Using XML Vocabularies to Exploit Changing Business Models: The NewsML Experience

Martin Fahy  
*National University of Ireland*, martin.fahy@nuigalway.ie

Joseph Feller  
*University College Cork*, jfeller@afis.ucc.ie

Patrick Finnegan  
*University College Cork*, p.finnegan@ucc.ie

Ciaran Murphy  
*University College Cork*, cm@ucc.ie

Follow this and additional works at: [http://aisel.aisnet.org/ecis2003](http://aisel.aisnet.org/ecis2003)

**Recommended Citation**
[http://aisel.aisnet.org/ecis2003/30](http://aisel.aisnet.org/ecis2003/30)

This material is brought to you by the European Conference on Information Systems (ECIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ECIS 2003 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
Using XML Vocabularies To Exploit Changing Business Models: The NewsML Experience

Martin Fahy
National University of Ireland, Galway, Ireland
Martin.Fahy@nuigalway.ie

Joseph Feller
University College Cork, Ireland
JFeller@afis.ucc.ie

Patrick Finnegan
University College Cork, Ireland
P.Finnegan@ucc.ie

Ciaran Murphy
University College Cork, Ireland
CM@ucc.ie

Abstract
In 1998, Reuters committed resources to exploring the use of XML (Extensible Markup Language) as the foundation for their next-generation news content management system. This was in response to changes in the news industry business model resulting from changing value webs and content delivery methods. The result was ReutersNewsML, an XML vocabulary for “wrapping” news content items from multiple sources into a single package. In 1999 Reuters submitted the concept to the International Press Telecommunications Council (IPTC), which managed the development of the NewsML specification. NewsML was ratified as an IPTC standard for the entire global news media industry in 2000. This paper explores the development of NewsML as an illustration of how XML standards can be exploited within complex business webs.

Keywords
New Business Models, XML, Case Study.

1. Introduction: Wind, Snow and Meta-Data
In February, 2002, at the Salt Lake City Winter Olympic Games, while the athletes on the ski slopes and in the skating arenas strove to make history, fifty-six Reuters employees and thirty-five third-party photographers in twelve different locations were breaking records of their own. Despite the challenges of weather and geography, the group produced approximately 1000 photos and hundreds of pages of text a day. They sent the information – quickly and wirelessly – from their digital cameras and laptops to the Reuters media centres, which then sent it on to Reuters’ worldwide client-base – publishers, broadcasters, and Internet content-aggregators, among others. They were able to do this, in no small part, thanks to a special data format, called News Mark-up Language (NewsML™).
NewsML has been called the “news wrapper,” and it is a fitting label. Essentially, NewsML is a tool for packaging multimedia news data in a way that makes it easier to transport, process and deliver – whether the destination is a publisher’s database or a consumer’s mobile phone. NewsML data packages come “wrapped” in rich descriptions about their contents (called meta-data), thus ensuring customers receive the most timely and relevant news, and in the appropriate format and language.

This paper presents the results of case study based research on the development of the XML vocabulary, NewsML. The paper begins by presenting the theoretical underpinnings of the study and the research method. The NewsML case study is presented in detail, followed by the findings from the research. The findings show that while technological considerations were an important factor in the success of NewsML, wider business model and value chain considerations within the industry offered powerful incentives for cooperation between the value chain participants. The paper concludes by providing recommendations for the deployment of other XML vocabularies where fundamental changes in business models are forcing firms to address new competitive dynamics.

2. Inter-organisational Systems and New Business Models

A business model is comprised of components (e.g. customer scope, product/service scope, customer value, pricing, revenue sources etc), linkages between activities, and dynamics (how a firm reacts to or initiates change to attain a new competitive advantage, or to sustain an existing one) (Afuah and Tucci, 2001). New development in IT and electronic business pose a significant threat to traditional business models. Such developments affect both how organisations relate to external parties (customers, suppliers, partners, competitors, and markets), and how they operate internally in managing activities, processes, and systems (Rayport and Jaworski, 2001).

Timmers (1999) sees new business models emerging from linkages between elements of the value chains of different organisations. Inter-organisational electronic trading systems are considered to be early examples of this as they represent linkages between purchasing and sales in independent organisations. More recent innovations focus on a broader range of value chain activities such as R&D, marketing etc.

Early inter-organisational systems (IOS) were based on Electronic Data Interchange (EDI) standards. Such systems were seen to positively affect inter-organisational transactions by reducing costs and improving efficiency (Swatman and Swatman, 1992). In addition, they reduced environmental uncertainty by facilitating communication and providing information (Henderson, 1990; Scala and McGrath, 1993). Finally, they affected the competitive positioning of organisations by encouraging closer relations with suppliers and customers (Cash and Konsynski, 1985; Swatman and Swatman, 1992; West, 1994). Early IOS tended to be developed internally and extended to, or imposed upon, others (Webster, 1995). As such system configurations become more complex, an approach where one organisation develops a system and simply extends it to other organisations will be inadequate, especially when business processes have to be altered (Finnegan et al., 1998).

XML (eXtensible Markup Language) is seen as having huge potential in facilitating the exchange of inter-organisational information in business-to-business environments (Lowry and Neumann 2001). XML has been identified as a disruptive technology in at least three key areas; EDI, Enterprise Application Integration (EAI) and content management (particularly
in heterogenous client environments) (Goldfarb and Prescod, 2001). XML requires organisations to develop specific vocabularies in order to exchange information. Such vocabularies tend to be industry specific, and their development must deal with complex inter-organisational dynamics (Nunes, 2000). Complexities that arise when activities cross organisational boundaries include: changes in business competencies and priorities; changes in ones’ sense of identity and sense of common purpose; an increased need for co-operative effort for establishing targets; the synchronisation of differing accounting, measuring and reward systems; and the establishment of trust (Ashkenas et al., 1995).

Much of the research on inter-organisational systems and business models has concentrated on well-established technology that has been implemented in simple supply chains (e.g. retailer and supplier). More sophisticated technology such as XML is proposed as a solution to the communication and collaboration needs of more complex business webs. However, for XML to be useful, an XML vocabulary must be developed. Nevertheless, the development of XML vocabularies to support new inter-organisational business models has received little attention in the research literature. It is thus important to explore the development and exploitation of XML vocabularies in complex value webs as a first step in understanding the challenges facing adopters.

3. Research Objective and Method

This study analyses the development of an XML vocabulary in a dynamic inter-organisational setting. The development of NewsML within the news industry was chosen as an appropriate focus for the following reasons:

1. NewsML is used to exchange content with organisations, integrate internal systems, and manage content delivery (key XML application spaces),
2. The news industry relies on the interchange of complex data between a myriad of players (O’Brien et al., 2001),
3. The news industry is heavily reliant on IT (Butler and Murphy, 1999), and
4. Data exchange within the industry is greatly affected by developments in electronic publishing e.g. mobile and Internet devices (Clarke, 1997).

Corbitt (2000) advocates the need for interpretative methods in studying IS issues, especially in electronic business environments. Interpretative studies focus on developing a greater understanding of social aspects of the research environment (Walsham, 1993), and are thus considered useful in the context of this study. Case studies are regarded as the most commonly used qualitative research method in IS, and are especially useful for studying organisational aspects of IS (Benbasat et al., 1987). Cases are most appropriate when the objective involves studying contemporary events, without the need to control variables or subject behaviour (Yin, 1994). The single case study method is considered to be a potentially rich and valuable source of data, while suited to exploring relationships between variables in their given context (Yin, 1994; Benbasat et al., 1987) and is appropriate where it represents a critical case (Yin, 1994).

Given the exploratory nature of this research and the need to obtain rich data in complex inter-organisational contexts, a case study approach is adopted. ‘A case study examines a phenomenon in its natural setting, employing multiple data collection methods to gather information from a few entities. The boundaries of the phenomenon are not clearly evident
at the outset of the research and no experimental control or manipulation is used’ (Benbasat et al., 1987).

The subject of the case study was chosen as it represented a critical case in the news industry. The researchers first conducted a thorough archival search to determine the existence of public domain material on Reuters and the development of NewsML. The accuracy of this material was then verified with key staff members at Reuters. This verification was conducted using conference calls and document exchange by email. As a result of this preliminary analysis, the researchers prepared a case study protocol as defined by Yin (1994), and sent it to the contact person at Reuters. Based on this protocol, interviews were arranged with key personnel at Reuters. These interviews were conducted during a site visit that took place in September 2002. In addition to interviews, the researchers were also given access to relevant documentation and system demonstrations. The site visit was followed by several conference calls with the contact person in order to clarify issues and confirm the accuracy of the data collected.

4. Reuters

Reuters was founded in 1851, and has grown into one of the world's largest and most successful news and financial information organisations. The company has over 16,000 employees in 220 cities in 94 countries, declared GBP3.6 billion in revenues for 2002, and services nearly 500,000 professional users in 52,900 client locations. Reuters provides data on more than 960,000 shares, bonds and other financial instruments, maintains more than 200 million data records (which contain over 3,000 billion record fields), and is read by users of more than 900 web sites around the world, with an audience reaching 50 million online users per month.

ii

 Reuters has a long history of technological and business innovation. Reuters’ innovations were generally built on proprietary technologies, and required suppliers and clients to use specialised tools to access Reuters’ news services. While this proprietary approach to technological innovation has proved very successful for Reuters, it also resulted in what the company has termed a ‘multiple feed – multiple boxes’ architecture for sending and receiving data. By the mid 1990s, it was clear that the number and diversity of content providers and clients was growing at a rate that made proprietary technology cumbersome for both acquiring and delivering content. In rethinking the ‘multiple feed – multiple boxes’ approach, Reuters’ goal was internal integration, so that they could combine the feeds from multiple suppliers and deliver content to clients in a more efficient manner. To achieve this objective, Reuters first began experimenting with Internet technologies in 1995.

Early efforts to “clean up” the company’s network infrastructure led to the development of an extranet solution. Taking the next step in its company-wide Internet strategy, Reuters then turned their attention to XML. Reuters believed that XML could be used to unify all of the various systems and processes operating throughout their organisation. Reuters began the development ReutersNewsML in 1998. In 1999, the specifications for the vocabulary were submitted to the International Press Telecommunications Council (IPTC). The IPTC then began the collaborative development of NewsML. NewsML has become the standard data format for multimedia content in Reuters Internet Delivery System (IDS) and was ratified as an IPTC standard for the entire global news media industry in 2000.iii
5. The International Press Telecommunications Council

The brief of the International Press Telecommunications Council (IPTC), established in 1965, is “to safeguard the telecommunications interests of the World's Press.” In practice, IPTC activities have been primarily focused on the development and dissemination of open industry standards for the exchange of news media data. IPTC membership includes the world’s major news media agencies, as well as smaller publishers, system vendors and digital media organisations. Membership of the IPTC is open to any organisation concerned with the collection, distribution and publishing of news media.

The IPTC has been developing standards for the interchange of news material since the late 1970s. IPTC standards include semantic frameworks for describing subject matter (the IPTC Subject Reference System), an XML based language for marking up text in news articles (the News Industry Text Format, or NITF), and low-level standards to facilitate the exchange of images and other multimedia objects. Although early work from the IPTC was based on a variety of technologies, the IPTC is committed to base all its future standards activity on XML. The IPTC standards developed prior to NewsML were limited to facilitating the exchange of single media objects, and furthermore lacked the flexibility built into XML-based data formats type to be exchanged in the containing ‘envelope’

6. How NewsML works

NewsML is an example of an XML vocabulary. NewsML leverages the strengths of XML to create a platform independent framework for packaging news media objects. It can be used by news providers to combine photos, audio, video, graphics and text in a form that can be exchanged between value chain participants and eventually rendered in hard or soft copy, through mobile phones, desktop computers, interactive television, handheld computers or other devices.

Figure 1 illustrates how Reuters’ publishing system combines various media formats into the standard NewsML format, which can then be distributed to various output devices. It is important to note that NewsML acts as a packaging mechanism for content of any media, including documents formatted using other XML vocabularies. NewsML itself does not address the formatting of individual content items. It provides a common structure, which enables news providers like Reuters to combine content items as they wish with very little effort. As such it helps publishers to create packages of information targeted at specific audiences and helps make the news personalised to the end user.

A NewsML package contains a number of components, or elements, that are used to structure the actual news content. These elements include metadata, which provide information about each News Item – such as physical characteristics and contextual data like author, publisher or owner. Each news item in a NewsML package is a discrete and manageable piece of news made up of several news media objects (the individual files that are combined together to make up the news item). For example, as shown in figure 2, a piece of text would be an object, and an accompanying picture would be another object. Each object contains or references a file (e.g. sound, text, image) and all of the metadata relevant to that file.
NewsML packages map out relationships between objects, such as equivalents and complements. As the name suggests, equivalents are alternative representations of the same news object. For example, they could be translations of the same piece of text into different languages or alternative resolutions or formats of the same image. Complements are descriptions of the functional relationships between related news objects. They complement one-another and should be used together to provide a complete news story. Typically this might involve text and an associated picture. NewsML thus has the flexibility to build up stories with multiple elements and then handle them as a single entity. For example, a single piece of news content could be translated from English into both French and Italian with each translation being carried in a separate object. The three objects would then be put together in a single package as equivalents. A photograph of the event described by the text would be included as another object.
7. Case Study Findings
A detailed analysis of the NewsML case study highlights some key findings for standard setters and those considering adopting XML vocabularies as a means of improving integration across changing business value chains.

7.1 Addressing the Needs of a Changing Industry Dynamic
One of the features of many technological innovations is the dearth of real world problems to which the fledgling technology can be applied. In the case of NewsML, genuine structural changes in the industry and its value chain created a pressing need for an integration standard. The case study findings indicate that Reuters sought to find a technological solution to meet their changing business requirements. According to Julien Dossier, Manager of the NewsML Program for Reuters, the company undertook the initial development of ReutersNewsML, and made a commitment to the development and use of the IPTC NewsML standard, in order to address three industry changes:

1. Changes within the news media market
2. Technological changes in the delivery and consumption of digital content, and
3. The need to address the enormous logistical challenges presented by Reuters’ media operations.

7.2 Supporting Emerging Business Models
Increasingly, news media organisations have found that they need to move from simply managing a physical network that specialised in distribution, to also facilitating a community of journalists and users. This community network is regarded as being circular rather than linear as consumers are often also producers. A key driver behind the collaborative development of NewsML was the increasing complexity of the media value web. This complexity takes many forms, for example:

- Proliferation of content providers: an increasing number of organisations whose core competencies lie elsewhere (e.g. banks) have begun to produce and publish content on the Internet (e.g. broker research).

- Multi-point positioning in the value chain: Organisations like Reuters find themselves positioned in a number of different points along the value chain, providing content to other publishers, content aggregators and redistributors as well as corporate and end-user consumers. A related factor was the need to offer increasingly customised services to the consumer market, and to address the related challenges created by the proliferation of new media devices. Dossier has noted that increasing connectivity has meant that there has been an increase in the number and variety of consumers, and of the modes in which they chose to consume content.

- Community-driven value: Organisations like Reuters rely on effective collaboration both within and between their community of journalists and community of users.

7.3 Supporting Inter-Organisational Business Processes
Many of the applications of XML vocabularies to date have focused on supporting systems integration and information exchange. A unique feature of NewsML was its support for the
business processes, which were a key element of the value chain. According to Julien Dossier, to understand the drivers behind the creation of NewsML, one must first understand the enormous quantity of data handled each day - all of which must be simultaneously consumable by both humans and computers. Each day, Reuters manages around 2500 editorial staff based in 190 locations in over 100 countries. Each day, Reuters produce and distribute about 14,000 headlines in 26 languages, over 1,000 pictures, and 23 hours of broadcast quality video.

Dossier has noted that Reuters realised that “content management is not only an issue in the packaging stage where you are combining content, but also at the content creation stage when you want to know who is doing what and when, and what has previously been done on the subject”. For Reuters, content management begins with content production using a ‘follow the sun’ approach, in which Reuters bureaus in various locations around the world collaborate over time to follow the unfolding of related news stories. The challenge continues with the internal aggregation and packaging of content, and extends further into the delivery and final presentation of media objects.

Reuters’ approach to content management required that stories be shared as they were being produced. The ‘follow the sun’ approach made it necessary to sketch out the elements of the story so that others could prepare for what was coming. Likewise, Reuters have a policy of internal peer-review. These practices are challenging when you consider that most journalists produce approximately ten stories per day, and are often required to produce items in different languages for global distribution. Reuters needed a method to facilitate multi-language collaboration in a global context. All of this needed to take place against a background where news items age rapidly, and timely production and distribution is a key value-added proposition.

The final stage in the value chain is consumption – the filtering, visualisation and rendering of content. Reuters’ aim is to create content only once and be able to reuse it in different contexts. The reuse of content for different forms of consumption is not only about different colours, and font sizes for different outlets. Colours and fonts are important for distribution across different platforms such as newspapers, the Web, and mobile phones. However, the content must also be presented slightly different in different contexts. As an example, Dossier explains that there is no need to explain to British audiences that Tony Blair is the Prime Minister. Thus a news items for the British public may begin; “Prime Minister Tony Blair wants to bring…”. However, the same piece appearing in a US publication would need to be modified slightly; “British Prime Minister Tony Blair wants to bring …”. In many cases, this altering may be a downstream activity, but Reuters needed a way to ensure that they managed content in a way that it would be easily achievable in an efficient manner.

Reuters sees content management as not simply an internal activity, but one which spans the entire value chain as shown in Figure 3. Reuters therefore needed a solution that began with content production, as it would facilitate more flexible packaging of the content. Irvine Levine, Senior Vice-President of Information Architecture, points out that Reuters’ services were traditionally “partitioned across different media types, and our textual services, video services, different media services, were for the most part sold as individual media distribution channels with no central integration on our part to provide a coherent multi-media delivery into a subscribers network.” Reuters recognized that by offering such an integrated distribution service, they would greatly increase their value offer. Such integration required a sophisticated meta-data driven framework to succeed, and Reuters saw XML as the most promising technology to address that requirement. Levine describes a sample NewsML package, which might contain “text from Reuters, an XBRL version of a company’s earning
statement, a report from an investment bank in another XML format, a video clip of the company’s CEO making a statement, photographs.”

<table>
<thead>
<tr>
<th>Supply Side</th>
<th>Internal Operations</th>
<th>Demand Side</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global coordination of Reuters journalists; Authentication of news items; Centralised collection of media objects; Aggregation of disparate sources</td>
<td>Editing and relationship building between objects; Packaging, retrieval and transmission.</td>
<td>Use and reuse of content objects; Filtering activities; format management for multi-agent, multi-lingual consumption; interaction with users to personalise the delivery of content.</td>
</tr>
</tbody>
</table>

**Figure 3. Content management across organisational boundaries.**

### 7.4 Technical Requirements for the XML Vocabulary

The technological platform required by Reuters and other news industry players needed to provide:

- **Content-neutrality**
  All formats and media types needed to be recognised equally, NewsML couldn’t make assumptions about the media type, format or encoding of news.

- **Rich relationships between media objects**
  NewsML needed to allow the construction of relationships between media objects (such as "Main Story", "Advisory", "Side bar", “Main Picture”, “Teaser”, “Picture Caption” etc). Relationships between objects represent a key intangible asset.

- **Historical sensitivity**
  NewsML needed to support the evolution of media objects over time, taking into account the need to update, add to, or replace earlier versions. A related requirement was the ability for clients to chose between versions of media objects.

- **Opportunities for cost reduction**
  NewsML content needed to support a model of “publish once, target multiple audiences and platforms.”
• **Extensible meta-data**

NewsML needed to support the creation of arbitrarily detailed meta-data for media objects. This in turn facilitates searching and filtering activities.\(^{vi}\)

### 7.5 From an Internal Solution to an Open Industry Standard

Looking at the representative drivers behind Reuters’ support of the NewsML standard, two things become clear. The first is that the challenges associated with news content management were strategic and operational, as well as technological. The second is that these challenges were not unique to Reuters, or to the internal processes of any one organisation, but were industry-wide. An industry-wide problem required an industry-wide solution, and this simple concept underpins the fact that NewsML was, necessarily, the product of a company-neutral, international organisation, rather than a unilateral effort by any one organisation.

According to Levine, “there have been various initiatives to employ XML, which have been driven unilaterally. Today’s world is perhaps a bit less amenable to having standards dictated from a single organisation.” According to Dossier, “one of the reasons [Reuters] chose the IPTC as a consortium is that is where our peers represented.” In fact, it is the ability of Reuters’ peers to use NewsML that makes the adoption of NewsML by Reuters’ suppliers and clients more likely. The value of adopting NewsML is clearer when many players in the industry use the standard.

NewsML, as an open standard, allowed Reuters to more effectively manage these processes, their relationships with other organisations and their community of journalists and users. Rather than threatening Reuters’ existing business model in any way, NewsML opened up new value creation opportunities for the company. Dossier saw it simply as “lowering the technical barrier in the exchange of information” to the benefit of both information providers and consumers.

### 8. Concluding Thoughts and Lessons Learned

The objective of the study was to explore the development and exploitation of XML vocabularies in a complex business web. The evolution of NewsML from an internal solution to an industry standard is illustrative. Organisations, like Reuters, find themselves competing in an increasingly complex technological and strategic environment. Rethinking the organisation’s value proposition means, among other things, addressing the need to leverage core intangible assets – such as the contextual data surrounding core processes and the capabilities for managing this data – to build up an offering based not only on content as a product, but on integration and flexible delivery as a service. More importantly, the NewsML experience points to the fact that in the “new economy,” organisations cannot effectively create value by remaining isolated. It is through cooperative, industry-wide platform building, not internal solutions development, that organisations can position themselves competitively.

Our analysis indicates that XML vocabularies need to be developed from the perspective of an industrial sector in order to meet requirements for new business models. This approach is likely to lead to more widespread adoption, and thus greater efficiency benefits for individual organisations. The NewsML experience illustrates the prevalence of network externalities in the development of XML vocabularies. Thus, developers and champions should be aware
that their efforts are unlikely to be a source of revenue or of comparative advantage. To exploit their investment, XML champions need to identify ways of exploiting the competencies that they acquire during the development process.

The development of XML vocabularies is likely to be an evolutionary process, with early versions evolving towards the industry standard. Thus investments by early adopters need to be regarded as R&D costs, as additional expenditure will be required to implement later standards. More mature (industry) standards are more likely to facilitate the development of new business models, but will require the adoption of new business methods. Overall, it is clear that the XML vocabularies must address genuine business needs in order to justify the required development effort and ensure the required co-operation. Finally, this case is an exploratory study, and the findings must be interpreted in that context. More research is required.

**Acknowledgements**

This research has been funded by the European Commission via IST project 2000-29665 ‘RESCUE’ (http://www.euintangibles.com), and was made possible through the generous cooperation of Reuters. This paper is presented as research and analysis by the authors, and is not to be interpreted as corporate communication by Reuters.

**References**


---

i XML is an international standard maintained by the World Wide Web Consortium (W3C) that allows organisations and entire industries to format content in a highly interchangeable format. XML adds rich metadata (data about data) to content so that it can be more effectively exchanged and used.


v Such devices include mobile phones, interactive television and handheld computers.