

## 報紙新聞數位典藏 Metadata 轉換系統之設計與應用

林信成 (Sinn-Cheng Lin)

淡江大學資訊與圖書館學研究所 副教授

sclin@mail.tku.edu.tw

康珮熏 (Pei-Shiun Kang)

淡江大學資訊與圖書館學研究所 研究生

693070038@s93.tku.edu.tw

### 摘要

利用 Metadata 格式描述新聞事件，可加強新聞的結構性且增加自我描述性，有利於更迅速的交換、傳遞與分享數位化新聞。本研究利用「臺灣棒球運動珍貴新聞檔案數位資料館之建置」資料庫中的新聞資料加以延伸、加值，建置可自動轉換新聞 Metadata 的系統平台，支援國內外普遍採行的數種新聞 Metadata 格式，包括 Dublin Core 格式、數位典藏聯合目錄 DAC 格式、RSS 格式和 NITF 格式等四種。其中，Dublin Core 不但簡單易用，亦為其餘若干格式之基礎；DAC 則為數位典藏計畫共通標準，為聯合目錄系統指定規格，旨在便於大眾檢索數位典藏計畫所產出之數位化資源；RSS 為線上新聞訂閱格式，可打破時地限制，讓新聞資訊傳播更無遠弗屆；NITF 可供給通訊社或報社典藏新聞之用。透過不同 Metadata 格式的轉換提供，可加速資料的散播、典藏、共用，更便於使用者利用。

關鍵字：數位典藏、詮釋資料、標示語言

## 壹、前言

科技與網路的蓬勃發展，使得越來越多新聞媒體利用電腦及網路來相互傳播新聞，數位化新聞遠比傳統新聞需要更強而有力的資訊組織方法，以便能更迅速有效的進行交換、傳遞與分享，因此對於新聞資料的保存及利用也就產生了新的技術與規格，以求能將新聞資源做最佳化的管理典藏，並且透過系統平台讓使用者快速且簡捷的獲得新聞資料，加速資料的散播。

為解決數位化新聞資訊組織的問題，許多專用於新聞事件的 Metadata 格式也就隨之產生，且各有不同用途。例如：由國際新聞通訊協會 IPTC (International Press Telecommunication Council)<sup>1</sup>所制訂的 NITF (News Industry Text Format)<sup>2</sup>著重在新聞內文的描述；NewsML (News Markup Language)<sup>3</sup>則著重封裝多種不同的媒體，用於描述電子出版、傳送、典藏的新聞檔；SportsML (Sports Markup Language)<sup>4</sup>則用於運動項目紀錄；ProgramGuideML (Program Guide Markup Language)<sup>5</sup>是廣播與電視新聞節目專用。再者，如 IDEAlliance 發佈的 PRISM (Publishing Requirements for Industry Standard Metadata)<sup>6</sup>主要是為滿足雜誌、新聞、目錄、書籍和期刊等平面媒體出版者的商業需求而設計。又如 XMLNews.Org 所研擬的 XMLNews<sup>7</sup>主要在描述新聞報導的實質內容，是借用 NITF 而來的。至於衍生自 Netscape 推播技術 (Push) 的 RSS 則是一種用於互通新聞和其他 Web 內容的資料交換規格，目前已普遍應用於入口引擎、新聞網站、Blog 和 WiKi 等系統中。此外，海峽兩岸也發展出數種新聞 Metadata 格式，以應用於中文新聞資源，方便加速新聞資源的管理及傳播利用。如台灣方面有文建會制訂的「新聞紀錄 Metadata 格式」(News Records Metadata Format, 本文簡稱為 NRMF)<sup>8</sup>；新聞業界則有聯合報系正進行中的「聯合新聞標示語言」(UDN Markup Language, 簡稱 UdnML)<sup>9</sup>。大陸方面則有新華社所發展的「新華標示語言」(Xinhua Markup Language, 簡稱 XinhuaML)<sup>10</sup>，以及由中國報業協會所制訂的「中國報業電子新聞文稿格式」(Chinese News Text Format, 簡稱 CNTF)<sup>11</sup>。除了上述數種新聞專用 Metadata 外，也有泛用型的 Metadata，例如由 OCLC (Online Computer Library Center) 和 NCSA (National Center for Supercomputing Application) 所推

<sup>1</sup> IPTC, "International Press Telecommunications Council", available from <<http://www.iptc.org/pages/index.php>> (2005/3/22)

<sup>2</sup> IPTC, "News Industry Text Format", available from <<http://www.nitf.org/>> (2005/3/20).

<sup>3</sup> IPTC, "News Markup Language", available from <<http://www.newsml.org/>> (2005/3/20).

<sup>4</sup> IPTC, "Sports Markup Language", available from <<http://www.sportsml.org/>> (2005/3/20).

<sup>5</sup> IPTC, "Program Guide Markup Language", available from <<http://www.programguideml.org/>> (2005/3/20).

<sup>6</sup> IDEAlliance PRISM Working Group, "PRISM: Publishing Requirements for Industry Standard Metadata", available from <<http://www.prismstandard.org/>> (2005/3/22).

<sup>7</sup> XMLNews.Org, "NewsML", available from <<http://xmlnews.org/NewsML/>> (2004/12/20).

<sup>8</sup> 行政院文化建設委員會國家文化資料庫知識管理系統，"News Records Metadata Format"，可來自 <<http://km.cca.gov.tw/download/rule.html>> (2005/3/24)。

<sup>9</sup> 聯合報系於 2004 年 8 月成立 XML 小組，對旗下各報社的新聞格式做規範與 Metadata 的制定。

<sup>10</sup> 新華社技術局標準工組，XinhuaMLv1.0 功能說明書，2003 年 1 月 18 日。

<sup>11</sup> 中國報業協會規範工作組，"中國報業電子新聞文稿格式"，2000 年 5 月。

動的都柏林核心集 (Dublin Core, 簡稱 DC)<sup>12</sup>, 可用於描述各種電子資源, 其格式簡單易懂, 具延伸性與互通性, 可因應不同的需求, 亦適用於數位化新聞的描述。國科會制訂的「數位典藏聯合目錄 Metadata 格式」(Digital Archive Catalog, 本文簡稱為 DAC)<sup>13</sup>則為數位典藏計畫共通標準。而由全球資訊網協會 (World Wide Web Consortium, 簡稱 W3C)所制訂的「資源描述框架」(Resource Description Framework, 簡稱 RDF)<sup>14</sup>, 更可作為各種 Metadata 整合機制。

本研究以參與國科會數位典藏計畫之「臺灣棒球運動珍貴新聞檔案數位資料館之建置」資料庫為基礎, 採用泛用型之 DC、專用型之 NITF 與 RSS、以及聯合目錄所制定的 DAC 格式等四種國內外普遍盛行之 Metadata 格式為標準, 開發支援此四種 Metadata 自動轉換之系統, 並衍生出 RSS 訂閱和聯合目錄大批匯出等應用, 以提升新聞的利用價值, 有助於新聞資源的管理、典藏與加值利用。

## 貳、研究方法與步驟

本研究先對四種新聞 Metadata 格式進行比對分析, 再以系統實作法建置整個 Metadata 轉換系統, 其系統乃利用「台灣棒球運動珍貴新聞檔案數位資料館」中之新聞資料加以轉換建立棒球新聞管理系統, 並建置四種 Metadata 輸出轉換介面, 以滿足使用者需求, 系統架構如圖 1 所示。研究步驟如下:

- (1) 比較、整理關於 RSS、DC、NITF 與 DAC 各欄位之意義, 製作欄位對照表。
- (2) 設計本系統之所需功能: 後端系統採用主從式架構, 以關聯式資料庫進行系統實作。
- (3) 應用 ASP 語法撰寫 Metadata 轉換程式, 於 WEB 介面輸出 XML 格式之 Metadata。
- (4) 由 RSS 格式之轉換衍生成 RSS 訂閱系統, 提供「棒球史上的今天」供使用者訂閱該 RSS Feed, 透過 RSS reader 閱讀。
- (5) 將新聞資料轉換成 DAC 格式, 以應用於大批匯出 DAC 格式之新聞資料於聯合目錄系統作為典藏之用。

---

<sup>12</sup> Dublin Core Metadata Initiative, "Recommendations," available from <<http://purl.org/DC/documents/recommendations.htm>> (2004/12/17).

<sup>13</sup> 國科會數位典藏國家型科技計畫, "數位典藏聯合目錄", 可得自 <<http://catalog.ndap.org.tw>> (2005/1/6)。

<sup>14</sup> W3C, "RDF Primer - W3C Recommendation 10 February 2004", 2004/2/10, available from <<http://www.w3.org/TR/rdf-primer/>> (2005/1/6)

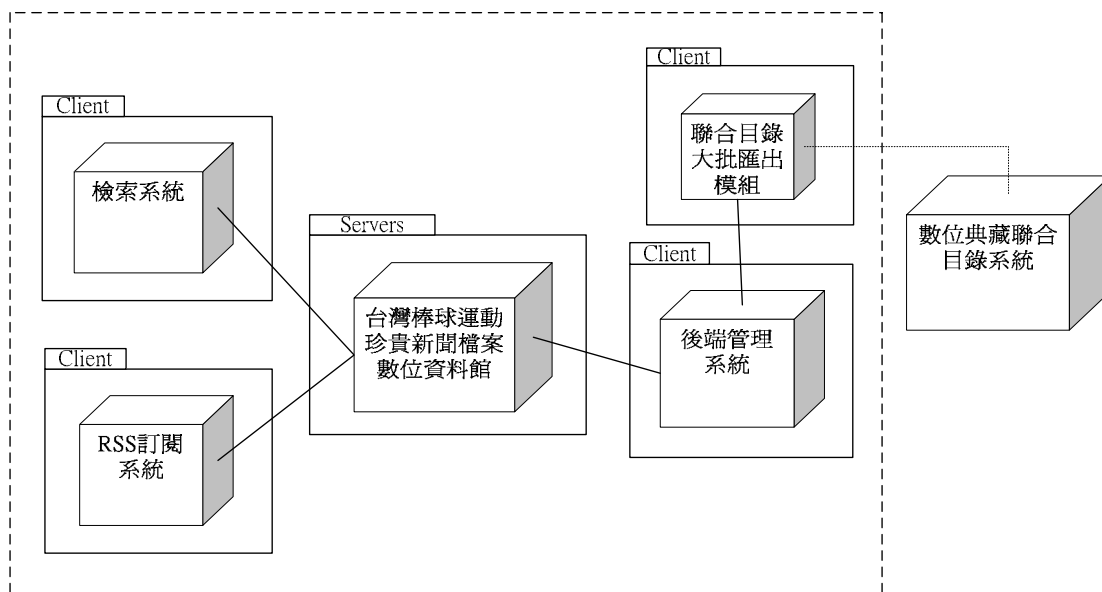


圖 1 「棒球新聞管理系統」之系統架構

## 參、四種 Metadata 比較分析

數位新聞相關之 Metadata 如前述所提有許多種不同格式，而本研究選用 DC、DAC、RSS 與 NITF 四種 Metadata 作為轉換，乃因這四種 Metadata 在不同領域各有其重要性，DC 簡單易用，不但是目前數位資源描述最普遍的 Metadata，且為其餘若干格式之基礎，其使用上具有豐富彈性，搭配 RDF 可融合各種 Metadata，將新聞資料詳加描述，更可配合 OAI-PMH 協定進行分散性資源檢索，以資源分享；DAC 為聯合目錄系統共通標準，乃國家數位典藏計畫指定規格，以整合數位典藏資源，供國人查詢利用；RSS 為線上新聞訂閱格式，大多用於個人用戶，經由訂閱步驟可讓使用者獲得即時新聞資訊，且資料內容以摘要方式呈現，便於使用者過濾所需新聞，挑選欲進一步閱讀者；而 NITF 則具豐富的內文語意，屬於內文標示格式，用於描述新聞內容資料，其元素規定嚴謹可詳細描述出新聞事件，以提供報社使用。至於其餘未於本研究中實作者，則留待作為未來研究方向。

### 一、 DC 規格探析

都柏林核心集 (Dublin Core) 的發展目的是希望建立一套描述網路上電子文件特色的方法，協助資訊檢索，用簡單的 Metadata 記錄來描述種類繁多的電子資源。DC 是一個簡單且具有彈性，各種專業人員皆可輕易瞭解、使用的資料描述格式，所以其只規範多數資源類型皆必須提及的資料特性，目前有 15 個欄位供資料描述，每個元素都為可選元素，也皆可重複。

將 DC 與 RDF 架構搭配使用，是描述數位化檔或網路資源的有效方式之一，利用 RDF 架構可便利於 DC 與其他 Metadata 加以整合，以豐富描述該文件之相關資訊，也可適用於數位化新聞資料，有助於數位典藏利用。

兩者搭配使用之規格架構可參考圖 2，其之根元素為<rdf:RDF>在此可宣告所引用之 XML Namespace，且<rdf:Description>中之屬性 rdf:about 為該篇新聞之 URI，其餘 15 個 DC 元素，則以<dc: title>形式注錄。

將新聞資料內文對應於 DC 欄位，除了因為 DC 為最普遍用於紀錄電子資源之 Metadata 外，未來更可應用與 OAI-PMH 機制，透過開放式檢索以助於新聞資料檢索與取用。

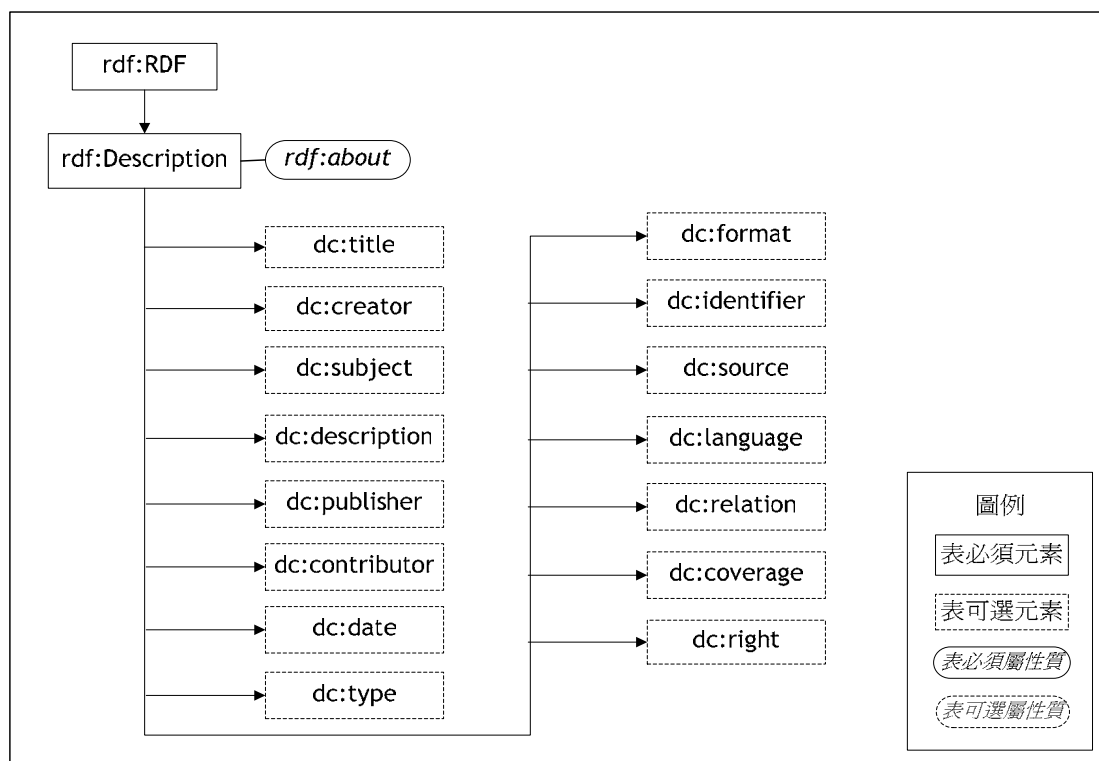


圖 2 RDF+DC 規格樹狀圖 (資料來源：本研究整理)

## 二、 DAC 規格探析

數位典藏聯合目錄 Metadata 格式 (Digital Archive Catalog)，主要功能乃將數位資源匯入聯合目錄系統之用，其根元素為<DACatalog>，包含<AdminDesc>與<MetaDesc>兩部份：<AdminDesc>為紀錄新聞相關管理資料，該部分由聯合目錄小組自行研訂；<MetaDesc>為紀錄新聞事件實質內容，遵從 DC 格式，可參閱圖 3。

<AdminDesc>中注錄該研究計畫之匯出機構相關資料，以及關於該計畫之數種分類，如內容架構、地理架構、時間架構等，以作為數位典藏分類檢索之用。其之子元素<DigiArchiveID>為紀錄唯一識別碼，<Hyperlink>則紀錄該資料之 link，<ICON>及記載藏品資料的縮圖，以 URL 表示。

<MetaDesc>則遵循 DC 的 15 個欄位，加以注錄該篇新聞之實質內容，如<Title>、<Creator>、<Publisher>、<Language>、<Type>等資料，聯合目錄之 DC 注錄採取首字大寫，如 title 改以<Title>為元素名稱。每一個 DC tag 僅包含一筆描述資料，而非多筆描述資料置入同一 DC tag 中。

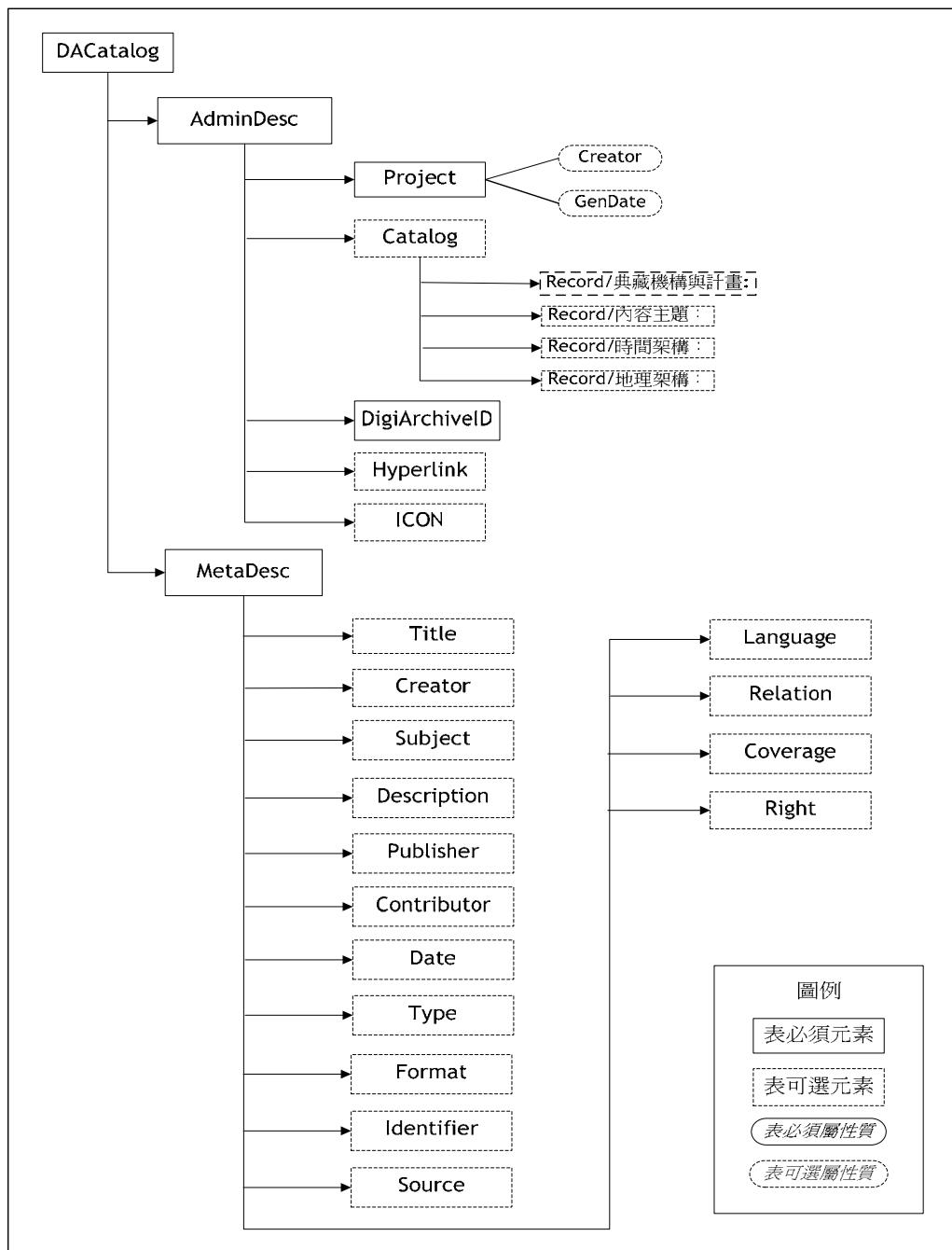


圖 3 數位典藏聯合目錄 Metadata 規格樹狀圖 (資料來源：本研究整理)

### 三、 RSS 規格探析

RSS 的規格與其發展歷史有很大的關係，從 1999 年 Netscape 公司所發展之 RSS 0.90 版為最早，此版本套用 RDF (Resource Description Framework) 框架，根元素為<rdf:RDF>，其規定單一 channel 只能有 15 個<item>，各個元素之內容長度也有各別規定，如<channel>之<title>長度只有 40 字元等規定，使用上較多限制，語法上也較為複雜，因此造成發展群意見產生分歧，令 RSS 技術衍生出許多不同版本。

當時，部份發展者則認為簡單的規格定義能有助於 RSS 的推廣與使用，因此 UserLand Software 在 2000 年 6 月衍生發展出 RSS 0.91 規格<sup>15</sup>，此版本強調使用上的便利與簡單性，故並無套用 RDF 框架。RSS 0.91 直接以<rss>為根元素，<rss>元素中強制版本屬性(version)用以表示所用的 RSS 的準確格式，但仍規定單一 channel 只能有 15 個 item，且<link>與<url>元素規定只能是 http 或 ftp。

2000 年 12 月 UserLand 公司又發展出 RSS 0.92 規格<sup>16</sup>，此版本可視為 RSS 0.91 之進階版，其規格多相似於 RSS 0.91。RSS 0.91 與 0.92 之差別在於新增了五個元素，將新聞資源描述更詳細，並可透過<item>的<enclosure>子元素描述多媒體附件之連結，且不再限制 channel 中 item 之數量，以及各元素內容長度之規定<sup>17</sup>。UserLand 公司其後又陸續推出 RSS 0.93、RSS 0.94 規格，但其修改不大，多依循 0.91 與 0.92 規格，故這些 UserLand 公司所主導之 RSS 版本可泛稱為 RSS 0.9x (Rich Site Summary)，0.9x 各版本間皆可相容。

除 UserLand 公司所發展標榜簡單易用的 RSS 0.9x 外，非營利網際社群 RSS-DEV 工作小組也在 2000 年 12 月發展出 RSS 1.0 規格<sup>18</sup> (如圖 4)，其語法上繼承 Netscape 公司發展之 RSS 0.90 規格，此版本亦套用 RDF 架構，並使用 XML Namespace 及延伸性 Module 語法，此版本之 RSS 與 Netscape 0.90 相同涵義，都代表 RDF Site Summary。RSS 1.0 的根元素為<rdf:RDF>在其中宣告 Module Namespace 的元素，其 RSS 1.0 規格中所規定之元素較其他版本少，需搭配 Module 才可豐富新聞事件之描述。RSS 1.0 目前有三種標準 Module，即 Dublin Core、Syndication 與 Content，除此尚有 16 種不同的 Propose Modules<sup>19</sup>。

---

<sup>15</sup> Dave Winer, "RSS 0.91", available from <<http://backend.userland.com/rss091>> (2005/03/22).

<sup>16</sup> Dave Winer, "RSS 0.92", available from <<http://backend.userland.com/rss092>> (2005/03/22).

<sup>17</sup> Hammersley, Ben, "Content Syndication with RSS", O'Reilly, 2003, p.18.

<sup>18</sup> RSS-DEV Working Group, "RDF Site Summary (RSS) 1.0", available from <<http://web.resource.org/rss/1.0/spec>> (2005/1/28).

<sup>19</sup> 同註 17, p.72.

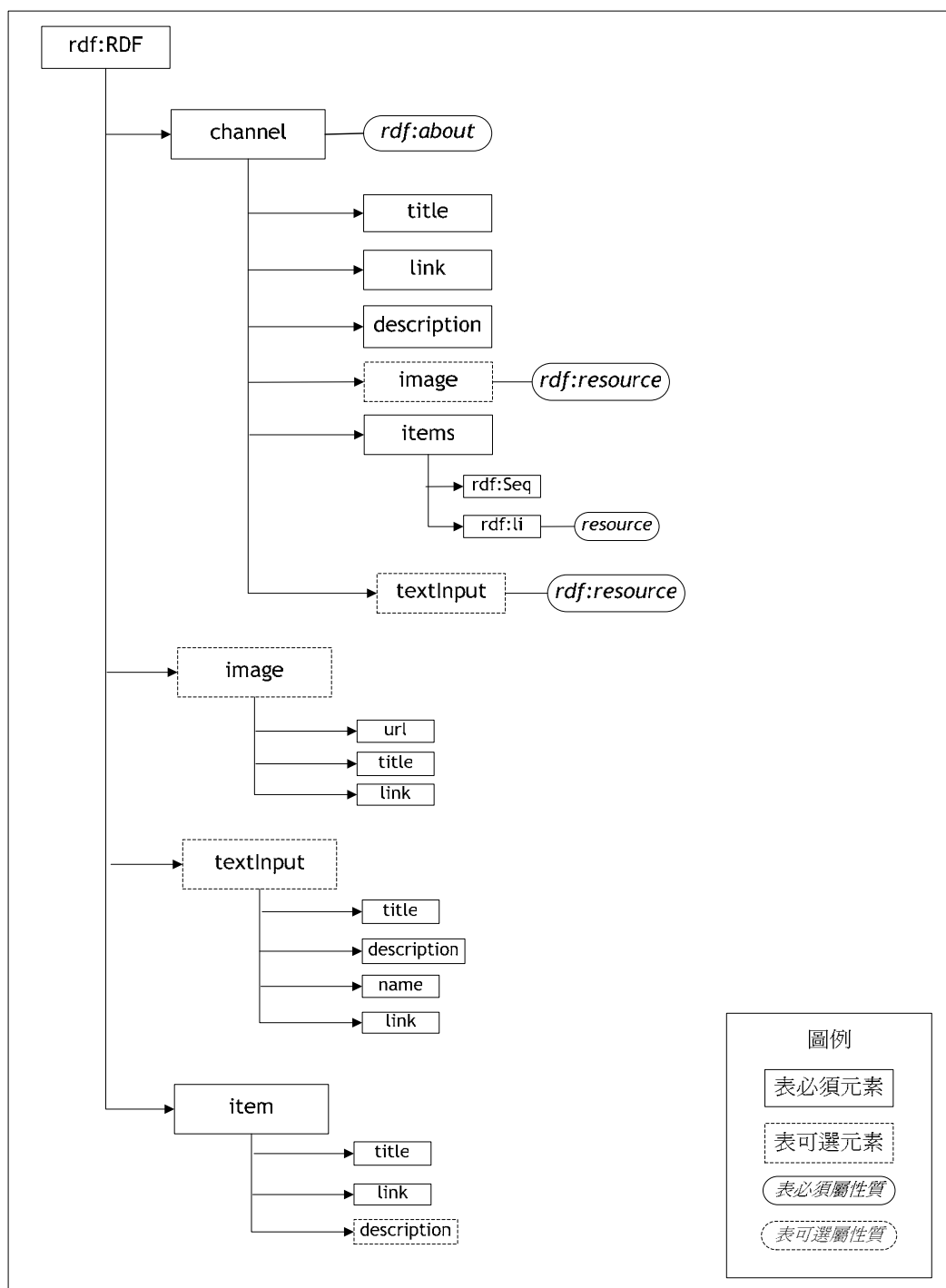


圖 4 RSS 1.0 規格樹狀圖 (資料來源：本研究整理)

2002年8月UserLand公司再次推出RSS 2.0規格<sup>20</sup>(如圖5)，此版本針對RSS 1.0的Module延伸架構而產生，基本元素皆繼承RSS 0.9x版本，但加以支援XML Namespace，允許可延伸Module的使用，為與RSS 0.9x版本相互區別，故RSS 2.0之RSS代表Really Simple Summary。其根節點<rss>元素中強制版本屬性(version)改以<rss version="2.0">表示，RSS 2.0以RSS 0.92為基礎，新增許

<sup>20</sup> Berkman Center, "RSS 2.0 Specification", available from <<http://blogs.law.harvard.edu/tech/rss>> (2004/10/29).

多新的元素，如<item>中增加唯一識別子元素<guid>，以解決<url>可能重覆之缺點，<url>不再限定只能為 http 或 ftp，而可支援任何符合 URI 的格式。RSS 2.0 目前已推出一個官方 BlogChannel Module，為使 RSS Reader 可藉以訂閱 Blog 文章，達到內容聯合。

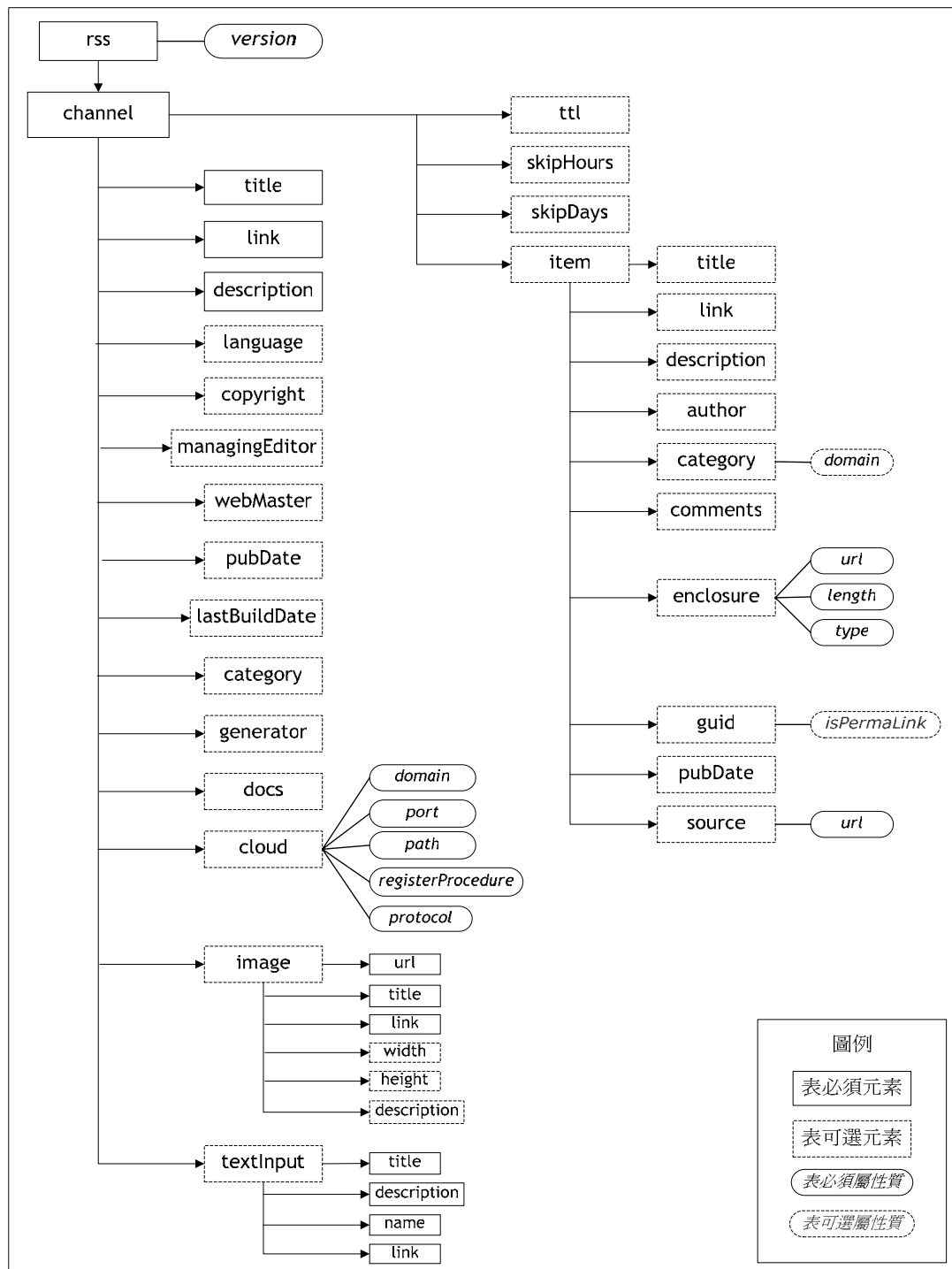


圖 5 RSS 2.0 規格樹狀圖（資料來源：本研究整理）

RSS 1.0 和 2.0 格式所包含的核心資訊相同，但其結構不一樣，RSS 1.0 的根

元素是<rdf:RDF>，RSS 2.0 的根元素是<rss>。另一個主要差別是 RSS 1.0 規格有 Namespace 限定，RSS 2.0 規格就沒有此一限定，但其實兩個版本的格式本質上是一樣的，兩個版本都包含<channel>元素，而<channel>元素又包含三個必須的元素：<title>、<description>和<link>。

目前盛行之 RSS 版本為具有 RDF 框架的 1.0 或簡單使用的 2.0 版，兩者皆被廣泛利用於 Blog、Wiki 和網上新聞頻道等，許多知名新聞社網站都提供 RSS 訂閱支援，而多數 RSS Reader 也皆可支援 RSS 1.0 與 2.0。

由於 Netscape 所發展之 RSS 0.90 版本目前已鮮少人使用，而 RSS 0.9x 與 RSS 2.0 一脈相承（如圖 6 所示），故本研究只採取 RSS 1.0 與 RSS 2.0 兩規格創建 RSS Feed，作為新聞 Metadata 格式轉換使用。

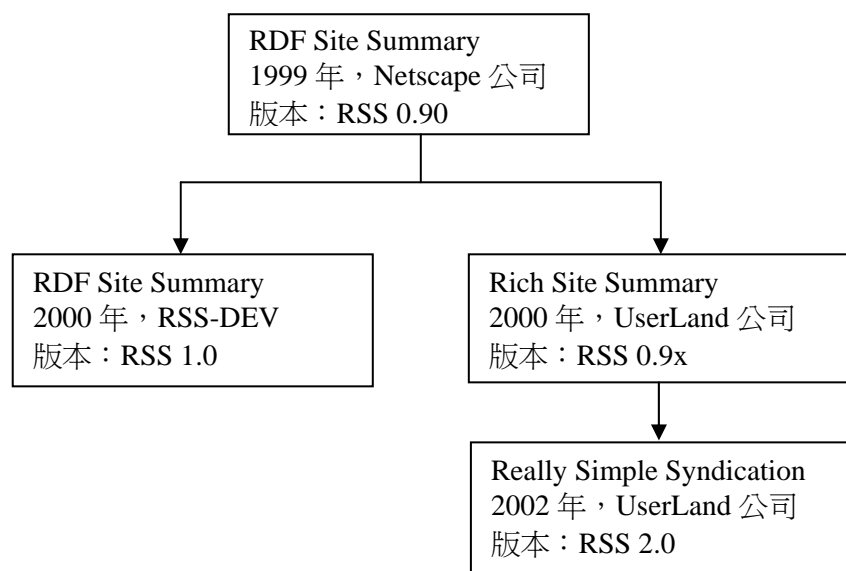


圖 6 RSS 版本發展歷史<sup>21</sup>

（資料來源：本研究整理）

#### 四、NITF 規格探析

NITF一開始採用SGML作為其語法規則，直到1998年XML頒佈後，即改採XML語法。NITF以XML為基礎標準，著重於新聞內文的描述，主要用於傳統平面媒體的標示，設計初衷在加強新聞檔重覆使用的便利性，讓通訊社、報業、網站等新聞同業可以將新聞稿件轉換成不同的資料格式，並呈現在各種新聞平臺上，達成多重發行的目的。

NITF規格最新版本為3.2版，包含138個元素，可分為五個層級，規定詳細縝密，將新聞事件完整描述紀錄，其之根元素為<nitf>其下可分為<head>與<body>兩部分，<head>用於紀錄關於該篇新聞相關之資訊，如<title>為該新聞之標題、<tbody>紀錄該新聞之類目、<docdata>可注錄該新聞相關之資料、<pubdata>則

<sup>21</sup> RSS 另有 3.0 版本，為純文字檔(text)型式，直接於 IE 瀏覽文字檔或以 mail 流通，該版本屏除 XML、名稱空間(Namespace)以及 HTML 格式，故不在本研究討論範圍。

以該新聞之發佈相關資訊為內容值等；<body>則紀錄該篇新聞實質之內容資訊，其又可分為<body.head>、<body.content>與<body.end>，<body.head>注錄文章刊頭資訊，如標題<headline>、作者<byline>等；而<body.content>則注錄新聞內容本文資料；<body.end>用於注錄一些與新聞資料相關之訊息，其下有兩個子元素<tagline>與<bibliography>，前者可作為該新聞的最後屬名，該元素之下又可加入許多子元素加強描述性；後者則注錄不限定格式之書目資料。因NITF之元素眾多，層屬關係複雜，因此圖 7只繪出三層元素結構，其所有元素的屬性值與第四、第五層元素架構可參考：

<http://www.nitf.org/IPTC/NITF/3.2/documentation/nitf.html>。

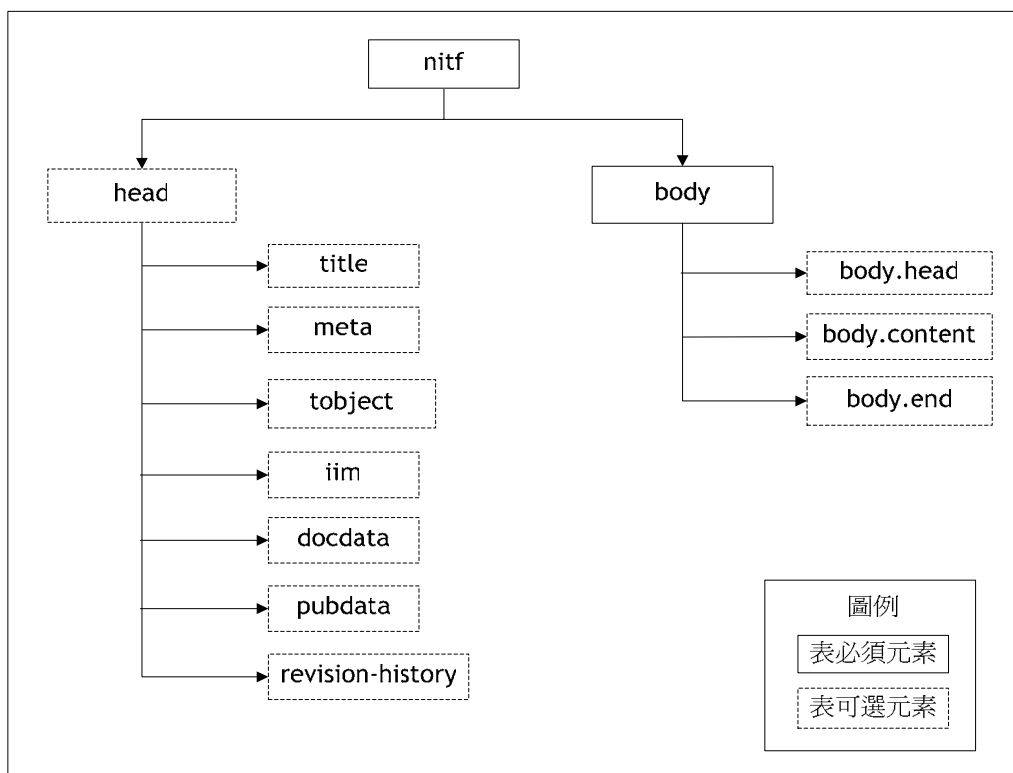


圖 7 NITF 簡易規格樹狀圖 (資料來源：本研究整理)

## 五、四種 Metadata 之綜合比較與對應

表 1 為上述四種 Metadata 之相關資料整理表，由該表可知其中最早發展者為 NITF。而四者之個別功能有其相異，如 RSS 與 Dublin Core 規格簡單，以描述該新聞資訊之基本資料欄位，使用多彈性，元素與規格較為簡易，且皆可支援延伸性 Module 的 Namespace 宣告；NITF 為滿足整體新聞媒體描述，使用規範詳細，語法較為複雜；DAC 較 Dublin Core 多了數個作為管理檔使用之元素，但仍繼承 Dublin Core 之使用彈性與簡易度。四者之共同性在於皆可以 text/xml 格式，透過通訊協定溝通傳輸，於 Web 介面呈現。

表 1 四種 Metadata 比較表

| 項目   | Dublin Core  | DAC   | RSS  |   | NITF  |
|------|--|---|--|---|---|
|      |  |   | RSS 1.0  | RSS 2.0   |   |
| 起始   | 1995 年   | 2002 年  | 最初 RSS0.90 為 1999 年                                      |   | 1990 年  |
| 發展組織 | OCLC 與 NCSA  | 行政院國家科學委員會                                      | RSS0.90—Netscape   |   | IPTC  |
|      |  |   | RSS1.0—RSS-DEV   | RSS2.0—UserLand   |   |
| 適用媒體 | 網路電子資源   | 數位化電子資源   | 網路新聞消息   |   | 平面新聞媒體  |
| 描述重點 | 定義資源之基本資料  | 定義資源之管理資料與內容                                    | 定義新聞事件之摘要資訊  |   | 定義新聞的內容與結構  |
| 應用目的 | 協助典藏描述電子資源   | 數位典藏資料記錄以供檢索                                    | 協助使用者快速便利獲得新聞資訊  |   | 協助內容識別  |
| 發佈對象 | 典藏機構   | 典藏機構  | 個人用戶   |   | 通訊社   |
| 規定元素 | 15 個元素   | 24 個元素  | RSS1.0 : 23 個元素  | RSS2.0 : 31 個元素   | 138 個元素   |
| 根元素  | <rdf:RDF>  | <DACatalog>                                     | <rdf:RDF>  | <rss>   | <nitf>  |
| 使用彈性 | 可搭配 RDF 架構，支援延伸性 Module 的 Namespace 宣告，所有元素皆為可選元素，彈性利用選取所需，皆可重複。 | 可分為兩部份，管理性資訊需依聯合目錄之規定注錄，數位資源內容描述則沿用 DC 為主要注錄欄位。 | RSS1.0—使用 RDF 架構，規定元素較少，支援延伸性 Module 的 Namespace 宣告。     | RSS2.0—富彈性，除三個必須元素外其餘皆為可選元素，支援延伸性 Module 的 Namespace 宣告，但並不支援 RDF 語法。 | DTD 規定詳盡，階層關係眾多，為求詳細表現新聞資料之內容描述，以滿足新聞增值，不支援延伸性 Module 的 Namespace 宣告與 RDF 語法。 |
| 難易   | 簡單易用   | 簡單易用  | 語法較複雜  | 簡單易用  | 元素層級架構繁雜  |
| 發佈方式 | 資料交換用，以 XML 於 Web 介面呈現。  | 資料典藏以作為檢索查詢之用。                                  | 拉式 (Pull-based) 為基礎，使用者主動訂閱 RSS Feed，利用 RSS Reader 瀏覽閱讀。 |   | 多做資料交換用，以 XML 於 Web 介面呈現。   |
| 格式   | 皆可透過 text/xml 呈現   |   |  |   |   |
| 環境   | 皆建立於網路作業環境，透過通訊協定溝通傳輸  |   |  |   |   |

(資料來源：本研究整理)

為建置 Metadata 自動轉換之系統，我們以 DC 為基礎，與 RSS、NITF 與 DAC 三種 Metadata 之元素，以及資料庫內部 Metadata 欄位進行比對分析歸納出常用元素與欄位對應關係，作為 Metadata 轉換之用。

### (1) Identifier (資源識別代號)

此元素乃 DC 用以紀錄該資源之唯一識別碼，故與內部 Metadata 之主索引相對應，而 DAC 記述實質內容元素乃以 DC 為架構，因此該元素與 MetaDesc 的子元素<Identifier>對應，除此 DAC 自行規定用以紀錄新聞相關管理元素中也設置<DigiArchiveID>為注錄該檔之唯一識別碼。

在 RSS 2.0 中<item>的子元素<guid>也是用以表示該 item 唯一識別字串，但其之格式並不限定使用 OSF GUID (Globally Unique Identifier) 演算法，故本研究參考現行之網路新聞發佈方式，給予該元素內容值為該 item 的 URI。而在 RSS 1.0 中本身規格並無規定唯一識別碼相關的元素，因此需搭配 RDF 引用 DC 標準 Module，以<dc: identifier >相對應。唯一識別碼屬於關於該篇新聞之相關資訊，而非新聞內容資訊，故在 NITF 中對應至<head>中紀錄該新聞資料的<docdata>子元素<doc-id>。

### (2) Date (出版日期)

Date 元素在於 DC 中並無規定其為該應為該新聞之發佈日期或該 XML 資源創建的日期，但因本研究著重於新聞本文之描述，所以將此元素對應於該新聞事件的發佈之日期，故與內部 Metadata 之新聞日期相對應。DAC 格式遵循 DC 架構，所以相同對應至<MetaDesc>的子元素<Date>，除此 DAC 自行規定用以紀錄新聞相關管理元素中為作為分類紀錄之用，設置<Catalog>以其之子元素<Record>紀錄不同類別之分類，其中包含「時間架構」之項目，也可相對應之。

RSS 2.0 的<channel>子元素<pubDate>是用以紀錄該 RSS Feed 發布之日期，而<item>的子元素<pubData>是用以表示該 item 之發佈日期，因此在本研究中 DC 的 Data 應與<item>的子元素<pubData>相對應。RSS 1.0 之規格中無規定日期相關元素，但因 DC 為其標準 Module 之一，故可搭配 RDF 與 DC 本身相同對應至<dc:date>。在 NITF 中 head 中紀錄該新聞資料的<docdata>子元素<data.issue>是用以紀錄該新聞發佈的日期；紀錄出版資料的<pubdata>子元素<date.publication>用以紀錄新聞出版日期；而新聞發佈日期也屬於新聞內容的一部份，所以在<body>的<body.head>中設置紀錄該篇新聞之發佈日之元素，於<hedline>的子元素<dateline>，此三者在本研究中皆可對應於 DC 的<date>，但其功用不盡相同。

### (3) Title (題名)

<title>元素對應至新聞事件，可視為新聞標題，但 DC 並無區別主標題或副標題於不同元素，雖 DC 欄位可重複但無法區分主標題與副標題之差異，副標題對於新聞事件又有相當重要性，故本研究參考現行網路新聞發佈規格，用單一<title>對應於內部 Metadata 之新聞主標題與副標題，注錄格式為「主標題：副標題」。故 DAC 也與<MetaDesc>的子元素<Title>相對應，採用相同注錄格式。

在 RSS 2.0 中記述新聞文件標題者為<item>的子元素<title>，而<channel>的<title>是用以紀錄該發佈 RSS Feed 頻道的標題，所以 DC 之<title>應和<item>的子元素<title>對應，採相同注錄格式。RSS 1.0 在<item>元素亦有定義紀錄文章標題用之<title>元素，所以不需引用 Module。而 NITF 為新聞專用內容描述的 Metadata，故其格式規定詳細，有不同的元素對應於主標題與副標題，<head>的子元素<title>以及<body>的<body.head>下層<headline>中區分子元素<h1>用以紀錄該篇新聞之主標題，<h2>用以紀錄該篇新聞之副標題。

#### (4) Creator (著者)

DC 的該元素是用以紀錄該資源的著者，因此在新聞事件中可對應於撰寫或採訪該新聞之記者，故與內部 Metadata 之新聞記者對應。DAC 記述內文部份以 DC 為架構，所以 DC 的<creator>與<MetaDesc>的子元素<Creator>對應。

在 RSS 2.0 中<author>為紀錄新聞作者的 mail，但因本資料庫之內容是報紙新聞，因此採注錄記者姓名而非 mail。RSS 1.0 本身架構並無規定相關元素，因此需引用 DC Module，對應於<dc:creator>。新聞作者也為新聞內容的一部份，所以在<body>刊頭資訊<body.head>之下層<headline>子元素<byline>用以表示該新聞之作者。

#### (5) Publisher (出版者)

出版者即表示該資源的發行組織，因此 DC 的<publisher>應注錄發佈該新聞之報別，與內部 Metadata 之新聞報別相對應。因此在以 DC 為主架構的 DAC 中該元素應與<MetaDesc>的子元素<Publisher>對應。

RSS 2.0 中並無紀錄出版者之相關元素，但新聞事件之出處應為重要之版權資訊，故本研究參考現行網路新聞發佈規格，將報別對應至 title 元素，其注錄方式為「主標題：副標題（報別）」。RSS 1.0 本身亦無相關版權資訊注錄元素，因此需引用 DC Module 對應注錄於<dc:publisher>。新聞報別應為擁有該篇新聞發佈之權利者，也為新聞內容的一部份，可讓讀者得知此一新聞的出處，所以可對應於 NITF 的<body>刊頭資訊<body.head>中的<headline>子元素<rights>。

#### (6) Contributor (其他參與者)

DC 中該元素用以紀錄除了著者外，對該資源創作有貢獻的其他相關人士或組織，在新聞資料中可對應至攝影者或繪圖者，但因本資料庫的記者欄位將文字記者以及攝影記者等相關作者，紀錄於同一欄位，故本元素與內部 Metadata 並無對應。但 DC 的<contributor>可與 DAC 中的<MetaDesc>的子元素<Contributor>對應。

在 RSS 2.0 中只有規定<author>一個元素，並無提供參與創作者之相關元素可與之對應。RSS 1.0 可引用 DC Module 加以對應於<dc: contributor>。NITF 規格中在<body>刊頭資訊<body.head>的子元素<contributor>用以紀錄資訊的貢獻者，與 DC 此一元素相對應之。

#### (7) Description (敘述)

該元素在 DC 中用以紀錄該文件的摘要或影像資源的內容敘述，對於新聞事

件而言，即指新聞之內文，可與內部 Metadata 之新聞內容相對應。因此在 DAC 中也與<MetaDesc>的子元素<Description>對應。

在 RSS 1.0 與 RSS 2.0 中記述新聞文件摘要或內容者皆為<item>的子元素<description>，而<channel>的<description>是用以紀錄該發佈 RSS Feed 頻道的相關內容資訊，故 DC 之<description>應與<item>的子元素<description>相對應，但在 RSS 1.0 此元素為可選元素，RSS 2.0 此元素則為必備元素，兩者規格有所差別。而 NITF 紀錄新聞內容者為<body>的子元素<body.content>。

#### (8) Subject (主題和關鍵詞)

藉由主題與關鍵詞可表示該資源的類目，所以用於報紙新聞資料，可從版名去分別出新聞之主題，因傳統報紙不同版面收集不同主題類目的新聞事件，因此可與內部 Metadata 之新聞版名對應，若該篇新聞無版名資訊，則由系統注入「體育新聞」作為內容值。DAC 格式乃以 DC 為架構，因此與<MetaDesc>的子元素<Subject>對應。

在 RSS 2.0 中<item>的子元素<category>是用以注錄該篇新聞的類目；<channel>亦有相同子元素<category>是作為紀錄該頻道的分類，其兩者用法相同，但紀錄內容意義不同，故本研究的 DC 之<subject>應與<item>的子元素<category>相對應。RSS 1.0 規格無類目相關的元素，需引用 DC Module 以強化其描述性，所以對應至<dc:subject>元素。NITF 有其規定之類目主題層次，NITF 的類目紀錄於<head>的子元素<tobject>，有其特別詳細之規定，因此在本研究 NITF 並無欄位對應於 DC 之<subject>。

#### (9) Source (來源)

在 DC 中該元素用以紀錄該資源的衍生來源，例如同一作品的不同媒體版本，或者是翻譯作品的來源等，因此本研究在該元素注錄該篇新聞之 HTML 格式資料，也就是該篇新聞之 URI。在 DAC 中則以聯合目錄小組自行規定用以紀錄新聞相關管理元素<AdminDesc>中已設置<Hyperlink>作為注錄該新聞之 URI，因遵循其典藏系統之規定，故不於<MetaDesc>的子元素<Source>中注錄相同值。

在 RSS 1.0 與 RSS 2.0 中紀錄該篇新聞之 URI 於<item>的子元素<link>，而<channel>的子元素<link>是用以紀錄該頻道之連結，故 DC 之<source>應與<item>的子元素<link>對應。NITF 並非用於網路新聞發佈，因此其之規格無提供相關的注錄元素。

#### (10) Language (語言)

語言在 DC 中表示該資源所使用的語言，因此本研究在該元素注錄該篇新聞所用之語言，可遵循 RFC1766 的規定來注錄，本研究之新聞資源皆為繁體中文，故其之固定值為「zh-tw」。而 DAC 紀錄新聞實質內容元素以 DC 為主要架構，但因其系統典藏之需求，而改固定值為「中文」，與<MetaDesc>的子元素<Language>對應。

在 RSS 2.0 中語文相關元素設置於<channel>之下層元素，用以紀錄該頻道的

語言，因 RSS 頻道所用之語言多與<item>所使用的語言相同，所以將 DC 的 <language>與<channel>的子元素<language>對應。RSS 1.0 並無語文之相關規定元素，因此可引用 DC Module 相對應於<dc:language>。該新聞使用之語言視為新聞內容的一部份，所以 NITF 之語文相關元素設置於<body>下層之<body.end>的<tagline>其可加入子元素<lang>加以對應，作為該新聞尾端補充資訊。

#### (11) Rights (版權規範)

<rights>在 DC 中可表示該資源版權聲明和使用規範，新聞資源而言，應在該元素注錄該篇新聞之版權擁有者，研究之新聞資料皆來自於聯合報系，不同於新聞報別，乃因聯合報系旗下擁有許多不同報別，各個報別為該新聞之出版者，而所有報別皆同歸於聯合報系所擁有其之版權，因此，此元素我們給予固定值為「聯合報系」。DAC 則注錄於<MetaDesc>的子元素<Rights>中。

在 RSS 2.0 中版權相關元素設置於<channel>之下，用以紀錄該頻道的版權資訊，而在本研究中發布頻道的版權內容即為所有新聞<item>之版權資訊，故該元素與<channel>的子元素<copyright>相對應。RSS 1.0 則因有 DC 為其標準 Module 之一，所以直接對應於<dc:rights>。如前所述，版權資訊屬於該篇新聞資源的版權擁有者，並非屬於新聞內容資料，故在 NITF 中對應至 head 中紀錄該新聞資料的<docdata>子元素<doc.copyright>。

#### (12) Format (資料格式)

該元素用以紀錄該資源的實體形式或者是數位特徵，此元素多以 MIME 格式表示，因此在本研究中將該元素視為注錄該篇新聞之資料格式，而新聞 Metadata 多以 XML 格式呈現，故此注錄之固定值為「text/xml」。DAC 依 DC 架構而注錄於<MetaDesc>的子元素<Format>，但依其系統之規定，將固定值改為「XML 檔」。

因 DC 與 DAC 皆為泛用型的 Metadata，各種資源注錄資料格式會有所不同，所以將有規定元素注錄其資料格式，而 RSS 2.0、RSS 1.0 與 NITF 格式，皆用於紀錄數位新聞事件，都以 text/xml 格式呈現，因此三種 Metadata 本身皆無規定用於紀錄其資料格式之元素。

#### (13) Type (資源類型)

在 DC 中該元素用以紀錄該資源的類型或所屬的抽象範疇，因此本研究在該元素注錄該篇新聞之資源類型，故此在此元素注錄固定值「text」。DAC 格式同 DC 架構，因此<MetaDesc>的子元素<Format>可與之對應，但因聯合目錄系統規定，改其固定值為「文字」。

因 DC 與 DAC 皆為泛用型的 Metadata，可通用於注錄各種類型資源，所以需將資源類型加以顯示，但 RSS 2.0、RSS 1.0 與 NITF 格式，皆為數位新聞事件 Metadata 專用格式，其資源格式都是 text 格式，所以三者本身皆無規定相關元素用以紀錄資源類型。

#### (14) Coverage (涵蓋時空)

DC 中以<coverage>作以紀錄關於該資源所涵蓋的時期和地理區域，新聞資

源而言，應在該元素注錄該篇新聞之涵蓋時空，而本研究之新聞資料皆為台灣棒球新聞事件，其時間範圍已在<date>中注錄，故該元素只紀錄涵蓋空間，設定固定值為「Taiwan, ROC」。在 DAC 新聞實質內容元素<MetaDesc>中以子元素<Coverage>對應，除此聯合目錄小組自行規定用以紀錄新聞相關管理元素<AdminDesc>中設置<Catalog>作為分類之用，以之下層子元素<Record>用於紀錄不同類別之分類，其中包含「地理架構」之項目，而依其系統之規定，改固定值為「台灣」與之對應。

RSS 1.0 與 RSS 2.0 皆以發布網路新聞為主，故對於時空範圍並無告知的需求，因此其之規格並無規定時空範圍的相關元素。涵蓋時空屬於該篇新聞的新聞資料，也可視為新聞內容的部分，故在 NITF 在<head>中紀錄該新聞資料的<docdata>子元素<doc-scope>告知該篇新聞之涵蓋範圍，並且在紀錄新聞實質內容<body>的尾端資訊<body.end>下層元素<tagline>中選擇加入子元素<location>以對應。

#### (15) Relation (關連)

在 DC 中該元素用以紀錄該資源與其他作品的關連，或所屬的系列和檔案庫，而本研究將其對應該新聞資源發佈的所屬資料庫，因此其固定值為「<http://ndap.dils.tku.edu.tw>」或「台灣棒球運動珍貴新聞檔案數位資料館」。而 DAC 除管理元素外，乃以 DC 為架構，因此該元素與<MetaDesc>的子元素<Relation>對應。

在 RSS 2.0 中<item>的子元素<source>用以表示該新聞之出處，也就是其發佈之來源資料庫，所以本元素相對應。RSS 1.0 則規定 DC 為其標準 Module 之一，故與 DC 本身相同對應至<dc: relation>。新聞所屬之發佈資料庫，應為關於該篇新聞之相關資訊，而非新聞內容資訊，故在 NITF 中對應至<head>中紀錄該新聞資料的<docdata>子元素<doc.right>。

表 2 四種 Metadata 與資料庫欄位對應分析 ( ▶ : 表下一層之元素 ; ⇨ 表該元素之屬性值 )

| 外部Metadata |   |                                      |   |   | 內部Metadata   | 說明         |
|------------|---|--------------------------------------|---|---|--------------|------------|
| DC         | DAC   | RSS 1.0                              | RSS 2.0                                 | NITF  | 資料庫欄位        |            |
| identifier | DACatalog ▶ AdminDesc<br>▶ DigiArchiveID                | rdf:RDF ▶<br>item ▶<br>dc:identifier | rss ▶<br>channel ▶<br>item ▶<br>guid    | nitf ▶ head ▶ docdata ▶<br>doc-id⇨id-string         | autoid (主索引) | 識別碼 (自動編號) |
|            | DACatalog ▶<br>MetaDesc ▶ Identifier                    |                                      |   |   |              |            |
| date       | DACatalog ▶<br>MetaDesc ▶ Date                          | rdf:RDF ▶<br>item ▶<br>dc:date       | rss ▶<br>channel ▶<br>item ▶<br>pubDate | nitf ▶ head ▶<br>docdata ▶ date.issue⇨norm          | date         | 新聞日期       |
|            | DACatalog ▶<br>AdminDesc ▶ Catalog ▶<br>Record / 時間架構 : |                                      |   | nitf ▶ head ▶ pubdata⇨<br>date.publication          |              |            |
|            |   |                                      |   | nitf ▶ body ▶<br>body.head ▶ headline ▶<br>dateline |              |            |
| title      | DACatalog ▶ MetaDesc<br>▶ Title                         | rdf:RDF ▶<br>item ▶ title            | rss ▶<br>channel ▶<br>item ▶ title      | nitf ▶ head ▶ title                                 | hotline      | 主標題        |
|            |   |                                      |   | nitf ▶ body ▶<br>body.head ▶ headline ▶ hl1         |              |            |
|            |   |                                      |   | nitf ▶ body ▶<br>body.head ▶ headline ▶ hl2         | subhotline   | 副標題        |
| creator    | DACatalog ▶<br>MetaDesc ▶ Creator                       | rdf:RDF ▶<br>item ▶<br>dc:creator    | rss ▶<br>channel ▶<br>item ▶<br>author  | nitf ▶ body ▶<br>body.head ▶ headline ▶<br>byline   | author       | 記者         |

| 外部Metadata  |                                       |                                       |   |   | 內部Metadata | 說明                      |
|-------------|---------------------------------------|---------------------------------------|---|---|------------|-------------------------|
| DC          | DAC                                   | RSS 1.0                               | RSS 2.0                                     | NITF  | 資料庫欄位      |                         |
| publisher   | DACatalog ▶<br>MetaDesc ▶ Publisher   | rdf:RDF ▶<br>item ▶<br>dc:publisher   | rss ▶<br>channel ▶<br>item ▶ title          | nitf ▶ body ▶<br>body.head ▶ headline ▶<br>rights | paperid    | 報別                      |
| contributor | DACatalog ▶<br>MetaDesc ▶ Contributor | rdf:RDF ▶<br>item ▶<br>dc:contributor | —   | nitf ▶ body ▶<br>body.head ▶ distributor          | —          | —                       |
| description | DACatalog ▶<br>MetaDesc ▶ Description | rdf:RDF ▶<br>item ▶<br>description    | rss ▶<br>channel ▶<br>item ▶<br>description | nitf ▶ body ▶<br>body.content                     | document   | 新聞內容                    |
| subject     | DACatalog ▶<br>MetaDesc ▶ Subject     | rdf:RDF ▶<br>item ▶<br>dc:subject     | rss ▶<br>channel ▶<br>item ▶<br>category    | —   | pagename   | 版名<br>(若無版名則置入固定值：體育新聞) |
| source      | DACatalog ▶<br>AdminDesc ▶ Hyperlink  | rdf:RDF ▶<br>item ▶ link              | rss ▶<br>channel ▶<br>item ▶ link           | —   | —          | 來源<br>(該篇新聞之URI)        |
| language    | DACatalog ▶<br>MetaDesc ▶ Language    | rdf:RDF ▶<br>channel ▶<br>dc:language | rss ▶<br>channel ▶<br>language              | nitf ▶ body ▶<br>body.end ▶ tagline ▶ lang        | —          | 語文<br>(固定值：zh-tw or 中文) |

| 外部Metadata |   |                                     |  |  | 內部Metadata | 說明   |
|------------|---|-------------------------------------|--|--|------------|--|
| DC         | DAC   | RSS 1.0                             | RSS 2.0                                | NITF   | 資料庫欄位      |  |
| rights     | DACatalog ▶<br>MetaDesc ▶ Rights                        | rdf:RDF ▶<br>channel ▶<br>dc:rights | rss ▶<br>channel ▶<br>copyright        | nitf ▶ head ▶ docdata ▶ do<br>c.copyrights⇒ holder | —          | 版權 (固定值：聯合報系)  |
| format     | DACatalog ▶ MetaDesc<br>▶ Format                        | —                                   | —                                      | —  | —          | 資料格式<br>(固定值：text/xml or XML檔)   |
| type       | DACatalog ▶ MetaDesc<br>▶ Type                          | —                                   | —                                      | —  | —          | 資料類型<br>(固定值：text or 文字)   |
| coverage   | DACatalog ▶ MetaDesc<br>▶ Coverage                      | —                                   | —                                      | nitf ▶ head ▶ docdata ▶<br>doc-scope⇒scope         | —          | 涵蓋時空<br>(固定值：Taiwan, ROC or 台灣)  |
|            | DACatalog ▶ AdminDesc<br>▶ Catalog ▶ Record / 空<br>間架構： |                                     |  | nitf ▶ body ▶<br>body.end ▶ tagline ▶<br>location  |            |  |
| relation   | DACatalog ▶ MetaDesc<br>▶ Relation                      | rdf:RDF ▶<br>item ▶<br>dc:relation  | rss ▶<br>channel ▶<br>item ▶<br>source | nitf ▶ head ▶ docdata ▶<br>doc.rights⇒owner        | —          | 關聯/來源<br>(固定值：<br><a href="http://ndap.dils.tku.edu.tw">http://ndap.dils.tku.edu.tw</a> or<br>台灣棒球運動珍貴新聞檔案<br>數位資料館) |

(資料來源：本研究整理)

## 肆、系統設計與實作

依表 2 的元素對應分析，接下來將進行棒球新聞管理系統規劃與實作，本系統包含屬於前端的檢索子系統、RSS 訂閱子系統以及後端管理子系統三部份，而檢索子系統中包括顯示與 Metadata 轉換模組；後端管理子系統則可分為聯合目錄大批匯出、棒球大事記及棒球名人錄管理三模組（如圖 8）。

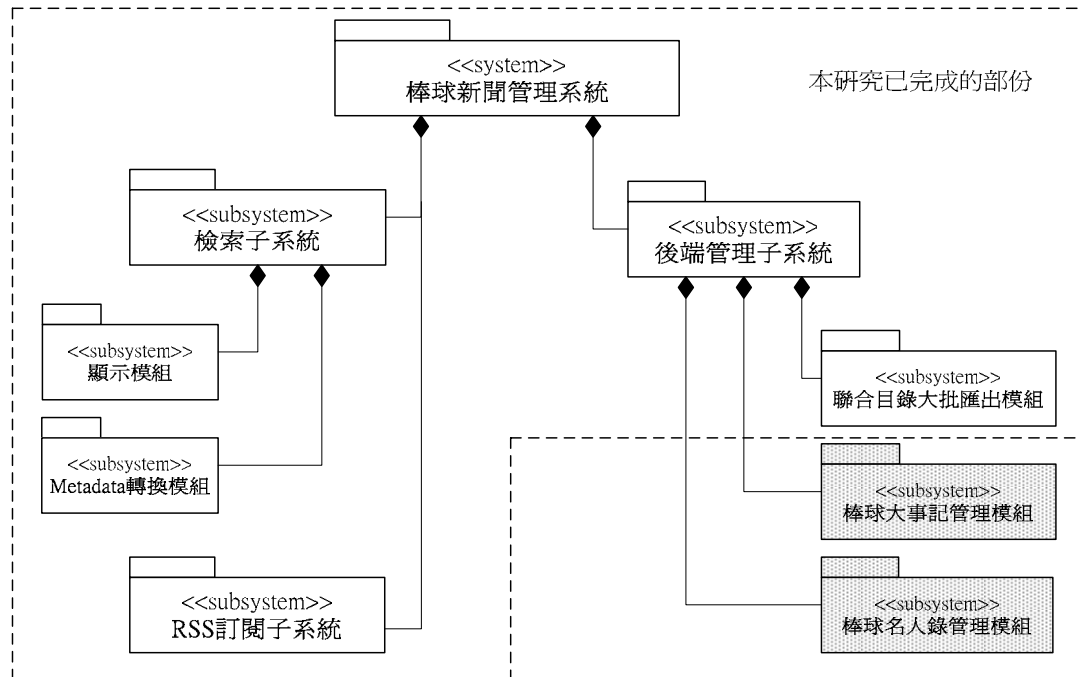


圖 8 棒球新聞管理系統

本系統乃利用 JSP 與 ASP 程式撰寫完成，圖 9 為檢索子系統與 RSS 訂閱子系統之活動圖，使用者可經由 Web 介面選取欲進行之動作，主要可分為：全文檢索、圖片檢索、訂閱歷史上的今天，以及結束作業。

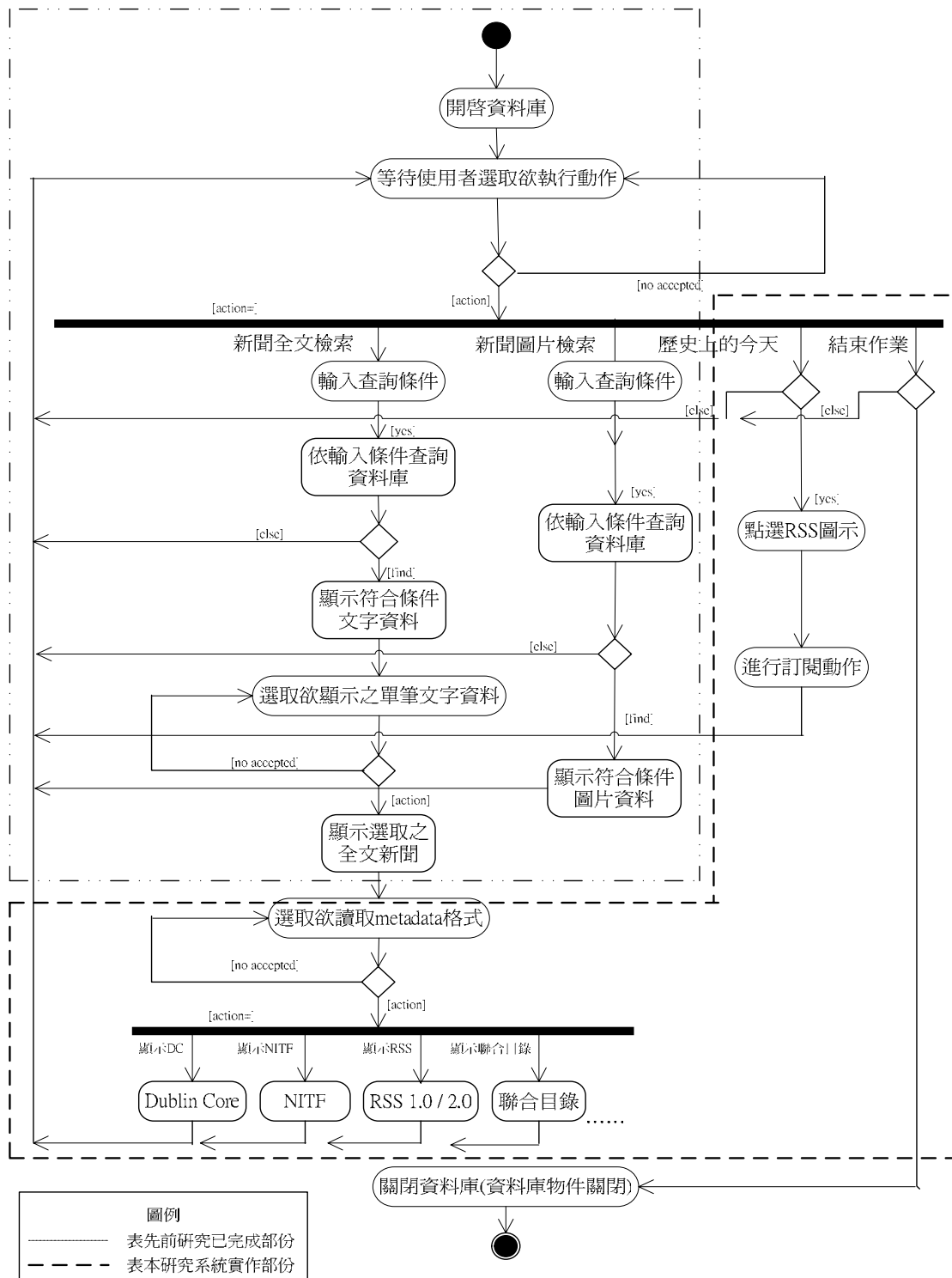


圖 9 檢索子系統與 RSS 訂閱子系統之活動圖

## 一、檢索子系統

此系統建置目的在於提供一般使用者檢索新聞資源，目前本系統提供文字新聞及圖片新聞兩種資源檢索利用，未來將提供棒球名人錄以及棒球大事記之檢索利用。如圖 9 活動圖所示，當使用者選擇全文檢索時，可在查詢介面輸入查詢條件，JSP 程式將解讀使用者輸入之條件組合成 SQL 查詢語法對後端資料庫進

行查詢動作，顯示符合條件的多筆文字資料，使用者在進行挑選所需資料全文顯示，再進一步選擇欲顯示之 Metadata 格式，ASP 程式將該篇新聞資訊以被選取之 Metadata 格式注錄顯示與 Web 介面上。故該系統之下又分為顯示模組，與 Metadata 轉換模組。

(1) 顯示模組：本模組是將使用者檢索結果加以顯示給予使用者瀏覽、閱讀，此部份之系統在之前的研究就已建置完成<sup>22</sup>，目前以可供使用者顯示所需之新聞資源。

(2) Metadata 轉換模組：本模組以表 2 所對應之欄位進行 Metadata 轉換，其下包含四種轉換子模組：DC 轉換子模組、DAC 轉換子模組、RSS 1.0/2.0 轉換子模組，以及 NITF 轉換子模組，詳細紀錄新聞資訊，讓使用者可利用此模組選擇欲顯示之 Metadata 格式，透過後端程式運作進行各子模組之轉換，滿足各種不同領域之使用者。各模組之轉換結果可參閱圖 10~圖 14。

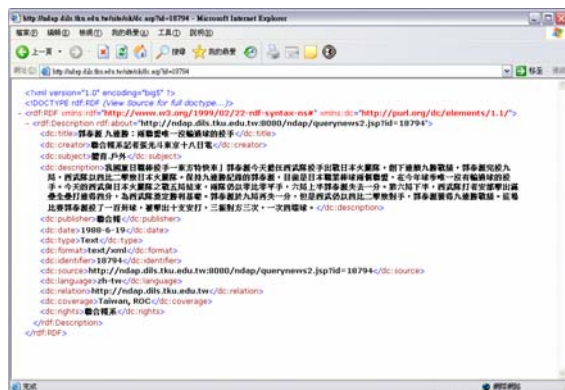


圖 10 RDF+DC 格式轉換結果

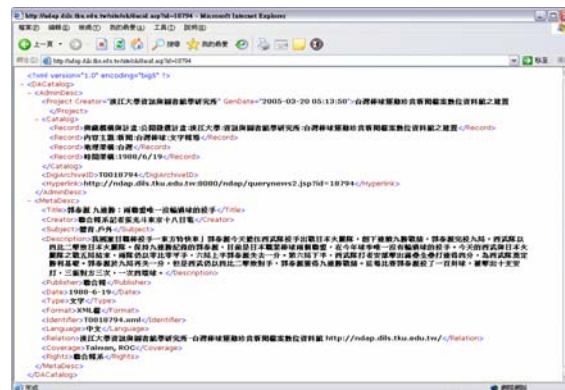


圖 11 DAC 格式轉換結果



圖 12 RSS 1.0 格式轉換結果

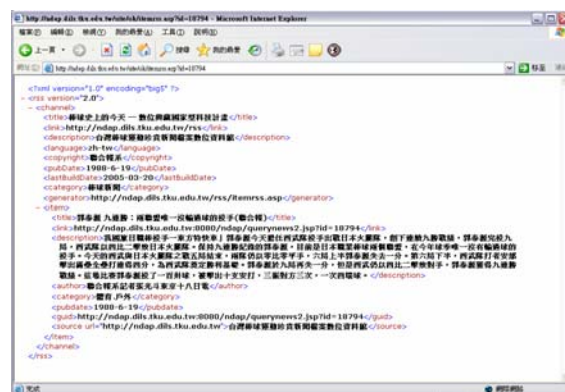


圖 13 RSS 2.0 格式轉換結果

<sup>22</sup> 其相關細節可參考：孫正宜，"新聞專卷的數位化與加值應用—以台灣棒球報紙新聞數位典藏為例"，淡江大學資訊與圖書館學系研究所碩士論文，民 93。

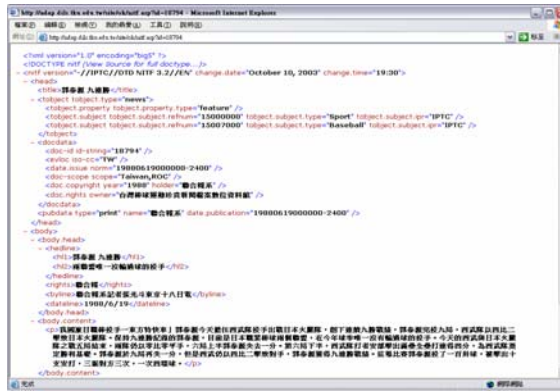


圖 14 NITF 格式轉換結果

## 二、RSS 訂閱子系統

本系統乃 RSS 格式之衍生應用而產生，RSS 規格除了可作為單篇新聞發佈的 Metadata 格式，也可包含多篇新聞資源於單一 RSS 文件，將「台灣棒球運動珍貴新聞檔案數位資料館」作為發佈頻道 (channel)，每日更新，提供棒球歷史上今天所發生過的新聞事件，每個 item 是一篇新聞，item 的新聞內容描述 (description) 只截取前 150 個字，使用者透過該系統提供之 RSS Feed 進行訂閱，利用 RSS Reader 即時獲取摘要型新聞資訊，使用者可從其摘要資訊選取欲進一步閱讀之新聞，以 RSS 提供之 URI 加以連結至新聞本文。

如圖 9 活動圖所示，若使用者選擇訂閱台灣棒球歷史上的今天，則需點選該 RSS 圖示，而在 Web 上將顯示 xml 格式之歷史上今天的棒球新聞，使用者則需將其 xml 頁面的 URL，也就是所謂的 RSS Feed，在 RSS Reader 軟體中進行加入該 RSS Feed 的完整 URL，即完成訂閱動作，此後使用者只要開啓 RSS Reader 就可獲得歷史上當日的棒球新聞事件。圖 15 為訂閱棒球史上的今天之畫面呈現，本研究採取 BlogLine 作為訂閱測試，BlogLine 乃經由瀏覽器閱讀，同一般 RSS Reader 之功用，但不需安裝軟體即可訂閱 RSS Feed。



圖 15 RSS Feed 訂閱結果

### 三、後端管理子系統

本系統是為管理者與維護者所建置，用以管理新聞資源之修改、更新，以提供前端使用者最佳之新聞資源，及台灣棒球相關之名人錄、大事記。該系統之下目前規劃有三個模組：聯合目錄大批匯出模組、棒球大事記管理模組、棒球名人錄管理模組。

(1)聯合目錄大批匯出模組：此模組之建置乃因數位典藏聯合目錄系統所提供將典藏資料匯入該系統的方式有兩種，一為匯出聯合目錄規定之 Metadata 格式，也就是前述之 DAC 格式；另外則建立遵循 OAI-PMH 協定，因本系統目前採用匯出 DAC 格式之 XML 檔予聯合目錄加以匯入該典藏系統，故本系統中設置本模組作為大批匯出新聞資料之用。而其匯出資料以年為單位區分資料夾，提供聯合目錄匯入典藏。

(2)棒球大事記管理模組：本模組之設計起因於本系統目前之棒球大事記乃為重要棒球新聞之整合而已，並非依照年代對台灣棒球事件加以整理、條目化，因此未來本系統將加以整理台灣棒球事件，以供使用者查詢利用，為此而設立本模組，作為管理者整理台灣棒球事件，管理者可經由此模組對棒球大事記相關類目進行增加、刪除、更新、查詢等動作，以提供使用者最佳棒球資訊。本模組目前尚為完成，屬設計規劃階段。

(3)棒球名人錄管理模組：本模組之設計乃因本系統目前之棒球名人錄，屬於該名人之相關棒球新聞之整合而已，並未對該人物進行加值，因此其之利用價值有限，因此未來本系統將加以整理台灣棒球名人之相關資料、經歷，建置人名權威檔，將棒球新聞人物加值、延伸，以供使用者查詢利用，為此而設立本模組，作為管理者整理台灣棒球名人資料，管理者可經由此模組對棒球名人錄相關類目進行增加、刪除、更新、查詢等動作，以提供使用者最佳棒球資訊。本模組目前尚為完成，屬設計規劃階段。

目前後端管理系統只有聯合目錄大批匯出作業模組已完成，其利用 ASP 程式，以年代為標準，將資料庫中的新聞資料，轉出成為 XML 檔，個別存放於不同資料夾（如圖 16），再經由聯合目錄工作小組將 XML 檔匯入數位典藏聯合目錄系統，於數位典藏網站上供使用者查詢檢索（如圖 17）。



圖 16 聯合目錄大批匯出作業



圖 17 數位典藏聯合目錄查詢結果

## 伍、結論

本研究目前將四種目前較為盛行之 Metadata 格式加以比較分析，綜合應用其之格式加以轉換，以利於新聞事件的發佈傳送，方便新聞資源的管理及傳播利用。而未來本系統可再加入格式的 Metadata，以滿足其他使用者獲取所需新聞資訊，並將 DC 搭配 OAI-PMH 機制提供新聞資訊之開放式檢索，將分散的資料整合查詢，強化電子化新聞文件典藏的取用，以實現新聞資源加值管理、傳播利用，並增進新聞傳播獲取率。

## 致謝

本文為 NSC 93-2422-H-032-001 研究計畫部分成果，感謝國家科學委員會經費補助，聯合報授權使用所需新聞資料，使本研究得以順利進行，特此致謝。

## 參考書目

1. Berkman Center, "RSS 2.0 Specification", available from <<http://blogs.law.harvard.edu/tech/rss>> (2004/10/29).
2. Dave Winer, "RSS 0.91", available from <<http://backend.userland.com/rss091>> (2005/03/22).
3. Dave Winer, "RSS 0.92", available from <<http://backend.userland.com/rss092>> (2005/03/22).
4. Dublin Core Metadata Initiative, "Recommendations," available from <<http://purl.org/DC/documents/recommendations.htm>> (2004/12/17).
5. Hammersley, Ben, "Content Syndication with RSS", O'Reilly, 2003.
6. IDEAlliance PRISM Working Group, "PRISM: Publishing Requirements for Industry Standard Metadata", available from <<http://www.prismstandard.org/>> (2005/3/22).
7. IPTC, "International Press Telecommunications Council", available from <<http://www.iptc.org/pages/index.php>> (2004/10/24)
8. IPTC, "News Industry Text Format", available from <<http://www.nitf.org/>> (2005/3/20).
9. IPTC, "News Markup Language", available from <<http://www.newsml.org/>> (2005/3/20).
10. IPTC, "Program Guide Markup Language", available from <<http://www.programguideml.org/>> (2005/3/20).
11. IPTC, "Sports Markup Language", available from <<http://www.sportsml.org/>> (2005/3/20).
12. RSS-DEV Working Group, "RDF Site Summary (RSS) 1.0", available from <<http://web.resource.org/rss/1.0/spec>> (2005/1/28).
13. W3C, "RDF Primer - W3C Recommendation 10 February 2004", 2004/2/10, available from <<http://www.w3.org/TR/rdf-primer/>> (2005/1/6)
14. XMLNews.Org, "NewsML", available from <<http://xmlnews.org/NewsML/>> (2004/12/20).
15. 中國報業協會規範工作組, "中國報業電子新聞文稿格式", 2000年5月。
16. 行政院文化建設委員會國家文化資料庫知識管理系統, "News Records Metadata Format", 可得自 <<http://km.cca.gov.tw/download/rule.html>> (2005/3/24)。
17. 吳政勸, "都柏林核心集的發展現況與其在圖書館的應用", 網際網路與圖書館發展研討會(台北市:中國圖書館學會, 1999年12月4日)。
18. 孫正宜, "新聞專卷的數位化與增值應用—以台灣棒球報紙新聞數位典藏為例", 淡江大學資訊與圖書館學系研究所碩士論文, 民93。
19. 國科會數位典藏國家型科技計畫, "數位典藏聯合目錄", 可得自

<<http://catalog.ndap.org.tw>> (2005/1/6)。

20. 新華社技術局標準工組，"XinhuaMLv1.0 功能說明書"，2003 年 1 月 18 日。

21. 嚴坤宗，"建構以 RSS 為基礎的新聞內容聯合機制-以《大學報》為例"，臺北大學資訊管理研究所碩士論文，民 93。