

教育資料與圖書館學

Journal of Educational Media & Library Sciences

<http://joemls.tku.edu.tw>

Vol. 54 , no. 1 (2017) : 5-29

圖書資訊學開放取用

期刊聯合目錄系統之設計與實作

Design and Implementation of a Library and

Information Science Open Access

Journal Union Catalogue System

林 信 成* Sinn-Cheng Lin*

Professor

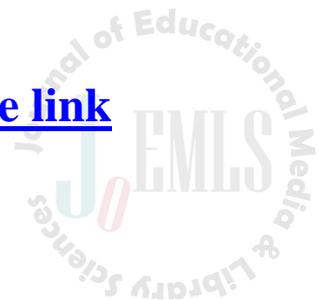
E-mail : sclin@mail.tku.edu.tw

周 庭 郁 Ting-Yu Chou

Postgraduate

[English Abstract & Summary see link](#)

[at the end of this article](#)





圖書資訊學開放取用 期刊聯合目錄系統之設計與實作

林信成^{a*} 周庭郁^b

摘要

開放取用是近年興起的學術傳播模式，但開放取用的學術資源分散於網路各處，使用上仍有所不便。本研究以圖書資訊學為學科主題，以OAIS參考模式為系統架構，以DOAJ與E-LIS兩個開放取用平台為資料來源，藉由系統實作法開發一套「圖書資訊學OA期刊聯合目錄」系統，並以OAI-PMH資料交換協定做為資料互通標準，以LAMP做為開發環境，完成攝入、典藏、管理與取用四大功能模組之設計與開發，並完成系統建置與整合。經實際實驗與驗證，本系統能順利匯集DOAJ與E-LIS中與圖書資訊學相關之OA期刊資源，且順利上線運行，提供圖書資訊學領域之研究者使用。

關鍵詞：開放取用，聯合目錄，OAIS參考模式，OAI-PMH，詮釋資料

前言

近年來全球圖書館皆面臨學術期刊訂費持續高漲，所需經費大幅增加的挑戰，即使逐年追加預算，所能訂購的期刊種數也僅微幅增加甚或不增反減，不但造成研究資源短缺，影響教學研究品質，且可能進一步引發學術傳播危機。有鑒於此，開放取用(Open Access, 簡稱OA)的觀念乃應運而生，且迅速發展。陳亞寧曾探究OA之發展歷史，最早可溯及1991年美國物理學家Paul Ginsparg建立arXiv伺服器，提供學者有關高能物理(high energy physics)方面的學術論文電子刊本(e-prints)服務開始；1994年Steven Harnad提出「顛覆式建議方案」(Subversive Proposal)，探討電子期刊對學術傳播的影響及願景，引起國際圖書館界重視；1999年Harold Varmus向美國國家衛生院(National Institutes of Health, NIH)提出E-biomed規劃案後，逐漸形成今日的OA學術傳播模式。

^a 淡江大學資訊與圖書館學系教授

^b 淡江大學資訊與圖書館學系碩士

* 本文主要作者兼通訊作者：sclin@mail.tku.edu.tw

¹ 2001年12月，開放社會學會(The Open Society Institute，簡稱OSI)於匈牙利首都布達佩斯召開會議，期望藉由國際間共同努力，達成所有學術文章都能免費開放公眾取用的目標，提出「布達佩斯開放取用先導計畫」(The Budapest Open Access Initiative，簡稱BOAI)²，並於2002年2月14日正式對外公佈，成為一項全球性推動開放取用之原則、策略及承諾聲明。如今，OA成為新興的學術傳播方式，利用不斷增加以及免費取得的線上OA資源，圖書館可擴展現存之圖書、期刊、資料庫等以外的資訊來源。

然而，開放取用的期刊論文由於分散於網路各處，若無有效整合，對讀者而言仍有所不便。Pinfield依據英國開放檔案方面的執行成果與經驗，認為「開放」至少有兩種不同意涵：第一種是指資料內容的無限制可得性(unrestricted availability)，如BOAI的開放取用；另一種是指技術方面的系統互通性(Systematic interoperability)，如開放典藏協會(Open Archives Initiative，簡稱OAI)所制訂的詮釋資料交換協定；Pinfield認為這兩者不必然相關。³不過我們認為：從資訊系統的角度而言，為了讓分散於網路各處的OA期刊資料能順利彙整、取用，OA期刊系統(包括匯入、典藏、取用、檢索……等功能模組)之設計應採用開放式架構，若能符合國際間通行的開放典藏資訊系統(Open Archival Information System，簡稱OAIS)參考模式(reference model)則更佳。⁴然而，許多OA期刊系統並未進行如此的系統規劃，以致於造成系統間資源共享與資料交換的困難。

建構於網路環境中的數位圖書館，為了因應電子資源四處分散的情形，經常會建置線上聯合目錄，供讀者進行瀏覽與查詢，並可依據不同的學科、關鍵字或作者……等欄位進一步篩選所需資料。據此，本研究以圖書資訊學做為學科主題，開發系統功能模組，嘗試建置一套符合OAIS參考模式之「圖書資訊學OA期刊聯合目錄」，並以「開放取用期刊目錄」(Directory of Open Access Journals, DOAJ)和「圖書資訊學電子刊本」(E-prints in Library & Information Science, E-LIS)兩個著名的OA平台作為原始的資訊生產者，探討聯合目錄類型及數位圖書館相關檢索協定之優劣與建置方式，接著討論做為資料來源的DOAJ與E-LIS中之詮釋資料格式，並選擇適用的檢索協定與詮釋資料格式。最終目的是讓使用者能夠更容易、更簡單的取得圖書資訊學OA期刊資源。

二、基於OAIS之系統架構

為了彙整圖書資訊學OA期刊資源，本系統將基於OAIS參考模式進行開發，以DOAJ與E-LIS作為原始的資訊生產者，並以開放典藏協會詮釋資料獲取協定(Open Archives Initiative Protocol for Metadata Harvesting，簡稱OAI-PMH)⁵做為系統間的通訊協定，進行資料交換與系統整合。

為了匯集圖書資訊學OA期刊資源，本研究基於OAIS參考模式，提出如圖2所示之系統架構圖。主要是以DOAJ與E-LIS兩個著名的OA資源作為原始的資訊生產者，並依據OAIS參考模式，開發「圖書資訊學OA期刊聯合目錄」所需之典藏、攝入、取用、管理等模組。本系統除做為資訊消費者，從DOAJ與E-LIS獲取資料外，也能做為資訊生產者，提供資料與檢索服務給一般使用者或其他系統。

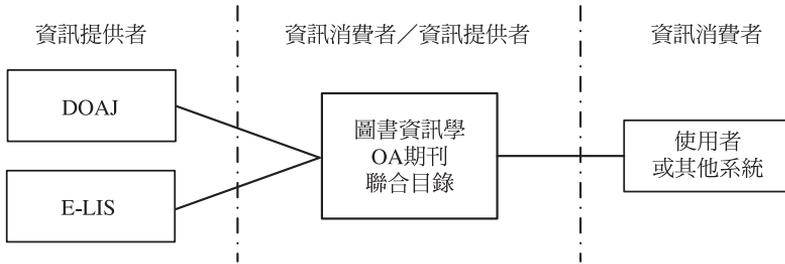


圖2 圖書資訊學OA期刊聯合目錄系統架構圖

資料來源：本研究繪製

本研究擬探討之研究問題如下：

1. 如何設計以DOAJ與E-LIS為資料來源之典藏模組？
2. 如何設計從DOAJ與E-LIS中取得詮釋資料之攝入模組？
3. 如何設計讓資訊消費者獲取詮釋資料之取用模組？
4. 如何設計讓管理員便於管理圖書學OA期刊聯合目錄資料之管理模組？
5. 如何整合上述模組，成為完整的圖書學OA期刊聯合目錄系統？

本研究與筆者2012年發表之「基於OAIS之開放取用期刊管理系統設計與實作」所探討之問題不同之處在於：2012年探討的是作為資訊生產者，開放取用期刊管理系統如何向DOAJ等系統提交資料的問題；而本文探討的是作為資訊消費者，開放取用期刊聯合目錄如何自DOAJ、E-LIS等系統匯集資料的問題。

此外，本研究僅著眼於開發一匯集DOAJ與E-LIS之聯合目錄，至於OAIS參考模式中數位物件的典藏管理、異地備援、長期保存策略等，並不在本文的研究範圍，有興趣的讀者可參閱：林信成、鄭國祥，「數位典藏資訊系統之長期保存規劃與實施」，教育資料與圖書館學，2009年3月，第46卷，第3期，頁323-349一文。

(三) DOAJ與E-LIS

由於開放取用的學術資源分散於網路各處，須要有效整合以方便讀者利用，OA學術資源聯合目錄便應運而生，本研究採用的DOAJ和E-LIS便是其中較知名者。

DOAJ創建於2003年，起因於OA期刊愈來愈多，而有經驗的圖書館員缺

乏時間編目，OA出版者也難以傳播期刊文章。為提升OA期刊的能見度、增加其使用性及影響力，2002年10月在瑞典蘭德大學(Lund University)召開的「第一屆北歐學術傳播會議」(First Nordic Conference on Scholarly Communication)，提出創建一個開放取用期刊目錄的構想，並由蘭德大學於2003年5月推出。⁷ DOAJ計畫目的是希望透過OPAC(Online Public Access Catalogue)支援OA學術出版品，並能夠遠端取用，不需分散找尋。⁸ DOAJ初期僅有300種OA期刊，至今已有近9,000種，學科主題包括科學、技術、醫學、人文與社會科學⁹，已成為匯集全球具備同儕評閱且確保品質的OA期刊整合平台，提供科學研究人員檢索利用，讓使用者能夠一次取足(one stop shop)。¹⁰

E-LIS則是一個圖書資訊學國際數位資料庫，於2003年由一批擁有圖書資訊學專業的志願團隊所建立，現由兩年一期的管理委員會、行政委員會與編輯委員會經營，並由AIMS(Agricultural Information Management Standards)、ULCC(University of London Computer Centre)、INRA(Institut National de la Recherche Agronomique)贊助，D&P Studio Legale提供免費法律諮詢。¹¹ 典藏的資料類型包含預印本(e-prints)、學位論文(thesis)、書籍(book)、書籍章節(book chapter)、書目(bibliography)、指南/手冊(guide/manual)、教程(tutorial)、圖書館教學材料(library instructional material)、會議論文集(conference proceedings)、會議論文(conference paper)、會議海報(conference poster)、演講(presentation)、專案/商業計畫(project/business plan)、報告(report)、部門技術報告(departmental technical report)、技術報告(technical report)、期刊文章(journal article)、評論(review)、報紙/雜誌文章(newspaper/magazine article)、館藏(in collection)、資料集(dataset)與其他(other)。¹²

DOAJ與E-LIS雖同為OA學術資源聯合目錄，但於典藏資料類型與資源描述形式、同儕評閱與資料更新方式上皆有所不同：在典藏資料類型與資源描述形式方面，因DOAJ的典藏資料類型僅包含期刊與期刊文章的詮釋資料，故資源描述形式只使用都柏林核心集(Dublin Core, 簡稱DC)及自訂的DOAJ Article作為其詮釋資料格式(以下簡稱「DOAJ詮釋資料格式」)；而E-LIS則因收錄了包括預印本、學位論文、期刊論文、會議論文、圖書……等多元資料類型，因此綜合採用了DC、數位項目宣告語言(Digital Item Declaration Language, 簡稱DIDL)、詮釋資料編碼與傳輸標準(Metadata Encoding and Transmission Standard, 簡稱METS)、英國電子學位論文(United Kingdom Electronic Theses and Dissertation, 簡稱UKETD)四種詮釋資料格式，可以對應使用在不同類型的資料上。¹³ 而在同儕評閱與資料更新方式上，DOAJ所典藏的資料皆為經同儕評閱之期刊出版者上傳，而E-LIS所典藏的資料皆由一般使用者註冊並上傳，因此不保證其資料經過同儕評閱。

換言之，DOAJ僅收錄OA期刊而不含其他類型資料，如圖書或學位論文

等，而在其所收錄的期刊文章中則包含圖書資訊學類與非圖書資訊學類；而 E-LIS 則僅收錄圖書資訊學類之 OA 資源，但除了期刊還包含圖書或學位論文等各式資源。本研究所擬開發的系統即試圖從上述兩個 OA 平台，自動擷取出與圖書資訊學相關之 OA 期刊資源，捨棄 DOAJ 非圖書資訊學 (Non-LIS)、E-LIS 非期刊 (Non-Journal) 的部分，完成一個「圖書資訊學開放取用期刊聯合目錄系統—LISOAJD」供圖書資訊學領域之研究者使用，如圖 3 所示。

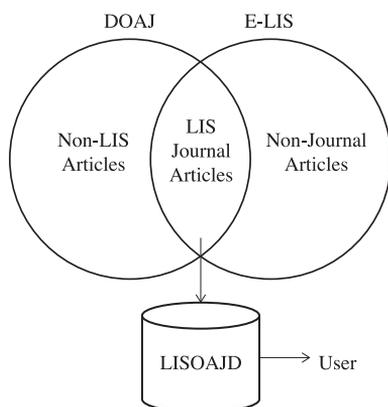


圖3 圖書資訊學OA期刊聯合目錄概念圖

(四) OAI-PMH 通訊協定

開放取用聯合目錄所需具備的重要功能之一是：支援標準化的通訊協定，以便在分散式的異質系統之間進行資料交換。早期線上聯合目錄普遍使用 Z39.50，之後逐漸發展出不同的檢索協定，如：開放典藏協會詮釋資料獲取協定 OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)¹⁴、簡易物件存取協定 SOAP (Simple Object Access Protocol)¹⁵、開放式統一資源定址協定 OpenURL¹⁶、SRU/SRW (Search and Retrieve URL / Search and Retrieve Web)¹⁷……等。¹⁸ 由於本研究擬連結的 DOAJ 與 E-LIS 皆支援 OAI-PMH，因此本系統也將基於 OAI-PMH 進行開發。

OAI-PMH 是由開放典藏協會 (Open Archives Initiative) 所制訂的詮釋資料交換協定。該協定自 1999 年開始發展，至 2002 年 2.0 版逐漸完備，為國際上數位化資源詮釋資料交換標準之一。¹⁹ OAI-PMH 最初的目的是作為學術性電子期刊預印本之互通性檢索，而後逐漸發展成為 Metadata 獲取 (Metadata Harvesting) 的整合機制，在資訊交換過程中提供標準化的通訊協定，以將分散的資源加以匯整，因此可作為數位圖書館、數位博物館、數位典藏庫之間的互通框架，達成分散式數位典藏整合檢索之目的。OAI-PMH 協定具實作容易、開放性，採用 XML 與 HTTP 等開放標準，相容性高……等優點，在歐美已有許多單位進行建置與使用。²⁰

在OAI-PMH架構中，主要是由資料提供者 (data provider)、服務提供者 (service provider) 和兩者之間共通的六個命令集 (verbs) 所構成。資料提供者主要作用是解讀並回應查詢需求，並將儲存器 (Repository) 之資料以符合OAI-PMH協定之 Metadata 格式傳送。服務提供者是一個客戶端應用程式 (client application)，主要任務是以獲取器 (harvester) 向資料提供者發送查詢需求，以便從儲存庫獲取所需數位典藏品的 Metadata，並以所獲得的 Metadata 建立加值服務。而兩者之間共通的六個命令集則分別為 Identify、ListMetadataFormats、ListSets、ListIdentifiers、ListRecords 及 GetRecord：²¹

- Identify：獲取器藉此命令取得儲存器之識別資訊。
- ListMetadataFormats：獲取器藉此命令取得儲存器所支援的 Metadata 格式。
- ListSets：獲取器藉此命令取得儲存器之資料集。
- ListIdentifiers：獲取器藉此命令取得儲存器於指定期間的典藏品識別資料。
- ListRecords：獲取器藉此命令取得儲存器於指定期間的 Metadata 資料。
- GetRecord：獲取器藉此命令獲取儲存器中單筆的 Metadata 資料。

(五) 系統開發環境

本研究基於 LAMP (Linux、Apache、MySQL、PHP) 環境開發並建置「圖書資訊學 OA 期刊聯合目錄」，詳細規格如表 1 所示。

表 1 系統開發環境

軟體環境	規格
Linux	作業系統：Centos 6.5 記憶體：31.2GiB 處理器：Intel(R) Xeon(R) CPU E5-2620 v2@2.10GHz
Apache	版本：2.2.15
MySQL	版本：5.1.73 協定版本：10 字元集：UTF-8
PHP	版本：5.3.3

三、系統模組設計與開發

本系統之設計與開發，共分為四大部分：(一)典藏模組與資料庫設計；(二)攝入模組與剖析器設計；(三)取用模組設計與流量控制；(四)管理模組設計。系統整合架構如圖 4 所示，讀者不難將其與 OAIS 參考模式對應。以下分述其開發過程與成果。

(一) 典藏模組與資料庫設計

典藏模組對應於 OAIS 參考模式之典藏儲存庫角色，本研究採用 MySQL 資料庫儲存 OA 期刊之詮釋資料。為了建置能夠儲存 DOAJ 與 E-LIS 之 OA 期

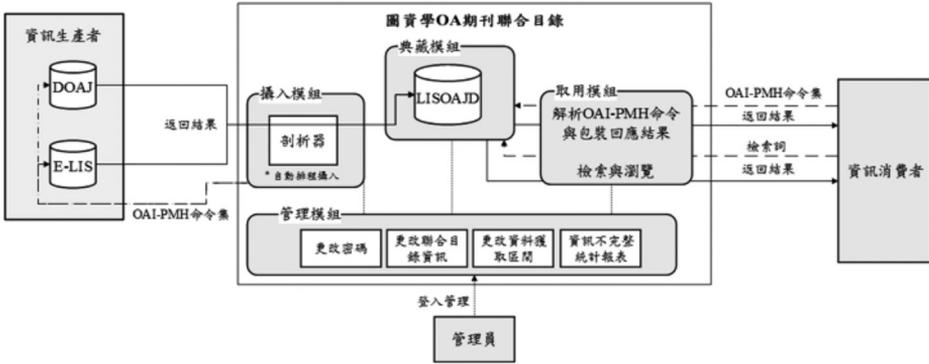


圖4 圖書資訊學OA期刊聯合目錄系統整合圖

刊詮釋資料的資料庫，需對兩者之詮釋資料格式及著錄方式有所了解。如前所述DOAJ採用了DC與自訂的DOAJ詮釋資料格式；而E-LIS則採用了DC、DIDL、METS與UKETD，除了DC之外，其他三者都應用於非期刊資料，不予討論。

經分析DOAJ資料庫中，以自訂的DOAJ詮釋資料格式(以下稱為oai_doaj)著錄期刊文章，所使用的標籤包括：language、publisher、journalTitle、issn、eissn、publicationDate、volume、issue、startPage、endPage、doi、publisher、RecordId、title、authors、abstract、fullTextUrl、keywords等。其中較特別的是：RecordId著錄該筆詮釋資料於原系統的唯一識別碼；而authors為複數標籤，因每篇文章可能不只一位作者，其下又以單數標籤author分隔每作者的基本資料，包括：name、email與affiliationId等，而affiliationId又對應到affiliationsList中分隔不同作者服務單位的標籤affiliationName之屬性affiliationId；同樣的，複數標籤keywords之下也是再以單數標籤keyword分隔不同的關鍵字。

而DOAJ資料庫中，以DC著錄的期刊文章詮釋資料(以下稱為oai_dc)，其標籤包括：title、creator、subject、description、publisher、date、type、source、rights、identifier、relation、language等。由於DC在使用上保有彈性，以致於DOAJ在著錄時，自訂了一些較特別的部分，如：以subject作為關鍵字欄位，並以屬性「xsi:type」著錄dcterms:LCSH，標示是否為美國國會圖書館標題表的詞彙；source則作為複合欄位，著錄期刊題名、卷、期、頁數與出版年份；identifier則分別著錄ISSN、EISSN、DOI與該文章於DOAJ中頁面的網址；relation則包含全文檔案網址與該文章的期刊於DOAJ中頁面的網址。

E-LIS雖然也採用DC著錄期刊文章(同樣稱為oai_dc)，但在使用方式上，自訂了一些與DOAJ不同之處，如：subject為E-LIS自行設定的主題分類；type為資料類型與是否經同儕評鑑；identifier標籤則重複著錄兩次，一為作者、期刊文章題名、期刊題名、文章出版年、卷期、頁數與資料類型，另一identifier則為全文檔案網址；relation為該文章於E-LIS中的頁面網址。

茲將上述分析整理成對應表，再依系統需求拆分為四個資料表，分別命名為 records、authors、keywords 和 rights，並規劃其欄位如表 2 所示：

表 2 DOAJ 與 E-LIS 詮釋資料欄位對應表

資料表欄位		DOAJ			E-LIS
資料表	欄位	oai_doaj		oai_dc	oai_dc
records	title	title			title
	abstract	abstract			description
	publisher	publisher			publisher
	publicationDate	publicationDate			date
	language	language			language
	journalTitle	journalTitle			source
	fullTextUrl	fullTextUrl			relation
	starPage	starPage			source
	endPage	endPage			source
	volume	volume			source
	issue	issue			source
	issn	issn			identifier
	eissn	eissn			identifier
	doi	doi			identifier
	format	format (fullTextUrl attribute)			format
PeerReviewed	N/A			N/A	
authors	name	authors	author	name	creator
	email			email	N/A
	affiliationId			affiliationId	N/A
	affiliationsName	affiliationsList		affiliationName	N/A
keywords	keyword	keywords		keyword	subject
	LCSH			N/A	xsi:type (subject attribute)
rights	right	N/A			rights

資料來源：本研究整理。

其中，records、authors、keywords 和 rights 資料表各以 recordid、authorid、keywordid 和 rightid 作為主鍵 (primary key, 簡稱 PK) 互相關聯，如圖 5 所示。此外，另有 admin、resumptiontoken、metadataformat、identify 四個獨立資料表作為其他模組之用。

(二) 攝入模組與剖析器設計

DOAJ 與 E-LIS 皆藉由 OAI-PMH 提供詮釋資料予資訊消費端。當本系統透過 OAI-PMH 從 DOAJ 與 E-LIS 獲取詮釋資料後，需藉由剖析器剖析，並將符合條件之圖書資訊學 OA 期刊詮釋資料存入資料庫，此即攝入模組之功能需求。

本模組之剖析器開發，需依據 DOAJ 與 E-LIS 詮釋資料著錄方式，再以 PHP 編寫剖析器程式碼，接著實際執行測試，若遭遇執行錯誤或無法成功剖析等問題，則對程式碼進行修改，直到能夠成功剖析詮釋資料為止。本模組係以 XML 的 DOM 方法，創建 DOMDocument 物件，並載入 XML 資料，再依據

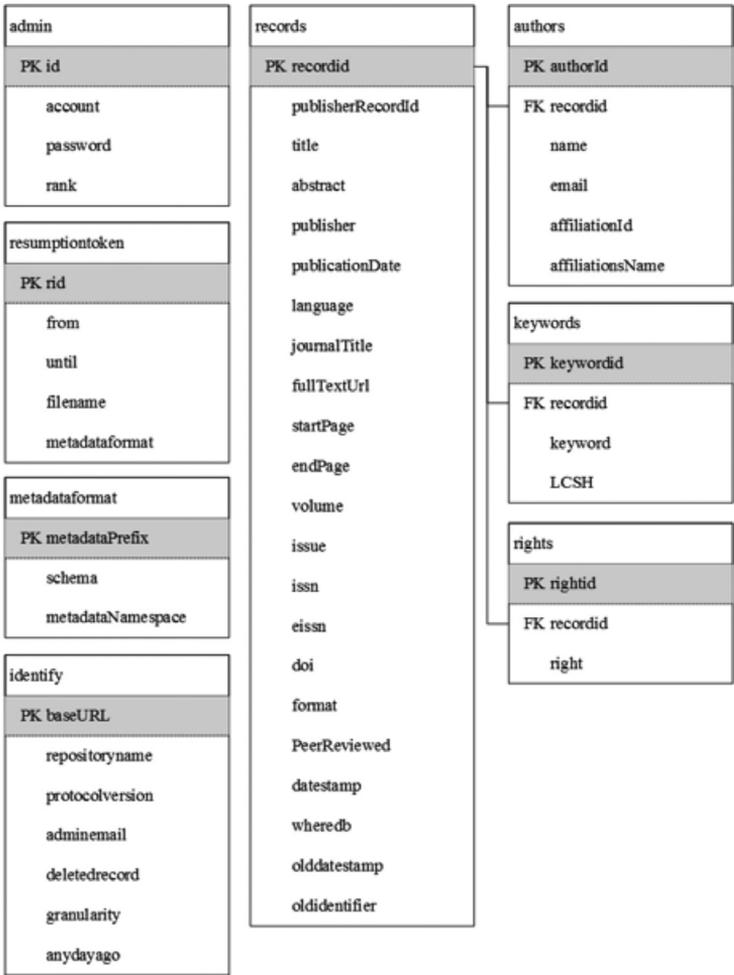


圖5 本系統之資料庫結構

XML 樹狀結構，將父標籤下的子標籤以陣列形式拆解後再逐項解析，最後取得標籤中的資料。

然而，在開發過程中不斷發現無論 DOAJ 或 E-LIS，在詮釋資料格式標準與詮釋資料的品質上皆存在許多差異與問題，甚至連 OAI-PMH 部分亦不符規格書的規範，以致造成攝入模組與剖析器開發複雜度與難度增加。這當中有些問題本研究已藉由撰寫程式一一克服並解決，但有些問題則是 DOAJ 與 E-LIS 資料品質問題。例如，來自 DOAJ 與 E-LIS 文章各項詮釋資料常有不一致情形，由於難以判斷何者正確，本系統以直接存入資料庫的方式處理，不另外清理不一致情形，且由於 DOAJ 與 E-LIS 兩者對於詮釋資料格式的定義不同，相同標籤中可能會存放類型迥異的資料，皆使得資料可能被存到不同欄位。此部分值得未來進一步對兩者資料品質進行探討，以下僅就本研究能處理之處加以說明。

1. 複合資料剖析問題

無論 DOAJ 或 E-LIS，採用 DC 所著錄的 oai_dc 詮釋資料，皆發現有單一標籤包含複合資料的情況：如 DOAJ 的 oai_dc 格式，在標籤 source 中包含了期刊題名、卷、期、頁數與出版年份，如「Journal of Information Science Theory and Practice, Vol 4, Iss 1, Pp 49-64 (2016)」，全部集中在一個標籤。E-LIS 的 oai_dc 格式，則在標籤 identifier 中包含作者姓名、期刊文章題名、期刊題名、出版年份、卷、期、頁數與資料類型，如「G. Figuerola, Carlos and Gomez Díaz, Raquel and López de San Roman, Eva Stemming and n-grams in Spanish: An evaluation of their impact on information retrieval. Journal of Information Science, 2000, vol. 26, n. 6, pp. 461-467. [Journal article (Print/Paginated)]」，包含了更多不同性質的資料。因此，除了將資料從標籤中剖析出來外，本研究另撰寫額外的程式碼來解析複合資料，以便順利將資料存入適當的資料表欄位中。

2. 詮釋資料格式標準不一問題

DOAJ 與 E-LIS 共同皆使用的詮釋資料格式為 DC，但包含之內容與格式卻存在著差異，雖然同為 DC，但 DOAJ 與 E-LIS 在 DC 的 15 項標籤中，便有 9 項包含的內容與格式有差異，分別為 subject、type、source、relation、rights、date、format、identifier 與 language，整理如表 3 所示。

表 3 DOAJ 與 E-LIS 使用 DC 標籤異同比較表

DC 標籤	DOAJ	E-LIS
title	文章題名	文章題名
subject	關鍵字(包含美國國會圖書館標題表)	E-LIS 自行設定的主題分類
description	文章摘要	文章摘要
type	資料類型(預設值為 article)	資料類型(期刊為 Journal article[線上或紙本/是否編頁碼])以及是否有通過同儕評鑑
source	期刊題名、卷期、頁數及出版年	未使用
relation	全文網址	文章資料於 E-LIS 中的連結頁面網址
coverage	未使用	未使用
creator	文章作者	文章作者
publisher	期刊出版者	期刊出版者
contributor	未使用	未使用
rights	文章授權方式	未使用
date	文章出版年月日(YYYY-MM-DD)	文章出版年月日(格式不定, YYYY 或 YYYY-MM 或 YYYY-MM-DD)
format	未使用	全文檔案格式
identifier	ISSN、EISSN、DOI 以及文章資料於 DOAJ 的連結網址	全文網址以及文章的期刊題名、卷期、頁數、ISSN 及出版年
language	文章出版語言(English、Chinese)	文章出版語言(en、zh)

本研究之攝入模組雖已依據DOAJ與E-LIS詮釋資料著錄方式，開發符合其格式的剖析器，但發現DC雖能廣泛的讓各主題擁有統一標準之詮釋資料格式，但在給予使用者有較寬裕的自由去詮釋典藏品資料的情況下，卻也因此造成各自解讀其定義及用法的情形，使得統一的標準形同虛設，也徒增剖析器設計的困難度與複雜性。

3. 資料缺漏問題

攝入模組開發過程中，亦發現DOAJ與E-LIS的資料有嚴重缺漏的情形，使得所獲取的詮釋資料有誤、重複或無法正確整併，也影響本系統的資料品質。截至2016年12月18日止，本系統共收錄期刊688種，期刊文章11,491篇。從攝入資料的過程中，整理出DOAJ與E-LIS的缺漏狀態共12項，分別為未填出版者、起始頁、結束頁、頁數、卷、期、ISSN、EISSN與作者姓名，以及作者單位、關鍵字與摘要資訊不全等，若將其各項缺漏狀態百分比繪製成圖，則如圖6所示。從圖6也可發現E-LIS在未填ISSN、EISSN、起始頁、頁數、卷、期與出版者等7項，缺漏百分比皆高於DOAJ，顯示E-LIS較無顧及資料品質。DOAJ與E-LIS資料品質問題並非本系統能解決，建議學術傳播領域研究者未來可進一步探討之。

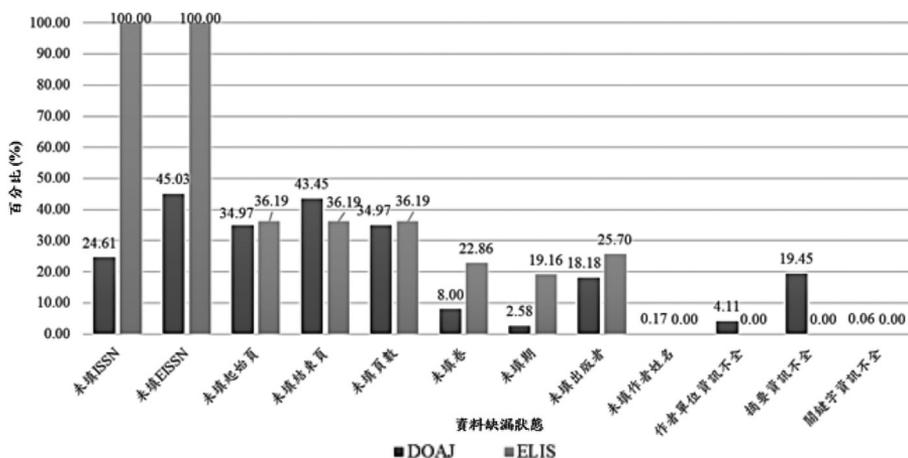


圖6 DOAJ與E-LIS各項缺漏狀態百分比

4. 未完全遵循OAI-PMH規範問題

OAI-PMH的規格書中，明確標示各個功能、要求、回應與屬性，但本系統在實際獲取DOAJ和E-LIS資料時，發現兩者並未完全遵循OAI-PMH規範。例如，標籤resumptionToken必須具備屬性completeListSize、cursor與expirationDate，但E-LIS並未具備上述三個屬性；另外，resumptionToken編號在超過expirationDate後，應立即失效，但DOAJ與E-LIS的resumptionToken並不會因超過expirationDate而失效，都還能夠再次使用。又例如，OAI-PMH為適用

於集中式聯合目錄的檢索協定，擁有整併及儲存重複資料為單筆紀錄的功能，但若DOAJ與E-LIS並沒有特別檢查重複與不完整的資料，即便OAI-PMH擁有去除重複與獲取最正確資料的特性，也會因為DOAJ與E-LIS原始資料的缺陷，造成仍然獲取重複與錯誤資料的情形。

雖然OAI-PMH不具強制性，不遵守部分規範實際上並不影響系統運作，但OAI-PMH標準的本意是期望能使資料更容易互通，加速資訊傳播的速度，使OA學術資源能見度更加提高，若不遵循其規範，則徒增系統開發的困擾。

(三) 取用模組設計與流量控制

本系統除了獲取DOAJ與E-LIS的詮釋資料之外，也能夠轉換角色，作為資料提供端，以OAI-PMH的提供資料予資訊消費端，此即取用模組之功能需求。

首先，取用模組最重要的是要能解析OAI-PMH六項命令集，亦即在接收資訊消費端以URL送來的指令後，開始判斷所使用的verb及引數是否合法，意即是否為Identify、ListMetadataFormats、ListSets、ListIdentifiers、ListRecords與GetRecord其中一項命令，並判斷其引數是否包含在命令集定義的引數中；若判斷為合法，便進一步查詢資料庫中是否包含符合條件的資料；若有則依據該命令，以相對應的詮釋資料格式回應所需資料並結束程序；而以上三個步驟若判斷為否，下一步皆顯示錯誤訊息，並且結束程序。

此外，值得注意的是：在提供詮釋資料的過程中，因DOAJ與E-LIS期刊文章的語言眾多，使用的編碼也不盡相同，故造成在提供詮釋資料時，會因特殊符號或未定義字碼而產生網頁錯誤，如期刊題名中包含「&」或題名與摘要中包含「，」或「”」。本系統使用「<![CDATA[[]]>」，將出現特殊符號的內容封裝於其中，以解決此一問題。

本系統取用模組之baseURL為「http://webserver.dils.tku.edu.tw/oai/oai.php」，六個命令集及引數可分別伴隨於「?verb=」之後。例如「http://webserver.dils.tku.edu.tw/oai/oai.php?verb=Identify」可取得本系統之識別資料；而「http://webserver.dils.tku.edu.tw/oai/oai.php?verb=ListMetadataFormats」則可得知本系統所支援之詮釋資料格式。取用模組在判斷命令及引數正確後，即以OAI-PMH所規範的格式回應所需的詮釋資料。

茲以GetRecord為例：

```
http://webserver.dils.tku.edu.tw/oai/oai.php?verb=GetRecord&identifier=5&metadataPrefix=oai_doaj
```

其中，「verb=GetRecord」表示資訊消費端想獲取單筆資料，而「identifier=5」則進一步指出想獲取識別碼5號之資料，「metadataPrefix=oai_doaj」則指定以oai_doaj格式回應詮釋資料。本系統之取用模組對該命令的回應如下：

```

<OAI-PMH xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.
openarchives.org/OAI/2.0/" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://
www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
<request verb="GetRecord" metadataPrefix="oai_doaj" identifier="5">http://webserver.dils.tku.
edu.tw/oai/oai.php</request>
<responseDate>2016-04-11T01:31:10Z</responseDate>
<GetRecord>
<record>
  <header xmlns:oai_doaj="http://doaj.org/features/oai_doaj/1.0/">
    <identifier>5</identifier>
    <datestamp>2016-03-31</datestamp>
    <setSpec>VOIEVOITQVdO=Mw==</setSpec>
  </header>
<metadata xmlns:oai_doaj="http://doaj.org/features/oai_doaj/1.0/">
  <oai_doaj:doajArticle xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.
openarchives.org/OAI/2.0/OAI-PMH.xsd http://doaj.org/features/oai_doaj/1.0/ https://doaj.org/
static/doaj/doajArticles.xsd">
    <oai_doaj:language>en</oai_doaj:language>
    <oai_doaj:publisher><![CDATA[ SAPIENS ]]></oai_doaj:publisher>
    <oai_doaj:journalTitle><![CDATA[ WIDWISAWN ]]></oai_doaj:journalTitle>
    <oai_doaj:publicationDate>2005</oai_doaj:publicationDate>
    <oai_doaj:volume><![CDATA[ 3 ]]></oai_doaj:volume>
    <oai_doaj:issue><![CDATA[ 2 ]></oai_doaj:issue>
    <oai_doaj:title><![CDATA[Mapping the Scottish Common Information Environment, with a
little help from Google]]></oai_doaj:title>
    <oai_doaj:authors>
      <oai_doaj:author>
        <oai_doaj:name><![CDATA[ Dunsire, Gordon ]]></oai_doaj:name>
      </oai_doaj:author>
    </oai_doaj:authors>
    <oai_doaj:abstract><![CDATA[This article describes how the simple incorporation of Google
Maps within information services can add value for users.]]></oai_doaj:abstract>
    <oai_doaj:fullTextUrl>http://eprints.rclis.org/3819/1/issue3_2.pdf</oai_doaj:fullTextUrl>
    <oai_doaj:keywords>
      <oai_doaj:keyword><![CDATA[ IZ. None of these, but in this section. ]]></oai_
doaj:keyword>
    </oai_doaj:keywords>
  </oai_doaj:doajArticle>
</metadata>
</record>
</GetRecord>
</OAI-PMH>

```

若改以「 metadataPrefix=oai_dc 」則指定以 oai_dc 格式回應詮釋資料：

```
http://webserver.dils.tku.edu.tw/oai/oai.php?verb=GetRecord&identifier=5&metadataPrefix=oai_dc
```

則本系統之取用模組的回應如下：

```

<OAI-PMH xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.
openarchives.org/OAI/2.0/" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://
www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
<request verb="GetRecord" metadataPrefix="oai_dc" identifier="5">http://webserver.dils.tku.edu.
tw/oai/oai.php</request>
<responseDate>2016-04-11T01:37:46Z</responseDate>
<GetRecord>
<record>
  <header xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:oai_dc="http://www.openarchives.
org/OAI/2.0/oai_dc/">
    <identifier>5</identifier>
    <datestamp>2016-03-31</datestamp>

```

```

<setSpec>V0IEV0ITQVdO>Mw==>Mg==</setSpec>
</header>
<metadata xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/">
  <oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/ http://www.openarchives.org/OAI/2.0/OAI-PMH.xsd">
    <dc:title>Mapping the Scottish Common Information Environment, with a little help from Google</dc:title>
    <dc:date>2005</dc:date>
    <dc:relation>http://eprints.rclis.org/3819/1/issue3_2.pdf</dc:relation>
    <dc:publisher><![CDATA[ SAPIENS ]]></dc:publisher>
    <dc:type>Journal article</dc:type>
    <dc:description><![CDATA[This article describes how the simple incorporation of Google Maps within information services can add value for users.]]></dc:description>
    <dc:creator><![CDATA[ Dunsire, Gordon ]]></dc:creator>
    <dc:subject><![CDATA[ IZ. None of these, but in this section. ]]></dc:subject>
    <dc:language>en</dc:language>
    <dc:source><![CDATA[ WIDWISAWN, Vol 3, Iss 2, Pp - (2005) ]]></dc:source>
  </oai_dc:dc>
</metadata>
</record>
</GetRecord>
</OAI-PMH>

```

在資料取用的過程中，由於 ListSets、ListIdentifiers 與 ListRecords 的資料量較為龐大，因此在提供資料給資料消費端取用時，可實施流量控制，以分批獲取資料。OAI-PMH 協定中控制流量之參數為 resumptionToken，但並未硬性規定此參數之格式，而由各系統設計者彈性運用。本研究依據 OAI-PMH 官網之說明，採用號碼牌方式，由資料提供端給予一個號碼牌作為註記，而資料消費端以該參數回傳號碼牌時，資料提供端則回應接續於上批資料之後的紀錄。

(四) 管理模組設計

本系統之管理模組主要提供更改系統資訊、更改資料獲取區間以及查看資訊不完整統計報表、管理員登入登出及修改密碼等功能。此外，因本系統之資料來源皆為 DOAJ 與 E-LIS，若資料有缺陷或重複乃 DOAJ 與 E-LIS 來源端問題，因此，本模組不提供新增、刪除與修改詮釋資料等功能。

1. 更改系統資訊

更改系統資訊的功能使管理員能夠登入修正 OAI-PMH 命令 Identify 中的 repositoryName、baseURL 與 adminEmail 等資訊。repositoryName 是本系統名稱；baseURL 則是本系統取用模組之網址；adminEmail 是管理員的電子信箱。其他資訊如協定版本、是否記錄刪除狀態及時間表示形式三項，則於程式內設定，不開放管理員更動。

2. 更改資料獲取區間

更改資料獲取區間可以讓管理員改變本系統每日執行的 ListRecords 命令之 from 與 until 間隔，因獲取資料的時間有時區限制，也因本系統時間表示形式僅使用年、月、日，未使用到小時與分秒以下單位，若資料攝入有所遺漏情況，

可自行調整區間，獲取排程於下一次執行時，以新的時間區間去獲取DOAJ與E-LIS中符合條件的詮釋資料。

3. 資訊不完整統計報表

資訊不完整統計報表可以讓管理員對資料缺漏情形一目了然，分別為未填出版者、起始頁、結束頁、頁數、卷、期、ISSN、EISSN與作者姓名，以及作者單位、關鍵字與摘要資訊不全。此報表允許另外點擊連結，進一步了解缺漏情況詳細資料。

透過修改密碼、更改系統資訊、更改資料獲取區間、資訊不完整統計報表與登出五個功能，給予管理員在不改變DOAJ與E-LIS詮釋資料的情況下，能夠管理典藏資料，修改聯合目錄資訊，了解典藏資料的缺陷。

四、系統整合與驗證

在上述攝入模組、典藏模組、管理模組與取用模組完成後，本研究依據OAIS參考模式完成系統整合，如圖4所示(系統網址：<http://webserver.dils.tku.edu.tw/>)。本系統之攝入模組能夠傳送OAI-PMH命令集至DOAJ與E-LIS；DOAJ與E-LIS回傳回應結果，經攝入模組的剖析器剖析及篩選後，將符合條件的中英文圖書資訊學OA期刊文章之詮釋資料，儲存至典藏模組之LISOAJD資料庫。本系統將資料獲取頻率設為每日一次，並以ListRecords取得DOAJ與E-LIS獲取資料日與獲取資料日前兩天，亦即前後三天之內所修改或新增的資料。為避免資料過時，若在獲取資料時發現該筆資料已擷取過，則不論是否有修改，一律將原有資料覆蓋過去，以確保資料為最新狀態。

另一方面，資訊消費端也能發送OAI-PMH命令集給本系統，透過本系統之取用模組取得符合其條件之詮釋資料，或透過本系統提供之Web介面檢索與瀏覽本系統由DOAJ及E-LIS所匯集之期刊文章詮釋資料；而管理員可以透過管理模組登入進行系統管理。

圖7為整合系統首頁，截至2016年12月18日止，本系統收錄來自DOAJ之期刊有36種、E-LIS有665種、兩者重覆的有13種，期刊總數共688種；收錄之期刊文章來自DOAJ者有9,514篇、E-LIS者有1,977篇，期刊文章總數共11,491篇。

進入本系統之查詢畫面後，可以選擇文章名稱、文章摘要、作者、關鍵字、期刊名稱、出版者、ISSN與EISSN進行檢索；查詢結果呈現期刊文章的題名、作者、期刊題名、是否經同儕評鑑、ISSN、EISSN、出版日期、卷、期、頁數、全文連結按鈕、全文格式、語言、出版者、摘要、關鍵字、授權方式及資料來自DOAJ或E-LIS，如圖8所示。

本系統之介面設計採簡約風格，主要以使用者能查詢與瀏覽期刊文章資料，並能明確知道該筆資料來自DOAJ或E-LIS；此外，本系統也能透過連結取

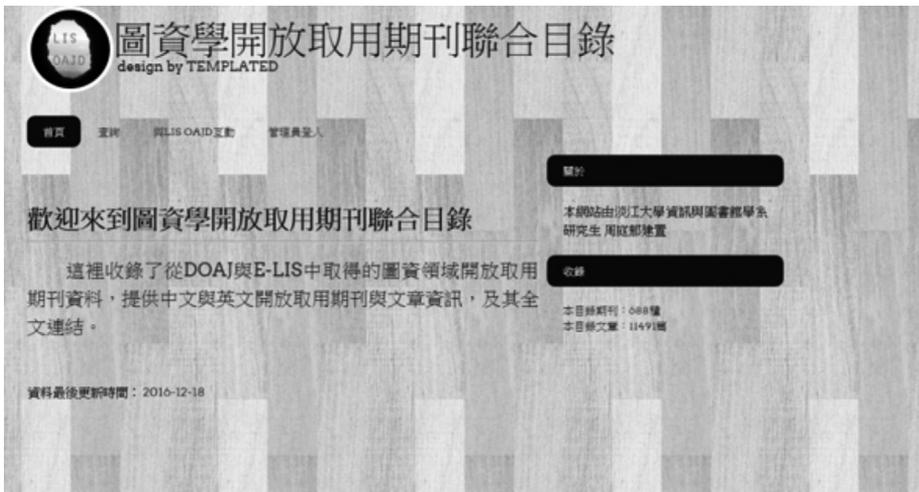


圖7 圖書資訊學OA期刊聯合目錄系統首頁

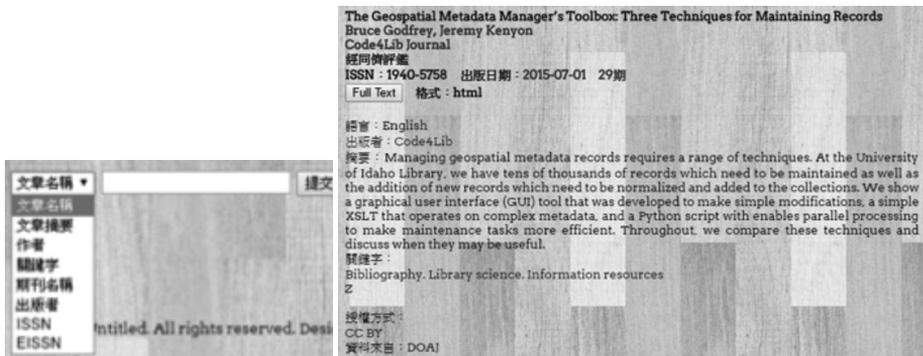


圖8 本系統查詢介面(左)與結果呈現(右)

得OA全文，並提供OAI-PMH命令集檢索，使本系統除了作為服務提供者外，也能變換角色，成為資料提供者，使開放取用資料能夠再次被傳播出去。

最後，為了驗證本系統是否符合OAI-PMH規範，本研究以OAI-PMH Validator & data extractor Tool (<http://validator.oaipmh.com/>) 進行系統驗證，該網站可驗證本研究所開發之系統是否符合OAI-PMH、DC與ESE (Europeana Semantic Elements) 三者之規定。驗證結果，六個命令集Identify、ListMetadata-Formats、ListSets、ListIdentifiers、ListRecords與GetRecord之HTTP狀態皆正常，內容類型、編碼與XML內容驗證結果皆為合法，並能夠正常連結OAI-PMH的XML綱要定義，回應時間也都於正常範圍。以ListSets驗證結果為例，如圖9所示。

從驗證結果顯示本系統具備正式上線運作的能力，從DOAJ與E-LIS中匯集圖書資訊學OA期刊資源，並提供合法且符合格式的OAI-PMH取用途徑，使需要的使用者能夠從該途徑以OAI-PMH獲取所需資料。



圖9 系統驗證結果—以ListSets為例

五、結 論

本研究以OAIS參考模式建置圖書資訊學OA期刊聯合目錄系統，兼顧資訊生產者、管理者、資訊消費者及典藏系統本身的需求，分別實作四大模組，並完成系統整合。茲依據研究問題歸納研究結論如下。

(一) 典藏模組之資料庫結構設計須配合DOAJ與E-LIS詮釋資料格式

由於DOAJ與E-LIS使用了不同的詮釋資料格式，因此，資料庫的資料表結構與欄位設計須依據兩者之異同，才能設計出同時容納DOAJ與E-LIS詮釋資料的資料表欄位。

(二) 攝入模組之剖析器設計須兼顧XML標籤之語法及語意

因為DOAJ與E-LIS在使用XML標籤的著錄方式上有部分差異，對於詮釋資料格式也有不同著錄方式，造成即使是相同名稱的標籤，也可能包含不同性質的詮釋資料。因此建置剖析器時，須要分別為DOAJ及E-LIS開發對應的剖析器，才能使詮釋資料被正確攝入到對應的資料表欄位。

(三) 取用模組須能解讀OAI-PMH命令集且將資料封裝為XML格式並管控流量

取用模組解讀OAI-PMH命令集並以相對應的詮釋資料格式回應結果，資訊消費者亦可依據需求指定詮釋資料之封裝格式。此外，本模組也能以resumption Token限定資料流量，若要獲取第二批以後的詮釋資料，則須以resumptionToken重新傳送。

(四) 管理模組設計應包含協助管理員管理本系統之功能

本系統之管理模組提供管理員修改密碼、系統資訊、獲取資料之基礎網址、管理員的電子信箱以及獲取資料的時間區間等功能，並隨時依據資料庫的詮釋資料情形，顯示詮釋資料資訊不完整統計報表，協助管理員了解資料品質。

(五) 本系統可依據 OAIS 參考模式進行整合

依據 OAIS 參考模式，本研究整合了攝入模組、典藏模組、管理模組與取用模組為網站形式，完成圖書資訊學 OA 期刊聯合目錄系統。經 OAI-PMH 驗證器驗證，結果顯示本系統符合 OAI-PMH 之規範。

最後，值得一提的是本系統雖未實做出所有 OAIS 架構的每一部分，但因 OAIS 係為資訊系統開發之參考模式，可採部分遵循方式為之，並非一定要將所有功能皆予實踐才能聲稱該系統乃依據 OAIS 參考模式架構設計。此外，本研究雖僅處理 DOAJ 和 E-LIS 的詮釋資料，對於 OAIS 參考模式中非常重要的數位物件典藏管理部分並未觸及。²² 而對於 OAIS 的保存規劃與管理部分，用來處理異地備援、各種長期保存策略的規劃與實施，亦未在本文研究範圍中。²³ 本研究雖為實務性論文，卻是後續研究之基礎建設，本文之重點不在於廣泛探討過去之相關研究並進行「創新」，而在於依現有的規範進行「實踐」。系統實作不僅是一種研究方法，實踐過程本身也有許多需要克服的困難與值得研究的問題，且完善的資訊工程建設實為資訊科學研究之基礎。未來在此系統之下，可進行開放取用期刊之大數據分析，此部分將於未來另以專文探討，並強化理論層次的論述，將資訊工程建設與資訊科學研究更緊密的結合。

註 釋

1. 陳亞寧，「開放式資訊取用之現況發展分析」，圖書與資訊學刊，51期（2004年11月）：91，檢索於2016年7月1日，http://www.lib.nccu.edu.tw/blis/fulltext/51/51_7.pdf。

2. Budapest Open Access Initiative, “Read the Budapest Open Access Initiative,” Budapest Open Access Initiative, February 14, 2002, accessed July 1, 2016, <http://www.soros.org/openaccess/read.shtml>.

3. Stephen Pinfield, “Open Archives and UK Institutions,” *D-Lib Magazine*, 9, no. 3 (March 2003), accessed July 1, 2016, <http://dlib.ejournal.ascc.net/dlib/march03/pinfield/03pinfield.html>.

4. 林信成、黃文彥，「基於OAIS之開放取用期刊管理系統設計與實作」，教育資料與圖書館學 49卷，4期（2012年7月）：508，檢索於2016年7月2日，<http://joemls.dils.tku.edu.tw/fulltext/49/49-4/507-532.pdf>。

5. Carl Lagoze et al., eds., “The Open Archives Initiative Protocol for Metadata Harvesting,” Open Archives Initiative, June 14, 2002, accessed July 1, 2016, <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

6. Consultative Committee for Space Data System, *Reference Model for an Open Archival Information System (OAIS)* (Washington: CCSDS, 2002), i-iii, accessed July 1, 2016, <https://public.ccsds.org/Publications/Archive/650x0b1.pdf#search=650%2E0%2DB%2D1>.

7. Linnéa Stenson, “Why all these directories? An introduction to DOAJ and DOAB,” *Insights* 25, no. 3 (November 2012), accessed June 23, 2016, <http://doi.org/10.1629/2048-7754.25.3.251>.

8. Anna Hood, "Bringing Open Access Journals into the Catalog," in *Growth, Creativity and Collaboration: Great Visions on a Great Lake: Proceedings of the North American Serials Interest Group, Inc, 19th Annual Conference, June 17-20 2004, Milwaukee, Wisconsin*, ed. Patricia Sheldahl French and Margaret Mering (New York: Haworth Information Press, 2005): 351.

9. Directory of Open Access Journals, "About DOAJ," Directory of Open Access Journals, 2016, accessed June 23, 2016, <https://doaj.org/about>.

10. Directory of Open Access Journals, "Directory of Open Access Journals (DOAJ)," Directory of Open Access Journals, 2011, accessed December 28, 2011, <http://www.doaj.org/>.

11. E-prints in Library & Information Science, "About Us," E-prints in Library & Information Science, n.d., accessed June 23, 2016, <http://eprints.rclis.org/information.html>.

12. E-prints in Library & Information Science, "Advanced Search," E-prints in Library & Information Science, n.d., accessed June 23, 2016, <http://eprints.rclis.org/cgi/search/advanced>.

13. E-prints in Library & Information Science, "OAI 2.0 Request Results," E-prints in Library & Information Science, n.d., accessed June 23, 2016, <http://eprints.rclis.org/cgi/oai2?verb=ListMetadataFormats>.

14. Carl Lagoze et al., "The Open Archives Initiative Protocol for Metadata Harvesting," accessed July 15, 2016.

15. W3C, "Web Services Activity," W3C, May 8, 2011, accessed July 15, 2016, <http://www.w3.org/2002/ws/>.

16. Herbert Van de Sompel, Patrick Hochstenbach, and Oren Beit-Arie, "OpenURL Syntax Description," Caltech Library, November 5, 2001, accessed July 15, 2016, http://library.caltech.edu/openurl/Record_Documents/OpenURL_Version_0.1.mht

17. The Library of Congress, "SRU: Search/Retrieval via URL," The Library of Congress, n.d., accessed July 15, 2016, <http://www.loc.gov/standards/sru/>

18. Jeroen Bekaert and Herbert Van de Sompel, "A Standards-based Solution for the Accurate Transfer of Digital Assets," *D-Lib Mag* 11, no. 6 (June 2005), accessed July 15, 2016, <http://dx.doi.org/10.1045/june2005-bekaert>; Jeroen Bekaert and Herbert Van de Sompel, "Access Interfaces for Open Archival Information Systems based on the OAI-PMH and the OpenURL Framework for Context-Sensitive Services," arXiv.org, September 28, 2005, accessed July 15, 2016, <http://arxiv.org/abs/cs/0509090>.

19. Open Archives Initiative, "The Open Archives Initiative Protocol for Metadata Harvesting," accessed July 15, 2016.

20. Open Archives Initiative, "OAI Community," Open Archives Initiative, n.d., accessed June 23, 2016, <http://www.openarchives.org/community/index.html>.

21. 林信成、黃文彥，515，檢索於2016年6月23日。

22. 有興趣的讀者可參閱筆者之前發表的論文：林信成、鄭國祥，「數位典藏資訊系統之長期保存規劃與實施」，*教育資料與圖書館學* 46卷，3期(2009年3月)：323-349。

23. 同上註。



Design and Implementation of a Library and Information Science Open Access Journal Union Catalogue System

Sinn-Cheng Lin^{a*} Ting-Yu Chou^b

Abstract

Open access is a mode of academic communication that has been on the rise in recent years, but open access academic resources are widely dispersed across the internet, making it occasionally inconvenient in terms of its use. This research is focused on library and information science, using the OAIS reference model as the system framework, two open access platform, DOAJ and E-LIS as the data sources, and through system implementation develop a “library and information science open access journal union catalogue” system. Using the OAI-PMH protocol as the data interoperability standard, and LAMP as the development environment, four major functionalities: ingest, archiving, management and access of information were designed, developed, and integrated into system build. Actual testing and verification showed this system is able to successfully collect data from DOAJ and E-LIS open journal resources related to library and information science. The system is now active and functional, and can be used by researchers in the library and science information field.

Keywords: *Open Access, OAIS reference model, OAI-PMH, Metadata*

SUMMARY

Open access is a new academic communication mode in recent years, but academic resources for open access are still scattered around the web, thus inconvenient for usages. This study focuses on the subject field of library and information science, adopting OAIS reference model as the system framework, and DOAJ and E-LIS open access directories as the data sources, to develop and implement a Library and Information Science Open Access Journal Union Catalogue System. This system uses Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) protocol as the data interoperability standard, and LAMP (Linux, Apache, MySQL, and PHP) as the development environment, to design and develop a model integrating the four functions of Ingest, Archival, Management and Access. Through empirical testing and verification, this system

^a Professor, Department of Information and Library Science, Tamkang University, Taipei, Taiwan

^b Postgraduate, Department of Information and Library Science, Tamkang University, Taipei, Taiwan

* Principal author for all correspondence. E-mail: sclin@mail.tku.edu.tw

has successfully integrated relevant library and information science open access journal resources in DOAJ and E-LIS directories, and operated smoothly online, providing services for researchers in library and information science.

Research Questions

The research questions of this study are listed below.

1. How to design an Archival Module that regards DOAJ and E-LIS directories as data provider?
2. How to design an Ingest Module that harvests metadata from DOAJ and E-LIS directories?
3. How to design an Access Module for data consumers to harvest metadata from this system?
4. How to design a library and information science open access journal union catalogue system that is convenient for system managers to operate the Management Module?
5. How to integrate the modules mentioned above to develop a comprehensive library and information science open access journal union catalogue system?

Research Results

This system is built under the Linux+Apache environment, using MySQL as the database of stored journal data, and adopting PHP programming language to develop and incorporate the four major modules.

As to the database structure, appropriate data tables and fields are designed for retrieving journal records from DOAJ and E-LIS directories, and for storing metadata of the two different formats.

The first module developed in this study is Ingest Module. When the system receives journal metadata sent from other systems, it should read and analyze the metadata, and then ingest into its database.

The second module developed in this study is Archival Module. After the metadata is ingested into Archival Module, the whole set of archival data including journal article metadata, electronic full text and backup files should be stored into the database for access and retrieval.

The third module developed in this study is Management Module. This module is for coordinating, maintaining and retrieving descriptive and management data. It should allow users such as researchers or journal publishers to register, and then determine different users' management authorities, for managing the metadata of journal articles.

The fourth module developed in this study is Retrieval Module. When receiving retrieval requests through OAI-PMH commands from other systems,

this module should retrieve data from its journal databases, and transcribe the data into packets in a format in accordance with OAI-PMH protocol, for the final delivery.

After developing the database and the four major modules, we incorporate and build up a simulation system, and successfully tested and verified that this system meets the requirements of the OAIS reference model, OAI-PMH standards, DOAJ and DC metadata formats.

Conclusions

In this study we develop a library and information science open access journal union catalogue system using the OAIS reference model, taking into consideration of the needs of information producers, managers, information consumers, and archival systems. Four major modules are developed and incorporated into an operating system. The conclusions are discussed below according to our research questions.

1. The design of the database structure of Archival Module should meet the requirements of metadata formats of DOAJ and E-LIS directories.

Since DOAJ and E-LIS adopt different metadata formats, the design of data tables and fields should take into consideration of the differences to design tables and fields that accommodate metadata of both DOAJ and E-LIS directories.

2. The design of the analyzer of Ingest Module should take into consideration of both syntax and semantics of XML tags.

Since there are partial differences in the transcribing methods of XML tags between DOAJ and E-LIS directories, the formats of transcribed metadata are also different, thus there should be two analyzers designed for each directory, for correctly ingesting the metadata into the corresponding tables and fields.

3. The Retrieval Module should read OAI-PMH commands, pack the data into XML format, and control the data flow.

The Retrieval Model reads OAI-PMH commands and responds with corresponding metadata formats; information consumers can also specify package formats of metadata they need. In addition, this module can control the data flow with Resumption Tokens.

4. The design of Management Module should consider assisting managers in managing the system.

The Management Module provides managers with services including changing passwords and system data, base URL of data provider, managing email accounts, and time interval of data harvesting. The module can also timely provide statistic reports of incomplete metadata, for managers to control data quality.

5. This system can be incorporated according to the OAIS reference model.

According to the OAIS reference model, we incorporate modules of Ingest, Archival, Management and Retrieval to develop and implement a system. The validating results indicate this system meets the relevant requirements.

Although this study discusses applications of a system development, it serves as the basic foundation for future studies, such as analyzing big data of the open access journals.

ROMANIZED & TRANSLATED NOTES FOR ORIGINAL TEXT

1. 陳亞寧，「開放式資訊取用之現況發展分析」，圖書與資訊學刊，51期（2004年11月）：91，檢索於2016年7月1日，http://www.lib.nccu.edu.tw/blis/fulltext/51/51_7.pdf。【Ya-Ning Chen, “A Review on Current Development of Open Access and Its Implications for Scholarly Communication,” *Journal of Library and Information Science*, no. 51 (November 2004): 91, accessed July 1, 2016, http://www.lib.nccu.edu.tw/blis/fulltext/51/51_7.pdf. (in Chinese)】

2. Budapest Open Access Initiative, “Read the Budapest Open Access Initiative,” Budapest Open Access Initiative, February 14, 2002, accessed July 1, 2016, <http://www.soros.org/openaccess/read.shtml>.

3. Stephen Pinfield, “Open Archives and UK Institutions,” *D-Lib Magazine*, 9, no. 3 (March 2003), accessed July 1, 2016, <http://dlib.ejournal.ascc.net/dlib/march03/pinfield/03pinfield.html>.

4. 林信成、黃文彥，「基於OAIS之開放取用期刊管理系統設計與實作」，教育資料與圖書館學 49卷，4期（2012年7月）：508，檢索於2016年7月2日，<http://joemls.dils.tku.edu.tw/fulltext/49/49-4/507-532.pdf>。【Sinn-Cheng Lin and Wen-Yan Huang, “Design and Implementation of an Information System for Open Access Journals,” *Journal of Educational Media & Library Sciences* 49, no. 4 (July 2012): 508, accessed July 2, 2016, <http://joemls.dils.tku.edu.tw/fulltext/49/49-4/507-532.pdf>. (in Chinese)】

5. Carl Lagoze et al., eds., “The Open Archives Initiative Protocol for Metadata Harvesting,” Open Archives Initiative, June 14, 2002, accessed July 1, 2016, <http://www.openarchives.org/OAI/openarchivesprotocol.html>.

6. Consultative Committee for Space Data System, *Reference Model for an Open Archival Information System (OAIS)* (Washington: CCSDS, 2002), i-iii, accessed July 1, 2016, <https://public.ccsds.org/Publications/Archive/650x0b1.pdf#search=650%2E0%2DB%2D1>.

7. Linnéa Stenson, “Why all these directories? An introduction to DOAJ and DOAB,” *Insights* 25, no. 3 (November 2012), accessed June 23, 2016, <http://doi.org/10.1629/2048-7754.25.3.251>.

8. Anna Hood, “Bringing Open Access Journals into the Catalog,” in *Growth, Creativity and Collaboration: Great Visions on a Great Lake: Proceedings of the North American Serials Interest Group, Inc, 19th Annual Conference, June 17-20 2004, Milwaukee, Wisconsin*, ed. Patricia Sheldahl French and Margaret Mering (New York: Haworth Information Press, 2005): 351.

9. Directory of Open Access Journals, “About DOAJ,” Directory of Open Access

Journals, 2016, accessed June 23, 2016, <https://doaj.org/about>.

10. Directory of Open Access Journals, "Directory of Open Access Journals (DOAJ)," Directory of Open Access Journals, 2011, accessed December 28, 2011, <http://www.doaj.org/>.

11. E-prints in Library & Information Science, "About Us," E-prints in Library & Information Science, n.d., accessed June 23, 2016, <http://eprints.rclis.org/information.html>.

12. E-prints in Library & Information Science, "Advanced Search," E-prints in Library & Information Science, n.d., accessed June 23, 2016, <http://eprints.rclis.org/cgi/search/advanced>.

13. E-prints in Library & Information Science, "OAI 2.0 Request Results," E-prints in Library & Information Science, n.d., accessed June 23, 2016, <http://eprints.rclis.org/cgi/oai2?verb=ListMetadataFormats>.

14. Carl Lagoze et al., "The Open Archives Initiative Protocol for Metadata Harvesting," accessed July 15, 2016.

15. W3C, "Web Services Activity," W3C, May 8, 2011, accessed July 15, 2016, <http://www.w3.org/2002/ws/>.

16. Herbert Van de Sompel, Patrick Hochstenbach, and Oren Beit-Arie, "OpenURL Syntax Description," Caltech Library, November 5, 2001, accessed July 15, 2016, http://library.caltech.edu/openurl/Record_Documents/OpenURL_Version_0.1.mht

17. The Library of Congress, "SRU: Search/Retrieval via URL," The Library of Congress, n.d., accessed July 15, 2016, <http://www.loc.gov/standards/sru/>

18. Jeroen Bekaert and Herbert Van de Sompel, "A Standards-based Solution for the Accurate Transfer of Digital Assets," *D-Lib Mag* 11, no. 6 (June 2005), accessed July 15, 2016, <http://dx.doi.org/10.1045/june2005-bekaert>; Jeroen Bekaert and Herbert Van de Sompel, "Access Interfaces for Open Archival Information Systems based on the OAI-PMH and the OpenURL Framework for Context-Sensitive Services," arXiv.org, September 28, 2005, accessed July 15, 2016, <http://arxiv.org/abs/cs/0509090>.

19. Open Archives Initiative, "The Open Archives Initiative Protocol for Metadata Harvesting," accessed July 15, 2016.

20. Open Archives Initiative, "OAI Community," Open Archives Initiative, n.d., accessed June 23, 2016, <http://www.openarchives.org/community/index.html>.

21. 林信成、黃文彥，515，檢索於2016年6月23日。【Sinn-Cheng Lin and Wen-Yan Huang, 515, accessed June 23, 2016. (in Chinese)】

22. 有興趣的讀者可參閱筆者之前發表的論文：林信成、鄭國祥，「數位典藏資訊系統之長期保存規劃與實施」，教育資料與圖書館學 46卷，3期（2009年3月）：323-349。【Sinn-Cheng Lin and Kuo-Hsiang Cheng, "Planning and Implementing Logn-tern Preservation for Digital Archive System," *Journal of Educational Media & Library Sciences* 46, no. 3 (March 2009): 323-349. (in Chinese)】

23. 同上註。【Ibid. (in Chinese)】

Sinn-Cheng Lin ORCID 0000-0001-5199-0073

Ting-Yu Chou ORCID 0000-0002-8649-7702

