

數位化新聞的資訊組織

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

孫正宜

聯合知識庫資管二組 組長

cy.sun@udngroup.com

摘要

本文基於我們在進行數位典藏計畫「臺灣棒球運動珍貴新聞檔案數位資料館之建置」的研究過程中，發現以「都柏林核心集」(DC)為藍本的「新聞紀錄詮釋資料格式」(NRMF)雖然簡單易用，但當其對應到整個新聞的流程、管理與傳遞時，尚有許多對新聞業者而言仍顯不足之處。因此，在本文中，我們除針對國際上較重要的幾個新聞專用的標示語言，如 NewsML、NITF、SportsML、ProgramGuideML、EventsML、RSS、PRISM 和 XMLNews 進行研究外，也將一併探討海峽兩岸中文新聞標示語言之發展，如 NRMF、UdnML、XinhuaML、CNTF 等。以瞭解其間的差異性、優缺點與現況，作為我們日後在系統中陸續加入支援其他規格的設計依據，以提升該系統之資料交換與傳遞能力。

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

Abstract

When planned to do the research of “digital library of Taiwan baseball valuable and memorable news spots”, we found that the DC-based approach “NRMF” is easy to use yet still insufficient when facing the reality of news processing, management, and delivering. Therefore, in this article, we inspect some of the most popular mark-up languages on earth such as NewsML, NITF, SportsML, ProgramGuideML, EventsML, RSS, PRISM, and XMLNews. And then we also do some study of the development of news mark language in-between two sides of the Taiwan strait such as NRMF, UdnML, XinhuaML, CNTF and so on. Understanding the difference, strength and defect, and the current status enables us to add the compatibility with other mark-up languages into our system, and also increase the capability in data transferring and delivering.

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

壹、前言

新聞是人類接觸、取得和傳遞資訊的重要媒介之一，具有大眾化、普遍性與易得性。在網路化媒體時代，數位化新聞的產出速度不但極快，數量又龐大，表達新聞資訊的文字、圖片、聲音、影像等，更可以使用各種方式重新組合。因此，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。以詮釋資料（Metadata）加強數位化新聞的結構性並增加自我描述性，不但能提升新聞的利用價值，使其更易記錄、探索、定位、評估、選擇，亦有助於新聞資源的管理與系統的建置。

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)¹、NewsML (News Markup Language)²、SportsML (Sports Markup Language)³、ProgramGuideML (Program Guide Markup Language)⁴、EventsML (Events Markup Language)⁵；IDEAlliance發佈的PRISM (Publishing Requirements for Industry Standard Metadata)⁶，XMLNews.Org所研擬的XMLNews；和目前普遍應用於入口引擎、新聞網站、Blog和WiKi的RSS (Really Simple Syndication)⁷等。這些標準皆本於可擴展標示語言XML (eXtensible Markup Language)，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。⁸

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。中文新聞標示的好處自不待言，一旦標準制訂、業界認同，中文新聞即可在不同系統、不同平台、不同新聞機構流通交換，不必每次遇到新的交換情況，就必需撰寫電腦程式、或在資料結構上更動，系統與資料才能順暢連接。現在的新聞涵蓋範圍很廣，包括文字、圖片、廣告、廣播、電視、網站、以及其他電子傳播方式，所以完整的中文新聞標示語言，也需包括多樣的新聞媒體。⁹海峽兩岸近年有感於此，也

¹ IPTC, "News Industry Text Format", available at <<http://www.nitf.org/>> (2004/7/18).

² IPTC, "News Markup Language", available at <<http://www.newsml.org/>> (2004/7/18).

³ IPTC, "Sports Markup Language", available at <<http://www.sportsml.org/>> (2004/10/8).

⁴ IPTC, "Program Guide Markup Language", available at <<http://www.programguideml.org/>> (2004/10/8).

⁵ IPTC, "Events Markup Language", available at <<http://www.iptc.org/EventsML/>> (2004/7/18).

⁶ IDEAlliance PRISM Working Group, "PRISM: Publishing Requirements for Industry Standard Metadata", available at <<http://www.prismstandard.org/>> (2004/10/9).

⁷ Berkman Center, "RSS 2.0 Specification", available at <<http://blogs.law.harvard.edu/tech/rss/>> (2004/10/9).

⁸ 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> (2003/9/5).

⁹ 那福忠, "新聞如何標示", 可得自

<<http://www.oiprint.com.tw/epaper/in2004web/frankna/na040530.htm>> (2004/5/30).

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)¹⁰，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)，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)等。

在我們目前正執行的數位典藏計畫中，以「臺灣棒球運動珍貴新聞檔案數位資料館之建置」為主題，¹¹藉由適度的資訊組織方法對數位化新聞進行加值處理，使成為更有價值之數位典藏資料，完整呈現新聞事件的發展歷程，並採用NRMF作為新聞資料交換之格式，讓使用者更深入、更有效率的探索相關新聞。¹²然而，我們在系統實作過程中發現，以DC為藍本的NRMF雖然簡單易用，但當其對應到整個新聞的流程、管理與傳遞時，尚有許多值得改進之處；再者DC較適用網路資源，對新聞業者而言仍有所不足，故而引發本文之研究動機。在本研究中，我們將針對以上所提及之各項新聞相關的標示語言進行探討，以瞭解其間的差異性與優缺點，作為我們日後在系統中陸續加入與其他規格之相容性設計的依據，以提升該系統之資料交換與傳遞能力。

貳、國際新聞標示之發展

一、IPTC 相關標準

國際新聞通訊協會 IPTC (International Press Telecommunication Council) 是推動和制訂新聞交換標準不遺餘力的機構之一。IPTC 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，多年來陸續制定了許多不同類型的新聞媒體交換標準，包含 NITF 和 NewsML、SportsML、ProgramGuideML、和 EventsML 等。

(1) NITF

早在 1979 年，美國報紙出版者協會 ANPA (American Newspaper Publishers Association) 和 IPTC 即分別發展了 ANPA 1312 和 IPTC 7901 兩種格式作為文字新聞的交換平台。1990 年 IPTC 根據此二標準制訂出 NITF，不久即成為眾多歐美新聞機構用於新聞資料交換的標準，適用於獨立的單篇新聞 (Independent news articles)。NITF 一開始採用 SGML 作為其語法規則，直到 1998 年 XML 頒佈後，

¹⁰ 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>> (2004/7/17)。

¹¹ 淡江大學資圖所數位典藏計畫小組，「臺灣棒球運動珍貴新聞檔案數位資料館之建置」，可來自 <<http://ndap.dils.tku.edu.tw/>> (2004/10/14)。

¹² 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。

即改採XML語法。¹³

(2) NewsML

為適應網際網路的興起，方便新聞完整的循環周期管理（包括從生產到大量傳遞與交換），IPTC於1999年開始NewsML的需求研擬及標準制訂等工作，歷經約一年的時間，在2000年10月正式公佈了NewsML 1.0規範。NewsML主要是把多媒體數位新聞納入，由於重視多媒體新聞封包（Multimedia news packages）傳輸的功能，亦成為許多國際大通訊社採用的標準，像法新社、與日本共同社都是NewsML的用戶。¹⁴

(3) SportsML

體育在新聞報導的領域中佔有一席之地，讀者遍及老中青的年齡層。體育資料進行交換時雖需跨越不同的載體、不同的體育項目，但報導之事卻是大同小異，交換標準的統一更是當務之急。有鑑於體育資訊的XML標準制定仍屬於空窗階段，IPTC乃於2001年3月制定了體育新聞專用的SportsML，便於運動項目紀錄的資料交換。¹⁵

(4) ProgramGuideML

廣播與電視亦是新聞媒介的重要一環，IPTC於2002年2月制訂了廣播與電視新聞節目專用的標示語言ProgramGuideML。ProgramGuideML亦用採用XML語法，主要元素為不同報紙的新聞目錄、評論節目表、評論節目所用的圖片、廣播新聞節目表單等，範圍包括了財經、天氣、事件清單、電視節目表、選舉結果、政府資訊等主題。2004年香港與日本的廣播與電視業也相繼加入這一標準行列。¹⁶

(5) EventsML

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，尤其是透過IPTC的主題參考系統（Subject Reference System）相互參照。其設計目標是希望可以輕易的應用NewsML來實行EventsML的相關需求，例如：出版事件、有價值新聞事件資訊的徵集、案發新聞事件整個過程和新聞事件機構系列資訊蒐集。¹⁷

¹³ IPTC, “NITF: Introduction”, available at <<http://www.nitf.org/intro.php>> (2004/9/30).

¹⁴ IPTC, “Introduction to NewsML”, available at <http://www.newsml.org/pages/intro_main.php> (2004/9/30).

¹⁵ IPTC, “SportsML: Introduction”, available at <<http://www.sportsml.org/intro.php>> (2004/9/30).

¹⁶ IPTC, “Introduction to ProgramGuideML”, available at <http://www.programguideml.org/pages/intro_main.php> (2004/9/30)。

¹⁷ 同註5。

二、其他相關標準

(1) RSS

RSS指的可能是以下三者之一的縮寫：¹⁸Really Simple Syndication（真正簡單的資訊組織格式）、RDF Site Summary（基於RDF的網站摘錄格式）或Rich Site Summary（含意豐富的網站摘錄格式），是一種用於共享新聞和其他Web內容的資料交換規格。因其特殊的發展歷程，以致造成如今名稱及內涵不一的情況。

RSS起源於網景公司（Netscape）的推播技術（Push），是一種將訂閱內容傳送至訂戶的通訊協定（Protocol）。Netscape最初所推出的RSS版本是 0.9 版，但因未能發展出合適的商業模型，以致無疾而終。後來，這項技術一方面由UserLand公司的Dave Winner接手，繼續開發新的版本，以適應新的網路應用。在Dave Winner的努力下，RSS陸續升級到 0.91、0.92 ... 等版本（通稱為 0.9x版），隨後在各種Blog工具中得到了應用，並被眾多的專業新聞網站所支持。另一方面，一個由若干專業人士共同組成的RSS-DEV工作小組，則根據W3C語義網（Semantic Web）架構的核心技術RDF（Resource Description Framework）¹⁹，對RSS 0.9重新定義為「RDF Site Summary」，並於 2000 年 12 月發布了RSS 1.0 規格，與RSS 0.9x並不相容。而Dave Winner則在 2002 年 9 月獨自把RSS升級為 2.0 版本，其中的定義也與RSS 1.0 不同。²⁰自此RSS分成兩個規格：即「非RDF-based規格的RSS 0.9x/2.0」²¹和「RDF-based 規格的RSS 1.0」²²，如圖 1 所示。

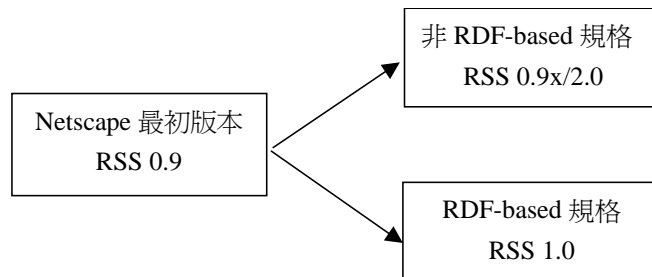


圖 1 兩種不同規格的 RSS

（資料來源：本研究整理）

雖然RSS目前廣泛用於blog、wiki和網上新聞頻道，世界許多知名新聞社網站也都提供RSS訂閱支持。²³但RSS規格爭議卻仍繼續進行著，並衍生出其他替

¹⁸ 維基百科，"RSS"，可得自 <<http://zh.wikipedia.org/wiki/RSS>> (2004/10/12)。

¹⁹ W3C, "Resource Description Framework (RDF) - W3C Semantic Web Activity", available at <<http://www.w3.org/RDF/>> (2004/10/12).

²⁰ Weblogpedia, 「Weblog 百科全書 — RSS」，可得自 <<http://www.cnblog.org/wiki/?p=RSS>> (2004/10/12)

²¹ Berkman Center, "RSS 2.0 Specification", available at <<http://blogs.law.harvard.edu/tech/rss>> (2004/10/12).

²² 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).

²³ 同註 18。

代規格，如Echo、Atom 和 Pie，²⁴造成許多開發人員和使用者的困擾。

(2) PRISM

PRISM是由IDEAlliance PRISM Working Group於2003年10月所發佈的新聞標示標準，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：其一為PRISM規格(PRISM Specification)²⁵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；其二為PRISM 聚合者DTD (PRISM Aggregator DTD)²⁶，一個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。

PRISM 建議在描述新聞出版資訊時，盡量採取符合XML、RDF、Dublin Core和ISO等標準，例如地點、語言和日期時間等格式。

(3) XMLNews

在1999年時，提供資料傳遞技術與服務的公司WavePhore，為了把新聞直接送到網站，在曾任W3C XML 資訊小組的召集人的David Megginson技術支援下，發表了以XMLNews規格為基礎的NewsPak 服務，能夠每天從34個著名新聞機構，合法取得2000則以上的新聞稿，供應一般用戶使用。²⁷

XMLNews分成兩部份：XMLNews-Story和XMLNews-Meta。²⁸ XMLNews-Story²⁹主要在描述新聞報導的實質內容，是借用另一個XML規格NITF而來的，事實上是NITF的精簡子集(Subset)。NITF的訂定是要取代以列印為主的ANPA 1312老舊規格³⁰，但Megginson認為NITF太龐大，許多Tag大多數人都不會用到，並且不夠靈活，要加一個Tag都不容易，所以才自行推出XMLNews。而XMLNews-Meta³¹則是一個可擴展的新聞詮釋資料詞彙集(Metadata vocabulary)，用以提供新聞報導的相關資訊和其他資源。XMLNews-Meta是依據W3C標準的規則RDF(Resource Description Framework)³²所制訂的，RDF是以物件導向的理念，用XML作規範，制訂出來一套描述資料的規則。

²⁴ 謝良奇，"Web 內容串聯技術 RSS 標準之爭"，自由軟體鑄造場計畫網站，2003/08/10，可得自 <<http://www.openfoundry.org/archives/000164.html>> (2004/10/12)。

²⁵ IDEAlliance PRISM Working Group, "PRISM Specification (Version 1.2)", available at <<http://www.prismstandard.org/specifications/#specifications>> (2004/10/12).

²⁶ IDEAlliance PRISM Working Group, "The PRISM Aggregator DTD (Version 1.0)", available at <<http://www.prismstandard.org/specifications/#aggregator>> (2004/10/12).

²⁷ 那福忠，"XML 傳遞新聞稿"，Apr 20, 1999，可得自 <http://www.brainnew.com.tw/Article/na1999/F_042099.htm> (2004/10/12).

²⁸ XMLNews.org, "XMLNews Specifications", available at <<http://www.xmlnews.org/XMLNews/>> (2004/10/12).

²⁹ XMLNews.Org, "XMLNews-Story Documentation", available at <<http://www.xmlnews.org/docs/xmlnews-story.html>> (2004/10/12).

³⁰ IPTC, "NITF: Introduction", available at <<http://www.nitf.org/site/intro.html>> (2004/10/12).

³¹ XMLNews.Org, "XMLNews-Meta Documentation", available at <<http://www.xmlnews.org/docs/xmlnews-meta.html>> (2004/10/12)..

³² W3C, "Resource Description Framework (RDF) - W3C Semantic Web Activity", available at <<http://www.w3.org/RDF/>> (2004/10/12).

XMLNews-Meta是描述新聞稿件的規格，其DTD中的Element有新聞機構代碼與名稱、版權、使用權、作者、新聞類別、產業名稱、語言名稱與代碼、人物名稱、發稿時間等多項。³³

參、NITF 與 NewsML

由以上所述可知各種組織數位化新聞資訊的標示語言，各有其應用特性，有些彼此之間可互通有無，交相利用；有些則互相競爭，各不相讓。其中，以 IPTC 的 NewsML 與 NITF 最為重要，應用最廣泛，整合性較佳。因此以下僅針對 NITF 與 NewsML 進行更進一步的說明，其他的標示語言則留待日後另以專文探討。

NITF與NewsML雖然同屬詮釋新聞的標示語言，但其作用並不相同。NITF 著重在單篇新聞的描述，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。NewsML的概念緣自於以詮釋資料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以NITF或其他標示語言嵌入其中，因此較適用大型的通訊社。由於此二種標示語言同在IPTC下發展，故可以互通或相互引用，彼此截長補短，其協作方式可參閱「NITF/NewsML Cooperation」³⁴。

一、NITF 結構概說

NITF 目前已發展至 3.2 版，這套標示規格主要用於傳統平面媒體的標示，設計的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透過 NITF 標示規格將新聞稿件轉換成爲不同的資料格式，並呈現在各種新聞平台之上，而達成多重發行的目的。NITF 語法遵循 XML 標準規格。NITF 特徵如下：

- ◆ NITF 是描述單一新聞媒體類型，是單一物件(single item)的詮釋，設計結構針對獨立的單篇新聞 (independent news articles)。
- ◆ NITF 對新聞的詮釋沒有邊欄的概念，對同一份新聞文件不同格式無二選一的功能。
- ◆ NITF 可讓出版者在新聞文件中標示出組織或人名等。
- ◆ NITF可以很清楚的描述出新聞物件的五個W。³⁵
- ◆ NITF 可描述新聞內容裡的表格、表單、協助出版者從短文或段落中區分出標題，亦可描述多媒體文檔格式。
- ◆ NITF 內容與呈現方式分離。NITF 標示新聞之後，新聞的資料內容與呈現即不相屬，標籤只描述新聞資料的內容與語意結構，而不處理新聞資料的呈現方式。因此編輯人員可以專注於資料處理，毋需重新關注格式

³³ 同註 27。

³⁴ IPTC, "NITF / NewsML Cooperation", available at <<http://www.nitf.org/nitf-newsml.php>> (2004/10/12)

³⁵ 5W指的是何人 (Who)、爲何 (Why)、何時 (When)、何地 (Where)、何物 (What)，乃構成任何新聞事件之要素。

問題，呈現工作則交由美工人員。

- ◆ 採用NITF標示新聞，使用者透過詮釋資料的描述，可提高檢索時的準確度。³⁶

目前使用NITF的著名機構有法新社（AFP）³⁷、美聯社（AP）、紐約時報（The New York Time）、義大利的ANSA通訊社等等。

NITF 的根元素為<nitf>，其下的子元素主要分為<head>與<body>兩部分：<head>是放新聞管理的 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，他可參考到其他專門領域的網址，<body.end>是描述新聞的結尾，包括了一篇文章的結尾註錄或書目資料。NITF 主要結構說明如下圖。

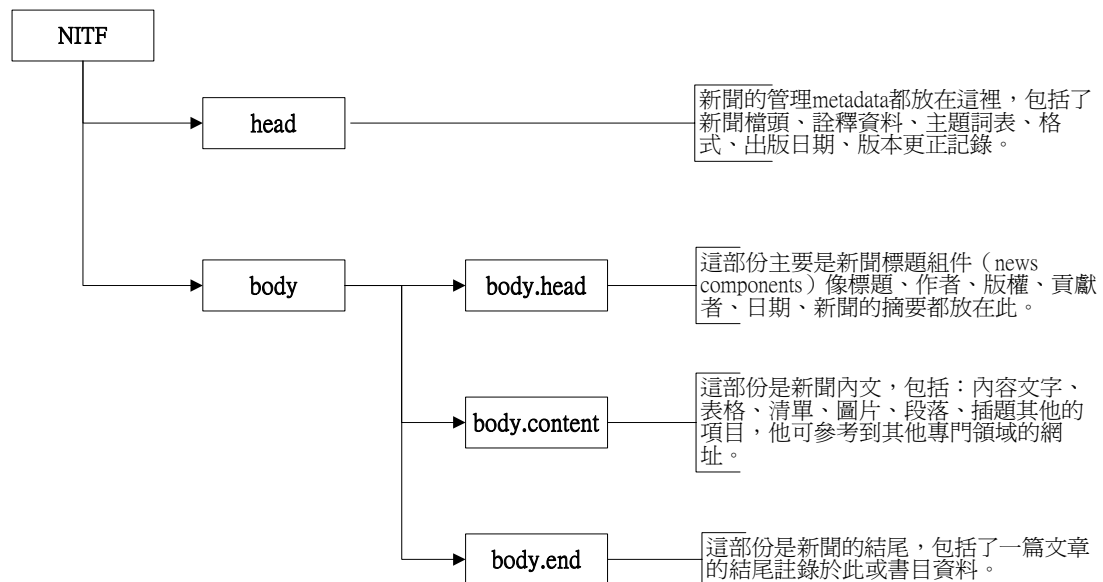


圖 2 NITF 主要結構圖

（資料來源：本研究整理）

二、NewsML 結構概說

NewsML 主要功用是新聞交換，像是通訊社發稿給報紙或電視台，或媒體之間的新聞傳遞，並可做為新聞的儲存格式，甚至用在製作、編輯、管理、出版上。因為其 XML 的邏輯結構，可與任何實體內容結合，而且是「媒體中立

³⁶ 陳百齡，〈讓電腦也能辨識新聞：新聞內容標示問題初探〉，《中華傳播學會 2002 年會論文資料》，取材自〈<http://ccs.nccu.edu.tw/oldccs/con2002/conworks/11B-1.doc>〉，民國 93 年 4 月 20 日。

³⁷ 法新社 2000 年 7 月公布正式使用 NewsML 來交換新聞。

(media-independent)」，對新聞內容的格式不預設立場，所以 NewsML 可以用在文字、圖片、語音、影視，以及未來發展的媒體型態。NewsML 特徵如下：

- ◆ NewsML 以目錄 (Catalog) 和主題集 (Topic Set) 進行控制詞彙 (Control Vocabulary) 的定義與引用，可以有效的管理整個標示語言的詞彙。
- ◆ NewsML 有新聞封包 (News Envelope) 可用來記錄新聞傳送過程中所需的各種訊息。
- ◆ 以描述的內容類型而言，NewsML 可以分成文字、結構特別的資料、與圖片三部份。文字採用原有的 NITF 標準及其 XML DTD；圖片影像則採用 JPEG 規格；結構特別的資料則另行設計，如體育新聞的結構就有 Sports ML1.0 版。
- ◆ NewsML 採用新聞組件 (News Component) 的概念，描述所包含的所有新聞物件在結構上之相互關係。
- ◆ NewsML 在同一件新聞主題裡支援多種媒體類型、多語言、多格式；同一個部分內，可選擇不同的呈現方式。
- ◆ NewsML 可詳盡包含所有新聞內容，或者透過參考，選擇包含部分內容。

NewsML 其下的子元素主要分為 <Catalog>、<TopicSet>、<NewsEnvelope>、<NewsItem> 四個部分；說明如下：<Catalog> 是 IPTC 所有定義控制詞彙集結於此的目錄，提供 NewsML 所使用不同詞彙的定義目錄區，包括了 TopicUse 和 Resource。<TopicSet> 是 Topic 的合集，ITPC 提供不同主題的詞彙控制表，可以不斷的增修，提供 NewsML 所使用詞彙的定義。<NewsEnvelope> 傳輸新聞的 Metadata 在此，包括了傳輸 ID、時間、對象、等級、提供者等等。<NewsItem> 分別包括了 <NewsManagement>、<NewsComponent>，這兩個子元素描述有關 <NewsItem> 的相關管理資訊、新聞權限、行政管理、新聞內容的 Metadata。

NewsML 結構如下圖所示。

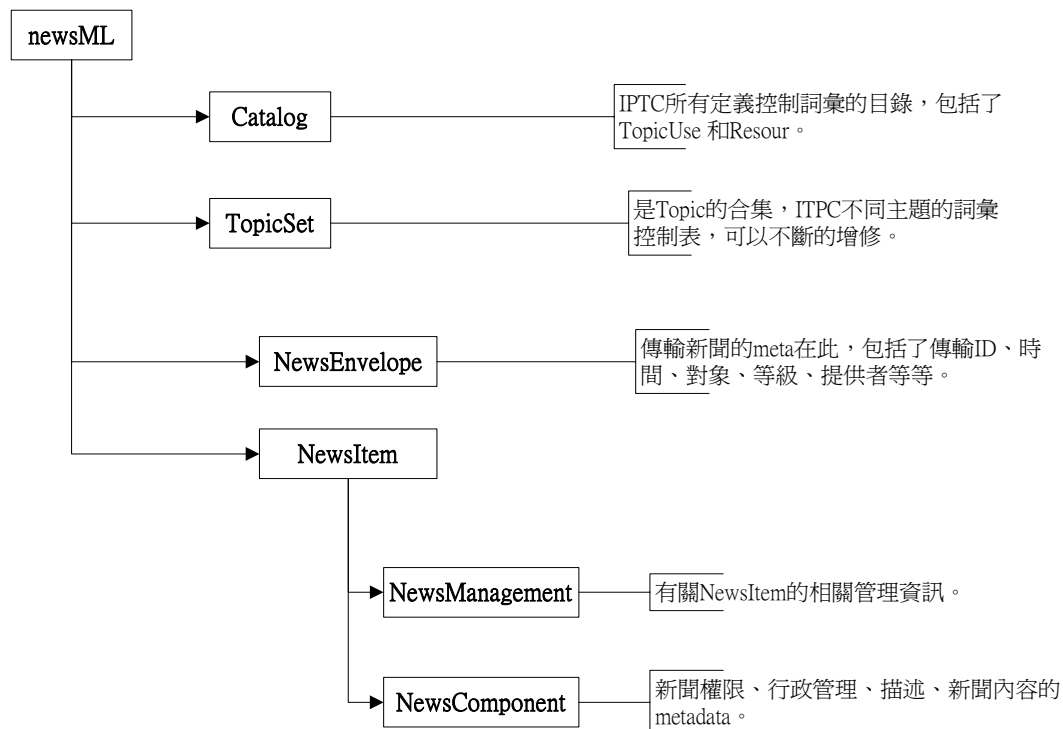


圖 3 NewsML 結構圖

（資料來源：本研究整理）

肆、中文新聞標示之發展

數位化新聞的資訊組織，不只國際上的標示語言格式眾多，中文化的標示規格亦相當多元，接著，我們便就兩岸三地的發展現況進行探討。

一、台灣方面

首先，我們分別從官、學、產三方面來瞭解台灣在數位新聞資訊組織方面的發展概況。

（一）NRMF

行政院文建會所頒訂的NRMF，是一個適合中文環境使用的新聞Metadata格式。³⁸NRMF主要是以都柏林核心集（Dublin Core Metadata Element Set，簡稱DC）³⁹的 15 個核心元素為基礎架構，再以子元素（Subelements）延伸出的欄位來著錄更詳盡的新聞資料。由於DC具有簡易、明確、彈性等特色，因此，NRMF也具備了相類似的性質。採用NRMF的優勢是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。NRMF目前的最新版本是 1.2 版，於 2004/5/24 修訂頒佈（前一版 1.1-1 版頒佈日期是 2003/12/18）。⁴⁰由於NRMF的設

³⁸ 同註 10。

³⁹ Dublin Core Metadata Initiative (DCMI), "Dublin Core Metadata Element Set, Version 1.1: Reference Description", 2003-06-02, available at <<http://dublincore.org/documents/dces/>>, 2004/7/18.

⁴⁰ 同註 10。

計是以DC為基準所擴展，因此也具備了DC的簡易、明確、彈性等特色。

NRMF 以單篇新聞的 Metadata 描述為重心，並未像 NITF 或 NewsML 有主題詞表、控制詞彙等嚴謹詞彙格式的規範，也未對新聞傳遞的封包描述特別著墨，是屬簡易型的新聞標示語言規格。

歸納 NRMF 的特徵如下：

- ◆ NRMF 是以 DC 做根據的新聞標示語言。
- ◆ 詮釋格式簡單，小規模業者很容易進入情況作業。
- ◆ 偏重多媒體的數位新聞描述。
- ◆ 對傳統報紙新聞描述的欄位不夠周全，例：版名、版次未有欄位可著錄。
- ◆ 新聞加值不易，不能將相關新聞彼此相結合成系列。

(二) 學界的發展

台灣學界對新聞標示的研究也不少，如謝瀛春教授等人共同發表的「新聞內容的標誌-XML之應用」⁴¹。該研究依據新聞學及新聞寫作的相關原理，運用XML標記新聞內容，並以純淨新聞體裁之科學新聞為樣本，進行標誌工作。在新聞事件(event)的陳述上，以人(who)、事(what)、時(when)、地(when)、如何(how)、為何(why)六大方向來描述，將新聞內容以此為表達的重點，讓使用者清楚、明白發生的新聞事件始末。林信成等人則於2002年發表了「基於XML之新聞管理與出版系統設計」一文⁴²，以電子新聞的管理與出版為例，藉由自訂的Metadata格式，以XML語法進行實驗性新聞資料庫之全文標誌，並自行設計一套新聞管理與出版系統，實際在Web環境中整合XML技術，探討與印證XML在電子文件管理與出版方面的優勢。此系統以XML為基礎，相較於傳統的資料處理模式來說，有著更彈性與更易加值處理的特點！再者，藉由XML優越的結構化與自我描述性，使得數位化新聞的「智慧化」程度得以提升，進而增進資訊檢索之精確度。

在國科會推動的「數位典藏國家型科技計畫」方面，今年度(2004年)正在執行的計畫中，屬於「新聞主題小組」⁴³的有兩件，分別是：(1)世新大學莊道明教授主持的「北平『世界日報』內容數位化開發計畫」；(2)淡江大學本文作者林信成所主持的「臺灣棒球運動珍貴新聞檔案數位資料館之建置」計畫。其中，「北平『世界日報』內容數位化開發計畫」因該報為早期之舊式報紙，在內容、版面、廣告等各方面之編排皆迥異於今日之報紙，以致於目前現有的Metadata格式不盡然適用，因此，該計畫預計制定一份有關舊式報紙的標示語言，目前正在研擬中。而另一計畫案「臺灣棒球運動珍貴新聞檔案數位資料館之建置」，則初步先採用與DC相容的NRMF作為Metadata格式，以便能儘速與同採DC格式的

⁴¹ 謝瀛春、黃學碩、維習安、雷約翰、謝清俊，「新聞內容的標誌-XML之應用」，海峽兩岸資料庫/數據庫與資訊/信息服務交流與合作論文集，頁205-212，民90年1月。

⁴² 林信成、陳勇任、楊翔淳，「基於XML之新聞管理與出版系統設計」，2002出版與圖書館研討會，台北淡水，頁14-29，民91年4月26日。

⁴³ 「數位典藏國家型科技計畫」依各計畫性質劃分為十二個主題小組，其中之一的「新聞主題小組」是以報紙、新聞影音、期刊、地方文獻等為主，詳見 <<http://www.ndap.org.tw/>>。

數位典藏聯合目錄接軌，未來擬再陸續加入與其他Metadata規格之設計，如NewsML、NITF或SportsML等，以提升其資料交換與傳遞能力。

(三) UdnML⁴⁴

台灣在新聞業界方面，因各報社之相關資料大都未公開，難以得知發展現況。本文作者之一因任職於聯合報，得以參與目前聯合報系成立的專案小組，專為旗下相關的五個不同型態的報紙設計一套業界適用的 UdnML，因此以下針對 UdnML 說明其概況。UdnML 小組的發展方針，是以國際標準的 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，將報系內編務管理的 Metadata 考慮在內，同時擬打破傳統由資訊管理員負責資料加值業務之作法，改由線上發稿記者、審稿編輯來進行資料加值作業，這項業務對系統人員、技術人員、採編人員均是一項挑戰。

UdnML 目前處於工作草案階段，尚未正式發表，預計於 2005 年春天制定完成。UdnML 的目標不僅是聯合報系內部使用，同時亦希望與全球中文的新聞標示規格能接軌，相互交換傳遞。

二、大陸方面

中國大陸目前有兩單位制定新聞標示語言，一是新華社所發展的 XinhuaML⁴⁵，另一是中國報業協會所制訂的CNTF⁴⁶。

(一) XinhuaML

大陸於 2002 年，由「新華社技術標準工作組」與「北大方正公司」合作研究，獲中共國家核准項目的共同成果。以新華社為實例發展出 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、儲存而制定的一套標準標示語言。在系統的設計上，借用了 NewsML 在通訊社的優勢，所以不到兩年時間，XinhuaML 即發展完成。

新華社基本的作業流程，是把外稿藉 XinhuaML 轉為標準 XML 格式，存放在龐大的「未審稿資料庫」，然後各個部門的編輯從中選稿，編輯、翻譯之後存放在自己的資料庫對外發佈，再移到典藏資料庫。同時因應時勢所趨，也將廣告標準規格放入。⁴⁷ 新華社是兩岸目前較廣泛使用標示語言的新聞機構，但即使如此，XinhuaML 的使用也僅限於社內，對外發佈新聞並未使用這一規格，因為外界尚無相對應的系統可以接收、解讀。所以專案負責單位希望兩岸三地都能共同努力，早日制定適用於全球中文新聞媒體的標準，同時也希望所訂出的標準

⁴⁴ 聯合報系於 93 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。

⁴⁵ 新華社技術局標準工組，XinhuaMLv1.0 功能說明書，2003 年 1 月 18 日。

⁴⁶ 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。

⁴⁷ 那福忠，那福忠訪新華社工作報告，2004 年 9 月 27 日。（聯合報系內部資料）

能與國際新聞媒體接軌。不過，新華社所屬的新聞網站「新華網」則於 2004 年 8 月開始提供RSS聚合新聞，是大陸大型新聞媒體業中最早提供此一服務者。⁴⁸使用者可使用RSS新聞聚合器直接上新華網訂閱各類聚合新聞。⁴⁹

綜合歸納XinhuaML主要特徵如下⁵⁰：

- ◆ XinhuaML 是獨立於媒體的，對媒體類型、格式或新聞對象的編碼可以通過控制詞表來規定和擴充。XinhuaML 的文件也可以包括內文、圖像 ... 等多媒體的聯合體。
- ◆ XinhuaML 對新聞的傳遞、管理、內容描述都有詳盡的規劃。

XinhuaML 元素是整個 XinhuaML 文檔的根元素，要讓新聞聞檔有效，可以包含一個「包」、一個或多個稿件、有零或多個主題詞表元素，這些元素包含 XinhuaML 文檔本身引用的主題詞表。

XinhuaML 主要元素大致分了七個子元素：<目錄>包括了資源元素和詞條使用元素，任何一個主要結構元素都可以包含一個目錄元素。<主題詞表> 包含主題詞元素，這些元素是指現實世界的事件。同時包含 XinhuaML 文檔中的某些新聞內容或詮釋資料的事物。<包> 元素含有關在一種業務工作流，或在新聞提供者和收受者之間的合約關係裡如何使用 XinhuaML 的文檔信息。包括了傳輸 ID、時間、對象、等級、提供者等等。<稿件>含稿件標識號、非正式標識號。<管理信息> 包括稿件類型、建稿時間、修訂時間、狀態、重要性等。<組件> 包括新聞稿作用圖解、等價列表、組件子元素。<內容項>為新聞實體內容的正文、圖像、視頻、音頻等媒體類型與格式。XinhuaML 結構如下圖所示。

⁴⁸ 同註 18。

⁴⁹ 新華網，”RSS 聚合新聞服務 - RSS 閱讀器下載 - 新華網 RSS 訂閱”，可得自 <<http://zh.wikipedia.org/wiki/RSS>> (2004/10/12)。

⁵⁰ 同註 45。



圖 4 XinhuaML 結構圖

(資料來源：本研究整理)

(二) CNTF

中國大陸另一個單位中國報業協會規範工作組則依據 NITF 3.0 版發展了 CNTF，於 2002 年 5 月正式公佈。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。

CNTF 主要元素分為<首部>與<主體>兩部分：<首部>元素包括文檔標題、變元、類型物件、iim、文檔資料、出版資料、修訂歷史記錄。<主體>元素包括又分：<主體首部>包括標題、注解、權利、首部署名、發行機構、日期、摘要、系列說明；<主體內容>是新聞主體的塊、塊內容。包括新聞段、表格、媒體、清單等等。<主體尾部>這部份是文稿的結尾署名、參考書目。

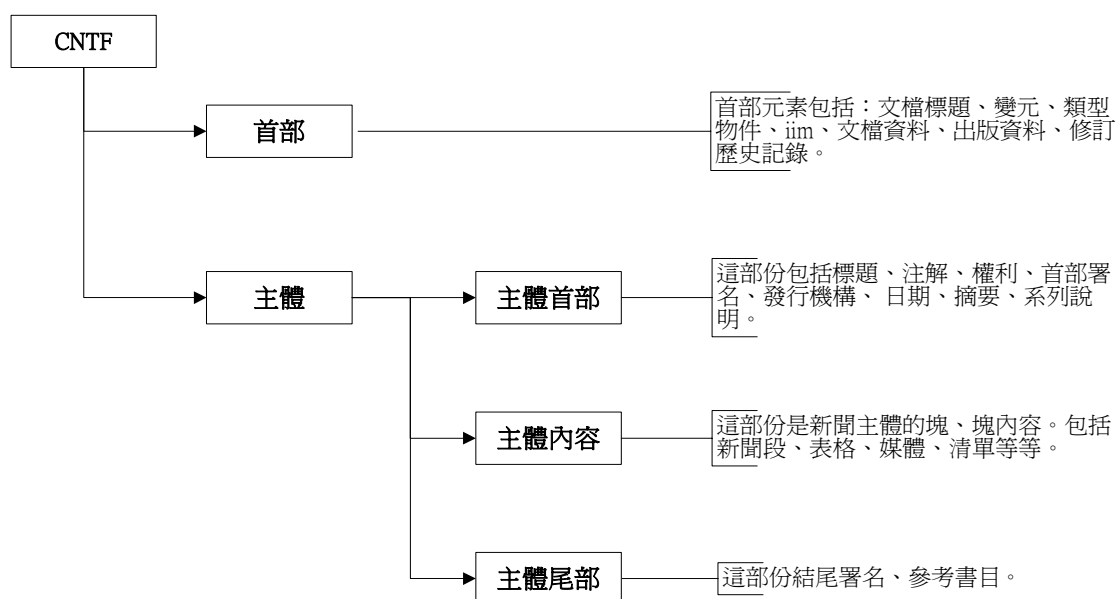


圖 5 CNTF 結構圖

(資料來源：本研究整理)

不過，目前 CNTF 這套規範的應用未若新華社 XinhuaML 廣泛，因而未再繼續發展下去。

三、香港方面

香港在中文新聞標示方面的研究與推廣亦不遺餘力，主要以香港中文大學設立的「中文新聞標示語言協會」⁵¹為主，該協會於 2002 年由香港特別行政區創新科技署贊助成立，並加入 IPTC 成為會員，亦是該會的唯一華文地區成員，致力於建立及推廣香港本地的新聞標示語言標準，開發支援軟體來加速香港電子商務的步伐及增強香港新聞媒體及出版業在國際上的競爭力，促進大中華區新聞媒體及出版業電子商務。

目前，中文新聞標示語言協會已開發了一套遵循 NewsML 標準的網上文檔處理及發佈系統 CIDAX⁵²，主要提供三方面的服務，包括：(1) 資訊編輯 — 使用者可製作以 NewsML 記錄的內容封包，包含文字及多媒體；(2) 檔案管理 — 儲存及分享檔案；(3) 資訊瀏覽 — 檢索、瀏覽及選取所需的內容封包，匯出至不同模式的媒體檔案，以便發佈及在不同媒介使用。同時，亦釋出 Java 開放源碼程式庫 CIDAX OpenLib，以協助程式開發人員利用 NewsML 處理多媒體、多語言的資訊內容。程式庫設定了一個表達包含多種語言及多媒體新聞內容的結構。而

⁵¹ 中文新聞標示語言協會，”中文新聞標示語言協會首頁”，可得自 <<http://cnewsml.org/>> (2004/10/14)。

⁵² 中文新聞標示語言協會，”應用演示 CIDAX 企業資訊數碼系統”，可得自 <<http://cnewsml.org/tradchi/cidax.shtml>> (2004/10/14)。

其中更特別集中討論 NewsComponent 的建立方法。⁵³

伍、結論

新聞標示語言之長期效益是以新聞機構之間資料交換為終極目標，但先決條件是所有新聞機構必須採用同一套標準。因此，在標準未全面普及前，短期內部使用所得的效益，將大於對外的功能，如法新社、共同社、新華社都先在機構內部使用獲益，而後才逐漸使其對外發佈新聞更加順暢。以法新社為例，從 2002 年的 100 多個 NewsML 顧客到今年的 300 多 NewsML 顧客，每天以六國語文發出七千則新聞，每則新聞以交換、管理、結構、內容四層的 NewsML 規格發出，文字稿用 NITF 規格、多樣媒體則包裹在一起，其中包含詮釋資料、自動控制詞彙、與文圖關連，並且使得每一種新聞產品規格統一。⁵⁴新華社自己也有傳播媒體，包括二十多家報紙，分佈在大陸各地。XinhuaML 最大的作用也是先利用此一標準標示語言從事內部新聞傳遞，對外交換尚未採用 XinhuaML，而是先以較簡易的 RSS 作線上聚合新聞的發佈，使個人用戶可以直接線上訂閱。新華社目前主導、規劃中的「全球華文新聞媒體標示語言標準」制定仍在積極進行，可能於 2004 年底完成初稿，然後到大陸各大媒體訪談、徵求意見，亦希望台港與海外的媒體參與，共同制定一個可以全球華人通用的中文標示語言，使中文新聞的通用交換格式早日定案。新華社的經驗值得台灣通訊社及媒體業者借鏡。

「新聞」是來幫助大眾觀察、了解、連接更寬廣的世界，網路的特殊力量，能讓我們達成此一任務。傳統報業必須改變，既無選擇、也不能回頭。新媒體則大家都參與，隨時更新內容，提供大部分的材料，成為一個豐富、多樣媒體。如此高度結構化的資料庫，更需要一套標準的標示語言供不同載體的媒介交換資訊。英文的新聞標示語言早在 1965 年即發展，至今已全球通用。華文人口眾多，但華文的中文新聞標示語言尚在起萌，實需產、官、學各方面齊頭並進努力。

誌謝

承蒙國家科學委員會經費補助（計畫編號：NSC 93-2422-H-032-001），使本研究得以順利進行，特此致謝。

⁵³ 中文新聞標示語言協會，「應用演示 CIDAX OpenLib」，可得自 <http://cnewsml.org/tradchi/cidax_openlib.shtml> (2004/10/14)。

⁵⁴ 同註 9。