

淺談開放取用學術電子期刊互連系統

林信成、周庭郁*

一、前言

學術期刊對於一般人的生活而言，也許不是那麼密切，但對於學術界而言，卻是命脈之所繫。包括為數眾多的大學教師、研究生及研究機構的研究員，皆需仰賴查找及研讀學術期刊當中的相關文獻，方能掌握最新研究趨勢，據以規劃並從事自己的研究，再將研究成果發表至相關期刊以享同儕。因此，大學及研究機構所屬圖書館，每年皆需編列大筆預算購買學術期刊資料庫，以供全校師生查詢使用。不過這些年來，學術期刊價格持續高漲，圖書館編列的預算即使逐年增加，但訂購的期刊種數卻未能隨之增加，甚至可能更少，造成研究資源短缺，嚴重影響教學研究品質，這也是學術界近年來面臨的重大考驗之一。為了因應此一趨勢，強調自由、免費且兼具品質的「開放取用」(Open Access, 簡稱 OA) 乃應運而生且迅速發展。2001 年 12 月，開放社會基金會 (The Open Society Foundations, OSF) 於匈牙利首都布達佩斯召開了一場小型會議，針對如何透過全球學術界的努力，促進研究論文全面開放取用的議題進行討論。會中提出「布達佩斯開放取用先導計畫」(The Budapest Open Access Initiative, 簡稱 BOAI)，並於 2002 年 2 月 14 日正式對外公布計畫內容，成為一項全球性推動開放取用之原則、策略及承諾聲明 (BOAI 詳情可參見：<http://www.budapestopenaccessinitiative.org/>)。

然而，這些開放取用的學術資源分散於網路之中，若無有效整合，對於讀者而言仍然需要到處查找，甚為不便。因此，為了提升 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) 中，開始有了創造一個全面性開放取用指南的構想。於是在 2003 年 5 月時，蘭德大學圖書館便創建了「開放取用期刊目錄」(Directory of

* 林信成，淡江大學資訊與圖書館學系教授兼文學院院長兼出版中心主任；周庭郁，淡江大學資訊與圖書館學系碩士。



Open Access Journals，簡稱 DOAJ）作為匯集全球具備「同儕評閱」且確保品質的 OA 期刊整合平臺，提供科學研究人員檢索利用。此處所謂「同儕評閱」指的是學術文章需先經相關領域的研究者審查通過後才能發表，是確保學術研究品質的重要過程與機制。如今，DOAJ 歷經十餘年的發展，已經成為一個收錄了 9 千餘種 OA 期刊的一站式服務（one stop shop）平臺（DOAJ 官網：<http://www.doaj.org/>）。

如今，OA 已成為新興的學術傳播方式，圖書館在經費有限的情況下，若能比照 DOAJ 的模式，自行開發足以匯集眾多 OA 期刊資源之資訊系統，便得以藉著不斷增加及免費取得的線上期刊，擴展現存館藏以外的學術資源供讀者使用。因此，筆者基於此動機，向科技部提交研究計畫「開放取用學術電子期刊互連系統之開發」，幸運的獲得補助並執行一年，本文共同作者周庭郁是筆者的研究生，擔任該計畫案之研究助理，協助撰寫程式，一同完成本研究案。在系統初步規劃階段，我們便認為此系統若要能順利運作，其設計必須採開放式架構，資料交換應遵循國際通用的通訊協定（Protocol）及詮釋資料標準。因此，我們採用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）為架構，以 OAI-PMH（Open Archives Initiative Protocol for Metadata Harvesting）作為資料交換協定，以 DOAJ 詮釋資料格式與都柏林核心集（Dublin Core，簡稱 DC）詮釋資料格式作為資訊封包，實際開發通用之 OA 期刊目錄彙整平臺，並自 DOAJ 資料庫擷取部分 OA 期刊書目紀錄為實驗資料，建構一套 OA 期刊開放互連整合模擬系統，作為圖書館有意建置類似系統之參考。

本文試著採用較為平易近人的寫法，盡量捨棄學理或規範當中較艱深難懂之處，以科普方式說明本研究採用之相關規範，以及使用系統實作法實作匯入模組、典藏模組、管理模組、取用模組四大模組，建置一個模擬之開放取用學術電子期刊互連系統之過程，期使讀者能從閱讀本文當中，了解我們所強調的系統開放互連及資源共享之核心精神，文中也穿插一些我們執行專案過程的特殊經驗、心得，以饗讀者。

二、相關規範簡介

首先介紹 OAIS 參考模式，這是一個提供開發者參考的資訊系統架構，強調開放互連之精神，是由太空資料系統諮詢委員會（Consultative Committee

for Space Data Systems，簡稱 CCSDS）所制訂，其規範書中定義了開放式系統的相關術語、概念、模型及過程。不過，為了能夠在各領域中廣為使用，OAIS 參考模式對於數位資訊類型、技術及處理方式都僅有原則性的規範而未作特別的限制（OAIS 詳細規範可參見：<http://public.ccsds.org/publications/archive/650x0m2.pdf>）。

這樣做的好處是彈性較大，壞處則是使得開發過程有時需自行揣摩，才能有所頓悟，如果再加上像筆者任教的系所是屬於文學院，在人文背景的學生當中要挖掘出具開發系統潛力者，已經不是太容易，何況還需自行揣摩規格書內容？幸好協助本案撰寫程式的周庭郁同學（本文共同作者）相當努力，一一克服困難，完成任務。在這過程中也展現學生只要不自我設限、肯投注心力，即使人文科系亦能擁有橫跨工程領域之技能。目前周同學已順利取得碩士學位，且學以致用順利於某科技公司擔任圖書館系統工程師，完全將所學貢獻於國家社會。

其次要介紹的是 Metadata 與 XML 標示語言。如果將 OA 期刊資料視同資訊系統中的數位物件，那麼加註不同型式的詮釋資料（Metadata），就相當於將其封裝成資訊封包，可在系統間傳遞，達成資源共享的目的。詮釋資料是一組人類或機器可理解的智慧型資料，具備結構性及自我描述性，可協助人們或資訊系統對於所描述的對象進行檢索、管理與利用。加註詮釋資料的實作方式也極其眾多。其中，XML 標示語言由於具有可擴展性、結構性、自我描述性，並採用資料和樣式分離原則，使其在資料的管理、交換上擁有極為卓越之性能，已成為著錄詮釋資料最重要、最通行的實作技術，可謂是當前著錄詮釋資料的標準規格；而在 OAIS 參考模式當中，系統間 OA 期刊資料交換的資訊封包，也以採用 XML 格式加以封裝為原則。

但詮釋資料也有各種標準與格式，在本案中我們採用的是 DC 與 DOAJ 詮釋資料格式。DC 是簡單且具有彈性、目前最為廣泛使用的詮釋資料標準格式，採用了 15 個簡單的欄位來儲存資料，適合應用在許多不同領域。這 15 個欄位分別是：題名（title）、創作者（creator）、主題（subject）、相關敘述（description）、出版者（publisher）、貢獻者（contributor）、日期（date）、資料型態（type）、資料格式（format）、識別碼（identifier）、資料來源（source）、語言（language）、相關連結（relation）、內容範圍（coverage）及使用權限（rights）。而 DOAJ 詮釋資料則專為期刊資料所設計，可自 DOAJ 官方網站查



詢相關資訊，並下載取得該份檔案 (<https://doaj.org/schemas/doajArticles.xsd>)。為了讓 OA 期刊系統順利介接、整合，DOAJ 詮釋資料又分為期刊 (Journal Feed) 與文章 (Article Feed) 兩個層次，比起 DC 複雜許多，但相對的，所能記錄的資料也更豐富。

最後介紹的是 OAI-PMH 資料交換協定，這是由開放典藏協會 (Open Archives Initiative) 所制訂的，為國際上數位資料交換標準之一，主要用途是在資訊交換的過程中提供標準化的通訊協定，以便將分散的資源加以彙整，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。OAI-PMH 協定的優點在於實作容易、具開放性，採用 XML 與 HTTP 等開放標準，相容性高，在國際上已有許多單位進行建置與使用，詳細規範可參見：<http://www.openarchives.org/OAI/openarchivesprotocol.html>。

三、研究成果與經驗分享

本系統在 Linux+Apache 的環境下，以 MySQL 作為儲存期刊資料之資料庫，以 PHP 程式語言開發四大模組再加以整合。之所以採用上述簡稱為 LAMP 的開發環境，是因為這是目前全球最普及、占有率最高的解決方案。

在資料庫結構方面，為了能夠從 DOAJ 資料庫擷取部分 OA 期刊書目紀錄作為實驗資料，因此設計符合 DOAJ 與 DC 詮釋資料格式的資料表與欄位，使資料表能夠同時儲存兩種不同格式的詮釋資料。

匯入模組為本系統第一個模組，當系統接收來自其他系統所傳遞的期刊詮釋資料時，需具備解讀與剖析資料的能力，以便將外來的詮釋資料轉入系統本身的資料庫中，並依照對應的欄位儲存起來。剖析器設計是本系統開發過程遭遇最多困難的，因為來源資料有各種不同語文版本，容易造成亂碼，甚至使程式無法順利運行；另外，DOAJ 來源端的資料品質也有很大的問題，我們原本以為它是一個全球知名且號稱經過同儕評閱的期刊資料庫，品質應該有保障。其實不然，在我們的實驗過程中，就發現了 DOAJ 的資料有許多缺漏與錯誤，雖然這並不在本計畫案的研究範圍內，但我們還是做了一些淺顯的統計分析，並決定日後可以針對 DOAJ 資料品質進行更深入的研究。

典藏模組為本系統第二個模組，匯入的資料在進入典藏模組後，本系統可將期刊文章的詮釋資料、電子全文與備份檔案等之典藏資料儲存進入資料庫中。同時提供檢索介面供一般使用者查詢資料，查詢欄位有：文章名稱、

文章摘要、作者、關鍵字與期刊名稱。

管理模組為本系統第三個模組，負責描述性資訊與管理性資訊的協調、維護與取用，可開放讓研究者或期刊出版社等使用者註冊，並且判別不同使用者的權限使其對期刊文章之詮釋資料進行新增、匯入、刪除、修改、查詢等動作，使資料庫之資料能夠更加完善，並維護系統的運作。

取用模組為本系統第四個模組，當接受來自其他系統以 OAI-PMH 命令集發出檢索需求時，必須從自身的期刊典藏庫中取出資料，再以符合 OAI-PMH 之協定格式，轉換成遞出資訊封包進行傳送。

開發上述資料庫及四大模組後，我們加以整合並建置出模擬資訊生產者的雛形系統，以及模擬資訊消費者的期刊匯集系統。該模擬系統成功驗證了能夠以 OAIS 參考模式、XML 標示語言、OAI-PMH 資料交換協定、DOAJ 與 DC 詮釋資料格式等相關技術之整合應用，並建置出一個通用的開放取用學術電子期刊互連系統。

四、結論與建議

本文簡要說明我們執行科技部研究計畫「開放取用學術電子期刊互連系統之開發」的過程，以 OAIS 參考模式、XML 標示語言、OAI-PMH 資料交換協定、DOAJ 與 DC 詮釋資料格式等相關技術，成功建置出學術期刊開放典藏資訊系統之雛形。就執行過程中所遭遇之問題及完成系統的運行狀況，給予以下兩項建議：

1. 圖書館因學術期刊價格持續高漲，預算無法趕上期刊漲幅，期刊訂購種數甚至不增反減，可考慮開發足以匯集眾多 OA 期刊資源之資訊系統，在有限的預算下仍能提供充足之研究資源，提升教學研究品質。
2. 因本研究僅建置模擬雛形系統，完成資料互通的功能部分，若未來欲使該系統能夠正式上線運作，並提供給學術研究人員或是特定族群使用，可再進行使用者需求分析，評估收錄內容哪些為宜，將架構與介面建置得更加完善，並於系統正式上線時，了解該系統是否符合使用者需求。