釋資料(metadata)資料錄(record)的機制。

OAI-PMH 提供選擇性擷取(selective harvesting)，讓擷取程式可以將擷取的範圍限制在資料庫的部分子集合，過濾掉其他不需要的部分。為了避免太過複雜的實作細節，並提供足夠的功能性，OAI-PMH 選擇以兩個較為簡單的條件作為選擇性擷取實的參數：時間戳記與集合。選擇性擷取使用在 ListRecords 和 ListIdentifiers 命令中，時間戳記與集合兩個參數可以獨立或合併使用。²²

(一)以時間戳記進行選擇性擷取

在 OAI-PMH 規範中規定每筆紀錄都必須包含時間戳記，用來表示資料物件最近一次新增、刪除、或修改的時間。擷取時可以由 from 和 until 決定完整的時間範圍，或是只使用 from 或 until 決定時間範圍的起點或終點。為了在擷取時，能同步世界各地不同時區的資料庫，時間的格式必須符合 UTCdatetime 格式(coordinated universal time)。

(二)以資料集進行選擇性擷取

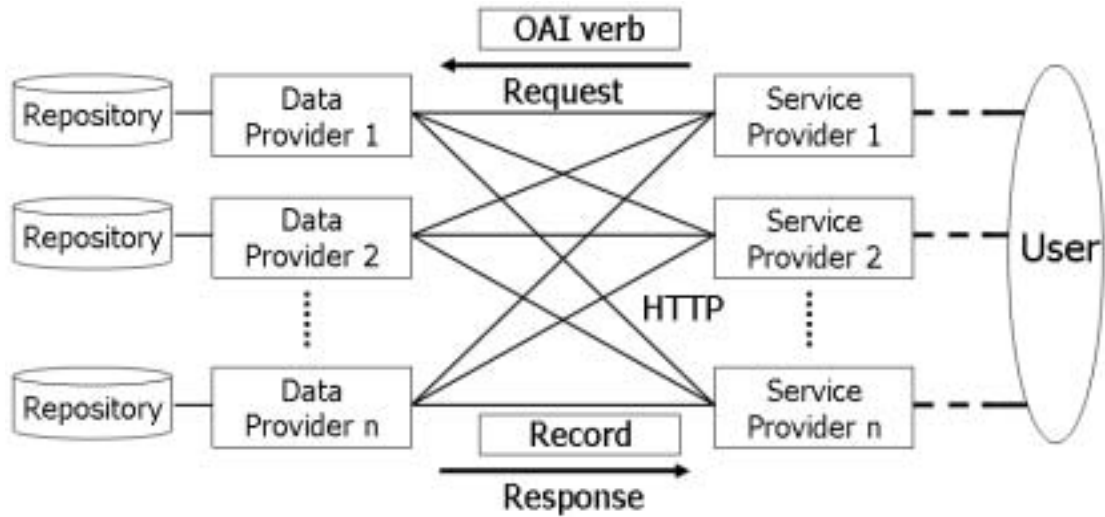
資料集(set)在 OAI-PMH 規範中只列為一個可選擇實作與否的條件。其功能在於將資料庫中的物件歸類為一個個子集合，以協助選擇性擷取指定其選擇的範圍。資料庫可以將其集合訂立為階層式的架構，即集合下還可以再分為若干子集

²² 同 5。

合，再往下依此類推。但特別要注意的是，一個資料物件可以被歸類在一個集合、多個集合中，或是完全不出現在任何集合中。每一個集合以其 setSpec 作為識別的名稱。

第三節 OAI 詮釋資料擷取協定之技術架構

OAI-PMH 整體的運作如圖二所示，核心主要是在 HTTP 協定上傳輸使用 XML 文件的協定，而前後端整個運作環境，包含下列五個主要的組成元件：



圖三：OAI-PMH 技術架構組成元件

(一) 資料提供者(data provider)

提供其文件內容，並以 OAI-PMH 作為發佈詮釋資料的協定。主要工作為維護一個或一個以上支援 OAI-PMH 來將其內容以詮釋資料發佈的儲存器(Web 伺服器)。

(二) 服務提供者(service provider)

透過 OAI-PMH 協定向資料提供者取得資料，並利用獲得的詮釋資料建立各種增值服務。

(三) 資料儲存器(Repository)

透過 HTTP，接受 OAI-PMH 所提出存取資料需求的伺服器。

(四)資料集(Set)

非必備功能，主要是為了方便取得部分範圍所需的資料。儲存器內可將不同類別的資料區分成不同的群組，並以階層式架構表示，以節點(node)作為各分類的區分，因此每一個節點即稱之為資料集。

(五)資料錄(Record)

一個資料錄是後端伺服器依據 OAI-PMH，從儲存器內將資料以 XML 編碼傳回前端的詮釋資料。

第四節 OAI 詮釋資料擷取協定指令說明

以下列舉 OAI-PMH 的命令(verb)，包括命令使用方式、所有參數、及錯誤處理，並舉例說明。其中 Identify、ListMetadataFormats、ListSets 屬於支援命令，負責提供與資料庫相關的資訊；ListRecords、ListIdentifiers、GetRecord 則屬於擷取命令，負責取回所指定的詮釋資料。

參數中若指明為必需(required)者，則一定要包含在要求(request)之中(除非要求中使用了具唯一性的參數)。若為非必需(optional)，則視要求的情況是否需要而選用。若為唯一的(exclusive)，則表示這項參數不能與其他參數合用。意即可以不使用，但若使用了，則它會是唯一使用的參數，不會同時出現其他參數。

(一)GetRecord

1.使用方式：當欲從資料庫(repository)中，單獨取得一筆特定的資料紀錄時使用。使用時必須指定其資料物件的識別碼(identifier)，及所要的詮釋資料格式。若資料庫能夠追蹤刪除資料，則在回應時，對於已刪除的資料可以在 header 欄加入 status 參數，其值為“deleted”。

2.可用參數：

- (1)identifier：必需，用以指示一個資料物件(item)在資料庫中的唯一識別碼。
- (2)metadataPrefix：必需，用以指示回應時使用何種詮釋資料格式。資料庫能提供的所有詮釋資料格式，可以藉 ListMetadataFormats 命令取得。

3.錯誤狀況：

- (1)badArgument：使用了不合法的參數，或缺失了必需的參數。
- (2)cannotDisseminateFormat：由 identifier 所指定的資料物件並不支援由 metadataPrefix 所指定的詮釋資料格式。
- (3)idDoesNotExist：指定的 identifier 不合法，或在資料庫中根本不存在。

(二)Identify

1.使用方式：用以取得資料庫的各項基本資訊。資料庫也可以將一些敘述性的資訊放在這個命令的回應裡。

2.可用參數：無。

3.錯誤狀況：

(1)badArgument：使用了不合法的參數。

(三)ListIdentifiers

1.使用方式：這個命令可以看作 ListRecords 的縮減版，與 ListRecords 的差別在於它只取回每一筆資料紀錄的標頭(header)部分，而不是完整的紀錄(一筆完整的紀錄還必須包括詮釋資料部分，以及非必需的相關(about)部分)。可以利用資料集(set)與時間戳記(datestamp)兩個參數，作選擇性擷取(selective harvesting)。若資料庫能夠追蹤刪除資料，則在回應時可以在標頭(header)部分加入 status 參數，並設定其值為“deleted”，來表示該筆資料已刪除。

2.可用參數：

(1)from：非必需，其值必須符合 UTC(Coordinated Universal Time)的標準格式。當我們依某一段日期時間範圍進行選擇性擷取時，用以指示開始的時間點。

(2)until：非必需，其值必須符合 UTC(Coordinated Universal Time)的標準格式。當依某一段日期時間範圍進行選擇性擷取時，用以指示終止的時間點。

(3)metadataPrefix：必需，用以指示回應時只取回詮釋資料格式符合 metadataPrefix 的標頭(若資料庫能夠追蹤刪除資料，則已刪除的資料標頭也會被取回)。資料庫能提供的所有詮釋資料格式，可以藉 ListMetadataFormats 命令取得。

(4)set：非必需，當進行選擇性擷取，想將範圍限制在某一個資料集內時使用。

(5)resumptionToken：唯一，其值為前一個 ListIdentifiers 命令未傳完全部的資料時，所傳回來的一個流量控制用的標記。在要求中放入該傳回來的 resumptionToken，則資料提供者會延續之前未完成的 ListIdentifiers 命令，繼續傳送接下去的資料。

3.錯誤狀況：

- (1)badArgument：使用了不合法的參數，或缺失了必需的參數。
- (2)badResumptionToken：所給予的 resumptionToken 是無效的，或是已經過期。
- (3)cannotDisseminateFormat：資料庫不支援要求中 metadataPrefix 所指定的詮釋資料格式。
- (4)noRecordsMatch：由 from，until，set 三個參數所構成的選取範圍裡沒有任何一筆紀錄。
- (5)noSetHierarchy：資料庫未支援資料集。

(四)ListMetadataFormats

1.使用方式：這個命令的用途是取得資料庫所有可用的詮釋資料格式。可以利用非必需的 identifier 參數，將傳回的結果限制為資料庫中某一個資料物件所有可用的詮釋資料格式。

2.可用參數：

(1)identifier：非必需，其值應該要是資料庫中某個資料物件的唯一識別碼，並且將傳回的結果限制為該資料物件所有可用的詮釋資料格式。當省略這個參數時，傳回的結果則是資料庫所有可用的詮釋資料格式。特別要注意的是：資料庫支援某種詮釋資料格式，並不表示資料庫裡的所有資料物件都能使用該種詮釋資料格式。

3.錯誤狀況：

- (1)badArgument：這個要求包含了不合法的參數，或缺失了必需的參數。
- (2)idDoesNotExist：指定的 identifier 不合法，或在資料庫中根本不存在。

(3)noMetadataFormats: identifier 所指定的資料物件目前並沒有可以使用的詮釋資料格式。

(五)ListRecords

1.使用方式：此命令用於從資料庫擷取大批資料紀錄。它提供一些非必需的參數，協助我們運用資料集、時間戳記，進行選擇性擷取。若資料庫能夠追蹤刪除資料，則在回應時可以在標頭(header)部份加入 status 參數，並設定其值為“deleted”，來表示該筆資料已刪除。已刪除的資料物件，其紀錄中不會包含有詮釋資料(metadata)部份。

2.可用參數：

(1)from：非必需，其值必須符合 UTC(Coordinated Universal Time)的標準格式。當我們依某一段日期時間範圍進行選擇性擷取時，用以指示開始的時間點。

(2)until：非必需，其值必須符合 UTC(Coordinated Universal Time)的標準格式。當我們依某一段日期時間範圍進行選擇性擷取時，用以指示終止的時間點。

(3)set：非必需，當進行選擇性擷取，想將範圍限制在某一個資料集內時使用。

(4)resumptionToken：唯一，其值為前一個 ListRecords 命令未傳完全部的資料紀錄時，所傳回來的一個流量控制用的標記。在要求中放入該傳回來的 resumptionToken，則資料提供者應該要延續之前未完成的 ListRecords 命令，繼續傳送接下去的資料紀錄。

(5)metadataPrefix：必需，用以指示回應的資料紀錄中，詮釋資料的部份應該要使用的詮釋資料格式，並且每筆紀錄只有能夠照 metadataPrefix 所指定之格式發佈詮釋資料，才會被取回。資料庫能提供的所有詮釋資料格式，可以藉 ListMetadataFormats 命令取得。

3.錯誤狀況：

(1)badArgument：這個要求包含了不合法的參數，或缺失了必需的參數。

- (2)badResumptionToken : 所給予的 resumptionToken 是無效的，或是已經過期。
- (3)cannotDisseminateFormat : 資料庫不支援要求中 metadataPrefix 所指定的詮釋資料格式。
- (4)noRecordsMatch : 由 from, until, set 三個參數所構成的選取範圍裡沒有任何一筆紀錄。
- (5)noSetHierarchy : 資料庫未支援資料集。

(六)ListSets

1.使用方式：這個命令的用途，在於取得資料庫內，資料集的階層架構，以協助選擇性擷取之處理。

2.可用參數：

(1)resumptionToken：唯一，其值為前一個 ListSets 命令未傳完全部的資料集架構時，所傳回來的一個流量控制用的標記。在要求中放入該傳回來的 resumptionToken，則資料提供者應該要延續之前未完成的 ListSets 命令，繼續傳送接下去的資料集架構。

3.錯誤狀況：

- (1)badArgument：這個要求包含了不合法的參數。
- (2)badResumptionToken : 所給予的 resumptionToken 是無效的，或是已經過期。
- (3)noSetHierarchy : 資料庫未支援資料集之使用。

(七)實例

下面的命令使用 OAI-PMH 來要求單一筆資料的詮釋資料，資料的識別碼為 oai:arXiv:cs/0112017，並要求詮釋資料必需以 Dublin Core 來呈現。

Request :

```
http://arXiv.org/oai2?verb=GetRecord&identifier=oai:arXiv:cs/0112017&metadataPrefix=oai_dc
```

```
<?xml version="1.0" encoding="UTF-8"?>

<OAI-PMH xmlns=http://www.openarchives.org/OAI/2.0/
  xmlns:xsi=http://www.w3.org/2001/XMLSchema-instance
  xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/
  http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
  <responseDate>2002-02-08T08:55:46Z</responseDate>
  <request verb="GetRecord" identifier="oai:arXiv:cs/0112017"
    metadataPrefix="oai_dc">http://arXiv.org/oai2</request>
  <GetRecord>
  <record>
  <header>
    <identifier>oai:arXiv:cs/0112017</identifier>
    <datestamp>2001-12-14</datestamp>
    <setSpec>cs</setSpec>
    <setSpec>math</setSpec>
  </header>
  <metadata>
    <oai_dc:dc
      xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/"
      xmlns:dc="http://purl.org/dc/elements/1.1/"
      xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
      xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
      http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
      <dc:title>Using Structural Metadata to Localize Experience
        of Digital Content</dc:title>
      <dc:creator>Dushay, Naomi</dc:creator>
      <dc:subject>Digital Libraries</dc:subject>
      <dc:description>With the increasing technical sophistication
        of both information consumers and providers, there is
        increasing demand for more meaningful experiences of
        digital information. We present a framework that separates
        digital object experience, or rendering, from digital object
        storage and manipulation, so the rendering can be tailored to
```

```
particular communities of users.  
</dc:description>  
<dc:description>Comment: 23 pages including 2 appendices,  
8 figures</dc:description>  
<dc:date>2001-12-14</dc:date>  
</oai_dc:dc>  
</metadata>  
</record>  
</GetRecord>  
</OAI-PMH>
```


第五節 OAI 詮釋資料擷取協定的優勢

OAI-PMH 之互通性架構規範具有許多優點，分述如下：²³

(一)提供學術溝通及交流的新模式

OAI-PMH 之架構使數位化文件能更容易、更廣泛的傳播。且採用詮釋資料蒐集 (metadata harvesting) 的方式，可涵蓋各種多媒體格式、資料型態、與內容等，擴展了數位化資料存取種類的範圍。

(二)實作容易

OAI-PMH 在設計時即以「簡單」為原則。一個具架設網路伺服器經驗的工程師，可以在極短的時間內架設起 OAI 伺服器。

(三)具開放性 (open)

任何人都能使用 OAI-PMH 定義的架構，來建構資料提供或服務提供的伺服器。

(四)採用 HTTP 及 XML 之開放性標準

OAI-PMH 利用 HTTP 通訊協定作為其基本的通訊協定。其優點在於：現今所有的網頁伺服器及瀏覽程式等，幾乎毫無例外的支援 HTTP；這使得 OAI 在先天上就已解決了跨平台及相容性等問題，也節省了另行架構的困難。另一方面，XML 亦漸漸成為全球共同的資料標準交換格式。由於 HTTP 及 XML 均為開放性的標準，採用 HTTP 及 XML 的組合不僅考慮了相容性的問題，也確保了 OAI-PMH 的開放性原則。

²³ 同 2。

第四章 以 OAI 互通性架構為基礎的相關應用

在歐美等地，已有許多機構與研究單位著手進行 OAI 等互通性架構之研究與實際系統的建置與使用。就目前發展的結果顯示 OAI-PMH 之規範的確具有易於建置及架構簡單的特性。各式的服務不斷地在發展，在 OAI 組織網站 (<http://www.openarchives.org>) 上，截至 2003 年 7 月已有 102 個資料提供者註冊，提供貯藏庫的詮釋資料檢索，如 OCLC(Online Computer Library Center)、美國國會圖書館(Library of Congress)，與大學、研究計畫都名列其中；服務提供者，則有 Arc、my.OAI、NCSTRL 等多項依據 OAI-PMH 所建立的加值服務系統。以下茲舉出幾項發展計畫，做為說明參考：

第一節 國外 OAI 相關計畫之發展

一、CIMI (Consortium for the Computer Interchange of Museum Information)：博物館界的資料提供者—提高文化資源在網路環境的 可得性

博物館擁有豐富的資源，但在網路環境中卻有許多內容無法被取得或者取得的資訊不具實用性與正確性。為此，CIMI 與 Digital Library Federation 及 Andrew W. Mellon Foundation 合作，提出 Metadata Harvesting Project，依據 OAI-PMH 之規範以解決資訊無法有效獲取的問題。此計畫目標共分三階段，依序為 CIMI 現有資料的測試，檢索機制的建立及發展入口網站服務。CIMI 採用 Dublin Core 及 XML 架構，欲使資料提供者與服務提供者間做良好的溝通，讓資訊容易被獲取。經由 CIMI 六個月的測試，發表了依據 OAI-PMH 1.0 版所建立的貯藏庫；初期評估顯示 OAI-PMH 具有簡單易建的優點，且其使用 Dublin Core 及 XML 的格式，在博物館界早已被認同其可用性。未來需要克服的問題在於：詮釋資料需有一致性的共識及其社群專指特性的滿足，並且各使用單位皆要能支援 XML 格式。CIMI 預期將會有越來越多的服務提供者接受 OAI 模式，並認為 OAI 的採用，將會使得博物館資源得以更廣泛地為其使用者所利用。CIMI 也將持續研究與測試

OAI-PMH 在博物館界的使用情形，並希望能執行更正式且具規模的測驗。²⁴

二、Arc：OAI-PMH 架構下的數位圖書館整合服務提供者

目前數位圖書館所遭遇的問題之一，在於互通性的缺乏，因而無法將資源以單一介面加以整合；而 OAI-PMH 的發展即強調檔案傳佈在技術上的互通性。Arc 為 Old Dominion University 所發展，是第一個根據 OAI-PMH 所建構的整合式檢索服務，根據原先的「國際預刊本伺服器」(Universal Preprint Server, UPS) 的設計理念，提供使用者數位圖書館形式的服務。其架構包括檢索者 (harvester)、階層式檢索 (hierarchical harvesters)、終端使用者檢索機制 (end-user search facility)；Arc 分別由不同的 OAI 貯藏庫檢索詮釋資料，將其標準化後儲存於關聯式資料庫中以提供服務，可同時檢索資料提供者及其它服務提供者的資料。²⁵

Arc 在發展過程中，所遭遇的問題主要有：

- (一)資料命名的不一致：不同的檔案具有不同的格式與命名方式；
- (二)資料提供者的安全保護機制可能阻礙檢索工作的進行；
- (三)資料現時性的維持等。

Arc 擬在未來逐漸擴增所涵蓋的資料提供者廣度，同時也擬增加其深度及服務的豐富性；並可做為研究詮釋資料在各類電子圖書館的品質及效用的工具。

三、Kepler：提供個人出版的 OAI 資料／服務提供者

基於多數個人期望有簡化的出版工具，並能保留對所屬資料的控制權；另一方面有鑑於 P2P (Peer to Peer) 系統的成功，Kepler 採用 OAI-PMH 發展所謂的「個人資料提供者」(personal data providers) 或稱「檔案資料小集」(archivelets)。其目標即在滿足一般大學研究學者自我出版、使用者資料獲取的需求。Kepler 架構包含四部份：²⁶

²⁴同 2。

²⁵同 2。

²⁶同 2。

- (一)依 OAI-PMH 架構建立的貯藏庫 (OAI compliant repository)；
- (二)出版工具 (publishing tool)：供出版者使用；
- (三)註冊服務 (registration service)：Kepler 除提供註冊服務外，並能持續追蹤已註冊的檔案資料庫的狀態；
- (四)服務提供者(service provider)：即 Kepler 本身。

Kepler 所提供的工具為：

- (一)能夠下載及自動安裝的檔案資料庫；
- (二)供出版者使用的自動註冊服務；
- (三)簡易的服務提供者，檢索來自於貯藏庫的詮釋資料。

Keple 目前所建立的標準系統，可以安裝在任何機器上，並能夠處理所有註冊伺服器及服務提供者之間的互動，希望能夠建立一個出版及資訊發掘的新典範。

四、Torii：開放型檔案庫下的入口網站

電子化及網路的發展，對學術性出版造成不小的衝擊；諸如電子與紙本型式、著作財產權等問題，皆引起許多的討論與爭議。TIPS (Tools for Innovative Publishing in Science) 即於此環境下因應而生，其認為未來網路的資訊傳播及出版在文獻交換上需要有較多面向的結構，以及一致性、桌面型式的網頁獲取服務及工具來管理資訊，還需要有發展完善的資訊檢索及過濾技術。Torri 為 TIPS 計畫下所建置的入口網站 (portal)，為 The European Union Fifth Framework Program Information Society Technologies Program (IST-1999-10419) 的一部份，以透過單一入口網站的方式提供終端使用者個人化的服務。而 OAI-PMH 則為入口網站與其各項服務間的基本溝通協定；透過 OAI-PMH，Torri 將更具擴張性。²⁷

²⁷同 2。

五、OLAC (Open Language Archives Community)：語言社群間的跨檔案檢索服務

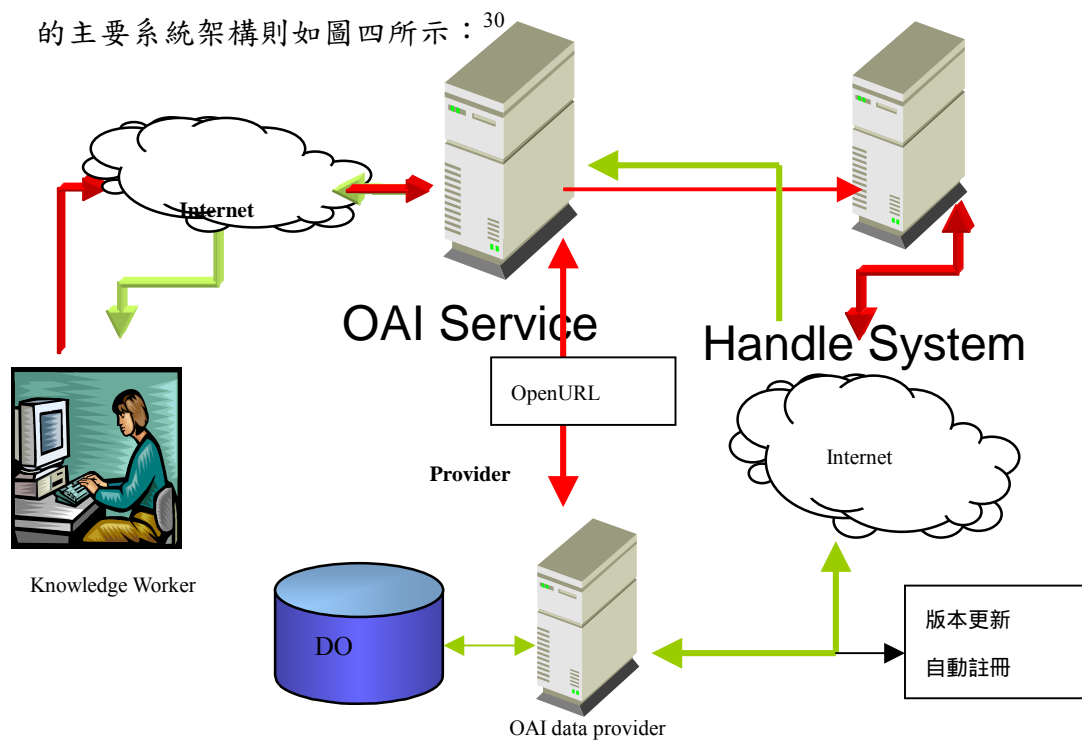
面對大量且多元的語言資源，如何建立語言資源及其相關使用者間的連結是必要的。OLAC 為一國際性組織，於 2000 年 12 月，由近百位語言學者發起。欲根據 OAI-PMH 及 Dublin Core 建立起全球性的整合語言檔案系統，發展一致性的數位語言資源檔案及其貯藏庫與服務的提供。由於 OAI-PMH 具有由上而下傳播的特性，同時是有效且結構化的資料庫，非常適用於資料成長迅速、使用者具有資源描述需求的語言資源環境。另一方面，OAI-PMH 的資料提供者可以根據諸如 Dublin Core 等詮釋資料標準，專家社群如語言資源社群即能依循其標準送出符合 OAI-PMH 的格式。即是使用 OAI-PMH 得以將社群檔案加以結合並提供終端使用者檢索使用。²⁸

²⁸同 2。

第二節 國內 OAI 相關計畫之發展

一、數位典藏國家型科技計畫聯合目錄原型系統設計

自 2002 年起展開的國科會「數位典藏國家型科技計畫」，提出以 OAI 建置「國家數位典藏聯合目錄」的構想。於九一年四月起邀請國內參與數位典藏機構之代表成立 OAI test-bed 小組，由陳昭珍教授主持，以 OAI 技術建構國家數位典藏之聯合目錄。經過四個多月的設計，已發展出數位典藏聯合目錄的原型系統。²⁹雖然 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，如與各單位之資料庫應如何連結、如何透過詮釋資料擷取數位物件、資料服務端之介面應如何設計等，都是 OAI 未訂定，也是無法訂定的部份，但在實際環境中，則是一定要處理的問題。聯合目錄的主要系統架構則如圖四所示：³⁰



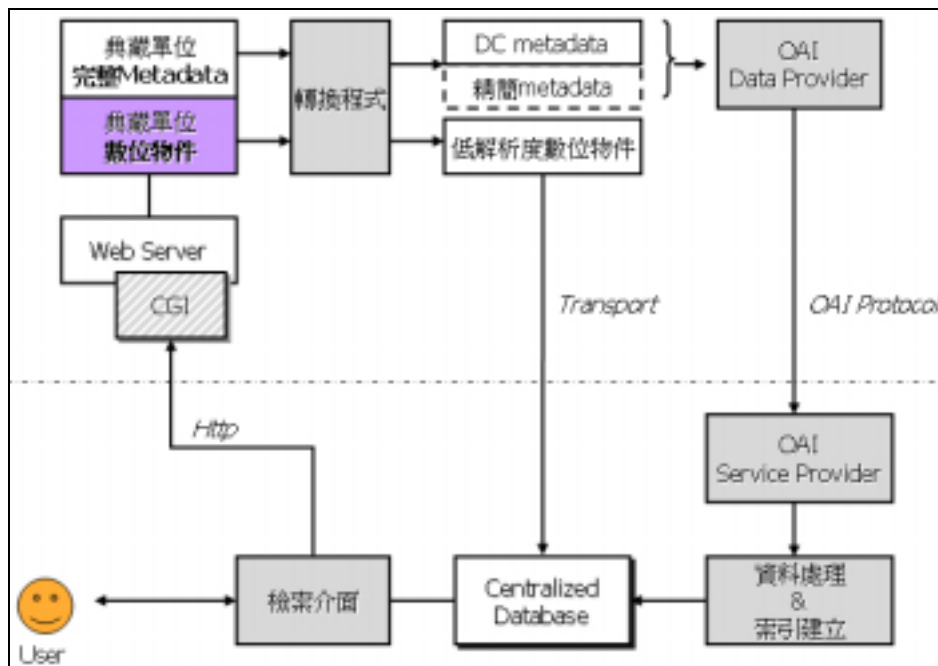
圖四：OAI 整體系統架構圖

²⁹陳昭珍、何佳欣，「數位化圖書館互通檢索機制之探討」，中國圖書館學會會報 70(民國 90 年) 頁 53-68。

³⁰陳昭珍，「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」，國家圖書館館刊 91(1)(民國 91 年 6 月)：頁 1-17。

二、「臺大典藏數位化計畫」之「臺灣大學數位典藏資源中心」建置

「臺大典藏數位化計畫」為國科會「數位典藏國家型科技計畫」中的一機構計畫，其下包含七個子計畫，有鑑於典藏單位眾多且獨立，為使各典藏單位所建置的系統得以分享，使用者藉由單一介面即能檢索到所有典藏單位的資料，並能讓數位典藏的全貌得以展現；「臺大典藏數位化計畫」擬規劃一「臺灣大學數位典藏資源中心」(NTU Digital Archives Resource Center; DARC)，以 OAI 架構為基礎的互通檢索系統規劃為其發展重點之一。期望使各個資料提供者與服務提供者之間的溝通更為容易，讓數位典藏的資料能夠保有詮釋資料的原始結構或 Dublin Core 格式，並透過標準且簡單的程序達到分享、使用與加值，有助於使用者更方便地檢索與獲取網路資源，滿足文獻資訊檢索的需求。整體系統架構環境則如圖五所示。



圖五：台大數位典藏資源中心架構圖

此系統為依據 OAI-PMH 2.0 版，藉由實作資料提供者及服務提供者兩端的程式，作為台灣大學數位典藏資源中心系統的運作核心。採用 Java 作為系統軟體之開發語言，而資料庫系統在開發時採用 Microsoft 的 SQL Server2000，不過只要符合 SQL92 標準之關聯式資料庫系統均可。綜觀實作過程的經驗如下：

(一)因各典藏單位詮釋資料性質的差異，必須另提供一共同欄位之精簡版詮釋資

料，而最適當的共同欄位合集便是 Dublin Core metadata。

(二)各典藏單位基於資料來源著作權等因素，部分資料並不適合公開，或只允許公開部分欄位內容，因此在 OAI Data Provider 運作上必須先經資料過濾的處理過程。基於前項與本項需求，在各典藏單位除了需建置 OAI 相關之程式外，尚需包含資料異動與轉換之處理程式。

(三)本質上，OAI 只是一個用來交換詮釋資料的協定，並不包含如文件、影像、聲音等全文資料(full-content)；因此其餘文件格式、內容均需透過其它程式應用技術輔助，並不在此協定處理的範圍。因此在整合資源中心的功能上必須要能連結回原典藏單位之數位物件所在。而在服務提供者端(資源中心)，除了提供整合性的資料檢索功能外，在呈現台大數位典藏資源中心架構圖時，亦需顯示簡易的數位物件內容，例如縮圖或部分多媒體檔案內容。整體而言，服務提供者端除了利用 OAI 協定向資料提供者獲取相關詮釋資料外，亦需下載所需的簡易數位物件資料。

第五章 結論與建議

近年來，由於世界各國數位圖書館與數位典藏貯藏庫的蓬勃發展，互通性的議題成為眾所注目且亟待解決的焦點，故有助於資訊系統間互通之協定與規範陸續被提出討論，例如 Z39.50、SOAP、OAI-PMH 等，其中 OAI-PMH 的簡單、易建特點，實為一項極具潛力的互通性協定。

回頭審視 OAI-PMH 的幾個主要目標：在跨系統、跨平台這個目標上，OAI 的架構用了最簡單最普及的 HTTP/XML 組合，雖然未必最簡潔、最有效率，但也因此，在實作的過程中，能利用許多成熟的程式元件處理網路服務及 XML 文件，兼顧易於實作及使用這個目標。此外，OAI 的架構中並不管資料是以何種方式儲存，無論是資料庫或檔案均可，詮釋資料是預先儲存或臨時產生也都是系統設計者的自由。由於只要求六道基本指令，系統設計者可以因應原本的資料型態，來巧妙的加入 OAI-PMH 服務。OAI-PMH 定時擷取的方式，也合乎數位典藏資料的特性，一筆數位資料在完成後，很少需要經常性的變動修改。不過這方面值得討論的問題在，詮釋資料的格式是經常更改的。在處理各資料來源時，常常遇到典藏單位因為不同的需要而更改其詮釋資料的欄位設計，甚至更改資料庫的設計。雖然 OAI-PMH 提供了支援多種詮釋資料格式的彈性，但若頻繁的更動詮釋資料格式，對服務提供者事實上是很難處理的。

使數位資料更有效率、更廣泛的傳播是 OAI 提出的另一項重大訴求，有幾個研究計畫，例如 DP9，其目標更特別專注於利用 OAI-PMH 來將隱藏在資料庫內的資料傳遞到一般搜尋引擎，使得平常的使用者不用特別經過某些 OAI 的服務提供者，而是依習慣在常用的搜尋引擎裡就可以找到 OAI-PMH 所傳遞的詮釋資料。而有的研究計畫，則是希望作出 OAI-PMH 與其他通訊協定的溝通介面。然而資訊的整合與流通，需要的不單單是資料流通的機制而已，即使 OAI-PMH 強制每個資料提供者都必須提供 Dublin Core 的詮釋資料，但是用於資料的分類、整合、使用，仍顯不足。因此在未來的一兩年，當 OAI 的機制普遍在數位典藏界建置完成後，資料的流通、使用模式，值得我們觀察分析。

對於數位典藏互通性理論架構的研究，歸納而言，有以下幾點建議：

一、由於網路環境具備全球性連結、開放與分散式的架構概念、異質的特性，數

位典藏的成果若欲讓使用者有效率的進行檢索、分享或執行其它工作，互通性的達成為其關鍵。

二、互通性指涉層面廣、觸及許多不同的面向，應加強跨領域間的合作與整合，並發展能提供一般層級的資訊分享架構。而在數位典藏互通性模式的發展與評估過程中，成本效益與功能的考量為其關鍵。

三、OAI-PMH 在許多的研究與實作結果中，顯示其具有易於建置與架構簡單的特性，且使用者能利用最經濟的時間，取得比先前更豐富且精確的資料，確實為數位典藏在互通性架構上提供一可行的解決方案。然有亦有一些重要但 OAI 未做規範的問題，如與各單位之資料庫應如何連結、如何透過詮釋資料擷取數位物件、資料服務端之介面應如何設計等，皆是在現實環境中需尋求解決的議題。

參考書目

1. Arms, W. Y. Digital Libraries. Cambridge, Ma.: MIT Press,2000.
2. Lagoze, C., Van de Sompel, H."The Open Archives Initiative : Building a Low-Barrier Interoperability Framework" 2001.
<<http://www.openarchives.org/documents/oai.pdf>> (10 January 2002)
3. Moen, W. E. "Assessing Interoperability in the Networked Environment : Standards, Evaluation, and Testbeds in the Context of Z39.50" September 2000.
<<http://www.unt.edu/wmoen/publications/InteropEvalPreprint.pdf> >(24 June 2003)
4. 余顯強。「淺談數位圖書館典藏資料互通之存取協定」。書藝 39(民國 92 年)。
5. 柯皓仁、黃夙賢、楊維邦。「詮釋資料與數位圖書館系統互通性之探討」。大學圖書館 5(1)(民國 90 年)：頁 49-78。
6. 徐代昕。數位典藏互通性架構之建置。國立臺灣大學資訊工程學研究所，碩士論文，民國 92 年。
7. 陳昭珍、何佳欣。「數位化圖書館互通檢索機制之探討」。中國圖書館學會會報 70(民國 90 年)：頁 53-68。
8. 陳昭珍。「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」。國家圖書館館刊 91(1)(民國 91 年 6 月)：頁 1-17。
9. 陳雪華、項潔、陳昭珍、洪筱盈、徐代昕、張懷文。「數位典藏互通性架構之探討」。中國圖書館學會會報 68(民國 91 年)：頁 1-13。
10. 國家圖書館。「數位圖書館分散檢索協定」。民國 90 年 10 月。 ，
<<http://www.ncl.edu.tw/bbs/pdf/6-3-3.pdf> >(民國 92 年 10 月 24 日)

附錄一：

出席國際學術會議心得報告

91年12月25日填

姓名	陳雪華	性別	女	年齡	年 月 日 生
籍貫	福建省 寧德縣	服務單位	台大圖書資訊學系		職稱 教授
會議名稱	(中文) 2002 亞洲數位圖書館國際會議 (英文) International Conference of Asian Libraries(ICADL)2002				
主辦單位	Division of Information Studies School of Communication & Information Nanyang Technological University (NTU)		會議地點	Grand Hyatt Singapore	
會 期	自 91 年 12 月 11 日 迄 91 年 12 月 14 日				
發表論文題目	(中文) (英文) Building an OAI-based Union Catalog for the National Digital Archives Program in Taiwan				
擔任職務	1. ICADL2002 Steering Committee member 2. ICADL2002 Program Committee member 3. ICADL2002 Session Chair				

參加國際學術會議活動概況簡表

姓 名		陳雪華				
服務單位及職稱		台大圖書資訊學系教授				
會 議 名 稱		中文：2002 年亞洲數位圖書館國際會議				
		英文：International Conference of Asian Digital Libraries, ICADL2002				
日 期		2002/12/11~ 2002/12/14	地 點	Grand Hyatt Singapore		
經 費 來 源		申請補助 補助單位：國科會				
國際 會議 資料	性質	<input type="checkbox"/> 國際組織 主辦	國際組 織	名稱	中文	
					英文	
				與我 關係	<input type="checkbox"/> 國家會員 <input type="checkbox"/> 團體會員 <input type="checkbox"/> 個人會員 <input type="checkbox"/> 贊助會員 <input type="checkbox"/> 觀察員 <input type="checkbox"/> 其它	
		地主國承辦單位				
		<input type="checkbox"/> 非國際組 織主辦	主辦單 位	性質	<input type="checkbox"/> 政府機關 <input type="checkbox"/> 學校 <input checked="" type="checkbox"/> 民間團體 <input type="checkbox"/> 其他	
				名稱	Division of Information Studies School of Communication & Information Nanyang Technological University (NTU)	
			協 辦 單 位	1. National Library Board of Singapore 2. Nanyang Technological University Library 3. National University of Singapore Libraries 4. Library Association of Singapore 5. Library Information Technology Association of Singapore		
重要性	涉及我與會之 <input type="checkbox"/> 會籍 <input type="checkbox"/> 名稱 <input type="checkbox"/> 權利 <input type="checkbox"/> 地位 請說明：					
參加國家或地區		亞洲、美洲、歐洲、澳洲 共三十個國家人員參與				
		中共有無參加人員	<input checked="" type="checkbox"/> 有 <input type="checkbox"/> 無 <input type="checkbox"/> 不確知			
其 他 說 明						

填表人簽章： 陳雪華

填表人電話 Tel：

2363-0231 ext 2296

填表日期： 2002/12/25

報告內容目次

一、參與會議簡介	p.47
二、會議議程	p.49
三、會議報告主題	p.53
四、建議事項	p.56

一、參與會議簡介

世界三大國際數位圖書館會議，分別為 JCDL (Joint Conference of Digital Library)、ECDL (European Conference on Digital Libraries) 及 ICADL (The International Conference on Asian Digital Libraries)。JCDL 為 ACM 及 IEEE-CS 兩機構聯合舉辦，以美國數位圖書館發展為主要探討對象的美國數位圖書館研討會，兩機構合辦歷史起於 2001 年，整合兩者所舉辦之數位圖書館相關技術研討之國際會議。ECDL 是歐洲各國探討數位圖書館技術、發展的主要論壇，2002 年 9 月在羅馬舉辦了第六屆的會議，第七屆於 2003 年 8 月在挪威舉行。ICADL (亞洲數位圖書館會議) 則為亞洲各國交流數位圖書館製作經驗、技術研究成果經驗的研討會，第一屆始於 1998 年 8 月，在香港舉辦，第二屆於 1999 年 11 月在台灣台北舉辦，第三屆於 2000 年 12 月在韓國漢城舉辦，第四屆於 2001 年 12 月，在印度班加羅爾舉辦，本屆為第五屆，於 2002 年 12 月在新加坡舉辦。

第五屆 ICADL 會議，由 Division of Information Studies, School of Communication, Nanyang Technological University of Singapore 所主辦，協辦單位則有 National Library Board of Singapore, Nanyang Technological University Library, National University of Singapore, Library Association of Singapore，及 The Library and Information Technology Association of Singapore。而本屆打破以往慣例，與第三屆的 WISE2002 (International Conference on Web Information System) 一同舉行，合辦主要意義在於鼓勵更多、更頻繁的跨領域互動，促進並推動數位圖書館研究在網際網路研究社群的發展。

ICADL 是一個讓參與者交流研究成果、創新意見，以及交流亞洲各國數位圖書館方面發展技術的一個國際會議，並自詡為亞太地區最大規模、亞洲國家自發之數位圖書館研討會。在新加坡舉辦的第五屆會議，目標在於強化 ICADL 成為首要數位圖書館研究領域的定位，能夠吸引全球具高學術價值的論文發表，以滿足亞洲太平洋地區數位圖書館研究機構需求。在會議進行規模制度上，不但廣納亞洲、美國及歐洲重要領域人士組成委員會，並設置電子論文註冊系統提升管理會議之效率及透明化、介紹正式議程及會議記錄於 LNCS 期刊，最後並計畫收羅本次會議的精華論文，出版 Information Processing and Management Journal 特刊。

今年會議的 Program Committee members 總共有 72 人，來自亞洲的有 41 人，美國有 14 人，歐洲國家有 17 人，雖為亞洲圖書館會議，但歐美國家 PC members 參加比率也將近四成。本屆 ICDAL2002 論文投稿方面，共來自 401 位作者，包含有 110 篇 full papers 及 61 篇 short papers 投稿，分別從 full papers 中，錄取 34 篇為 full papers 和 10 篇為 short papers；另外，從 short papers 中採用了 10 篇為

short papers 及 16 篇為 poster papers，總共收錄了 70 篇論文。其中以印度投稿最為踴躍（37 篇），但只錄取 5 篇。其次是美國與新加坡（均為 19 篇），分別錄取 12 篇與 10 篇。大陸居於第三（18 篇），錄取 9 篇。在台灣方面，雖然投稿並不多，投稿了七篇，錄取四篇，錄取比率算是很高的。各國論文發表情況請見下表：

表 1：各國發表論文統計

	F-S	F-AF	F-AS	S-S	S-AS	S-AP
Australia	5	1	0	1	0	0
Austria	0	0	0	2	1	0
Belgium	0	0	0	1	0	0
China	12	3	4	6	0	2
Czech	0	0	0	1	0	1
Greece	1	0	1	0	0	0
Germany	1	0	0	1	0	0
HK China	1	0	0	0	0	0
India	29	1	1	8	0	3
Indonesia	1	0	0	2	0	0
Iran	1	0	0	0	0	0
Italy	4	3	0	0	0	0
Japan	4	2	0	2	0	0
Korea	3	0	0	11	1	4
Malaysia	3	1	1	1	0	0
Nepal	1	1	0	0	0	0
Netherlands	1	1	0	0	0	0
New Zealand	2	1	1	1	1	0
Palau	0	0	0	1	0	0
Philippines	0	0	0	1	0	1
Russia	0	0	0	1	0	0
Singapore	15	7	1	4	1	1
Spain	2	1	0	0	0	0
Sri Lanka	2	0	0	0	0	0
Switzerland	1	0	0	1	0	0
Taiwan	4	2	1	3	0	1
Thailand	5	2	0	0	0	0
UK	6	5	0	0	0	0
USA	6	3	0	13	6	3

23 Countries	110	34	10	61	10	16
--------------	-----	----	----	----	----	----

F-S: Number of submitted full papers

F-AF: Number of submitted and accepted full papers

F-AS: Number of submitted full papers accepted as short papers

S-S: Number of submitted short papers

S-AS: Number of submitted and accepted short papers

S-AP: Number of submitted short papers accepted as poster papers

二、會議議程

● Wednesday 11 December, 2002 - Tutorials

0800	Registration	
0900-1215	<u>Tutorial 1a</u> Evaluating Digital Libraries for Usability (Christine Borgman: Univ. of California, Los Angeles, USA)	<u>Tutorial 1b</u> Knowledge Management Systems: A Text Mining Perspective (Hsinchun Chen: Univ. of Arizona, USA)
1330-1700	<u>Tutorial 2a</u> Digital Libraries: Theory and Practice (Edward Fox: Virginia Tech, USA)	<u>Tutorial 2b</u> Dublin Core (Stuart Weibel: OCLC, USA)
1800-2100	Registration	
2000-2200	Pre-conference reception	

● **Thursday 12 December, 2002 - Paper Sessions**

0800	Registration	
0900-1030	<u>Opening Ceremony / Joint Keynote Address 1</u> Web Services and Web Integration Serge Abiteboul, INRIA, France	
1030-1100	Break	
1100-1200	<u>ICADL Keynote Address</u> Progress on Educational Digital Libraries Jane Prey, National Science Foundation, USA (Session chair: Hsinchun Chen, University of Arizona, USA)	
1200-1300	Lunch	
1300-1500	<u>Paper Session 1a</u> Information Retrieval Techniques (Session chair: Jieh Hsiang, National Taiwan University, Taiwan)	<u>Paper Session 1b</u> Multimedia Digital Libraries (Session chair: Tat-Seng Chua, National University of Singapore)
1500-1530	Break	
1530-1615	<u>ICADL Invited Talk 1</u> From Digital Library to Digital Government: A Case Study in Crime Data Mapping and Mining (Hsinchun Chen, University of Arizona, USA) (Session chair: Christopher C. Yang, Chinese University of Hong Kong)	<u>ICADL Invited Talk 2</u> Challenges in Building the Digital Library for the 21st Century (Christine Borgman, UCLA, USA) (Session chair: Sung Hyon Myaeng, Chungnam National University, Korea)
1615-1700	<u>Paper Session 2a</u> Data Mining in Digital Libraries (Session chair: Christopher Khoo, Nanyang Technological University)	<u>Paper Session 2b</u> Special Purpose Digital Libraries (Session chair: Li Jianzhong, Harbin Institute of Technology, China)
1700-1830	City Tour enroute to Gala Dinner	
1830-2130	Gala Dinner (Alkaff Mansion)	

● **Friday 13 December, 2002 - Paper and Poster Sessions**

0900-1000	<p><u>Joint Keynote Address 2</u></p> <p>Data Mining Technologies for Digital Libraries and Web Information Systems</p> <p>Srikant Ramakrishnan, IBM Almaden Research Lab, USA</p> <p>(Session chair: Raju Buddharaju, National Library Board, Singapore)</p>	
1000-1030	<p>Break</p>	
1030-1215	<p><u>Paper Session 3a</u></p> <p>Digital Library Services</p> <p>(Session chair: Ian Witten, University of Waikato, New Zealand)</p>	<p><u>Paper Session 3b</u></p> <p>Digital Libraries for Community Building</p> <p>(Session chair: Erich Neuhold, Darmstadt Univ. of Technology, Germany)</p>
1215-1315	<p>Lunch</p>	
1315-1500	<p><u>Paper Session 4a</u></p> <p>Information Retrieval in Asian Languages</p> <p>(Session chair: Hsueh-Hua Chen, National Taiwan University, Taiwan)</p>	<p><u>Paper Session 4b</u></p> <p>Building and Using Digital Libraries</p> <p>(Session chair: Christine Borgman, UCLA)</p>
1500-1530	<p><u>Poster Session and Break</u></p>	
1530-1615	<p><u>ICADL Invited Talk 3</u></p> <p>Dublin Core: Process and Principles</p> <p>Shigeo Sugimoto (ULIS, Japan)</p> <p>Stuart Weibel (OCLC, USA)</p> <p>Thomas Baker (GMD, Germany)</p> <p>(Session chair: Robert Allen, University of Maryland, USA)</p>	<p><u>ICADL Invited Talk 4</u></p> <p>Building Digital Libraries Made Easy: Toward Open Digital Libraries</p> <p>Edward Fox (Virginia Tech, USA)</p> <p>(Session chair: Shalini Urs, University of Mysore, India)</p>
1615-1700	<p><u>Paper Session 5a</u></p> <p>Metadata Issues</p> <p>(Session chair: Vilas Wuwongse, Asia Institute of Technology, Thailand)</p>	<p><u>Paper Session 5b</u></p> <p>Algorithms and Protocols</p> <p>(Session chair: San-Yih Hwang, National Sun Yat-Sen University, Taiwan)</p>
1700-1900	<p>Library Tours</p>	

● **Saturday 14 December, 2002 - Paper Sessions/Post-Conference Workshop**

0900-1000	<p><u>ICADL Invited Talk 5</u></p> <p>Examples of Practical Digital Libraries: Collections Built Internationally Using Greenstone</p> <p>Ian Witten (University of Waikato, New Zealand)</p> <p>(Session chair: Lau Kai Cheong, National Library Board, Singapore)</p>	<p><u>ICADL Invited Talk 6</u></p> <p>Introduction of China Digital Library Project</p> <p>Zhang Xiaoxing (National Library of China, PRC)</p> <p>(Session chair: Zawiyah Baba, National Library of Malaysia)</p>
1000-1030	<p>Break</p>	
1030-1215	<p><u>Paper Session 6a</u></p> <p>Human Computer Interfaces</p> <p>(Session chair: S. Sadagopan, Indian Institute of Information Technology, India)</p>	<p><u>Paper Session 6b</u></p> <p>Digital Library Infrastructure</p> <p>(Session chair: Shigeo Sugimoto, University of Tsukuba, Japan)</p>
1215	<p>Closing of Conference</p>	
<p>Post-Conference Workshop</p> <p>Library and Information Science Education in Asia (LISEA)</p>		
1400-1515	<p><u>LISEA Workshop Session 1</u></p> <p>Roles and Opportunities for E-Learning in the LIS Curriculum</p>	
1515-1545	<p>Break</p>	
1545-1700	<p><u>LISEA Workshop Session 2</u></p> <p>Standards for Accreditation: Thinking About LIS Program Accreditation in Asia</p>	

三、會議報告主題

本屆 ICADL 的主題為：「數位圖書館：人、知識與科技」，反映出 ICADL 成員的信念，也就是說，數位圖書館成功的發展與應用，倚賴三個方面的成就：內容的豐富程度能夠滿足被服務社群的需求；科技的應用要能夠建立起一個以「使用者為中心 (user-centered)」的系統環境；在人的因素方面，對於內容必需要有良善的管理維護措施，並且能跟上內容、技術，以及使用者需求改變的步調。

本次會議結合第三屆的國際網路資訊系統工程會議 (International Conference on Web Information Systems Engineering, WISE) 的舉辦，為了促進 ICADL 與 WISE 參與者的互動，兩個會議共同舉行開幕典禮以及與三場主題演說 (keynote addresses)。2002 年 ICADL 會議包含 4 場 Tutorials、6 場邀請性的演講 (invited talks)、54 篇論文發表。另外，還有 16 篇的海報展示等。內容涵蓋數位圖書館社群的建立 (digital libraries for community building)、資訊檢索技術 (information retrieval techniques)、人機介面 (human computer interfaces)、以及數位圖書館服務 (digital library services) 方面。

ICADL2002 議程內容包含了一天的 Tutorial 專業入門課程，以及兩天半主要的論文研討會議。各主題列舉如下：

(一) 四場專業入門課程 (Tutorials)：

- Metadata Standards: Dublin Core
演講者：Stuart Weibel (OCLC, USA)
- Digital Libraries: Theory and Practice
演講者：Edward Fox (Virginia Tech, USA)
- Evaluating Digital Libraries for Usability
演講者：Christine Borgman (Univ. of California, Los Angeles, USA)
- Knowledge Management Systems: A Text Mining Perspective
演講者：Hsinchun Chen (Univ. of Arizona, USA)

(二) 三場主題演說 (Keynote Addresses)：

- Joint Keynote Address 1
Web Services and Web Integration
演講者：Serge Abiteboul (Ecole Polytechnique, France)

■ Joint Keynote Address 2

Data Mining Technologies for Digital Libraries and Web Information Systems

演講者：Srikant Ramakrishnan (IBM Almaden Research Lab, USA)

■ Keynote Address 3

Progress on Educational Digital Libraries: Current Developments in the National Science Foundation's (NSF) National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL) Program

演講者：Jane Prey (National Science Foundation, USA)

(三) 六場邀請性演講 (Invited Talks)：

■ From Digital Library to Digital Government: A Case Study in Crime Data Mapping and Mining

演講者：Hsinchun Chen (University of Arizona, USA)

■ Challenges in Building the Digital Library for the 21st Century

演講者：Christine Borgman (UCLA, USA)

■ Dublin Core: Process and Principles

演講者：Shigeo Sugimoto (Univ. of Tsukuba, Japan)

Stuart Weibel (OCLC, USA)

Thomas Baker (GMD, Germany)

■ Building Digital Libraries Made Easy: Toward Open Digital Libraries,

演講者：Edward Fox (Virginia Tech, USA)

■ Examples of Practical Digital Libraries: Collections Built Internationally Using Greenstone

演講者：Ian Witten (University of Waikato, New Zealand)

■ Introduction of China Digital Library Project

演講者：Zhang Xiaoxing (National Library of China, PRC)

(四) 十二個論文發表時段 (54 篇論文) (Paper Sessions)：

共計有 34 篇 long papers，每篇發表 20 分鐘；20 篇 short papers，每篇發表 10 分鐘。

(五) 海報展示 (Poster Session)：

共計有 16 篇的海報展示。

綜觀上述所列項目整理，在專業入門課程部分，探討課題包含了「Metadata 相關標準」、「數位圖書館的理論及實作」、「評價數位圖書館的用途」及「知識管理系統—內容探勘」等。三個主題演講則涵蓋了「網路服務及整合」、「數位圖書館及網路服務系統的資料探勘」、「教育數位圖書館的推廣」。邀請性演講的六場，為「數位圖書館到 e 化政府」、「21 世紀數位圖書館的挑戰」、「Dublin Core」、「輕鬆打造數位圖書館」、「中國數位圖書館計劃的介紹」，以及「數位圖書館實際應用實例」等。

另外，本次會議安排 12 個時段的論文發表時間，論文主要探討方向大概是圍繞著數位圖書館的各項應用技術，諸如：檢索技術、資料探勘、特殊目的亞洲語言資訊檢索、人機介面、社群建立、協定及演算法等，當然還有 metadata 的相關標準等。以上論文將會在 Information Processing and Management Journal 發表。在海報展示方面，內容包括建造數位圖書館的對話系統、和讀者需求的服務平台、有效的檢索策略等，與該場次進行主題相關之討論。

四、建議事項

此次參加 ICADL2002 會議，與國際其他數位圖書館相關之研討會相較，同時身為 ICADL Program Committee Member 的一員，深感本研討會未來之規劃執行宜加強對「亞洲地區」數位圖書館議題之觀摩探討，不應偏重於歐美各國的觀點。針對此項重點及本次研討會情形，對於 ICADL 主辦單位及亞洲數位圖書館學域有下列之建議事項：

1. 為穩固亞洲數位圖書館社群之發展，以及達到長期 ICADL 研討會之目標，宜正式成立常設性指導委員會，負責 ICADL 之規劃發展。
2. 論文審議的過程尚有很多進步的空間，特別是在論文審查的過程中，可用 paper assignment 的方式允許 PC 委員選擇審查之文章，未來並由審查小組議決文章審查結果，以強化審查過程的透明度。
3. 亞洲地區由於語言的限制，導致有很多重要文獻資訊，不如歐美主流國家的文獻容易搜尋查檢。為解決亞洲地區數位圖書館著述文獻的公開與交流，宜藉由 ICADL 之機會廣徵亞洲各國 DL 研究論文，及其在重要期刊上的發表，以達到研究分享的目的。
4. ICADL (The International Conference on Asian Digital Libraries) 之定位為「亞洲地區」數位圖書館會議，應該以亞洲各國數位圖書館發展之情形及困難為會議重點。然本次研討會受邀者卻普遍以歐美人士為主，探討面向多從歐美發展經驗的角度切入，缺乏亞洲地區發展概況之交流，及共同所遭遇問題（諸如中文字碼檢索、亞洲地區相關技術標準）之協商解決，本次的研討會的邀請性演講部分甚至只獨邀請一場講演—“Introduction of China Digital Library Project,” Zhang Xiaoxing, (National Library of China, PRC)—探討關於亞洲的數位圖書館情況，且獨獨針對大陸發展做介紹。今年會議獨重西方技術發展的情況，使得亞洲地區國家無法藉由亞洲數位圖書館研討會做實務上的經驗交流，並從會議中獲得最適切於針對亞洲區域數位圖書館發展的實質幫助。
5. 已建議下一屆 ICADL 主辦單位，宜加強 ICADL 在亞洲各國數位圖書館發展執行所扮演的角色及功能，應將亞洲地區學者專家列為演講受邀人的優先人選，並將「亞洲各國數位圖書館發展之探究」增闢為系列講題，與論文發表的議程並列舉行，以確實達到 ICADL 促進代表亞太文化之亞洲各國數位圖書館交流的意義。也建議台灣地區的學者踴躍投稿相關的論文。

附錄二：

出席國際學術會議發表之論文

Building an OAI-based Union Catalog for the National Digital Archives Program in Taiwan

Chao-chen Chen

Associate Professor, Department of Adult and Continuing Education
National Taiwan Normal University

cc4073@cc.ntnu.edu.tw

Hsueh-hua Chen

Professor, Department of Library and Information Science
National Taiwan University

sherry@ccms.ntu.edu.tw

Abstract

On January 1st 2002, the National Science Council (NSC) of Taiwan launched a National Digital Archives Program (NDAP). The first phase of the NDAP was planned as a 5-year program, from 2002 to 2006. Many universities and research organizations participated in this program, and major content holders in Taiwan, including Academia Sinica, National Taiwan University, National Central Library, National Palace Museum, National Museum of Natural Science, National Museum of History, Historica Sinica, Taiwan Historica, etc. have also been involved.

In order to share and conserve all valuable collections, retrieve the digital collections of these content holders via a union interface, and allow the general public to access the digital collections in Taiwan, it is urgent to build a union catalog of National Digital Archives. To design the union catalog of digital libraries, in addition to the efficient mechanism of harvesting metadata, we must consider the naming principle for digital objects and a persistent connection method between metadata and the digital objects. In this article, we define the functions and the system architecture of the OAI-based union catalog and explain the reason why the OAI (Open Archives Initiative) protocol for metadata harvesting was chosen as the mechanism for creating the Union Catalog of National Digital Archives Program in Taiwan.

Keywords : National Digital Archives Program、 Open Archives Initiative (OAI)、
Naming、 Handle System、 Interoperability、 Metadata

1 Introduction

National Digital Archives Program (NDAP) is a major policy concerning digital content after the proposal of knowledge economy by the Taiwan government. Current participating units of the initiative include Academia Sinica, National Taiwan University, National Central Library, National Palace Museum, National Museum of Natural Science, National Museum of History, Historica Sinica, Taiwan Historica and dozens of other academic groups. Therefore, how the resources within the digital collection built by these units should be shared, how to allow users to search for all data from the archive institutions from one single interface, and how to show the public the entirety of the digital archive, remain very important issues.

In order to share the digital resources built by the respective units, the construction of a union catalog is of priority importance. The second question is how to show the digital objects of full text, image, sound, and visual through metadata. Union catalog can be built on two models: one is a collective union catalog; the other is a distributed virtual union catalog. The former has the advantage of offering better search results, but also has the disadvantage of a high construction cost.[1] The advantage of adopting a virtual union catalog is the low construction cost, but with poor search results. In order to maintain the advantages of both models and avoid their drawbacks, a new protocol for distributed harvesting of metadata – Open Archives Initiative (also known as OAI) was born in the digital era. Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) provides a simple solution for harvesting the metadata of different databases automatically, in batches, or in different distributions, as well as constructs a collective union catalog.

2 Introduction to OAI-PMH

OAI was born in the meeting of Universal Pre-print Service that took place in October, 1999 at Santa Fe by Paul Ginsparg, Rick Luce, and Herbert Van de Sompel. The motivation for this creation was because different databases and systems were not interoperable. Therefore, related data or data of different fields were stored in different locations and not integrated, which limited the flow of data to be imperfect. Representatives participating in the meeting regarded it necessary to develop an interoperable standard structure for academic electronic pre-print and related digital archives. Thus OAI was established. [2]

The original purpose of the OAI was for the interoperability between academic electronic pre-prints. However, the interoperability problem encountered by digital libraries was very similar to the aforementioned purpose, therefore in early 2000, OAI expanded its applicability to digital libraries. In the Cambridge Meeting held by the

Digital Library Federation and the Andrew W. Mellon Foundation in Harvard University, the issue of how to send the archive data of digital libraries to the Internet search engines was discussed. Representatives in the meeting (including professionals from libraries and museums) reached the consensus that the key factor would be the spread of metadata through interoperability.

In January 2001, OAI announced the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to provide a feasible solution for interoperability of network resources. OAI-PMH used the technology of Internet and metadata with excellent results in function strengthening and simplifying the implantation procedure. [3]

OAI is a simple, easily designed protocol. Its main function is to provide the front-end with protocol required for harvesting the necessary information from the back-end repository. OAI elements include OAI Service Provider and Data Provider. Therefore, the union catalog architects based on OAI-PMH mainly obtains metadata from the data provider periodically by the service provider, and builds a collective union catalogue. As OAI-PMH is an application protocol established in HTTP, so its commands supply the back-end corresponding server programs via the variable names and contents transmitted between the front-end and the back-end, based on the results transmitted back after processing the variable contents, which must obey the XML format regulated by the OAI protocol XML schema. Service Provider mainly concentrates on maintaining the metadata harvested from respective systems, and add value-add services on the obtained metadata. Data Provider mainly maintains the data repositories, and supports the OAI protocol to obtain the data in the repositories. The relationship is indicated in Figure 1: [4]

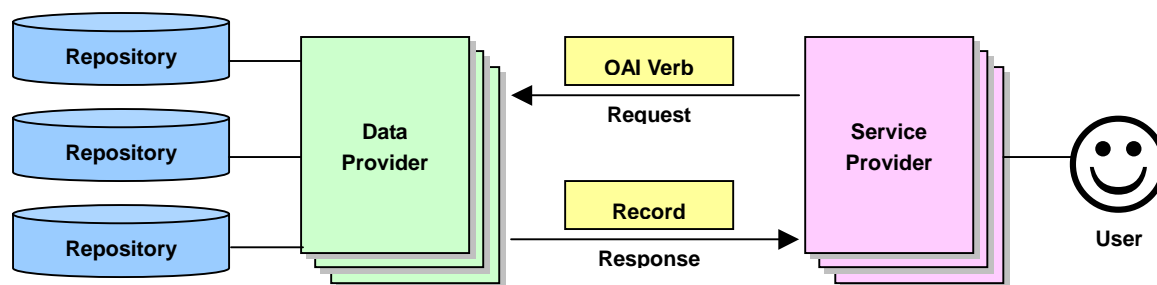


Figure 1. Component of OAI-based technical architecture

3 Management of the Metadata and digital object links

Traditional union catalog does not handle metadata and digital resource links. However, in the digital era, for union catalogue of a digital archive system, it is not sufficient to find related titles, but must further link to the relevant digital files. In

the Web environment, these links are mostly based on the URL mechanism. Though URL is an easy and flexible mechanism, it is also unstable and unreliable, because URL is composed of hostname, path, filename and the communication protocol that obtains this file, such as http, ftp, gopher, etc. Once the server shifts, or path or file name changes, URL cannot position accurately. Since the physical storage space for the data is limited, a need will arise one day to move the data to other servers as the data keep growing constantly. When we browse through web pages, HTTP v1.0/1.1 Error404 error messages often appear. This indicates that the intended connection resource has been removed to prevent further employment. Therefore, the link mechanism of network resources has always been emphasized, especially since the digital libraries, digital museums and digital archives have become the research direction. In related projects, this issue will definitely be handled. Links of digital resources usually involve two questions, the first is the naming of digital resources, and the other is connecting to the system where data is stored via the resource name. The reason why naming resources becomes an important issue in network environment, except for the aforementioned reasons, the following perspectives offer several reasons justifying naming digital resources as an issue that cannot afford to be neglected in a digital information system that wishes to archive data for an extension period of time: [5]

- 3.1** Naming is an important factor from the viewpoint of digital data storage management. For example, within any institution, more than one department would undertake the task of digitalizing the data. Different departments generate different types of files. If the naming method for digital resources cannot be unified but based on independent assumptions, problems of overlap, or different systems would arise. Identification would become impossible after an extensive period of time, hardware resources would not be easily distributed.
- 3.2** Also, as the digital resources of a region or a country become richer and richer, the public would not be satisfied with searching databases one by one, but prefer for a unified operation method. By this time, data of similar fields may need to be gathered. If institutions all have a unified naming method, problems of overlap, illegibility, deviation, difficulty in linking titles to digital resources could be avoided.
- 3.3** Moreover, from the perspective of electronic commerce, a unique identification code should be assigned to each digital resource. Take the books, periodicals or audio/visual data for instance, unless there are ISBN, ISSN, and ISRC codes, international selling of books, periodicals, or audio/visual data will be extremely difficult. Now, when the data unit to be

sold is a chapter, a figure, or a chart of a book, instead of the entire book, the coding method requires details down to chapter, section, figure, chart, and so forth.

If network data can provide a persistent naming method that cannot be changed by modifications to the external storage devices, much convenience would benefit applications of large amounts of network data. Every book on the library shelves has its call number, which is given by classification number and author number, instead of its shelf position. Every book has a unique call number. It is the link point between collection records and books on the shelf. After a reader has found the collection record, books can be located by its call number. Librarians do not have to change the call number on the collection record even if the book is placed at another location, because the code expresses a relative location, instead of an absolute location. If the network resources have the unique identifier, or otherwise known as the file name, instead of the absolute position, resource manager would be exempted from the need to change the file name and path of the metadata due to changes of data storage computers. Also, as a digital information system usually is not a closed linking system, or the only database system, there will be more than one way to search the digital resources. Therefore, each digital resource must have a unique name to facilitate file builders and users to establish a link or access data.

As mentioned above, if the URL of the full text or multimedia files of the metadata is recorded, then when the server name is changed, or the full text and multimedia data location is changed, the found metadata would not be linked to the digital file. Therefore, generally speaking, records in the metadata are usually the file names of its digital file name (or called identifier), instead of the actual storage position. Then, digital resource naming and analysis management system (or what is called the identifier parser) manage the file names and the corresponding URL data. This is what we call a resolution or a handle system, a mandatory part of the digital library union catalog system.

Handle system analyses identifiers based on the distributed model, with several possibilities at the client's end. It would be a client end with this type of functions, or a web browser with special plug-ins, or a regular client end that connects through a certain proxy. Regardless of the circumstances, the client end will communicate with the handle system by using the communication protocol defined by the handle system. These communication protocols have already been formally defined, even produced. The flow steps of the analysis can be briefly represented in the following chart:

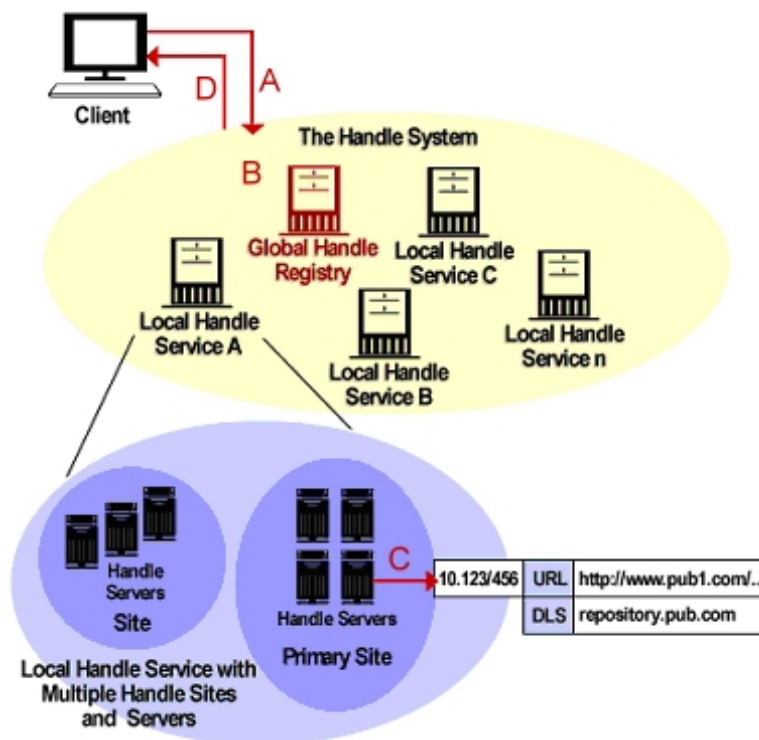


Figure 2. Work flow of handle system

As indicated in Figure 2, we can define the procedure of analyzing the identifier into the following four steps (indicated as A through D on the figure above):

- A. When a certain client-end (for example, the web browser) has encountered a need for analyzing identifiers (for example 10.123/456). When this circumstance occurs in a particular hyperlink, or the reference situation of a particular link, the client end will send the requirement for analyzing the identifier particularly concerning the handle system regardless on the Internet or individual intranet. As mentioned above, this task can be directly completed at the client-end, or through a particular proxy.
- B. The handle system contains a collection formed by identifier services. Every service includes at least a main server group and a certain number of sub-server groups. Every server group is also formed by a certain number of identifier proxies. The procedure of the analysis shows that each server group duplicates all the identifiers that belong to the identifier service. Among which, there is a unique service, the global handle registry (GHR), which is responsible for indicating the naming service of each region. Services of each region would also know how to access the GHR. Such a relationship allows the request for identifier

analysis to take place at anywhere, as well as directing the request to the certain service responsible for analyzing the identifier.

C & D Every identifier is connected to one or many data values of a certain type. In this case, the identifier 10.123/456 is connected to two types of data value, one of which is the URL type data, the other a new type of protocol named DLS. Such data would finally be transmitted back to the client end for its reference. Yet, please note that the so-called various data values of certain types could also be the same data type, for example various URLs. In other words, a single identifier datum can correspond to many URL data values for subsequent operations. However, at the same time, the handle system would not make any assumptions about the subsequent usage of these data. In other words, the employment of these data is entirely based on client end needs. In such circumstances, applications can be allowed to have maximum flexibility to use the naming service provided by the handle system. In this case, the client end can use the communication protocol of the data values to position the network resources. Again, everything depends on the usage purpose of the client end.

4 Function of OAI-based Union Catalog for the National Digital Archives Program

In order to build the union catalog of the National Digital Archive Program (NDAP), the program has invited several representatives that participate in NDAP to form the OAI test-bed team, and build the union catalog of the national digital archive with OAI-PMH and handle system technology. The initial phase will be participated by National Taiwan University, Academia Sinica, National Palace Museum, and National Museum of Natural Science. OAI-PMH is a simple and easily designed protocol. However, some problems have not yet been considered by the OAI-PMH in the actual union catalogue system design, for example how should the databases of different units be connected, how to harvest digital objects through metadata, and how to design the data service end interface. Here we will introduce the system provision and system structure that we have defined for the national union catalog, including: integrative function, service provider, data provider, handle system, system register, system testing, and main system structure as shown in Figure 3.

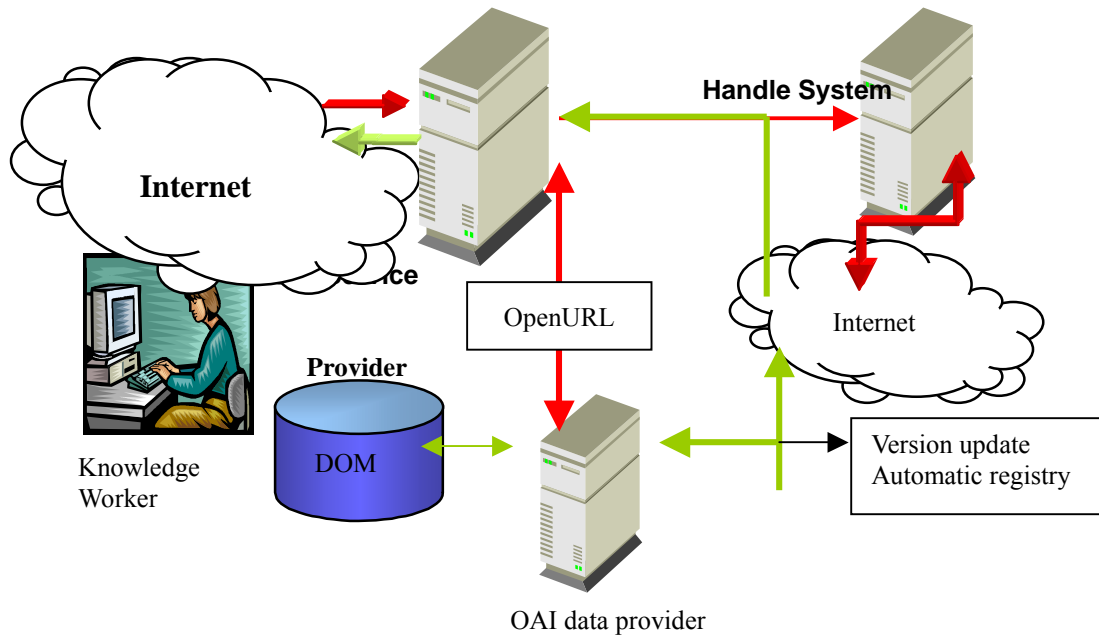


Figure 3. Architecture of the OAI-based union catalog for the National Digital Archives Program

As shown above, the main functions of the union catalog include:

4.1 Metadata Harvesting and Maintaining

- A. Harvest metadata periodically from each data provider to update the metadata in the union catalog. Download criteria include set, date, identifier and metadata type. Update frequency is subject to parameters.
- B. It is equipped with data provider registration function, which maintains basic information, upload metadata and schema through a friendly interface.
- C. Service provider is a system of web services, providing user search and browse functions.
- D. Service provider can save and maintain a more detailed metadata of each archive institution with metadata cross-mapping function.

4.2 Browse

- E. Access the type, creator, subject, date, place and repository in the metadata and select this information as the category to browse.
- F. Catalogue editing function should be a module that can discern processed and unprocessed data.
- G. Save the image file of every institution as thumbnails that are assembled as a picture index allowing users to conduct searches based on Type, Creator, Subject, Date, Place and Repository. The selected results will be shown in thumbnails that link to the archiving institution, where a larger

picture can be called up.

4.3 Search

- H. The system provides field search, fuzzy search (including words with identical pronunciation, synonyms, traditional and simplified characters), and full text search.
- I. Common search methods include simple search and advanced search. Advanced field search includes title, type, creator, subject, date, place, repository, etc.
- J. The system has Boolean search, truncate search, and full search. Date, type, place, format, repository or other criteria may be used to limit the search range. Restrictions may be processed by clicks to free the users from key-in the data.
- K. Apart from a common interface in the integrative search, users can choose to search data in only selective repositories. The found search result will be presented in a common field, and can be configured for linking with richer metadata.
- L. The system allows the user to select the sorting sequence of data presentation, which can be shown in pages. Available sorting methods include relativity, subject, creator, date, and repository. The data can also be presented in either text or graphs.

4.4 Full Record and Multimedia link management

- M. The system should provide accurate full text or multimedia linking, showing, or playing. Link mechanism includes a link that is transformed into a URL through the handle system.

4.5 Authority and System Management

- N. Service providers can limit the authority in various stages based on the hierarchy of service provider, catalogue maintainer, general user, and data provider log-on editors.

4.6 Other Expansion Function

- O. The service provider can also be a data provider, providing Dublin Core format record to communicate with foreign institutions.
- P. With CCCII , big-5 , UTF-8 transformation function that transforms data sent by different repositories into UTF-8.
- Q. The front page can search and browse as well as integrate a search function for Internet websites as a portal website.

5 Conclusion

Many digital libraries overseas have participated in the OAI-based union catalog,

and many museum systems support the OAI services, for example the Bibliotheque National de France, Great Britain Library, UKOLN, American Digital Library Federation, UC Berkeley, etc. Therefore, OAI-based union catalog will be a foundation for cooperating and communicating with foreign systems and our digital archives initiative. Multi-language support should also be a focal point for future development of our union catalogue system.

Acknowledgement

We would like to acknowledge the support for this project from the National Science Council (NSC) of Taiwan under contract NSC91-2413-H-002-023 and NSC91-2422-H-001-3203. Without this support, this project would not have been possible.

We also thank Mr. Ya-ning Chen and Ms. Shu-jiun Chen for their advice, and all the research assistants of this project: Kuo-shen Su, Ming-jie Liu, Huei Oyang, Chen-ji Chang, Hsing-ron Pan, Hsueh-yun Yian, Jao-ru Chen, Hsiao-ying Hung, Huai-wen Chang, and Tai-hsin Hsu.

References

- [1] Chao-chen Chen (2000), Integrated search theory and practice of digital library (In Chinese). Taipei: wen-hua, 31-34.
- [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. WWW= <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). WWW= <http://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>
- [4] NCL, National Central Library (2001). Meeting of draft technical standard of book information (In Chinese). Taipei: NCL, 3-1.
- [5] Chao-chen Chen, Li-yuan Chen and Wen-hsi Chang (2001). "Principle of digital record naming (In Chinese)" National Central Library Newsletter 90(3). WWW= http://www.ncl.edu.tw/pub/c_news/89/01.html