大學圖書館
8 卷 1 期（民國 93 年 3 月） 頁 107-129

數位物件辨識碼系統之實作

An Implementation of Digital Object Identifier Systems

葉冠志 **Kuan-chih Yeh**
國立臺灣大學圖書資訊學系碩士
Master, Department of Library and Information Science,
National Taiwan University
E-mail：kcyeh@mail.lis.ntu.edu.tw

陳光華 **Kuang-hua Chen**
國立臺灣大學圖書資訊學系副教授
Associate Professor, Department of Library and Information Science,
National Taiwan University
E-mail：khchen@ntu.edu.tw

【摘 要】

數位物件辨識碼系統（Digital Object Identifier System, DOI System）為一可在數位環境下獨特辨識、描述與連結智慧產權物的系統。本研究參考DOI系統架構，依據國內之環境實際建置一套<LIPS-DOI>系統，包含列舉、描述與解析模組。以延伸標記語言（Extensible Markup Language, XML）為基礎發展詮釋資料架構，以簡單都柏林核心集（Simple Dublin Core）作為核心應用描述檔案格式，並提供代理解析服務。

【Abstract】

The Digital Object Identifier (DOI) system is a system for identifying and exchanging intellectual property in the digital environment. With reference to the DOI's framework, an experimental system named <LIPS-DOI> based on Taiwan's environment is reported in this article. This <LIPS-DOI> system mostly resembles the DOI system in its three modules: numbering, metadata and resolution. The system framework is based on XML technology, the Dublin Core is used as the Application Profile, and a resolution proxy service is provided as well.

關 鍵 詞：數位物件辨識碼；智慧產權；詮釋資料；可延伸標記語言
Keywords：Digital Object Identifier; Intellectual property; Metadata; XML

臺灣大學學術期刊資料庫

壹、前言

圖書館資源傳統上以圖書或期刊為主。但隨著新興工業技術與資訊科技的發展，今日圖書館管理的資源已經涵蓋圖書期刊以外的各種媒體資料，如錄音帶、錄影帶、影音光碟、資料庫乃至於數位化的典藏等等。當全球網際網路形成之後，各式各樣的網路資源更成為圖書資訊研究關注的焦點。由網際網路存取各種數位化內容已成為發展數位圖書館（Digital Library）的中心概念，透過數位加值方式建立電子式聯合目錄並提供分散式檢索服務更能進一步提升圖書館組織整理各種資源物件的成效，令使用者獲得可利用、有價值的知識，進一步提供協助決策與促進產能的知識管理（Knowledge Management）之目標。雖然應用資訊技術對將資源數位化可便利地進行組織、整理、連結、檢索等加值工作，提升各種資源的可用性與價值，但是網路上的各種數位資源不但更新迅速、種類繁多，而且經常容易發生連結不存在、存取到未更新的版本、存取到非同一語系的譯本或為檢索得到的是錄音或影片資料等其他形式的媒體展現等各種錯誤情況。同一資源的複本可能以同名稱放在不同的存取位置，也可能以不同的名稱存在於不同的存取位置，而有時同樣名稱的物件其實指的並不是相同的資源。這樣渾沌混亂的環境使得大大減損資源的可利用性，取用這些容易混淆的資源不僅耗時耗力，嚴重時還可能因為資訊錯誤，造成更大損失。為了使數位網路環境下資訊組織管理的工作能夠更順利進行，吾人首先必須明確辨識每一個欲組織管理的資源物件，為其分配一個持續且全球唯一的辨識碼，繼以資訊加值工作（如線上分類編目、整合為電子式聯合目錄等）進而透過全球資訊網提供使用者瀏覽、查詢、檢索以及正確連結。

「數位物件辨識碼系統」（Digital Object Identifier System，以下簡稱 DOI 系統）是一種可以在數位環境下辨識資源物件的編碼系統，利用數位物件辨識碼系統可以對各種物件分配全球唯一且持續不變的辨識碼，此可解決網路環境中各種物件資源難以辨識的問題，同時透過完整數位物件辨識系統提供的組織管理機制，吾人可以進一步在全球網路環境的架構下有效存取與組織整理各種類型的智慧產權資源物件。

本文將說明數位物件辨識碼的起源、架構與應用，並依據國內的環境實作一套<LIPS-DOI>系統。第貳節說明DOI的起源與特點；第參節詳細探討DOI的架構；第肆節描述我們實作<LIPS-DOI>的想法、架構、與系統。第伍節討論互通性的議題；第陸節則是簡短的結論。

貳、DOI 的起源與特點

數位物件辨識碼是一種持續辨識碼，此辨識碼一旦被分配後不會再變

臺灣大學學術期刊資料庫

動，因此具有持續、永久、全球唯一的辨識特性。數位物件辨識碼目前主要被應用在電子出版業方面，特別是應用在具有智慧產權性質的資源。美國出版商協會（Association of American Publishers，簡稱 AAP）於 1994 年建立了「技術實現委員會」（Enabling Technologies Committee）開始設計一種既能保護智慧產權又能保障出版商利益的系統。1996 年起，美國國家研究先導團體（The Corporation for National Research Initiatives，簡稱 CNRI）便開始著手研發此系統，並於 1997 年法蘭克福圖書博覽會首次展示此數位物件辨識碼系統。由次年的博覽會起，此系統陸續亮相引起了國際出版界廣泛的重視。（註 1）因應網路時代智慧產權物數位化管理的需求，國際數位物件辨識碼基金會（International DOI Foundation，簡稱 IDF）於 1998 年正式成立，負責管理數位物件辨識碼系統的運作。（註 2）

IDF 並在 2003 年 8 月宣布透過 CrossRef 組織（由 1999 年開始發展，提供以科學與學術專業為主的文獻連結服務），DOI 系統所發佈出的數位物件辨識碼已超過一千萬個。（註 3）此足見數位物件辨識碼系統對出版界、圖書館界、學術界已經產生國際性的影響力。

DOI 具有以下的特點：

1. DOI 為獨特的（unique）辨識碼

在功能上，DOI 為一獨特的字串，其可唯一辨識某一實體（entity）。DOI 系統確保同一個 DOI 不會被發行兩次，而一個 DOI 只用在一個被辨識的物件上。因此，一個被辨識的物件不會分配到兩個 DOI。但需注意的是，一個被辨識的物件雖然只有一個 DOI，但是同一個物件仍可擁有兩種個以上的辨識碼，如一本書可以擁有一個 DOI 和一個 ISBN 及一個系統編號等，這也是數位物件辨識碼系統可以整合其他編碼系統的主要原因。此外，DOI 亦可以應用在辨識內容片段（fragments of content），原本沒有編碼系統的機構更可以直接應用 DOI 服務來組織整理其資源。（註 4）

2. DOI 為隱匿的（opaque）辨識碼

DOI 辨識碼為「隱匿字串」（opaque string），或說是隱匿的數字（dumb number），也就是說這些號碼（number）的表面並不提供資訊，而是必須透過與這些 DOI 關聯的詮釋資料方能得知描述此物件的資料。此特色令一般使用者不易僅由辨識碼本身就獲得註冊者及註冊物件的資訊，是較具資訊保全性的作法。（註 5）

3. DOI 為持續的（persistent）辨識碼

DOI 可用來辨識任何智慧產權物的實體物件載體（manifestations），如書、CD、期刊文章等。DOI 亦可以用來辨識較難以感官感知的載體，如表演、抽象觀念（abstraction）等。即使實體的擁有權或版權改變了，變更的只有實體的詮釋資料，實體的辨識碼仍然維持不變。（註 6）

4. DOI 為行動的（actionable）辨識碼

臺灣大學學術期刊資料庫

DOI 編碼是系統最基礎的部份，根據DOI可以辨識物件，管理者還可透過系統改變物件呈現或提供存取的方式。這樣的行動性使應用DOI管理的物件隨時能夠更新呈現方式與存取方式。（註7）

5. DOI為互通的（interoperable）辨識碼

數位圖書館和傳統圖書館之間最大差別之一就是資料來源和使用機制不再限於紙本，由單一轉向多元。因此，這種新形式的「亞歷山卓圖書館」必須要考慮到各種互通性的問題。

DOI 系統是一個互通性良好的系統，設計為可與過去、現在、未來科技整合運用。DOI 系統能夠將所謂的「舊有的」（legacy）辨識碼，也就是將曾經從過去沿用至今原有的合法辨識碼整合至其編碼系統之中。（註8）透過 DOI 系統能夠產生新的、全球唯一的辨識碼，其亦可選擇性保留舊有的命名原則或註冊號碼。應用數位辨識碼系統不僅不會破壞原有的編碼系統，還可進一步協助整合原本編碼系統不明確的資源。DOI 可以克服實體（因為擁有者改變或因為管理因素）在網路空間移動而造成解析不持續（如broken link）的辨識問題。兩種異質且編碼不同的資源也可透過DOI的整合輕鬆結合在一起。DOI 系統的各種核心系統技術（如 Handle）之開發也以開放性標準為基礎（open standards based）以達到完全的互通性。因為DOI系統互通性良好，許多研究者目前正討論應用其作為全球電子商務交易的基礎架構。

參、DOI 系統之架構

DOI 系統分配的 DOI 號碼是合法的URN，符合RFC1737的規定，具有下列特性：（註9）

- 全球範疇（Global scope）
- 全球唯一性（Global uniqueness）
- 持續性（Persistence）
- 延展性 (Scalability)
- 舊有支援（Legacy support）
- 擴充性（Extensibility）
- 獨立性（Independence）
- 解析性（Resolution）

DOI 系統主要的功能在分配全球唯一且持續的辨識碼，但完整的DOI系統並非只是一個分配編碼的命名系統，其實際上包含許多整合性功能。完整的 DOI 系統包含列舉、詮釋資料、解析等功能（註10），而由 IDF 負責維護其發展的政策，以下逐一探討。

一、列舉（Numbering）

列舉就是分配一個號碼給DOI所辨識的智慧產權實體。比較正確的說法應該是分配「字母數字字串」（alphanumeric string），因為 DOI 亦可包含字母，但以下為了避免太過繁複，僅以辨識碼稱之。每一個DOI均為一獨特的辨識碼，可用來唯一辨識一個實體。雖然 DOI 系統會確保同一個 DOI 不會被發行兩次，但 DOI 字首（prefix）的註冊者（公司或個人）基

臺灣學術期刊資料庫

本上應負責保持物件命名的獨特性。

由於同一個物件可能原本就擁有數種不同的辨識機制，如一本書可以擁有一個索書號、一個 ISBN 及一個圖書館自動化系統的系統編號，DOI 系統的列舉機制可以將這些辨識碼整合在字尾或詮釋資料中，因此舊有的辨識碼仍可以繼續使用。此外，DOI 可以用來擴充原有的辨識功能，若要將一本有 ISBN 的書或有 ISSN 的期刊再細分更小的章節甚至到段落，只要為每一章節或段落各編一個 DOI，整個圖書或期刊典藏在編碼上仍然可以保持持續統一。許多類型不同的資料（如文字、圖畫、語音、影視、軟體等）只要經過 DOI 系統編碼，再透過詮釋資料資料庫（Metadata Database，簡稱 MDDB）管理，各種類型的資料就可被方便地組織整理。尚無統一編碼規範的任何類型物件均可利用 DOI 系統簡單而快速地建立具有唯一性且全球通行的數位辨識碼。（註 11）

DOI 為兩個成分所組成：字首（Prefix）以及字尾（Suffix），以正斜線（/）分開。（註 12）DOI 字串的語法如圖一所示。

1. 字首（prefix）

DOI 字首為數字型態（numeric），包含目錄編號與註冊者編號兩個部分。字首的目錄編號為必要項，目前唯一的有效值為 "10"，也就是 Handle 系統中保留給 IDF 的預設目錄編號，以作為與其他 Handle 系統的應用區別之用。（註 13）DOI 的字首初期僅以單純方式使用，其未來可能視需求進一步再細分到「次字首」（sub-prefixes），如圖二所示。（註 14）字首的註冊者編號為必要項，其與目錄編號中間以 "." 分隔，為 IDF 所分配的 DOI 註冊者編號。

2. 字尾（Suffix）

DOI 字尾字串亦為必要項，為 DOI 註冊者管理分配。

3. 字串長度

在字串長度方面，DOI 字串沒有刻意規定長度限制。因此字首（prefix）或字尾（suffix）的長度並無技術上的限制，所以 DOI 數量具有擴充彈性。

4. 字串編碼

DOI 的編碼以 Unicode 2.0 字集為原則，可使用 Unicode 字集中所有的字元，結合任何環球字集（Universal Character Set, UCS-2, ISO/IEC 10646）中可列印的字元。但是"."具有分隔目錄編號和註冊者編號的功能且未來可能作為分隔次字首使用，而"/"分隔字尾和字首，因此這兩個字元均被保留使用。

IDF 鼓勵同一機構申請多個 DOI 字首，其目的在於同一組織可使用多重式的字首（multiple prefix）達成分散式的管理，這也是 DOI 系統可以發展為電子商務模型的重要基礎。此外，由於 DOI 的字尾為註冊者自訂，因此 DOI 系統能夠整合舊有的編碼系統，同時亦可更新詮釋資料及獲得更多的存取選擇性。

臺灣大學學術期刊資料庫

DOI 的語法

<DIR>.<REG> / <DSS>

圖一、DOI 的語法

<資料來源：
http://www.doi.org/handbook_2000/enumeration.html>

10.10.1000/123456

次字首（sub-prefix）

圖二、DOI 的次字首

<資料來源：
http://www.doi.org/handbook_2000/ enumeration.html>

二、詮釋資料（Metadata）

因為 DOI 本身具有隱晦性，為 DOI 所辨識的實體建立描述性資料（即詮釋資料）可令使用者獲得辨識物件的相關資訊，因此詮釋資料對 DOI 系統來說非常重要，DOI 本身也可說是詮釋資料的一部份。DOI 系統的詮釋資料架構採用<indecs>計畫所發展的資料模型（註 15）要求所有註冊 DOI 的物件都使用語法結構完整（well-formed）的詮釋資料，而且要求詮釋資料必須符合「核心詮釋資料」（"kernel" of metadata）的規範。這並非強迫所有註冊DOI的智慧產權物件都用同一種詮釋資料架構來進行描述，而是要求所有的詮釋資料的描述架構必須先合於「核心」的規範。以此核心詮釋資料架構為基礎，不同類型物件之間複雜的詮釋資料才能夠對照相通、交換互動。如果沒有核心詮釋資料的規範，詮釋資料之間便容易產生「交換障礙」（trade barriers）。（註 16）

為了上述的目的，DOI 對每一筆 DOI 註冊時使用的詮釋資料有所控制，要求所有註冊者註冊DOI時其詮釋資料的宣告必須包含核心詮釋資料，也就是每個DOI被建立時必要的宣告項目，任何使用者均可透過核心詮釋資料初步瞭DOI所辨識的實體的基本資訊。

不同的類型（types）或類別（classes）的智慧產權物需要不同的詮釋資料，如期刊文章、音樂錄音或攝影相片資源的描述方式（Description）不同，因此需要不同的詮釋資料。DOI系統在詮釋資料方面要求註冊者（物件擁有者）描述資源時必須採用DOI應用描述檔（DOI Application Profile，簡稱DOI-AP，在DOI系統發展初期其被稱為Genre），每一DOI所辨識的實體至少要被分配一種以上或更多的DOI-AP。基礎應用描述檔（Base-AP）是最簡單基本的DOI-AP，也就是包含核心詮釋資料中所有必要宣告項目的應用描述檔。（註 17）

事實上，Base-AP的功能就好比都柏林核心集（Dublin Core, DC）試圖為網際網路上的各種資源建立共通詮釋資料產生出來的最簡化形式（minimum

臺灣大學圖書館 期刊資料庫

form）—簡易都柏林核心集的15個欄位（elements）。Base-AP是所有AP的交集。也就是說，所有物件的描述至少應包含AP-Kernel的每一個欄位。所有註冊DOI的物件，其DOI-AP選擇上最低程度的要求為Base-AP。（註18）

<indecs>計畫曾討論DOI有所謂的「功能階層性」（Functional granularity）之需求，也就是說當有理由應辨識時則應該能夠辨識此實體。（註19）功能階層性展現在AP的管理上。換句話說，不同的DOI使用者團體（DOI User Community，簡稱DOI-UC）要求某類型的物件詮釋資料能更加詳確、標準化、符合其使用需求時，其可對IDF提出建議，然後在IDF的管理之下，共同發展描述更為詳盡而適用的各種應用描述檔。管理者註冊DOI的物件可以依其功能階層性來決定所適用的AP，而以不同APs描述、具有不同的功能階層性的物件仍然可以進行整合的組織整理工作（如查詢、檢索等）。

DOI詮釋資料系統（DOI Metadata System，簡稱DMS）是基於<indecs>資料辭典（indecs Data Dictionary，簡稱iDD），iDD也是ISO 21000（即MPEG-21）第六部分「權利資料辭典」（Rights Data Dictionary，簡稱RDD）的基礎。所有DMS中使用的詞彙集（TermSet）最後也將對照至iDD。（註20）DOI DMS中，Base-AP宣告的詮釋資料為所有AP的核心基礎，其核心欄位如表一所示。（註21）

DMS的發展依據為iDD、DOI Term-Set與核心詮釋資料的XML Schema，而前兩者的版本與後者的草案在2003年七月隨MPEG組織Mpeg-21的發展而確定。目前，DMS的核心宣告採用XML Schema，並且也有非XML格式的表示方式。（註22）

三、解析（Resolution）

DOI可持續辨識數位物件。URL所辨識的是網際網路上存取資源的網址。雖然有時DOI所提供的解析與URL相似均是指向某個網路位址，但這兩者並不完全相同。因為DOI所辨識的是物件實體，而URL所辨識的是物件的位址。DOI的解析是透過解析系統建立使用者與物件之間存取使用的橋樑。DOI能夠具有「行動性」就是因為解析系統的功能。目前，DOI的解析功能是採用CNRI所發展的Handle系統技術。

Handle系統的解析有兩種：單一轉址（single redirection）與多重解析（multiple resolution）。當狀態資料（state data）只有一個，為單一轉址；選擇性更多時則為多重解析。以多重解析為例，其狀態資料如圖三所示。（註23）

狀態資料為DOI註冊者的基本責任。Handle系統上的資料類型可以繼續擴充，而使DOI可以解析至任何網際網路上可存取的資源。DOI亦可解析至Java applet、CGI script或其他機制。

臺灣大學圖書館術期刊資料庫

表一　DOI 詮釋資料核心欄位

核心欄位（Kernel Element）	說明	數量	結構/允許的值
DOI 應用描述檔 DOIApplicationProfile	應用描述檔的名稱	至少一個以上	註冊的 AP 名稱
DOI	數位物件辨識碼	唯一	任何 DOI.
辨識碼 Identifier	DOI 以外的獨特辨識碼（如舊有的識別碼）	若有至少一個以上	任何辨識的字母數字字串及其辨識碼類型（如 ISBN、ISSN、UPC、PPN 等或另一
題名 Title	實體為人所知的名稱	若有至少一個以上	任何字母數字字串
結構類型 Structural Type	實體的主要結構類型	一個	允許的值：（種類或品質） • 可感知的固定物（實體）Tangible Fixation (Physical) • 不可感知的固定物（數位式）Intangible Fixation (Digital) • 表演（時空）Performance (SpatioTemporal) l 抽象概念（抽象）Abstraction (Abstract)
模式 Mode	智慧產權實體希望被感知的主要模式	至少一個以上	允許的值：l 視覺的（Visual）l 聽覺的（Audio）l 視聽的（Audiovisual）l 抽象的（Abstract）
首要代理 PrimaryAgent(s)	首要代理者的名稱或辨識碼，通常但非必定是實體的第一名稱創作者	至少一個以上	為 iDD 認可名稱空間中的命名或辨識碼。說明首要代理之構成時使用何種 AP 必須被宣告，但此可由註冊代理自行決定
首要代理角色 PrimaryAgentRole(s)	首要代理所扮演的角色	一至 N 個	一致的名稱空間中的編碼或名稱（如作者、出版者、編者等）

<資料來源：http://www.doi.org/handbook_2000/metadata.html#5.4>

臺灣大學學術期刊資料庫

DOI	資料類型	DOI狀態資料
doi:10.1004/123456	URL	http://www.pub.com
	URL	http://www.pub2.com
	DLS	loc/repository
	XYZ	1001110011110
DOI的值		關連資料

圖三　DOI 的狀態資料

<資料來源：http://www.doi.org/handbook_ 2000/resolution.html#6.7>

四、政策（Policies）

政策為整個DOI系統運作的管理規則，掌管列舉、詮釋資料與解析系統的發展、協調與統合。如圖四所表示。政策確保DOI系統的三個子系統功能完備。整個DOI系統的政策執行由IDF理事會成員所負責。數位物件辨識碼的註冊主要是由註冊代理工作小組（Registration Agency Working Group，簡稱 RAWG）負責，並討論DMS中應用描述檔的規範，解析服務方面則由Handle系統掌管規範政策。（註24）

肆、實作<LIPS-DOI>系統

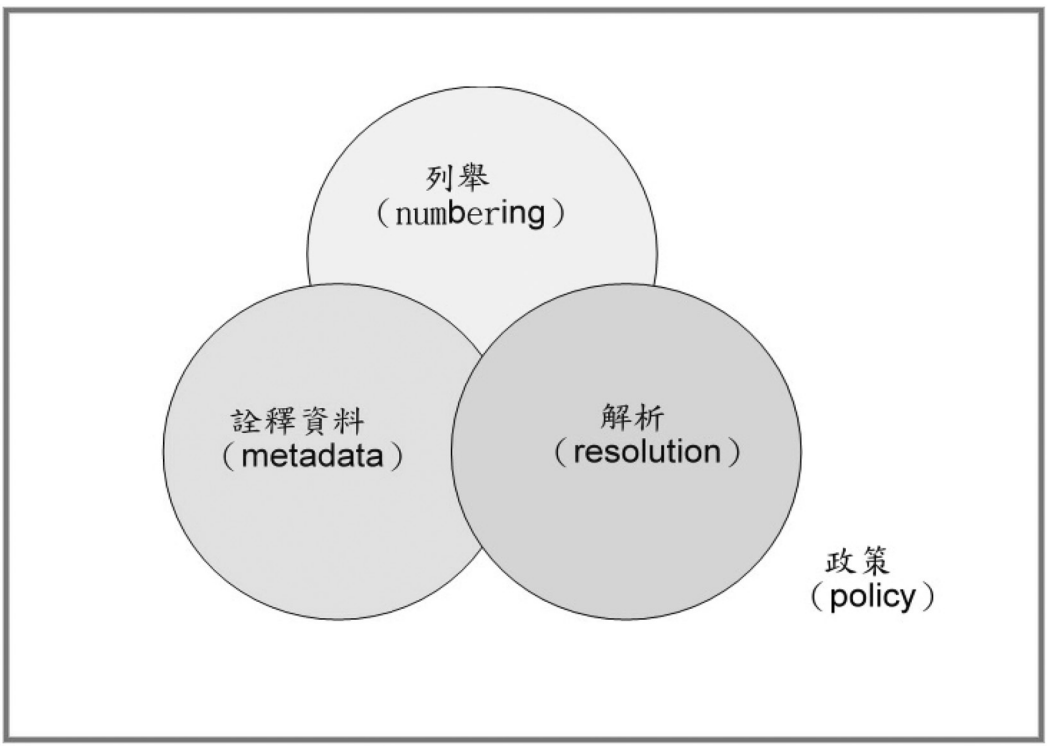
為了測試DOI系統在中文資源方面的應用成效，本研究建置了中文環境的實驗系統。經過構思與初步實踐之後，正式定名為<LIPS-DOI>，LIPS為本文作者之一陳光華教授指導的語言資訊處理系統實驗室（Language & Information Processing System Lab，LIPS）之簡稱，DOI 表示此系統是模擬DOI系統所建立，沿用DOI系統與<indecs>系統命名的標記符號（mark-up）。實驗系統未涉及政策方面的實作，為本研究的限制與範圍。

一、列舉模組

在DOI System中，列舉系統的{字首 / 字尾}兩者均為隱匿的，不過由系統管理的觀點上來看，其{字首 / 字尾}代表的是{註冊者資訊 / 註冊物件原有辨識碼}的原則。也就是說，DOI字串僅以字首來區別註冊使用者，字尾部分則完全由使用者自行定義，其語法含義如圖五所示。

在系統設計上，<LIPS-DOI>也開放使用者自行定義字尾，但不同的是，使用者也可以選擇不定義字尾，而只採用<LIPS-DOI>系統定義的時間字串作為辨識碼，<LIPS-DOI>系統語法結構將註冊時間字串定義為必要

臺灣大學圖書館 期刊資料庫



圖四　DOI 系統政策概念圖

<資料來源：http://www.doi.org/handbook_2000/components.html>

<DirectoryCode>.<Registrant Number>/<LocalName>	
字首	字尾

圖五、DOI 的語法含義

項，置於字首與字尾中間，而字尾部分為選擇性的，其語法含義如圖六所示。<LIPS-DOI>系統選擇時間作為辨識字串的理由如下：

1. 智慧產權物件的註冊宣告與時間有絕對關係。
2. 時間字串按照年月日時分秒排列，格式為 YYYYMMDDhhmmss，長度是 14 個字元，除了在機讀上明確表現出物件的註冊時間和順序之外，字串仍保持高度隱匿性，難以人讀的方式由字串得知物件內容。
3. 沒有辨識碼架構的使用者可依賴<LIPS-DOI>時間字串著錄資源，不

臺灣人文學術期刊資料庫

[<DirectoryCode>.]<Registrant>-<RegistrationTime>[-<LocalName>]		
字首	時間字串	字尾

圖六　<LIPS-DOI>的語法意涵

需另外規劃字尾命名規則。

4. 字串可表現出同一時間註冊的物件為單件或多件（批次註冊物件時，使用者僅需將原有的資料庫主鍵值或特別選定的辨識碼欄位定義為字尾即可轉入系統）。
5. 不使用字尾的情況下，由字串僅可獲得註冊時間資訊，字串的隱匿性更高，但在批次處理或使用者有需求時仍然保持使用字尾的彈性。

字首方面，Handle系統目前將DOI的目錄編號（Directory Code）統一訂為 10，同時有次字首（sub-prefix，參見圖二）設計，其目的乃為了在全球環境下以分散式架構提供較高的效率，也就是說Handle系統中的全球註冊伺服器可將使用者端傳來不同的字首或次字首之DOI字串傳遞給不同子系統處理，子系統（當地的Handle服務）再將字串傳遞給子系統內負責此組 DOI 字串的網站（Handle Servers Site）下的伺服器（至少有一台主要伺服器），由此伺服器將與此DOI字串關連的數位物件詮釋資料與解析傳遞給使用者端。換句話說，透過對使用者傳遞過來的 DOI 字串進行剖析（parsing）工作，可以有效率地將使用者的DOI查詢轉至負責此筆紀錄的Handle伺服器以提供使用者回應。由於<LIPS-DOI>系統的架構僅屬實驗性質，尚不需如同Handle系統使用目錄編號，因此字首部分的目錄編號暫時以括號表示保留未來之發展。

時間字串方面，由於列舉模組採用了時間字串，為保持時間的準確性，<LIPS-DOI>採用時間通訊協定（Time Protocol）下的全球標準時間架構。伺服器使用時間與頻率國家標準實驗室的NTPClock軟體（1.21版）進行校時工作。（註25）時間通訊協定目前較常用的為 NTP（Network Time Protocol，RFC-1305）及 SNTP（Simple Network Time Protocol，RFC-2030）兩種，SNTP 算是 NTP 的一個子集，由於<LIPS-DOI>系統的實作屬於實驗性質尚不需同時和多台伺服器進行校時，因此僅採用後者。（註26）使用時間通訊協定校時的好處是當<LIPS-DOI>系統如Handle系統一樣發展至全球架構時，時間字串透過全球標準時間的轉換可以明確標示出任何物件註冊的先後順序。

二、詮釋資料模組

<LIPS-DOI>實驗系統採用都柏林核心集的十五個核心欄位（Simple DC）

臺灣大學學術期刊資料庫

作為詮釋資料模組的核心詮釋資料，並非使用DOI系統定義的核心應用描述檔，主要目的在使詮釋資料模組更具有描述實體類型資料的能力且更符合目前國內外數位典藏計畫使用的詮釋資料架構。事實上，DOI也承認可以使用DC作為系統初建的選擇，但其說明核心應用描述檔對物件描述更嚴謹，且Kernel AP在功能上可提供任何詮釋資料對照的基礎。（註27）因為<LIPS-DOI>系統採用Simple DC作為Kernel AP，資料類型（Type）元素的值遵照DCMI Metadata Terms Section 5規定的字彙來著錄。

此外，<LIPS-DOI>系統和DOI系統一樣也是以XML作為描述物件的資料結構。詮釋資料模組是以HTTP POST的方式將XML資料存入資料庫，建置為資料庫式的索引（Database Indexing）並透過微軟的XMLHTTP物件剖析XML字串提供編輯與更新的操作。

三、解析模組

DOI系統在解析方面發展了許多工具，包括了伺服器端工具與使用者端工具，使用者端的工具目前有下列幾種：（註28）

1. Handle/DOI Plug-in for Adobe Acrobat and Acrobat Reader
2. Server Side Resolution
3. Protozilla
4. DOI Button
5. HDL/DOI Protocol Handler for Mozilla
6. CNRI Handle System Resolver

其中最主要者為CNRI Handle System Resolver，其為適用於Internet Explorer與Netscape（6.x以前版本）的外掛程式，加裝後可以使瀏覽器能夠辨識DOI與handle進而解析之。另外，由Handle系統發展出來伺服器端的系統工具包括了：（註29）

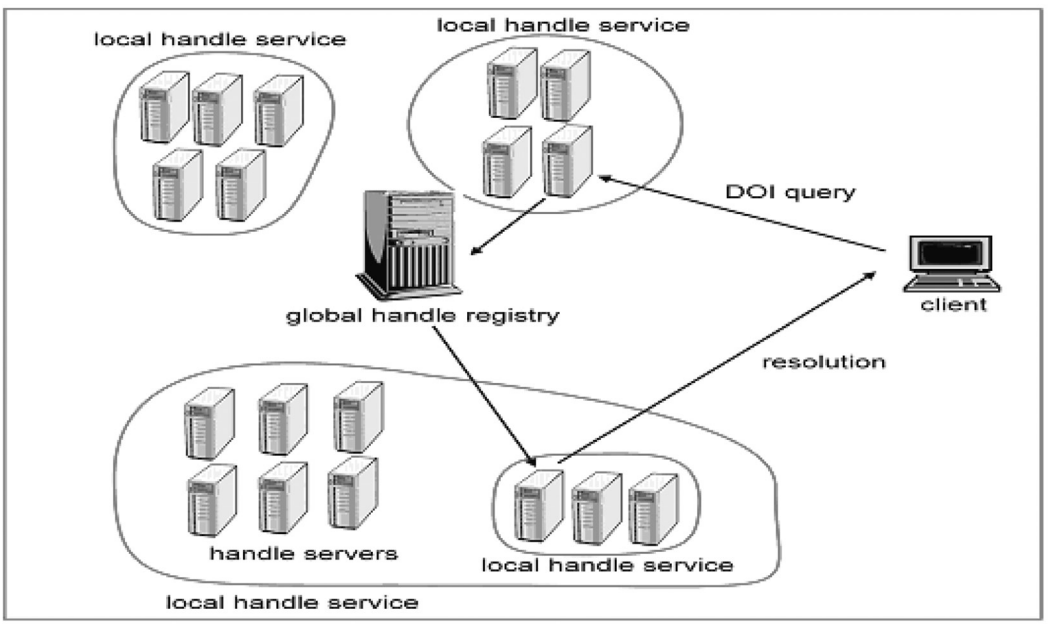
1. HSj (Local Handle System Server Version 5.3.1) JAVA Version.
2. Handle System Client Library (JAVA Version 5.3.0)
3. Handle System Client Library (C/Perl Version 5.3.0)

其中HSj為Local Handle Service中Handle伺服器用來管理使用者端的Handles所需要的軟體，Library則為不同語言的所需的函式庫。

有了列舉模組和詮釋資料模組之後，再加入解析模組的功能，才能使DOI系統真正有行動性。Handle系統的全球解析架構如圖七所示。

在DOI系統的Handle架構中，當使用者端使用瀏覽器在網路上瀏覽或使用到一組DOI時，這組DOI號碼將被送到Handle系統進行解析。在此情況中，使用者端的瀏覽器必須裝有Handle外掛程式才能理解Handle解析的通訊協定，或者使用者端也可透過一台瞭解此通訊協定的代理伺服器（proxy server）負責解析工作。Handle系統包含許多handle服務站台，每一服務至少包含一個主要站台（primary site）以及若干次要站台（secondary sites），而站台又包含若干的Handle伺

臺灣大學學術期刊資料庫



圖七　Handle 系統解析示意圖

服器。為了達到解析的目的，每一站台複製所有 Handle 服務中的全部 Handle。因此，在全球架構下，全球註冊服務（Global Handle Registry Service）負責記載本地服務（Public Local Services）的位置與名稱空間，而不屬於本地服務的解析查詢則可透過本地服務轉到全球註冊服務而由全球註冊服務轉至另一個本地服務。

在實作方面，<LIPS-DOI>系統已經完成提供單一解析代理服務，單一的解析也就是轉址服務，透過主要資料表中的解析網址，實驗系統可以將使用者端查詢的<LIPS-DOI>字串對照其值達成解析功能。使用者透過代理服務網頁鍵入<LIPS-DOI>字串之後就可以獲得轉址服務。多重解析方面<LIPS-DOI>系統規劃出兩個解決方案，第一個方案是建立解析資料表並以列舉字串與主要資料表關連以根據使用者的查詢由資料庫轉出呈現多重解析超連結的網頁。第二個方案是以XML描述多重解析（包含通訊協定與網址）置於主要資料表的解析欄位，然後透過XSL樣式語言進一步操作即可達到多重解析功能。

四、資料建置與系統測試

詮釋資料模組功能完備之後，<LIPS-DOI>系統建置初期以單筆方式著錄國立台灣大學圖書館出版的大學圖書館期刊電子全文。單筆著錄測試

臺灣大學學術期刊資料庫

完成後開始進入批次處理階段。首先進行整合的資源是國立台灣大學圖書資訊學系實習圖書室館藏中含有國際標準號碼（ISBN、ISSN）的中文圖書與期刊（實體物件）。批次處理的工作主要是參考 DC 與中國機讀編目格式的轉換對照表如表二所示，選擇符合的中國機讀編目格式欄位轉出符合的書目資料，透過<LIPS-DOI>的批次處理模組處理轉換成 batch.xml 檔案，再以批次方式將資料

表二、中國機讀編目格式轉出 Simple DC 欄位對照選取表中國機讀編目格式欄位標記與意義 Simple DC 欄位 001 系統控制號 DC: Identifier010 $a 國際標準書號 (ISBN)DC: Identifier011 $a 國際標準叢刊號（ISSN）DC: Identifier101 $a 作品語文 DC: Language122 $a 作品涵蓋時間 DC: Coverage200 $a 正題名 DC: Title200 $f 第一著者敘述 DC: Creator204$a 資料類型標示 DC: Type210 $c 出版者，經銷者等名稱 DC: Publisher210 $d 出版，經銷等日期 DC: Date300 $a 一般註 DC: Source330 $a 摘要註 DC: Description681 中國圖書分類號 DC: Subject702 $a 其他著者 DC: Con-

表二　中國機讀編目格式轉出 Simple DC 欄位對照選取表

中國機讀編目格式欄位標記與意義	Simple DC 欄位
001 系統控制號	DC: Identifier
010 $a 國際標準書號 (ISBN)	DC: Identifier
011 $a 國際標準叢刊號（ISSN）	DC: Identifier
101 $a 作品語文	DC: Language
122 $a 作品涵蓋時間	DC: Coverage
200 $a 正題名	DC: Title
200 $f 第一著者敘述	DC: Creator
204$a 資料類型標示	DC: Type
210 $c 出版者，經銷者等名稱	DC: Publisher
210 $d 出版，經銷等日期	DC: Date
300 $a 一般註	DC: Source
330 $a 摘要註	DC: Description
681 中國圖書分類號	DC: Subject
702 $a 其他著者	DC: Contributors
856 $u URL	Resolution

臺灣大學學術期刊資料庫

tributors856 $u URL Resolution

匯入詮釋資料庫。在台灣大學總圖書館編目組的協助之下，<LIPS-DOI>系統由 TULIPS 系統轉出符合需求的中國機讀編目格式欄位進行批次處理。共有37個中國機讀編目格式欄對照到15個 Simple DC 欄位，經過篩選決定取用的欄位。

由於<LIPS-DOI>系統使用簡易都柏林核心集作為核心詮釋資料，為了補足使用 Simple DC 不足之處，在轉入資料時將部分欄位的修飾語（qualifier）以括號方式置入說明編碼規則（Encoding Schemes），如使用美國國會圖書館標題（LCSH）或中國圖書分類號（CCL）來說明「主題和關鍵詞」（DC: Subject），以求詮釋資料更為詳確。（註30）

批次處理時，ISBN、ISSN或系統號等原有系統的唯一辨識碼均可作為字尾，不過為了保持更高的隱匿性，<LIPS-DOI>系統統一以系統控制號作為字尾。ISBN 和 ISSN 則使用在解析部分作為檢索 TULIPS 系統之用，提供存取該筆館藏的解析資訊。

在轉入資料時我們遭遇的問題是，即使以一對一的方式對照欄位，仍有欄位無法正確匯入，如300 $a一般註會同時對照到 Format、Source 與 Rights三個元素（但以功能來說，DC: Description 才是設計來放一般註的地方，不過此欄直接轉入時是取CMARC的330 $a 摘要註）。<LIPS-DOI>系統採取的解決方法是先將300 $a一般註全部轉入某一元素（<LIPS-DOI>系統的選擇是Source）再追溯修改。而856 $u URL 一欄則暫時置於 DC: Description，待多重解析的功能建置完成時再追補。

五、系統展示

<LIPS-DOI>系統的網址為：http://lipsdoi.lis.ntu.edu.tw，系統的主畫面如圖八所示，其中左列為中文選單，上列為英文選單。以下系統展示部分均以中文操作說明之。

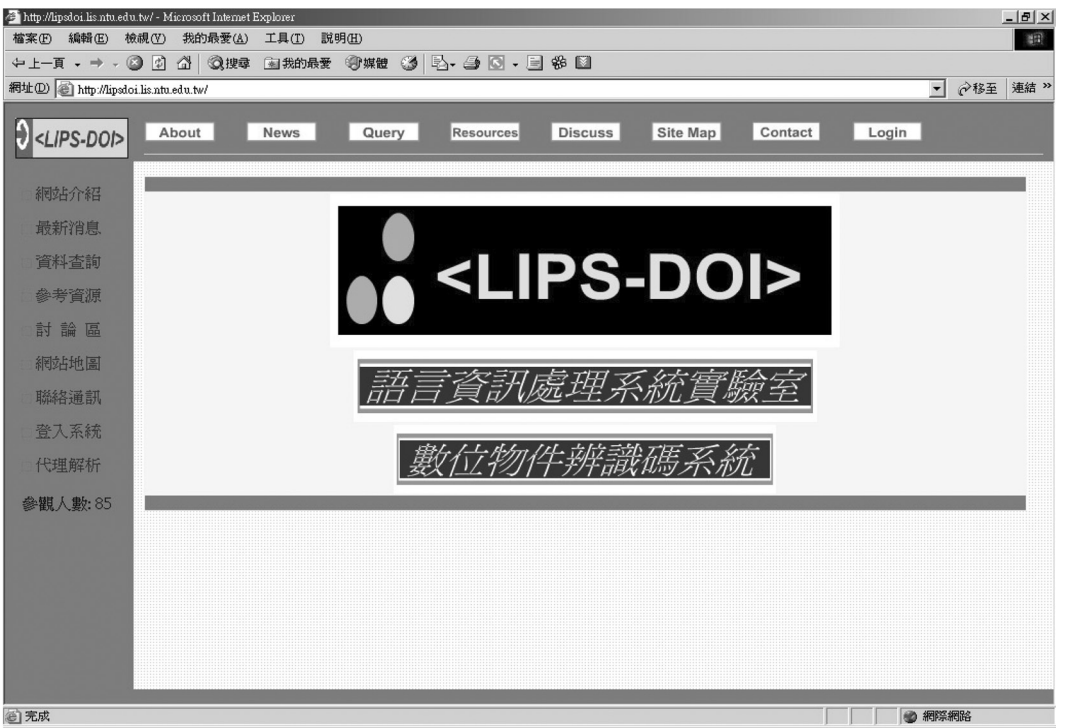
在選單中，網站介紹部分說明有關<LIPS-DOI>系統產生的緣起以及系統目前整合的資源數量與範圍。最新消息部分提供了與本研究相關的技術或文件之連結。資料查詢的部分可提供一般使用者檢索<LIPS-DOI>系統中已註冊的物件，其畫面如圖九所示。

欄位查詢包含了<LIPS-DOI>字串查詢、物件詮釋資料查詢、物件解析查詢三個部分。使用者可根據檢索需求鍵入部分或完整查詢語，然後按下 find 按鈕即可查詢。例如，使用者可以000005-來查詢系統中<LIPS-DOI>字首符合此值的註冊物件，或者，可輸入關鍵詞「圖書館」來查詢註冊物件的詮釋資料以尋找符合的物件。

跨欄位整合查詢的部分則令使用者能夠以不分欄位的方式同時查詢<LIPS-DOI>系統中的註冊物件，其實也就是整合了主要欄位查詢的三個欄位同時進行查詢。

由於所有註冊物件的詮釋資料是

臺灣大學學術期刊資料庫



圖八　<LIPS-DOI>系統主畫面

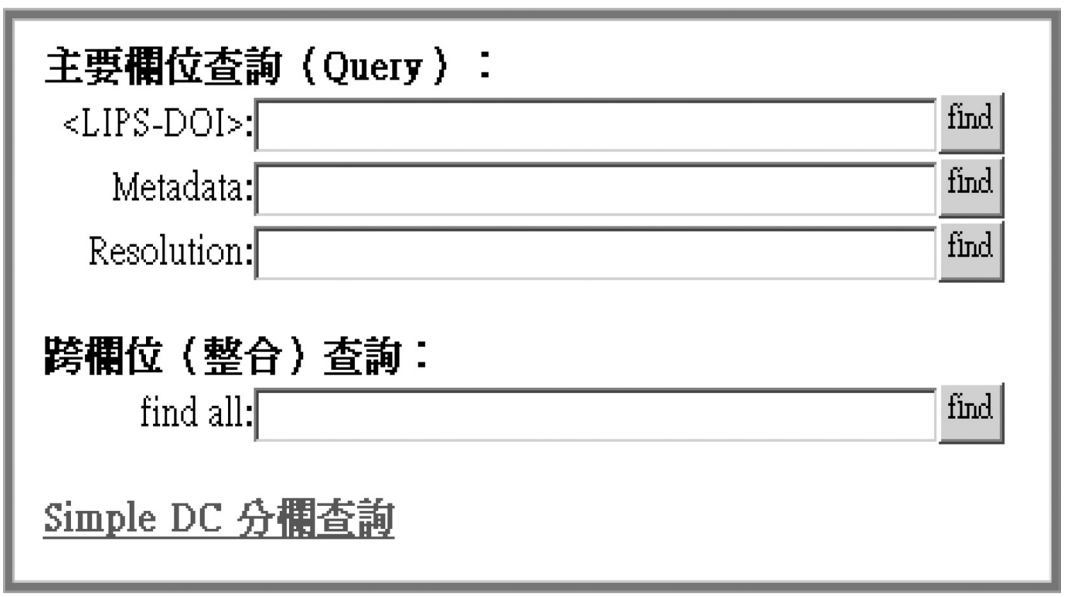
以 Simple DC 作為核心詮釋資料，因此為了進一步提供查詢的功效，<LIPS-DOI>系統也提供使用者以 Simple DC 元素名稱作為欄位來查詢註冊物件的詮釋資料，使用者只要按下「Simple DC 分欄查詢」即可連結至此查詢頁，如圖十所示。

在 Simple DC 查詢畫面中，使用者同樣只要在欲查詢的SimpleDC元素名稱分欄中鍵入部分或完整的查詢語即可檢索出符合的註冊物件。如查詢language為en或zh的註冊物件或format為 text/pdf 或 text/doc 的註冊物件。

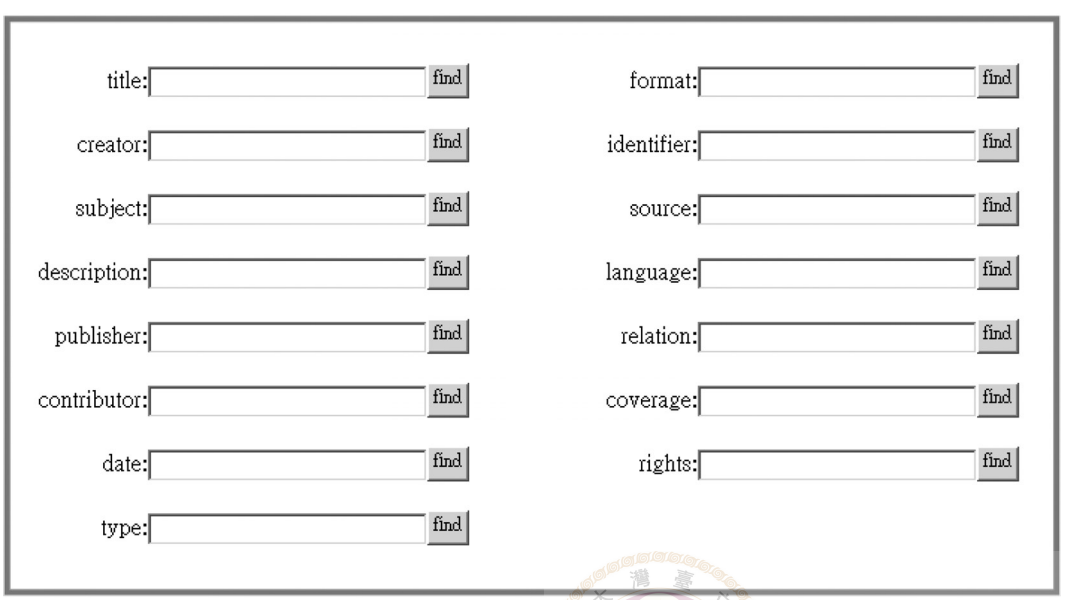
當使用者按下登入系統連結時則可進入<LIPS-DOI>的註冊使用者登入畫面，使用者輸入字首、登入名稱與密碼經過驗證無誤之後即可被導引至管理介面以進一步管理其註冊物件。<LIPS-DOI>系統註冊使用者管理介面中包含了單筆註冊、批次註冊、查詢註冊物件詮釋資料與列出註冊物件清單等操作功能，其中註冊物件清單部分包含編輯（edit）與刪除（delete）兩種操作的按鈕。如圖十一所示。

註冊使用者欲以單筆方式註冊新的物件時，可以在管理介面中按下「註冊新的 lipsdoi」新增按鈕，管理介面會轉至單筆註冊畫面，註冊使用

臺灣大學圖書館 期刊資料庫



圖九　<LIPS-DOI>系統資料查詢畫面

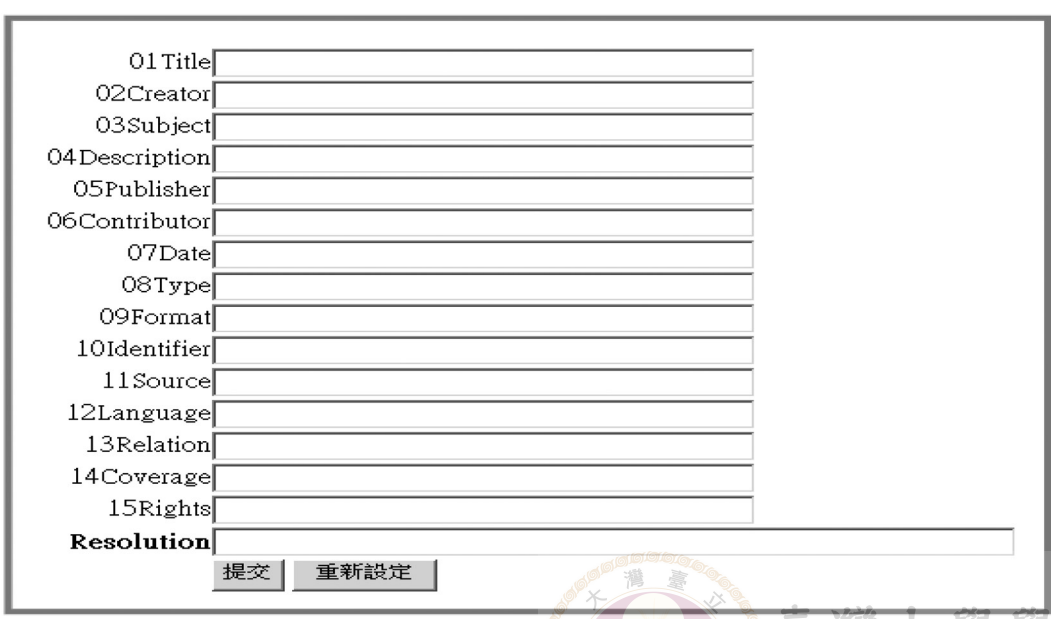


圖十　Simple DC 查詢畫面

臺灣大學學術期刊資料庫



圖十一　<LIPS-DOI>系統註冊使用者管理介面



圖十二　<LIPS-DOI>系統註冊使用者管理介面單筆註冊畫面

臺灣大學學術期刊資料庫

者按照 Simple DC 欄位著錄註冊物件的詮釋資料與解析網址後，即可按下提交按鈕註冊此物件。其中，identifier 可以著錄為原本物件資源的原有辨識碼，系統會將輸入的字串置於<LIPS-DOI>辨識碼的字尾部分。如註冊者不用原有的辨識碼，則字尾部分將省略。單筆註冊畫面如圖十二所示。

若註冊使用者欲以批次方式註冊物件，可按下瀏覽按鈕尋找批次處理檔案，進行批次新增動作。為了方便註冊使用者，<LIPS-DOI>系統的批次處理模組可以對 Microsoft Access 資料庫轉出的 XML 檔案進行批次處理。如果註冊使用者本來已有記載於資料庫軟體中的物件資料，只要將欄位名稱變更為與 Simple DC 相同的 15 個欄位名稱並將解析網址命名為 resolution 置於第 16 欄，便可轉出批次處理使用的 XML 檔案。批次處理 XML 檔案如圖十三所示。

在註冊使用者管理介面中，註冊使用者也可以透過<LIPS-DOI>字串、註冊物件詮釋資料、註冊物件解析資料以及跨欄位整合查詢的方式快速找到已經註冊的物件；也可以透過註冊物件清單逐次尋找註冊物件。除了已經註冊的<LIPS-DOI>辨識碼字串不會再變更之外，註冊物件的詮釋資料經過編輯之後便可以送出更新資訊，但是其辨識碼是不會變動的，如果註冊的數位物件在網路有所變動，註冊者只需要透過管理介面更新網址，則透過<LIPS-DOI>字串仍可以解析到正確的網址。

最後，使用者按下「代理解析」連結則可連到<LIPS-DOI>系統所提供的代理解析服務。如圖十四所示。使用者只要在服務網址後面鍵入某物件的<LIPS-DOI>字串，即可獲得此物件的正確解析，亦即連結到該物件。目前，<LIPS-DOI>系統的解析模組只達到單一轉址功能，在代理服務網址後鍵入欲解析的<LIPS-DOI>字串將馬上轉址到此物件最新的解析位址。

伍、問題討論

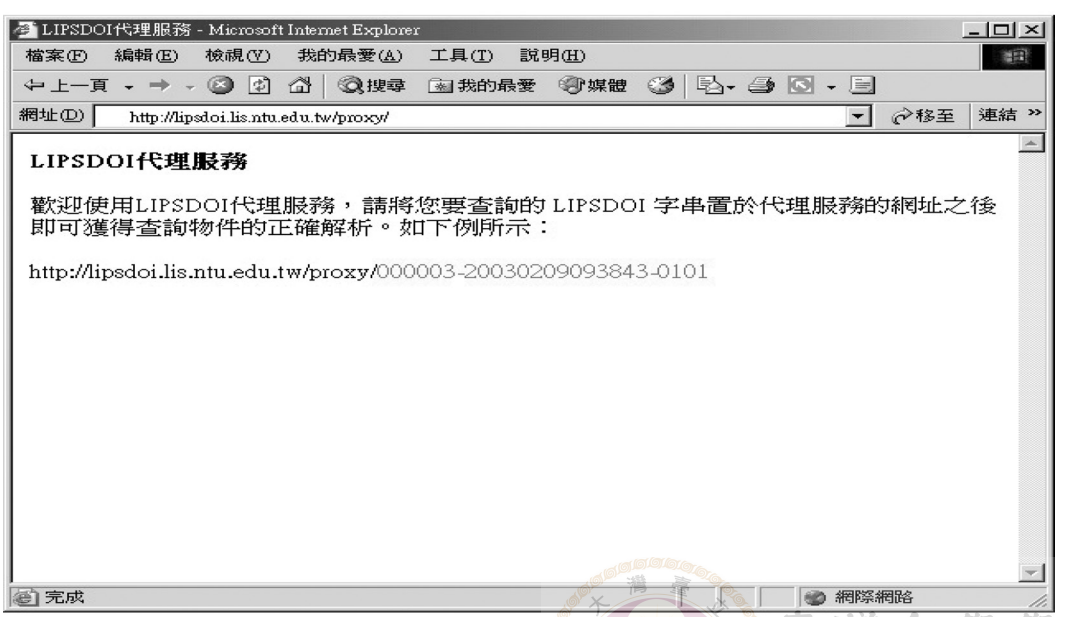
一、詮釋資料的互通

由<LIPS-DOI>系統描述模組建置批次處理工作階段的經驗來說，吾人可發現將複雜的 CMARC 詮釋資料轉換為<LIPS-DOI>Simple DC 核心詮釋資料，頗有削足適履的感覺。雖然說，當初<LIPS-DOI>系統採用 Simple DC 作為核心詮釋資料的主要原因在於 DC 已經為許多數位圖書館計畫所採用（如OAI-PMH等）。將實體圖書期刊的書目資料以 Simple DC 十五個欄位為標準批次轉入<LIPS-DOI>系統的主要目的是在印證<LIPS-DOI>系統可以兼容數位式資源與實體資源而形成聯合目錄。但為了保持原本物件的詮釋資料詳盡的描述資訊，建置 Metadata Crosswalk 功能可說是極為重要的後續發展方向。

臺灣大學學術期刊資料庫

```
<?xml version="1.0" encoding="UTF-8" ?>
<dataroot xmlns:od="urn:schemas-microsoft-com:officedata">
  <lipsdoi-dccore>
    <title>市場調查完全手冊</title>
    <creator>赫格、傑克生</creator>
    <publisher>小知堂</publisher>
    <date>2000[民89]</date>
    <type>text</type>
    <format>book</format>
    <identifier>957-03-4975-1</identifier>
    <language>zh</language>
    <rights>國立台灣大學圖書館</rights>
    <resolution>http://140.112.113.1/search*chi/i?SEARCH=957-03-4975-1</resolution>
  </lipsdoi-dccore>
  <lipsdoi-dccore>
    <title>執行力組織</title>
    <subject>csh 組織(管理)、csh 決策管理</subject>
    <publisher>麥格羅希爾</publisher>
    <date>2003[民92]</date>
    <type>text</type>
    <format>book</format>
    <identifier>957-493-728-3</identifier>
    <language>zh</language>
    <rights>國立台灣大學圖書館</rights>
    <resolution>http://140.112.113.1/search*chi/i?SEARCH=957-493-728-3</resolution>
  </lipsdoi-dccore>
</dataroot>
```

圖十三　批次處理 XML 檔案範例



圖十四　<LIPS-DOI>系統代理解析服務

臺灣大學學術期刊資料庫

二、DOI 系統的互通

由於<LIPS-DOI>系統是實驗性的架構，使用者可能會問，如果用<LIPS-DOI>系統著錄資源，將來有需要時能夠轉到DOI系統嗎？答案是肯定的。因為<LIPS-DOI>系統是開放式的架構，在列舉方面則只要將字首換成DOI系統註冊的字首，時間與字尾字串可以當作舊有的辨識碼，一樣可以整合到DOI的列舉模組。詮釋資料模組方面，在DOI架構之下，使用不同AP 的詮釋資料之間仍可以透過 iDD（indecs Data Dictionary, RDD, DOI Term-Set）互通。因此，吾人只要將<LIPS-DOI>系統中以 Simple DC 描述的詮釋資料對照到DOI的核心詮釋資料，仍然可以將描述模組的詮釋資料整批轉移到DOI系統。解析模組則只要將完整資料匯出，則可完全轉移到DOI系統的解析模組。

陸、結論與建議

<LIPS-DOI>系統設計之初，即將系統定位在整合具有公共存取性的中文資源的應用上，目的是要避免像DOI系統一樣僅針對大型的電子期刊出版商提供服務。不過由於DOI系統已漸形成國際間電子期刊出版的辨識碼標準，因此，<LIPS-DOI>系統的辨識碼語為[<DirectoryCode>.]<Registrant>-<RegistrationTime>[-<LocalName>]的結構，其中不用正斜線（/）作為字首與時間、字尾字串的分隔，目的在於與DOI系統的<DirectoryCode>.<Registrant-Number>/<LocalName>語法有所區別。換句話說，當<LIPS-DOI>系統整合的資源日益增多時，甚至形成重要的註冊碼發行機構時，吾人只要向IDF註冊字首，將<LIPS-DOI>辨識碼字串整個當作字尾或是向IANA申請URN使用，仍可以使<LIPS-DOI>系統與國際接軌。（註31）

如果<LIPS-DOI>系統成為國內提供中文資源數位物件辨識碼分配的機制，則吾人可以獲得經濟上的效益與使用上彈性。因為向IDF註冊每一個字首每一年都需要付出一筆相當的經費，申請越多字首，所必須付出的費用也就越龐大，這對中小型圖書館或其他規模不是很大的智慧產權工業從事者將是很大的經濟負擔。如果在台灣能夠透過<LIPS-DOI>系統來為國內的中文資源分配全球唯一且持續的辨識碼，則可省下許多經費。而當與全球使用DOI系統的資源接軌的需求提高時，由<LIPS-DOI>系統統一向 IDF 申請一個DOI字首，再由<LIPS-DOI>系統下的使用者共同負擔註冊DOI字首每年需繳的費用是比較合乎經濟效益的作法。同時，因為<LIPS-DOI>系統的辨識碼的語法在設計上即避免使用正斜線（/），所以可以將整個<LIPS-DOI>辨識碼作為DOI的字尾字串，再將合適的詮釋資料與解析資料轉入DOI系統，仍符合DOI系統架構的規範。<LIPS-DOI>系統的註冊使用者在使用上也能夠在需要的時候自行向

臺灣學術期刊資料庫

DOI系統註冊字首，然後由<LIPS-DOI>系統轉出資料匯入DOI系統，獨自付費擁有不同的DOI字首，因此其使用上具有極大的彈性。

整合各種資源類型確實是數位物件辨識碼系統重要的功能，尤其在詮釋資料互通方面，應用各種合於核心詮釋資料的應用描述檔將可更詳細地描述註冊物件，同時達到以同一入口檢索不同類型資源的能力。雖然<LIPS-DOI>系統已經針對國立台灣大學圖書館出版中心提供的電子期刊全文書目資料進行了基本的詮釋資料互通測試，但這方面的探討與測試顯然還有許多尚待努力之處。

近幾年來，我國的數位圖書館博物館計畫與檔案數位化計畫如火如荼地展開，數位資源在數量上不斷增加。雖然各個數位典藏專責機構系統內部有統一的命名機制或檢索功能，但隨著整合的需求日增，我們必須以數位物件辨識碼系統建立跨系統、整合各種類型資源的統一命名機制，然後以此基礎發展聯合目錄，提供各種資源的詮釋資料檢索、存取與註冊管理。<LIPS-DOI>系統提供了中文資源應用數位物件辨識碼系統合理而可行的架構，期待此研究能夠有更完善且持續的發展。

註釋

註 1：陳燕。「電子出版控制的一項工具 - DOI 系統」。<http://www.cgan.com/book/books/print/netpub/link/doi.htm> (22 Apr. 2003) .

註 2： DOI. “IDF member list," <http://www.doi.org/idf-member-list.html> (25 Apr. 2003) .

註 3：DOI News August 2003 <http://www.doi.org/news/03aug-news.html> (29 Aug. 2003) .

註 4：DOI Handbook (Version 2.7.0, February 2003). “What is a Digital Object Identifier?" <http://www.doi.org/handbook_2000/what_is_a_doi.html> (21 Apr. 2003) .

註 5：同註 4。

註 6：同註 4。

註 7：同註 4。

註 8：同註 4。

註 9：RFC1737, <http://www.ietf.org/rfc/rfc1737.txt> (12 Apr. 2003) .

註 10：三個功能模組在初期被稱為列舉（Enumeration）、描述（Description）、解析（Resolution），2003 年中更動為列舉（Numbering）、詮釋資料（Metadata）、解析（Resolution）。

註 11：DOI Handbook (Version 2.7.0, February 2003). “4 Enumeration," <http://www.doi.org/handbook_2000/enumeration.html> (29 Apr. 2003) .

註 12：DOI Handbook Appendix 1. “ANSI/NISO Z39.84-2000 Syntax for the Digital Object Iden-

臺灣大學學術期刊資料庫

tifier," <http://www.doi.org/handbook_2000/appendix_1.html> (29 Apr. 2003) .

註 13：同註 9。

註 14：DOI Handbook (Version 2.7.0, February 2003). “5 Metadata," <http://www.doi.org/handbook_2000/metadata.html> (29 Apr. 2003) .

註 15：DOI Handbook (Version 2.7.0, February 2003) Appendix 4. “An Introduction to <indecs> R," <http://www.doi.org/handbook_2000/appendix_4.html> (30 Apr. 2003) .

註 16： DOI Handbook (Version 2.7.0, February 2003). “3.3 Description," <http://www.doi.org/handbook_2000/components.html#3.3> (30 Apr. 2003) .

註 17： 同註 12。

註 18：在 DOI 系統發展初期註冊 DOI 的物件並沒有宣告詮釋資料，因此這些物件的 AP 以「沒有應用描述檔」（Zero-AP）統稱之，但 DOI 系統將逐漸以適用的 AP 為其重新建立詮釋資料以使其功能不會受到限制。

註 19：indecs 計畫。<http://www.indecs.org/project.htm> (30 Apr. 2003) .

註 20：同註 12。譯自 Table 1: DOI Metadata Kernel elements。

註 21：DOI Handbook (Version 3.0). “Metadata," <http://www.doi.org/ handbook_2000/metadata.html> (27 May. 2003) .

註 22：DOI Handbook (Version 2.7.0, February 2003). “6.4 Handle System technology," <http://www.doi.org/handbook_2000/resolution.html#6.4> (30 Apr. 2003) .

註 23：DOI Handbook (Version 2.7.0, February 2003). “6.7 The maintenance of DOI state data," <http://www.doi.org/handbook_2000/resolution.html#6.7> (30 Apr. 2003) .

註 24：DOI Handbook (Version 2.7.0, February 2003). “7 Policies," <http://www.doi.org/handbook_2000/policies.html> (30 Apr. 2003) .

註 25：時間與頻率國家標準實驗室。<http://www.stdtime.gov.tw/> (5 May 2003) .

註 26：吳翠鳳。「如何讓您的 Windows 2000 的 AD 架構與真實世界的時間同步」。<http://edu.uuu.com.tw/ucomtips/010706.htm> (7 May 2003) .

註 27：DOI Handbook. “DOI Metadata System and other metadata initiatives" <http://www.doi.org/handbook_2000/metadata.html#4.1.8> (16 May 2003) .

註 28：DOI Tools。<http://www.doi.org/tools.html> (16 May 2003) .

註 29：Handle System Software Download Instructions. <http://www.handle.net/ download.html> (17 May 2003) .

註 30：Dublin Core Metadata Initiative. “DCMI Metadata Terms Section 4: Encoding Schemes," <http://dublincore.org/documents/dcmi-terms/#H4> (27 May 2003) .

註 31：RFC3406. “Information about how to register a URN Namespace" <http://www.ietf.org/rfc/rfc3406.txt> (30 May 2003) .

臺灣大學學術期刊資料庫