

## 主題式報紙新聞數位典藏之研究—以台灣棒球運動為例

林信成

淡江大學資訊與圖書館學研究所 副教授

sclin@mail.tku.edu.tw

### 摘要

傳統上，報紙數位典藏系統大多採取將所有新聞內容納入資料庫的作法，其優點是涵蓋面較廣，可以收錄各式各樣的主題；缺點則是資料庫龐大，深度加值不易。因此，本研究提出以「主題式」方式進行新聞的數位典藏與加值利用，並以聯合報的台灣棒球運動歷史性新聞為例，先進行系統分析與建置，再探討內容語意描述及 Metadata 轉換，使其成為更具利用價值的數位典藏庫。

關鍵字：數位典藏、詮釋資料、新聞標示語言、新聞文檔格式、資源描述框架

## Thematic Digital Newspaper Archive: A Case Study of Taiwan Baseball

Sinn-Cheng Lin

Associate Professor

Department of Information and Library Science,

Tamkang University, Taipei, Taiwan, R.O.C.

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

### Abstract

Traditionally, most of the digital newspaper archive systems used all of the report contents to create a complete database. One of the advantages is the collection can cover a broad range of subjects; however, the database could be too huge to be made value-add deeply. Therefore, this paper proposes a thematic approach for constructing a digital newspaper archive, and uses the historical reports of Taiwan baseball, which is provided by United Daily News, as case study. First, we work on the system analysis and establishment, then we study on the semantic description and the metadata conversion, such that the resulting digital archive system can be more useful.

Keyword: Digital Archives, Metadata, NewsML, NITF, RDF

## 壹、前言

報紙反映人們生活的縮影，記錄社會變遷的軌跡，也承載人類文化的資產，更可做為歷史的佐證與研究的參考。今日的新聞乃明日的歷史，歷史性的報紙新聞藉由資訊科技進行數位典藏後，可創造許多應用價值。為了彰顯新聞數位典藏之重要性，在國科會主導的數位典藏國家型科技計畫<sup>1</sup>中，特別成立了新聞主題小組<sup>2</sup>，旨在：「以電子通訊出版、報紙、期刊、新聞影音為主要數位化典藏內容 ... 將新聞媒體裡的新聞事件 5W1H加以分析，並編制新聞標題索引目錄，建立新聞資料的加值利用 ... 製作一個內容充實的數位典藏資料庫 ...」<sup>3</sup>。歷年來與新聞主題相關的計畫有：「北平世界日報內容數位化開發計畫」<sup>4</sup>、「國家圖書館期刊報紙典藏數位化計畫」<sup>5</sup>、「電視新聞多媒體資料館」<sup>6</sup>、「後設資料動態管理系統研發計畫」等。新聞主題小組亦曾召開「中文新聞內容數位化研討會」，廣泛討論各界研究成果<sup>7</sup>。業界近年來也有許多報社與資訊公司合作，將其報紙全文影像數位化或資料庫化以利檢索利用<sup>8</sup>。

從現有成果觀之，在報紙數位典藏方面大都針對報紙的所有內容，建立全面性之資料庫，再以全頁影像、標題索引、欄位檢索或全文檢索方式提供使用者利用。此種作法的優點是涵蓋面較廣，可以收錄政治、經濟、體育、娛樂、社會 ... 等多樣化的主題；缺點則是新聞資料庫龐大，深度加值不易，極難達成「將...新聞事件 5W1H加以分析...建立新聞資料的加值利用」<sup>9</sup>之目標。因此，我們乃提出「台灣棒球運動珍貴新聞檔案數位資料館之建置」<sup>10</sup>計畫，研擬以專卷分類方式提供「主題式」新聞的數位典藏與加值利用，並進行內容語意描述之構想。此計畫與聯合報合作，取得自光復以來與台灣棒球運動相關之新聞專卷檔案，進行內容數位化之工作，並加註適當的Metadata，使其成為極具參考價值之數位典藏庫，以協助棒球界人士、棒球運動研究者以及所有對棒球有興趣的社會大眾，能更深入、更有效率的探索台灣棒球運動與社會發展的脈動。本文乃是此研究計畫執行過程中的階段性成果，包括棒球新聞數位典藏系統之建置與棒球新聞Metadata之分析與探討。

<sup>1</sup> 國科會，"數位典藏國家型科技計畫"，可得自<<http://www.ndap.org.tw/>>（上網日期：2005/1/25）。

<sup>2</sup> 國科會，"數位典藏國家型科技計畫：內容發展分項計畫--新聞主題小組"，可得自<[http://content.ndap.org.tw/main/vision\\_brief.php?class\\_vision=16](http://content.ndap.org.tw/main/vision_brief.php?class_vision=16)>（上網日期：2004/1/25）。

<sup>3</sup> 同註 2。

<sup>4</sup> 世新大學資訊傳播學系，"北平世界日報內容數位化開發計畫"，可得自<<http://icd.shu.edu.tw/lipo/>>（上網日期：2004/11/4）。

<sup>5</sup> 同註 12。

<sup>6</sup> 國科會數位典藏國家型科技計畫，"「電視新聞多媒體資料館」簡介"，電子通訊第四期，2002年06月，可得自<[http://www.ndap.org.tw/1\\_newsletter/content.php?uid=239](http://www.ndap.org.tw/1_newsletter/content.php?uid=239)>（上網日期：2004/11/29）。

<sup>7</sup> 國科會數位典藏國家型科技計畫新聞主題小組，"數位典藏國家型科技計畫內容發展分項計畫"，中文新聞內容數位化研討會論文集，台北：中央研究院，民92年3月7日。

<sup>8</sup> 孫正宜、林信成，"中文報業數位化技術與現況探討－聯合知識庫數位化經驗"，2003年資訊科技與圖書館學術研討會，台北·淡江大學，頁73-93，2003/5/29。

<sup>9</sup> 同註 2。

<sup>10</sup> 淡江大學資圖所數位典藏計畫小組，"台灣棒球運動珍貴新聞檔案數位資料館之建置"，可得自<<http://ndap.dils.tku.edu.tw/>>（上網日期：2004/11/29）。

## 貳、相關研究或計畫

### 一、棒球運動方面

本研究選擇以棒球新聞作為主題，主要是棒球運動已被公認為我國的「國球」，不但是台灣文化不可分割的一部份，更是老、中、青不同時代的共同記憶！2004 年中華棒球代表隊進軍雅典奧運雖未獲得獎牌，但在舉國關注的氛圍中掀起一波波棒球熱潮，再度顯示棒球運動在台灣紮根之深、影響之遠。

台灣棒運的發展已有百年歷史，從日據時期的殖民色彩，到戰後時期的蓬勃發展，再到外交困頓時期的三冠王狂熱，最後走上國際舞台及職業棒球的多元化階段，就如同一世紀以來台灣史的縮影！為此，近年來已陸續有研究者開始投入研究台灣棒球史的行列，試圖為台灣棒球的輝煌歷史留下記錄。不過，大多數的資料都散在一些坊間書刊、報章雜誌或網站當中，並未有系統的進行數位典藏；民間建置的許多棒球相關網站，也往往偏向職棒比賽之動態報導<sup>11</sup>。國家圖書館的全國報紙資訊網<sup>12</sup>，雖提供完整的報紙新聞全頁影像與標題檢索，但並未另外進行專門議題的加值處理。官方網站如中華民國棒球協會<sup>13</sup>、行政院體育委員會<sup>14</sup>等，曾致力於部分棒球史料和文物的數位化；國內著名入口網站「蕃薯藤」，則自 2003 年起陸續由曾文誠先生執筆撰寫一系列有關台灣棒球史的相關文章，內容相當豐富，目前仍持續進行中<sup>15</sup>。在書籍方面，由於台灣棒球發展受日、美影響深遠，因此，市面上甚早即有專門探討日本棒球<sup>16</sup>和美國棒球的專書<sup>17</sup>；而討論台灣棒球史的書，早期有《東昇的旭日—中華棒球發展史》<sup>18</sup>和《嘉義棒球史話》<sup>19</sup>等；近期有謝仕淵與謝佳芬合著之《台灣棒球一百年》<sup>20</sup>，該書中許多寶貴資料也都是作者辛辛苦苦從舊報紙中翻找取得；最近的著作則有徐宗懋的《三冠王之夢》<sup>21</sup>；曾文誠與孟峻瑋合著的《台灣棒球王》<sup>22</sup>。此外，還有一些是關於特定人、事、物的專書<sup>23</sup>。從近期棒球相關出版品的趨勢

<sup>11</sup> 中華職棒聯盟，"中華職棒全球資訊網"，可得自<<http://www.cpbl.com.tw/>>（上網日期：2004/11/2）。

<sup>12</sup> 林淑芬，"期刊文獻資訊網新服務—「全國報紙資訊網」及「國家圖書館期刊影像資料庫」上線服務"，可得自<[http://www.ncl.edu.tw/pub/c\\_news/92/05.html](http://www.ncl.edu.tw/pub/c_news/92/05.html)>（上網日期：2004/11/4）。

<sup>13</sup> 中華民國棒球協會，"中華民國棒球協會官方網站"，可得自<<http://www.ctba.org.tw/index.asp>>（上網日期：2004/11/5）。

<sup>14</sup> 行政院體育委員會，"體育數位博物館 > 棒球 > 棒球文物"，可得自<[http://media.justsports.net.tw/spo\\_demo/digital\\_baseball.asp](http://media.justsports.net.tw/spo_demo/digital_baseball.asp)>（上網日期：2004/11/2）。

<sup>15</sup> 蕃薯藤，"蕃薯藤->運動->棒球->文史館"，可得自<<http://sports.yam.com/list.php/baseball/history>>（上網日期：2004/11/1）。

<sup>16</sup> 高正源，《日本棒球發展史》，聯經，1994/08/15。

<sup>17</sup> 陳筱玉，《美國棒球發展史》，聯經，1994/08/15。

<sup>18</sup> 高正源，《東昇的旭日—中華棒球發展史》，民生報，1994/6/1。

<sup>19</sup> 蘇錦章，《嘉義棒球史話》，聯經，1996 年 7 月。

<sup>20</sup> 謝仕淵、謝佳芬，《台灣棒球一百年》，果實，2003 年 8 月。

<sup>21</sup> 徐宗懋，《三冠王之夢》，大地，2004/6/21。

<sup>22</sup> 曾文誠、孟峻瑋，《台灣棒球王》，創智，2004/6/21。

<sup>23</sup> 超級圖書館，"棒球書-中華職棒及評論"，可得自<<http://www.superstadium.net/first/book/base/1.htm>>（上網日期：2004/11/8）。

來看，陸續有許多人開始努力紀錄下台灣棒球運動的發展史，的確是可喜的現象。在棒球博物館的建置上，民國八十五年動工興建的澄清湖棒球博物館<sup>24</sup>，目前由那魯灣公司行銷企劃部負責經營，館藏中西棒球文物約五千件，是國內較有系統珍藏棒球文物的機構，但這批文物也尚未進行數位典藏；而爲了凸顯棒球對台灣的重要，行政院體委會正在籌設一座國家級規模、展覽場地可達 3000 坪以上的「台灣棒球博物館」<sup>25</sup>，目前尚在規劃中。

## 二、新聞專卷方面

所謂「新聞專卷」是針對不同的新聞主題，把相關的資料加以收集、組織、整理，再分門別類建檔的歷史性新聞文獻。最初只是報社爲了提供採編同仁在新聞事件發生時，能有快速、完整且可靠的資料來源可供參考，於是由資料整理人員每日剪報、分類、建檔，供編輯、記者使用，不必再另花時間搜尋或查證。未料長期下來，竟建立了一個又一個完整且可信度高的資料檔，累積成一座資料豐富的專門圖書館（如圖 1 所示），記錄了整個社會發展的軌跡。



圖 1 聯合報新聞專卷（資料來源：聯合報提供）

本研究選擇與聯合報合作，是因為聯合報資料中心內部珍藏了近三萬冊「新聞專卷」，其中有一部份便是棒球新聞史料。聯合報這些新聞專卷乃是從民國 45 年開始收集，詳述近五十年來台灣人民生活的點滴與重大新聞事件。每個新聞專卷都是專業人員經過一定嚴謹的蒐集、整理與加值過程，才成爲一個完整事件的報導，非常具有參考價值，尤其對想了解台灣發展史的使用者而言是不可或缺的一個新聞資料庫。不同於每日出刊的報紙全文，每個新聞專卷都是分別針對單一主題長期追蹤與整理所得，比原始報紙全文更具參考與利用價值。在目前資訊不患寡而患不精的時代，若能有系統的將新聞專卷加以數位化，建構成主題式新聞資料數位典藏庫，

<sup>24</sup> 人間福報電子報，「澄清湖棒球博物館 中西文物豐富」，2003.06.25，可得自 <http://www.wfdn.com.tw/9206/030625/01-08/062508-5.htm>（上網日期：2004/11/1）。

<sup>25</sup> 陳志祥，「棒球博物館 有影」，《中時電子報》，2003 年 11 月 12 日，可得自 <http://tw.news.yahoo.com/2003/11/12/sports/ctnews/4363057.html>（上網日期：2004/11/1）。

必然更適合於對某個專門議題感興趣的人士使用。於是，長期累積的新聞專卷檔案，正好可以提供一個迅速且完整探索台灣棒球發展過程的珍貴史料館。

### 三、新聞 Metadata 方面

國際上為了解決數位化新聞的資訊組織問題，已發展出數種新聞專用的Metadata格式，例如，IPTC制訂的NITF (News Industry Text Format)<sup>26</sup>、NewsML (News Markup Language)<sup>27</sup>、SportsML (Sports Markup Language)<sup>28</sup>、ProgramGuideML (Program Guide Markup Language)<sup>29</sup>、EventsML (Events Markup Language)<sup>30</sup>；IDEAlliance發佈的PRISM (Publishing Requirements for Industry Standard Metadata)<sup>31</sup>，XMLNews.Org所研擬的XMLNews<sup>32</sup>；和目前普遍應用於入口引擎、新聞網站、Blog和WiKi的RSS (Really Simple Syndication)<sup>33</sup>等。這些標準皆本於可擴展標示語言XML，資訊管理者可依需求定義新聞文件的DTD或XML Schema，既規範文件結構又保留擴充彈性；經過XML標示的新聞文件，也可利用程式轉換到其它格式或平台，有助於新聞文件的儲存、檢索以及重複使用。<sup>34</sup>海峽兩岸近年有感於數位化新聞以標示語言進行資訊組織已是舉世公認的共同作法，也開始著手此方面之研究。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>35</sup>，新聞業界則有聯合報系的「聯合新聞標示語言」(UDN Markup Language，簡稱UdnML)；大陸方面則有新華社的「新華標示語言」(Xinhua Markup Language，簡稱XinhuaML)<sup>36</sup>，中國報業協會的「中國報業電子新聞文稿格式」(Chinese News Text Format，簡稱CNTF)<sup>37</sup>等。

## 參、類目分析與系統建置

### 一、類目分析與規劃

本研究界定之資料範圍是聯合報長期篩選、整理，與棒球運動相關之報紙新聞

<sup>26</sup> IPTC, "News Industry Text Format", available at <<http://www.nitf.org/>> (2004/7/18).

<sup>27</sup> IPTC, "News Markup Language", available at <<http://www.newsml.org/>> (2004/7/18).

<sup>28</sup> IPTC, "Sports Markup Language", available at <<http://www.sportsml.org/>> (2004/10/8).

<sup>29</sup> IPTC, "Program Guide Markup Language", available at <<http://www.programguideml.org/>> (2004/10/8).

<sup>30</sup> IPTC, "Events Markup Language", available at <<http://www.iptc.org/EventsML/>> (2004/7/18).

<sup>31</sup> IDEAlliance PRISM Working Group, "PRISM: Publishing Requirements for Industry Standard Metadata", available at <[http://www.prismstandard.org/PAM\\_1.0/](http://www.prismstandard.org/PAM_1.0/)> (2004/10/9).

<sup>32</sup> XMLNews.org, "XMLNews Specifications", available at <<http://www.xmlnews.org/XMLNews/>> (2004/10/4).

<sup>33</sup> Berkman Center, "RSS 2.0 Specification", available at <<http://blogs.law.harvard.edu/tech/rss>> (2004/10/9).

<sup>34</sup> Hall, Richard, "Why XML is Important for Printing and e-publishing?" Online Technology, available at <[http://www.newsandtech.com/issues/2000/05-00/ot/05-00\\_hall.htm](http://www.newsandtech.com/issues/2000/05-00/ot/05-00_hall.htm)> (2003/9/5).

<sup>35</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", 可得自 <<http://km.cca.gov.tw/download/rule.html>> (上網日期: 2004/7/17)。

<sup>36</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>37</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。

專卷及圖片，詳細的將五十年來台灣棒運的發展，依不同主題、事件、人物 ... 等歸納成「棒球人事」、「棒球比賽」、「棒球場地」、「職業棒球」、「職棒賭博」... 等專卷，是研究台灣棒運發展的珍貴文獻。由於資料量龐大，必須詳加分類方能有效利用。因此我們乃著手規劃一個「棒球專卷類目表」，最終分為 14 大類，再細分為 174 小類，組成樹狀資訊結構。此份類目表在本研究進行過程中，扮演了數位典藏系統建構與網站規劃設計的重要依據。

## 二、系統設計與建置

為了儘快建立一個可供大眾使用之主題式「台灣棒球運動珍貴新聞檔案數位資料館」，我們在類目表規劃的同時亦同步進行典藏系統的分析與設計、資料庫的建置，並開發親和之使用介面與便捷之檢索系統。從使用者層面而言，目前已完成的第一期系統，約有資料 2 萬多筆，檢索功能則有傳統的「全文/欄位檢索功能」和針對專卷特性設計的「加值檢索功能」；而加值檢索則又分為「專卷類目表」、「棒球大事記」和「棒球名人錄」等。系統首頁如下圖所示。



圖 2 台灣棒球運動珍貴新聞檔案數位資料館首頁 (<http://ndap.dils.tku.edu.tw>)

(1) 全文/欄位檢索：此功能採用傳統式的字詞比對方式，以供使用者進行廣泛搜尋，可得到較高的回現率，此乃全文檢索以字詞比對為導向的特性使然。例如，以查找郭泰源的新聞為例，只要文中有郭泰源三字者皆會被檢索出來，以目前已入庫的資料而言，共有 383 筆。

(2) 加值檢索-專卷類目表：依據預先規劃的類目表，先完成如下圖左方之樹狀組織瀏覽介面，再設計一分類程式將數位化後的新聞一篇篇歸檔到適當的專卷類目表中，最後由管理者透過管理介面檢覈，完成篩選、整理工作。使用者可以很方便的點選感興趣的主題，以調閱相關之棒球專卷新聞。如圖 3 所示，視窗左側以樹狀

結構完整呈現棒球新聞專卷組織表，點選特定專卷即可列出該專卷所屬新聞於視窗右側中，再點選標題即可閱讀該篇數位化新聞。相較於尚未數位化前堆積如山的專卷卷宗，數位化類目表搭配數位典藏資料庫系統，因查找方便，可顯著提高專卷調閱效率。

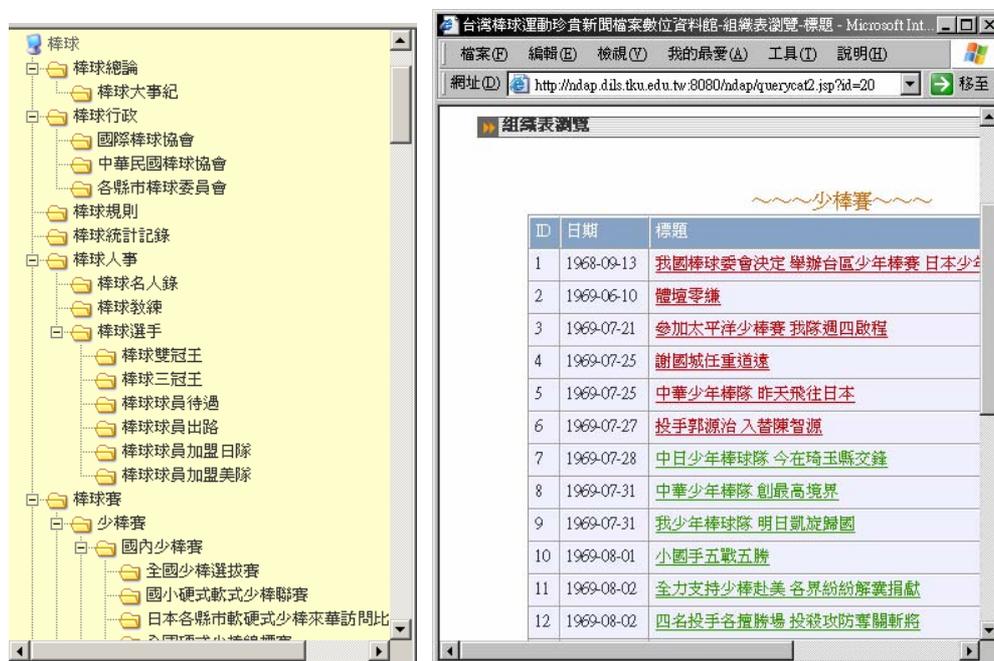


圖 3 數位化之棒球新聞專卷類目表瀏覽頁面

(3) 加值檢索-棒球大事記：棒球大事記指的是在此新聞資料庫中的重大新聞事件。由於不見得每篇新聞都是重要新聞，因此，我們在後台管理系統中加入可讓管理者選取重大新聞作為棒球大事記的功能；而前端使用者介面則將歷年來棒球重要事件列表呈現，使用者點選大事記中某筆新聞後，可瀏覽該則新聞內容，同時也可進一步連結至所屬專卷，進一步調閱相關新聞。

(4) 加值檢索-棒球名人錄：棒球名人是棒球運動中極重要的元素。然而，若僅以人名進行全文檢索，必將得到該名人所有新聞而非重要新聞。因此，我們在後台管理系統中加入可讓管理者以人名檢索後選取較重大的新聞作為棒球名人錄的功能，又分為「大事年表」、「詳細新聞」與「相關專卷」三個部分。「大事年表」主要是該棒球名人歷年所發生的重大新聞事件。同樣以郭泰源為例，目前已建立的大事年表中僅有經篩選過的 26 筆重要新聞，遠比全文檢索的 383 筆精練許多。點選大事年表中某筆新聞後可瀏覽該新聞內容，同時也可進一步連結至所屬專卷。「詳細新聞」是與某棒球名人真正有關之較重要新聞，其重要性次於大事年表。同樣若以郭泰源為例，目前已建立的詳細新聞僅列出真正有關的新聞 168 筆，也遠比全文檢索的 383 筆精練許多。「相關專卷」主要是列出與某棒球名人有關的相關專卷，並由這些相關專卷再連結出去，如此可形成網狀的資訊組織結構。同樣以郭泰源為例，其相關專卷有「張瓊姿」、「張純芳」、「二郭一莊」... 等。



圖 4 郭泰源的相關專卷

## 肆、新聞 Metadata 探討

本系統中亦提供新聞Metadata轉出功能，不過由於在系統開發之初，我們採用NRMF作為新聞資料交換之格式，<sup>38</sup>雖簡單易用，但當其對應到整個新聞的流程、管理與傳遞時，卻有許多不足之處，故而引發進一步研究各種新聞相關Metadata的動機，以瞭解其特性，以便於下一階段的第二期系統中完成更完整的Metadata轉出功能，將新聞轉換成更具價值之數位典藏資料，供進一步的加值利用。

### 一、國際通用的新聞 Metadata 格式

國際上為了解決數位化新聞的資訊組織問題，已發展出許多專用的 Metadata 格式。例如：

(1) 由國際新聞通訊協會IPTC (International Press Telecommunication Council) 所制訂的NITF (News Industry Text Format)<sup>39</sup>、NewsML (News Markup Language)<sup>40</sup>、SportsML (Sports Markup Language)<sup>41</sup>、ProgramGuideML (Program Guide Markup Language)<sup>42</sup>、EventsML (Events Markup Language)<sup>43</sup>等：(a) NITF著重在新聞內文的描述，主要用於傳統平面媒體的標示，設計初衷在加強新聞文件重覆使用的便利性，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

<sup>38</sup> 林信成、鄭國祥、孫正宜，「報紙新聞專卷的數位典藏與加值利用」，2004年海峽兩岸圖書資訊學暨教育發展研討會論文集，頁11~27，大連理工大學，2004/8/23~2004/8/24。

<sup>39</sup> IPTC, "News Industry Text Format", available at <http://www.nitf.org/> (2004/7/18).

<sup>40</sup> IPTC, "News Markup Language", available at <http://www.newsml.org/> (2004/7/18).

<sup>41</sup> IPTC, "Sports Markup Language", available at <http://www.sportsml.org/> (2004/10/8).

<sup>42</sup> IPTC, "Program Guide Markup Language", available at <http://www.programguideml.org/> (2004/10/8).

<sup>43</sup> IPTC, "Events Markup Language", available at <http://www.ipc.org/EventsML/> (2004/7/18).

在各種新聞平台上，達成多重發行的目的。目前使用NITF的著名機構有法新社（AFP）、美聯社（AP）、紐約時報（The New York Time）、義大利的ANSA通訊社等。<sup>44</sup>(b) NewsML的主要作用則是以Metadata來封裝多種不同的媒體(如文字、照片、圖像、錄影資料等)，主要用來描述電子出版、傳送、典藏的新聞文件，新聞內容的相關描述則藉助NITF或其他標示語言嵌入其中，因此較適用於大型的通訊社；(c) SportsML主要用於運動項目紀錄的資料交換，以凸顯體育在新聞報導領域的重要地位。SportsML的特色在於提供了識別與描述大量體育資料的能力，強調的重點在於：得分(Scores)、賽程(Schedules)、名次(Standings)、紀錄統計(Statistics) 和新聞報導(News)等。此外，SportsML也可先描述單一體育事件，再引用NITF描述該事件的相關或詳細新聞內容，最後再以NewsML進行新聞封裝，以求整體的結構性與完整性。(d) ProgramGuideML是為廣播與電視新聞節目專用的標示語言。(e) EventsML目前尚處於初步的需求研擬階段，其目的在建立一個發佈新聞事件的XML格式，可與NewsML、SportsML和NITF等標準互通。

IPTC 是推動和制訂新聞交換標準不遺餘力的機構之一，於 1965 年成立於英國，早年即研究以 SGML 格式作新聞交換，後改以 XML 格式為主。IPTC 成立了許多專業內容工作小組（Specialised Content Working Party），負責旗下相關標準的發展和維護。

(2) 由網景公司（Netscape）的推播技術（Push）衍生而來的RSS（Really Simple Syndication）<sup>45</sup>。這是一種用於共享新聞和其他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。但因其特殊的發展歷程，RSS有可能是以下三者之一的縮寫：<sup>46</sup>(a) Really Simple Syndication（真正簡單的資訊組織格式）、(b) RDF Site Summary（基於RDF的網站摘錄格式）或(c) Rich Site Summary（含意豐富的網站摘錄格式）。Netscape最初推出的是RSS 0.9 版，後來這項技術一方面由UserLand公司的Dave Winner接手開發新的版本（通稱為 0.9x版）；另一方面則由若干專業人士共同組成的RSS-DEV工作小組，根據W3C語義網（Semantic Web）架構的核心技術RDF（Resource Description Framework）<sup>47</sup>，對RSS 0.9x重新定義為「RDF Site Summary」，成為RSS 1.0，並與RSS 0.9x不相容。而Dave Winner也自行把RSS 0.9x升級為 2.0，亦與RSS 1.0 不同。<sup>48</sup>自此RSS分成兩個規格：「非RDF-based的RSS 0.9x/2.0 規格」<sup>49</sup>和「RDF-based的RSS 1.0 規格」<sup>50</sup>。

(3) 由IDEAlliance發佈的PRISM (Publishing Requirements for Industry Standard

<sup>44</sup> IPTC, "NITF: Introduction", available at <<http://www.nitf.org/intro.php>> (2004/9/30).

<sup>45</sup> Berkman Center, "RSS 2.0 Specification", available at <<http://blogs.law.harvard.edu/tech/rss/>> (2004/10/9).

<sup>46</sup> 維基百科, "RSS", 可得自 <<http://zh.wikipedia.org/wiki/RSS>> (2004/10/12)。

<sup>47</sup> W3C, "Resource Description Framework (RDF) - W3C Semantic Web Activity", available at <<http://www.w3.org/RDF/>> (2004/10/12).

<sup>48</sup> Weblogpedia, 「Weblog百科全書 — RSS」, 可得自 <<http://www.cnblog.org/wiki/?p=RSS>> (2004/10/12)

<sup>49</sup> Berkman Center, "RSS 2.0 Specification", available at <<http://blogs.law.harvard.edu/tech/rss/>> (2004/10/12).

<sup>50</sup> RDF Site Summary (RSS) 1.0, "RDF Site Summary (RSS) 1.0 - Official Specification", available at <<http://web.resource.org/rss/1.0/>> (2004/10/12).

Metadata)<sup>51</sup>。IDEAlliance一向致力於出版業的XML Metadata標準之制訂，PRISM包括了兩個部分：(a) PRISM規格（PRISM Specification）<sup>52</sup>本身，定義了PRISM的整體框架，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，如管理、聚集、後製作和多元利用等；(b) PRISM聚合者訊息（PRISM Aggregator Message，簡稱PAM），這是一個藉由PRISM 聚合者 DTD（PRISM Aggregator DTD）<sup>53</sup>整合，可讓出版者於線上傳送內容給聚合者和報業組織的標準格式，也可與XHTML搭配使用。

(4) 由XMLNews.Org所研擬的XMLNews<sup>54</sup>。XMLNews分成兩部份：XMLNews-Story和XMLNews-Meta。<sup>55</sup> XMLNews-Story<sup>56</sup>主要在描述新聞報導的文字內容，是借用NITF而來的，事實上是NITF的精簡子集（Subset）；XMLNews-Meta<sup>57</sup>則定義了一組關於新聞物件的Metadata，用以傳送不同形式的新聞物件，如文字格式的XMLNews-Story文件，或非文字格式的多媒體影像、聲音檔。

## 二、海峽兩岸中文新聞 Metadata 格式

中文新聞的標示語言大多也是遵循XML規範，海峽兩岸近年來也開始著手此方面之研究。例如：

(1) 台灣方面有國科會制訂的「數位典藏聯合目錄Metadata格式」(Digital Archive Catalog，本文簡稱為DAC)<sup>58</sup>；文建會制訂的「新聞紀錄Metadata格式」(News Records Metadata Format，本文簡稱為NRMF)<sup>59</sup>。DAC和NRMF都是可適用於中文環境的新聞Metadata格式，兩者皆以DC的15個核心元素為基礎架構，再依需求延伸出其他欄位來著錄更詳盡的資料。由於DC具有簡易、明確、彈性等特色，因此，DAC和NRMF也具備了相類似的性質，極容易便可與大量採用DC作為Metadata格式的系統進行資料互通與共享。

此外，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>60</sup>。然而，UdnML目前尚處於工作草案階段，僅知是以國

<sup>51</sup> IDEAlliance PRISM Working Group, "PRISM: Publishing Requirements for Industry Standard Metadata", available at <<http://www.prismstandard.org/>> (2004/10/9).

<sup>52</sup> IDEAlliance PRISM Working Group, "PRISM Specification (Version 1.2)", available at <<http://www.prismstandard.org/specifications/#specifications>> (2004/10/12).

<sup>53</sup> IDEAlliance PRISM Working Group, "The PRISM Aggregator DTD (Version 1.0) ", available at <<http://www.prismstandard.org/specifications/#aggregator>> (2004/10/12).

<sup>54</sup> XMLNews.Org, "XML and the news industry", available at <<http://xmlnews.org/>> (2005/1/20).

<sup>55</sup> XMLNews.org, "XMLNews Specifications", available at <<http://www.xmlnews.org/XMLNews/>> (2005/1/20).

<sup>56</sup> XMLNews.Org, "XMLNews-Story Documentation", available at <<http://www.xmlnews.org/docs/xmlnews-story.html>> (2005/1/20).

<sup>57</sup> XMLNews.Org, "XMLNews-Meta Documentation", available at <<http://xmlnews.org/docs/xmlnews-meta.html>> (2005/1/20)

<sup>58</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>59</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", 可得自 <<http://km.cca.gov.tw/download/rule.html>> (2004/7/17)。

<sup>60</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。

際標準的NewsML規格為主，大陸新華社的XinhuaML（後述）為輔，同時也融合了NITF可描述新聞內容段落與插題的特色，再配合聯合報系本身的報業作業流程一併考慮在內的規格。

(2) 大陸方面目前有兩單位制定新聞標示語言，一是新華社所發展的「新華標示語言」(Xinhua Markup Language，簡稱XinhuaML)。XinhuaML是根據新華社內部需求進行研發，以NewsML為藍本，參照NewsML經驗，針對新華社新聞的採集、加工、發布和評估等涉及新聞資訊的處理、交換、儲存而制定的一套新聞標示語言。<sup>61</sup>另一個是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)。CNTF是依據NITF所發展的，制定目的在提供中國大陸報業電子新聞文稿基本結構，以及描述新聞文稿詮釋資料和內容的一套完整的標示語言。採用CNTF標示語言編寫的新聞文稿可以方便的檢索、分類、歸檔和多次反復引用和出版。<sup>62</sup>

### 三、泛用型的 Metadata 格式

除了上述新聞專用的 Metadata 格式之外，有些泛用型的 Metadata 亦可適用於數位化新聞的環境中。例如：

(1) 由OCLC (Online Computer Library Center) 和NCSA (National Center for Supercomputing Application) 所推動的都柏林核心集 (Dublin Core，簡稱DC)<sup>63</sup>，它雖不是專對數位化新聞而設計的，卻也能廣泛適用。DC創始於 1995 年 3 月，是一個簡易有彈性並且連非專業人員也能輕易使用的Metadata格式，最初訂定的核心資料項只有 13 項，後來修訂為 15 項。目前DC已經進入發展成熟與穩定的階段，在DCMI (Dublin Core Metadata Initiative) 的網頁上，可以找到最新的建議規範 (Recommendations)。

(2) 由全球資訊網協會 (World Wide Web Consortium，簡稱W3C) 所制訂，作為各種Metadata整合機制的「資源描述框架」(Resource Description Framework，簡稱RDF)<sup>64</sup>。現今由於各種Metadata格式眾多，各有其應用領域，不可能有適用於任何領域的Metadata，因此在選用Metadata格式時，應先分析自身的資料特性，再選用最適合者，若該規格不敷使用，再結合其他的Metadata，完成資源的整合。W3C便在此一前提下提出了RDF，以便進行各種電子化資源之整合，在其資料模型 (Data Model) 中，被描述的物件一律稱為「資源」(Resource)，每個資源都可以使用唯一的URI來加以識別，並以特性 (Property) 與值 (Value) 的關係來描述。從人工智慧的角度觀之，RDF相當於「物件-屬性-值」三元式 (Object-Attribute-Value Triples) 知識表示法

<sup>61</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>62</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。

<sup>63</sup> Dublin Core Metadata Initiative, "Recommendations," available at <<http://purl.org/DC/documents/recommendations.htm>> (2004/12/17).

<sup>64</sup> W3C, "RDF Primer - W3C Recommendation 10 February 2004", 2004/2/10, available at <<http://www.w3.org/TR/rdf-primer/>> (2005/1/6)

(Knowledge Representation)<sup>65</sup>。如今，RDF已發展成爲建構語意網的核心技術了。

## 伍、棒球新聞 Metadata 著錄範例

由於新聞 Metadata 格式眾多，本研究進一步分析了各種 Metadata 之間的對應、轉換與整合，做爲未來系統改進時的參考。在此舉一則「台灣棒球運動珍貴新聞檔案數位資料館」中的新聞報導爲範例（見表 1），以數種不同方式進行對應轉換，分別是：(1) DAC、(2) RDF+DC、(3) RSS、(4) NITF、(5) SportsML+NewsML+NITF、(6) RDF+DC+PRISM、(7) PAM+PRISM+DC。

表 1 範例新聞：金龍少棒榮獲一九六九年世界少棒錦標賽冠軍（資料來源：聯合報）

欄位	值
識別碼	2126
新聞日期	1969-08-24
主標題	廿三屆世界少年棒球賽 我隊榮膺冠軍
副標題	五 A 比 0 大挫美西隊
記者	本報駐美記者施克敏台北時間廿四日凌晨威廉波特電話
報別	聯合報
新聞內容 (節錄)	<p>台中金龍少年棒球隊，今天以五 A 比零擊敗美國西區代表隊，榮獲一九六九年世界少年棒球賽錦標。...</p> <p>...</p> <p>此場比賽第一局雙方均掛零，第二局金龍隊發動兇猛攻勢，余宏開擊出二壘安打，陳鴻欽也得一支一壘安打，美西隊捕手慌張中屢現失誤，金龍小將一口氣攻下三分，以三比零領先。第三局，美西隊曾獲一支安打，這是陳智源在此次大賽中首次被擊出安打。美西隊曾佔領一、三壘，但後繼無力，仍然掛零。金龍隊本局也沒有建樹。第四局前半，美國攻勢無力發動，金龍小將又靠純熟的打擊與盜壘，獲得兩分，以五比零領先。第五局雙方無建樹，金龍隊似已贏定了。第六局美西隊反攻得勝希望已微，力攻不下。</p> <p>...</p> <p>兩隊列出的陣容爲：金龍隊：首棒莊凱評（游擊手）、二棒黃正一（一壘手）、三棒林建良（右翼）、四棒陳智源（投手）、五棒陳弘丕（二壘手）、六棒郭源治（中堅）、七棒余宏開（三壘手）、八棒陳鴻欽（左翼）、九棒蔡松輝（捕手）。美國隊：首棒威福、二棒科恩、三棒第沙里歐、四棒卡汀、五棒史密斯、六棒巴第納、七棒波特、八棒沙肯、九棒威爾。</p>
版名	體育 (or 體育新聞/棒球)
來源	<a href="http://ndap.dils.tku.edu.tw:8080/ndap/querynews2.jsp?id=2126">http://ndap.dils.tku.edu.tw:8080/ndap/querynews2.jsp?id=2126</a>
語文	zh-tw (or 中文)
版權	聯合報系
資料格式	text/xml (or XML 檔)
資料類型	text (or 文字)
涵蓋時空	Taiwan, ROC
關聯	<a href="http://ndap.dils.tku.edu.tw">http://ndap.dils.tku.edu.tw</a>

<sup>65</sup> Giarratano and Riley, Expert Systems: Principle and Programming, PWS-KENT, 1989, Page 73.

## 一、DAC 格式

國科會數位典藏聯合目錄 DAC 格式分為 AdminDesc 和 MetaDesc 兩部分：AdminDesc 區段內包含計畫管理與分類用的若干欄位；而 MetaDesc 區段內則是採用 DC 的 15 個欄位來描述數位典藏品內容。

參照DC格式再配合聯合目錄需求<sup>66</sup>，可將表 1 的範例新聞轉成如下的數位典藏聯合目錄DAC格式：

表 2 DAC 格式

```
<?xml version="1.0" encoding="big5"?>
<!DOCTYPE DACatalog SYSTEM "DAC.dtd">
<DACatalog>
  <AdminDesc>
    <Project Creator="淡江大學資訊與圖書館學研究所" GenDate="2005-01-05 06:50:21">台灣棒球運動珍貴新聞檔案數位資料館之建置</Project>
    <Catalog>
      <Record>典藏機構與計畫:公開徵選計畫:淡江大學:資訊與圖書館學研究所:台灣棒球運動珍貴新聞檔案數位資料館之建置</Record>
      <Record>內容主題:新聞:台灣棒球</Record>
      <Record>地理架構:台灣</Record>
      <Record>時間架構:1969/8/24</Record>
    </Catalog>
    <DigiArchiveID>2126</DigiArchiveID>
    <Hyperlink>http://ndap.dils.tku.edu.tw:8080/ndap/querynews2.jsp?id=2126</Hyperlink>
  </AdminDesc>
  <MetaDesc>
    <Title>廿三屆世界少年棒球賽 我隊榮膺冠軍：五 A 比 0 大挫美西隊</Title>
    <Creator>本報駐美記者施克敏台北時間廿四日凌晨威廉波特電話</Creator>
    <Subject>體育</Subject>
    <Description>台中金龍少年棒球隊，今天以五 A 比零擊敗美國西區代表隊，榮獲一九六九年世界少年棒球賽錦標。...
    ...
    (新聞內文，略)
    ...
  </Description>
  <Publisher>聯合報</Publisher>
  <Date>1969-8-4</Date>
  <Type>文字</Type>
  <Format>XML 檔</Format>
  <Identifier>2126</Identifier>
  <Language>中文</Language>
  <Relation>淡江大學資訊與圖書館學研究所-台灣棒球運動珍貴新聞檔案數位資料館
  http://ndap.dils.tku.edu.tw/</Relation>
  <Coverage>Taiwan, ROC</Coverage>
  <Rights>聯合報系</Rights>
</MetaDesc>
</DACatalog>
```

<sup>66</sup> 數位典藏聯合目錄系統建置計畫工作小組，「數位典藏國家型科技計畫—數位典藏聯合目錄系統」簡介：數位典藏各子計畫資料庫與聯合目錄系統（2004.11.01 版），可得自 <<http://catalog.ndap.org.tw/System/Info/Intro.doc>> (2004/12/27)。

## 二、RDF + DC 格式

RDF 結合 DC 是描述數位化文件或網路資源的有效方式之一，因此也極適用於數位化新聞。表 1 的範例新聞可先以 DC 格式加以標示後，再嵌入 RDF 框架中，轉成如下的 RDF+DC 格式：

表 3 RDF+DC 格式

```
<?xml version="1.0" encoding="big5"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
<rdf:Description rdf:about="http://ndap.dils.tku.edu.tw:8080/ndap/querynews2.jsp?id=2126">
  <dc:title>廿三屆世界少年棒球賽 我隊榮膺冠軍：五 A 比 0 大挫美西隊</dc:title>
  <dc:creator>本報駐美記者施克敏台北時間廿四日凌晨威廉波特電話</dc:creator>
  <dc:subject>體育</dc:subject>
  <dc:description>台中金龍少年棒球隊，今天以五 A 比零擊敗美國西區代表隊，榮獲一九六九年世界少年棒球賽錦標。...
  ...
  (新聞內文，略)
  ...
</dc:description>
<dc:publisher>聯合報</dc:publisher>
<dc:date>1969-8-4</dc:date>
<dc:type>text</dc:type>
<dc:format>text/xml</dc:format>
<dc:identifier>2126</dc:identifier>
<dc:source>http://ndap.dils.tku.edu.tw:8080/ndap/querynews2.jsp?id=2126</dc:source>
<dc:language>zh-tw</dc:language>
<dc:relation>http://ndap.dils.tku.edu.tw</dc:relation>
<dc:coverage>Taiwan, ROC</dc:coverage>
<dc:rights>聯合報系</dc:rights>
</rdf:Description>
</rdf:RDF>
```

## 三、RSS 格式

RSS 1.0 和 0.9x/2.0 的結構雖然不同，RSS 1.0 遵循 RDF 和 XML Namespace 標準，而 RSS 0.9x/2.0 則強調簡易性而無此限制，但兩者的核心元素在本質上其實是一樣的，兩個版本都以 rss 為根元素，以下包含 channel 元素，而 channel 元素又包含三個必須的元素：title、description 和 link。除了這些必須的元素外，RSS 1.0 定義了三個附加元素：image、items 和 textinput，其中，image 和 textinput 是可選的，另允許通過定義在單獨的 XML Namespace 來擴展詮釋資料。另一方面，RSS 2.0 則提供了 16 個附加元素，其中也包括 image、items 和 textinput，此外還有 language、copyright、managingEditor、pubDate 和 category 等。

表 1 的範例新聞可以轉換成如下表的 RSS 格式：

表 4 RSS 格式

```

<?xml version="1.0" encoding="big5"?>
<rss version="2.0">
  <channel>
    <title>棒球史上的今天 — 數位典藏國家型科技計畫</title>
    <link>http://ndap.dils.tku.edu.tw/rss</link>
    <description>台灣棒球運動珍貴新聞檔案數位資料館</description>
    <language>zh-tw</language>
    <copyright>聯合報系</copyright>
    <pubDate>1969/8/24</pubDate>
    <item>
      <title>廿三屆世界少年棒球賽 我隊榮膺冠軍：五 A 比 0 大挫美西隊(聯合報)</title>
      <link>http://ndap.dils.tku.edu.tw:8080/ndap/querynews2.jsp?id=2126</link>
      <description>台中金龍少年棒球隊，今天以五 A 比零擊敗美國西區代表隊...
        ...
        (新聞內文，略)
        ...
      </description>
      <author>本報駐美記者施克敏台北時間廿四日凌晨威廉波特電話</author>
      <category>體育</category>
      <pubdate>1969-8-4</pubdate>
      <source url="http://ndap.dils.tku.edu.tw">台灣棒球運動珍貴新聞檔案數位資料館</source>
    </item>
  </channel>
</rss>

```

#### 四、NITF 格式

NITF 主要分為 head 與 body 兩部分：head 是放置新聞管理的 Metadata，包括了新聞檔頭、詮釋資料、主題詞表、格式、出版日期、版本修訂記錄。body 則又分為 body.head、body.content、body.end 三部份；body.head 這部份主要是描述新聞標題組件 (news components) 像標題、作者、版權、貢獻者、日期、新聞的摘要；body.content 是描述新聞內文，包括內容文字、表格、清單、圖片、段落、插題其他的項目；body.end 是描述新聞的結尾，包括了附註或書目資料。此外，NITF 還具備豐富的內文語意標籤，可加註到文章中，用以描述新聞事件的 5W1H。如：person、location、event、chron、object.title、org、function、byttl ... 等。

表 1 的範例新聞若加入豐富的 NITF 內文語意標籤，可轉換成如下表所示之結果：

表 5 NITF 格式

```

<?xml version="1.0" encoding="big5"?>
<!DOCTYPE nitf SYSTEM "..\dtd\nitf-3-2.dtd">
<nitf version="-//IPTC//DTD NITF 3.2//EN">
  <head>
    <title>廿三屆世界少年棒球賽 我隊榮膺冠軍</title>
    <docdata management-status="usable" management-doc-idref="2126">
    <doc-id id="2126" />
    <doc.copyright id="2126" year="1969" holder="台灣棒球運動珍貴新聞檔案數位資料館" />
    <doc.rights owner="聯合報系" />
    </docdata>
  </head>

```

```

<body>
  <body.head>
    <headline>
      <h1><event>廿三屆世界少年棒球賽</event>我隊榮膺冠軍</h1>
      <h2>五 A 比 0 大挫<org>美西隊</org></h2>
    </headline>
    <byline><byttl>本報駐美記者</byttl><person>施克敏</person>
      台北時間廿四日凌晨<location>威廉波特</location>電話</byline>
    <dateline>1969/08/24</dateline>
  </body.head>
  <body.content>
    <p><location>台中</location><org>金龍少年棒球隊</org>，今天以五 A 比零擊敗<org>美國西區代表隊</org>，榮獲<chron>一九六九年</chron><event>世界少年棒球賽</event>錦標。...</p>
    <p>... 此場比賽第一局雙方均掛零，第二局<org>金龍隊</org>發動兇猛攻勢，<person>余宏開</person>擊出二壘安打，<person>陳鴻欽</person>也得一支一壘安打，<org>美西隊</org>捕手慌張中屢現失誤，金龍小將一口氣攻下三分，以三比零領先。第三局，<org>美西隊</org>曾獲一支安打，這是<person>陳智源</person>在此次大賽中首次被擊出安打。<org>美西隊</org>曾佔領一、三壘，但後繼無力，仍然掛零。<org>金龍隊</org>本局也沒有建樹。第四局前半，<location>美國</location>攻勢無力發動，金龍小將又靠純熟的打擊與盜壘，獲得兩分，以五比零領先。第五局雙方無建樹，<org>金龍隊</org>似已贏定了。第六局美西隊反攻得勝希望已微，力攻不下。</p>
    <p>... 兩隊列出的陣容為：<org>金龍隊</org>：首棒<person>莊凱評</person>（<function>游擊手</function>）、二棒<person>黃正一</person>（<function>一壘手</function>）、三棒<person>林建良</person>（<function>右翼</function>）、四棒<person>陳智源</person>（<function>投手</function>）、五棒<person>陳弘丕</person>（<function>二壘手</function>）、六棒<person>郭源治</person>（<function>中堅</function>）、七棒<person>余宏開</person>（<function>三壘手</function>）、八棒<person>陳鴻欽</person>（<function>左翼</function>）、九棒<person>蔡松輝</person>（<function>捕手</function>）。<org>美國隊</org>：首棒<person>威福</person>、二棒<person>科恩</person>、三棒<person>第沙里歐</person>、四棒<person>卡汀</person>、五棒<person>史密斯</person>、六棒<person>巴第納</person>、七棒<person>波特</person>、八棒<person>沙肯</person>、九棒<person>威爾</person>。</p>
  </body.content>
  <body.end>
    <tagline type="std">
      <org>聯合報</org>
      <lang>zh-tw</lang>
      <location>Taiwan, ROC</location>
    </tagline>
  </body.end>
</body>
</nitf>

```

## 五、SportsML + NITF + NewsML 格式

SportsML 的根元素為 sports-content，子元素除了必要的 sports-metadata 外，sports-event 和 article 都是可選擇性的。sports-event 用以描述體育事件，並與 team 和 player 元素進行組合，以記錄體育活動中的競賽隊伍；而 article 則是作為該運動項目的新聞報導之描述，在規格書中建議結合 NITF 使用。再者，SportsML+NITF 更可再進一步和 NewsML 整合。NewsML 分為新聞封裝 NewsEnvelope 與新聞條目 NewsItem

兩部份：NewsEnvelope 主要是用來記錄新聞傳送過程中所需要的各種訊息，就如同一個將新聞打包的過程，將要傳遞的新聞內容裝在一個信封裡，並在信封註明何時要由何地將新聞資料送往何處等資訊。NewsItem 則記錄了新聞內容、新聞素材及相關資訊，也可說是包裹在信封裡，實際要傳送的東西，包含新聞的實際內容、各種管理訊息和識別資訊等。NewsML 結構的最底層為內容素材 ContentItem，SportsML 和 NITF 所描述的內容被視為 ContentItem 的一部份，再與其他相關的內容素材進行結合，而成為新聞組件 NewsComponent 的一部份，並可與其他的 NewsComponent 再進行嵌合，其次再註明何時要傳遞到何地，以一層層包裹的方式進行新聞封裝 NewsEnvelope，最後再進行傳遞。

將前例以 NITF 所著錄的資料，再整合 SportsML 和 NewsML 即可成為如下表的格式：

表 6 SportsML + NewsML + NITF 格式

```

<?xml version="1.0" encoding="big5"?>
<NewsML>
  <NewsEnvelope></NewsEnvelope>
  <NewsComponent>
    <ContentItem>
      <DataContent>
        <sports-content>
          <sports-metadata event-status="post-event"/>
          <sports-event>
            <team>
              <team-metadata><name first="中華" last="金龍少棒隊" /></team-metadata>
              <team-stats score="5" event-outcome="win" />
            </team>
            <team>
              <team-metadata><name first="美國" last="西區少棒隊" /></team-metadata>
              <team-stats score="0" event-outcome="loss" />
            </team>
          </sports-event>
          <article format="NITF">
            <nitf version="-//IPTC//DTD NITF 3.2//EN" >
              :
              :
              新聞內文插入於此，參見前例（略）
              :
              :
            </nitf>
          </article>
        </sports-content></sports-content>
      </DataContent>
    </ContentItem>
  </NewsComponent>
</NewsML>

```

由於本研究僅著眼於數位新聞的典藏利用，並未探討有關新聞業者的作業流程，如新聞的採集、加工、評估、編排和發布等層面。因此，此處僅將 NewsML 用於單一新聞內容素材的封裝，並未充分發揮 NewsML 在整個新聞產出與消費的生命週期中所擁有的強大管理能力，可謂大材小用。

## 六、RDF + DC + PRISM 格式

PRISM 採用 RDF 架構，透過名稱空間（Namespace）以模組化（Modularized）方式整合不同的 Metadata 詞彙集（Vocabulary），以利彈性擴充與應用。除了採用 DC 的<dc:identifier>、<dc:title>、<dc:creator> ... 等作為基本元素外，另有適用於描述新聞報導 5W1H 的<prism:event>、<prism:location>、<prism:person> ... 等。表 1 的範例新聞若以 DC、PRISM 等元素加以標示，再嵌入 RDF 框架中，即可轉成如下的 RDF+DC+PRISM 格式：

```
<?xml version="1.0" encoding="big5"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:prism="http://prismstandard.org/namespaces/1.2/basic/">
  <rdf:Description rdf:about="http://ndap.dils.tku.edu.tw:8080/ndap/querynews2.jsp?id=2126">
    <dc:title>廿三屆世界少年棒球賽 我隊榮膺冠軍：五 A 比 0 大挫美西隊</dc:title>
    <dc:creator>本報駐美記者施克敏台北時間廿四日凌晨威廉波特電話</dc:creator>
    <dc:subject>體育</dc:subject>
    <prism:event>金龍少棒隊首獲世界少棒賽冠軍</prism:event>
    <prism:location>威廉波特</prism:location>
    <prism:person>陳智源</prism:person>
    <prism:person>郭源治</prism:person>
    <prism:person>余宏開</prism:person>
    ... (依需要增加，略)
    <dc:description>台中金龍少年棒球隊，今天以五 A 比零擊敗美國西區代表隊 ...
      ...
      (新聞內文，略)
      ...
    </dc:description>
    <dc:publisher>聯合報</dc:publisher>
    <dc:date>1969-8-4</dc:date>
    <dc:type>text</dc:type>
    <dc:format>text/xml</dc:format>
    <dc:identifier>2126</dc:identifier>
    <dc:source>http://ndap.dils.tku.edu.tw:8080/ndap/querynews2.jsp?id=2126</dc:source>
    <dc:language>zh-tw</dc:language>
    <dc:relation>http://ndap.dils.tku.edu.tw</dc:relation>
    <dc:coverage>Taiwan, ROC</dc:coverage>
    <dc:rights>聯合報系</dc:rights>
  </rdf:Description>
</rdf:RDF>
```

## 七、PAM + PRISM + DC 格式

PAM（PRISM Aggregation Message）主要用於出版者將數位內容重整、聚合，作為二次授權或加值應用，可結合 XHTML 進行 Web 數位發行；也可如同 NITF 般，利用豐富的內文語意標籤加註於文章中，如：<pim:event>、<pim:location>、<pim:person>、<pim:industry> ... 等，以描述新聞事件的 5W1H，稱為 PRISM Inline Markup（簡稱 PIM）。

```

<?xml version="1.0" encoding="big5"?>
<!DOCTYPE pam:message PUBLIC "-//PRISMstandard.org/DTD Aggregation with XHTML v1.0/EN"
"./DTD/pam-norm.dtd">
<pam:message
  xmlns:pam="http://prismstandard.org/namespaces/pam/1.0/"
  xmlns:prism="http://prismstandard.org/namespaces/basic/1.2/"
  xmlns:pim="http://prismstandard.org/namespaces/pim/1.2/"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <pam:article>
    <head>
      <dc:identifier>2126</dc:identifier>
      <dc:title>廿三屆世界少年棒球賽 我隊榮膺冠軍：五 A 比 0 大挫美西隊</dc:title>
      <dc:creator>本報駐美記者施克敏台北時間廿四日凌晨威廉波特電話</dc:creator>
      <prism:publicationName>聯合報</prism:publicationName>
      <dc:publisher>聯合報系</dc:publisher>
      <prism:coverDate>1969-8-4</prism:coverDate>
      <prism:coverDisplayDate>中華民國 58 年 8 月 4 日</prism:coverDisplayDate>
      <dc:subject>棒球</dc:subject>
      <prism:event>金龍少棒隊首獲世界少棒賽冠軍</prism:event>
      <prism:location>威廉波特</prism:location>
      <prism:category>體育新聞</prism:category>
      <prism:copyright>聯合報系</prism:copyright>
    </head>
    <body>
      <h1>廿三屆世界少年棒球賽 我隊榮膺冠軍：五 A 比 0 大挫美西隊</h1>
      <p><pim:location> 台 中 </pim:location><pim:organization> 金 龍 少 年 棒 球 隊
      </pim:organization>，今天以五 A 比零擊敗<pim:organization>美國西區代表隊
      </pim:organization>，榮獲一九六九年<pim:event>世界少年棒球賽</pim:event>錦標。…
      </p>
      <p>… 此場比賽第一局雙方均掛零，第二局<pim:organization>金龍隊
      </pim:organization>發動兇猛攻勢，<pim:person>余宏開</pim:person>擊出二壘安打，
      <pim:person>陳鴻欽</pim:person>也得一支一壘安打，<pim:organization>美西隊
      </pim:organization>捕手慌張中屢現失誤，金龍小將一口氣攻下三分，以三比零領先。
      第三局，<pim:organization>美西隊</pim:organization>曾獲一支安打，這是<pim:person>
      陳智源</pim:person>在此次大賽中首次被擊出安打。<pim:organization>美西隊
      </pim:organization>曾佔領一、三壘，但後繼無力，仍然掛零。<pim:organization>金龍隊
      </pim:organization>本局也沒有建樹。第四局前半，<pim:location>美國</pim:location>攻
      勢無力發動，金龍小將又靠純熟的打擊與盜壘，獲得兩分，以五比零領先。第五局雙方
      無建樹，<pim:organization>金龍隊</pim:organization>似已贏定了。第六局美西隊反攻得
      勝希望已微，力攻不下。</p>
      <p>… 兩隊列出的陣容為：<pim:organization>金龍隊</pim:organization>：首棒
      <pim:person>莊凱評</pim:person>（游擊手）、二棒<pim:person>黃正一</pim:person>（一
      壘手）、三棒<pim:person>林建良</pim:person>（右翼）、四棒<pim:person>陳智源
      </pim:person>（投手）、五棒<pim:person>陳弘丕</pim:person>（二壘手）、六棒
      <pim:person>郭源治</pim:person>（中堅）、七棒<pim:person>余宏開</pim:person>（三
      壘手）、八棒<pim:person>陳鴻欽</pim:person>（左翼）、九棒<pim:person>蔡松輝
      </pim:person>（捕手）。<pim:organization>美國隊</pim:organization>：首棒<pim:person>
      威福</pim:person>、二棒<pim:person>科恩</pim:person>、三棒<pim:person>第沙里歐
      </pim:person>、四棒<pim:person>卡汀</pim:person>、五棒<pim:person>史密斯
      </pim:person>、六棒<pim:person>巴第納</pim:person>、七棒<pim:person>波特
      </pim:person>、八棒<pim:person>沙肯</pim:person>、九棒<pim:person>威爾
      </pim:person>。</p>
    </body>
  </pam:article>
</pam:message>

```

## 陸、結論

數位典藏與電子出版時代，優質化資訊的重要性與日俱增，而 Metadata 則是提升數位化新聞品質的有效途徑之一。本文除了完成台灣棒球報紙新聞數位典藏與加值系統之第一期建置外，更探討了國內外較重要的新聞 Metadata，也嘗試對若干常用的新聞 Metadata 進行對應分析，並以多種不同形式的 Metadata 組合，對數位化新聞進行著錄實作，不但能增進我們對各種新聞 Metadata 之特性與優缺點的瞭解，也有助於提升日後在數位典藏系統中陸續加入支援其他 Metadata 規格之能力。

## 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，鄭國祥老師協助程式設計，研究生孫正宜、游忠諺、康珮熏協助資料分析與系統建置，使本研究得以順利進行，特此致謝。