

NewsML

— enabling a standards-led revolution in news publishing?

Tony Allday

Reuters Media Group

NewsML has been developed and ratified as an open standard by the International Press Telecommunications Council (IPTC), for the structuring of multimedia news.

This article provides an introduction to NewsML, which is a derivative of the rapidly-spreading XML standard.

The emergence of a global standardized network for distributing information has truly vast potential for publishing. Over the last decade, the Internet's exponential growth has reflected the enthusiasm of both producers and consumers of content to embrace the medium as quickly as possible. Publishers suddenly have access to an ultra-cheap and flexible medium for delivering and displaying their wares, whilst consumers have free access, wherever they are, to an unprecedented depth and breadth of information.

Many commentators predicted a rapid industry revolution, with new advertising-based business models levelling the playing field and opening the way for a flood of new publishers to undermine the traditional players, swiftly and for ever. At the same time, newly-empowered information consumers would enjoy the "Daily Me", a perfectly-personalised and utterly relevant media experience, enabling them to save time and to work smarter.

In the aftermath of the Internet boom, we know that neither the publisher nor the consumer "revolutions" have happened. Rather than the "Daily Me", consumers got information overload: a rich glut of excessive media choice instead of the nirvana of total relevance. And new publishers came, and went away again, the advertising-based model failing largely because there is a limit to how much time consumers will spend reading online, and because it is an enormously complex and expensive task to create intelligent, personalised websites which achieve the "stickiness" that publishers needed.

However, both publishers and consumers still accept the potential of digital media to improve the publishing process and radically enhance the media experience. Nowadays it is perhaps seen less in terms of an industry revolution, and more in terms of the potential that technology offers to streamline and improve the whole process of publishing and consuming information.

Firstly, bandwidth limitations are slowly lifting, meaning that more and more users of computers – and, ultimately, mobile devices – will be able to access graphic, audio, and video data as smoothly as they can now access text.

Technology convergence means that many different platforms will be able to carry the same digital information, with all of its potential to be stored, tagged, and structured. PCs, phones, TVs, even in-car systems, will soon be accessing the same streams of data.

And surrounding this technology convergence is the growth of common, open standards – such as XML – which rapidly enable different types of data to be described, managed and distributed without the need for the enormous

editorial efforts of, for example, early worldwide web publishing, and without the barriers of incompatible systems and standards.

These three trends, of *increasing bandwidth*, *technology convergence* and *open standards*, are creating an environment in which many of the early dreams of publishers and consumers – for streamlining publishing and creating highly relevant, personalised content – can start to come true.

In the field of news, the enabling development is *NewsML*, an emerging industry standard for structuring multimedia news and which itself is a derivative of the rapidly-spreading XML standard.

Why NewsML?

As we have seen, both consumers and publishers want to take advantage of the digital medium's potential, whilst avoiding the problems of complexity and information overload.

News publishers have a wide range of business issues and challenges driving them. They need to assemble news in different formats – graphics, text, audio, and video – in a way which is easy to use and which results in compelling content for readers. News needs to be integrated from different sources, each with its own potential set of proprietary standards. As stories develop, different versions need to be tracked in a way which is easy to manage accurately. As more and more delivery and distribution happens automatically, news now needs to be optimized for manipulation by machines. Finally, news needs to be easily packaged for a wide range of devices and delivery channels, from mobile phones to digital interactive televisions.

Genesis of NewsML

NewsML has been developed and ratified as an open standard by the International Press Telecommunications Council (IPTC) – www.iptc.org

For the uninitiated, the IPTC was established in 1965 to safeguard the telecommunications interests of the world's press, and one of its primary functions is as a standards-setting body for the interchange of news data. For example, among other standards, the IPTC also manages the NITF (News Industry Text Format) standard. This provides a vocabulary for defining the content and structure of news articles, thus making them more flexible and adaptable. Members of the IPTC include major news agencies, newspaper publishers and system vendors.

In essence, NewsML is an XML language for the news industry. It is intended, as the IPTC says, to “represent and manage news throughout its lifecycle”. It is designed to provide a media-independent structural framework for news, so that publishers can work quickly and efficiently in the new digital environment, to the benefit of those who consume their content. Crucially, NewsML is an open standard, which means it can be easily adopted and used by any publisher.

The development of NewsML has been spurred by continuing growth in the production, use and re-use of news throughout the world – with the Internet and its potential as perhaps the key driver. The IPTC identified a set of formal requirements which have shaped the standard that NewsML has now become. These requirements, detailed below, demanded that the new standard must:

- **Support the representation of different electronic news entities** – the standard must be able to reflect the fact that news may be delivered as single items, or as packages of items, and contain metadata which allow efficient production, delivery, and use.
- **Be usable throughout the news lifecycle** – although the standard's primary use is for news interchange, it must also be usable in archiving and for publishing in networked systems.

- **Allow news items to consist of arbitrary mixtures of media types, languages, and encodings** – under the standard, news packages should be able to consist of different content types. For example: as text, graphics or video; in different languages; or as images in low- and high-resolution formats.
- **Replace or allow the transport of existing news formats and encodings** – the new standard should replace existing ones where appropriate, or include standards which offer additional features (such as the text-formatting capabilities of NITF).
- **Support different physical constructions of the same data** – the standard should allow, for example, the same news item to be delivered as a full multimedia package to one user, or as a short text summary to another.
- **Support the management and development of news items over time** – the standard should reflect the fact that news stories develop gradually, and should assist the management of news over time.
- **Be simply extensible and flexible** – the standard should be able to change to reflect market developments, and should also allow users to add their own features and extensions.
- **Allow for authentication and signature of metadata and news item content** – so that the reliability, and hence value, of news content can be assured.
- **Not be unduly verbose** – whilst bandwidth is increasing globally, transmission systems vary in capacity throughout the industry. The standard should reflect this.
- **Use XML and other appropriate standards and recommendations** – building on a proven and fast-growing technology will help to ensure acceptance by the wider information industry, and will mean that tools and development expertise are available.

A standard which adheres to these requirements will allow much greater efficiency and flexibility in news production and consumption.

Structure of NewsML

Fig. 1 (next page) illustrates the richness of NewsML. There is no need to use all of the features in what is, on the face of it, a complex and layered structure. The underlying logic of NewsML is, however, straightforward.

NewsML simply takes the form of an XML document which has a series of components – or elements – that are used to structure and process the actual news content.

For example, consider the way in which a single piece of news content – such as a text story, a digital image, or an audio or video clip – would be handled. NewsML is media-neutral, so no assumptions are made about the media type, format or encoding of the data. As can be seen in the diagram, the news content itself is carried as a data element within a **ContentItem**. If encoded, this may have an **Encoding** element, and an optional **Characteristics** element that can specify the requirements for the system which is to carry the data; for example, the file size.

The **ContentItem** may also contain a wide range of metadata elements, a few examples of which are media type, format and notation. At the next level up, the **ContentItem** is carried by a **NewsComponent**. This exists mainly to provide a way of establishing relationships between multiple items of news content. The **NewsComponent** enables, for example, the specification of whether two pictures are *equivalent* (i.e. the same image at different resolutions) and whether the picture is a *complement* to a news story.

At the next layer, the **NewsItem** is the central component of NewsML. It is a standalone, manageable, piece of news that must have both **Identification** and **News Management** elements and which contains a single **NewsComponent**.

At the top level, the NewsML element contains one or more **NewsItems** and a **NewsEnvelope**. The latter contains workflow-related data such as the date and time of issue, sender, recipient, and priority information: the **NewsEnvelope** can even specify whether the contents of a transmission form part of a particular product.

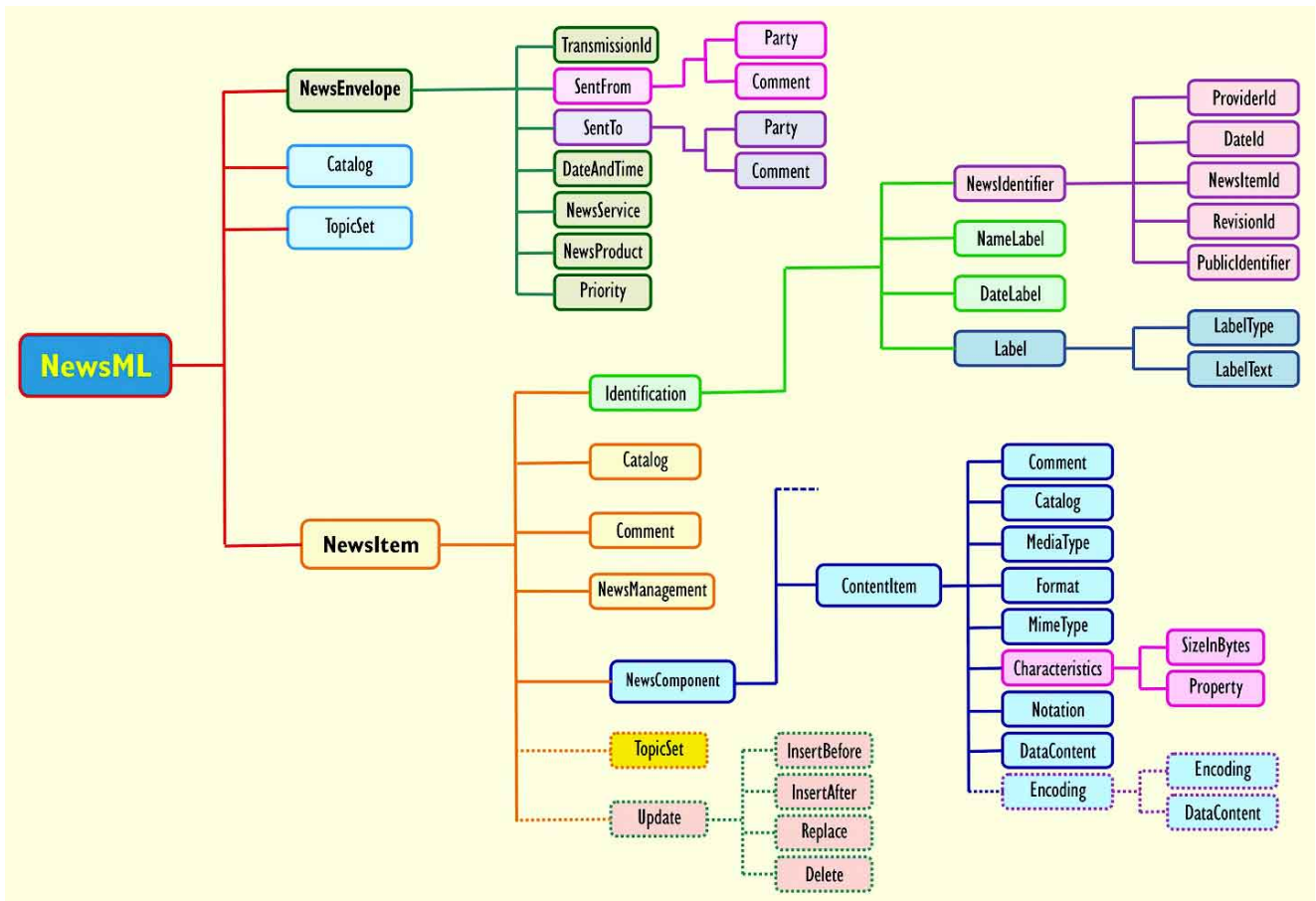


Figure 1
The layered structure of NewsML.

The two figures shown below offer another way of looking at the structure of NewsML and illustrate how the different NewsML components fit together.

Fig. 2 shows the **ContentItem**, or actual news information, held within the **NewsComponent**. The **NewsComponent** adds information about the news, for example, how it relates to other **ContentItems** within the same package: is it a *complement* – for example an image relating to a news story? Or is it an *equivalent* – for example a German translation of an English story? The **NewsItem** is the published package, itself containing information about when it was published, by whom, its destination, and priority.

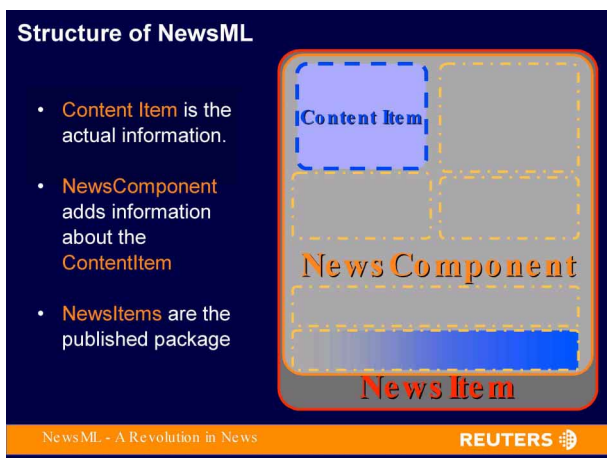


Figure 2
The actual news information held within the NewsComponent.

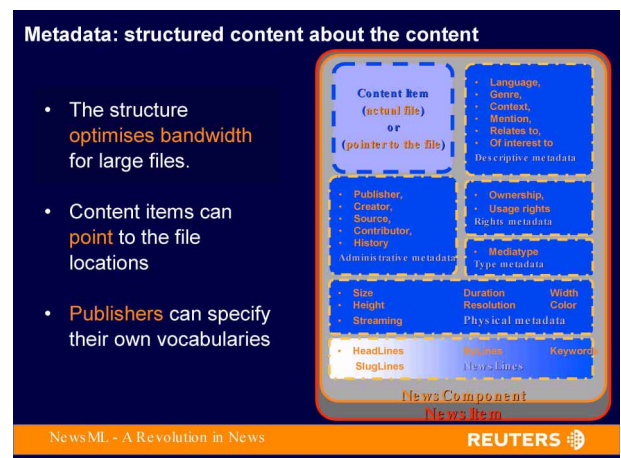


Figure 3
The rich range of metadata held within a NewsML document.

Fig. 3 shows some of the rich range of metadata – or “content about the content” that can exist within a NewsML document. This structure optimizes bandwidth for large files by telling systems what size and type of data they can expect. Systems that are scaled for low-bandwidth could therefore choose to reject the rich-media elements described by the metadata: content items can point to the file locations. Publishers can also specify their own metadata vocabularies, giving them additional flexibility.

Key features of NewsML

NewsML offers a highly structured environment for publishing, which immediately offers important key features for publishers.

The standard:

- **Enables all formats and media types to be recognized equally** – NewsML can equally represent text, video, audio, graphics and still photographs which means that it offers a publishing structure for different types of publishing.
- **Facilitates the development of evolving news** – NewsML provides strong revisioning support which makes it easy for publishers to modify and update stories, a crucial capability in a digital environment.
- **Allows for named relationships between NewsItems** – such relationships may stand, for example, for “see also” or “related news” or “for more detail”. This makes it easy to build effectively cross-referenced and compelling news packages.
- **Supports collections of NewsItems** – it is a characteristic feature of news items that they are often presented as collections. NewsML enables news items to be grouped and seen as collections regardless of whether or not that was the original journalistic intent.
- **Allows for alternative representations within the same NewsComponent** – NewsML allows for news content to exist in different media types in the same space.
- **Permits explicit inclusion and exclusion of NewsComponents** – NewsML provides the capability to include material when it is transmitted or merely to include references to material. This makes it easy to manage bandwidth if large files are likely to be involved.
- **Creates a structure consisting of parts and named relationships between parts** – in NewsML, NewsItems consist of parts which have a named role in relation to the NewsItems; for example, a NewsItem may consist of a text part and several photos as secondary parts.
- **Allows for the attachment of metadata from standard and non-standard controlled vocabularies** – this allows for an enormous range of description of the NewsComponent and ContentItems (author, publisher, subject, importance, generality, etc) without the need for prescribed vocabularies.

For publishers, this opens up a rich seam of new possibilities as well as offering many ways to improve efficiency in the publishing process.

Benefits of NewsML

If and when NewsML becomes widely adopted throughout the news publishing industry, it has the potential to offer significant benefits for all involved in the publishing process, from content originators through to news consumers.

Benefits for content originators and publishers

NewsML is of particular relevance for any organization involved in the generation of original content, and especially when that content is allowed to be distributed by many different publishers and platforms.

Firstly, it enables easy publishing to all platforms. Historically, news publishers struggled with the challenge of re-purposing their content for the web, then for mobile devices and PDAs. As access devices proliferate in the future, this complexity could become impossible to manage. However, in a world where all content is originated digitally, the existence of a broadly adopted news publishing standard – NewsML – means that news is automatically and instantly re-edited depending on the platform which is receiving it. This offers huge savings for content originators, as well as creating a rich pool of content for aggregators and publishers.

NewsML also makes it easy to create packages of targeted, multimedia information. Again, the business of interactive news publishing has been historically held back by the technical complexity of creating targeted and personalised information. Producing such personalisation in different media adds another layer of complexity which has proved enormously difficult to overcome. With NewsML, highly targeted packages of information can simply be created “on the fly”. Packaged in NewsML, each news story carries so much information about itself that it can automatically slot into, for example, a consumer’s personalised webzine or a PDA-delivered newsletter. Building applications to deliver this level of targeting becomes a simple matter of ensuring the application asks for the correct NewsML elements.

NewsML enables publishers easily to combine content from multiple content originators, without having to wrestle with different standards and formats. In news, this opens up the possibility of even further targeting and personalisation. For example, publishers will be able to take content from multiple sources on one particular subject in order to create, for example, a highly specialised and powerful news resource.

And as well as achieving highly effective levels of targeting, NewsML could help bring about the elusive “stickiness” that web publishers have long been seeking. One of the web’s greatest features is hyperlinking and cross-referencing, yet managing that process has proven enormously complex and time-consuming. By automating the process in which related news items group together, NewsML can ensure, for example, that every web page is always peppered with genuinely related articles, prompting readers’ interest and their propensity to return to a site.

The revision management capabilities of NewsML also aid the streamlining process, making it easy for content originators to manage multiple versions of rapidly-evolving news stories, and to add new material.

In summary, the existence of this standard offers content originators and publishers greater efficiency, allowing them to publish at reduced cost and automatically target multiple audiences and platforms.

Benefits for consumers

For news consumers, NewsML brings the mythical “Daily Me” a step closer. The standard will make it much easier for the consumer to define what information they would like to receive and when – potentially saving huge amounts of time in trawling the web for relevant news, or visiting multiple sites. It also has the potential to harmonize news use across different platforms, so that consumers receive news according to their preferences for subject, media, timeliness, etc. whether they happen to be looking at a mobile phone, PC, digital TV, or even listening to a digital car radio broadcast.

Abbreviations

IPTC	International Press Telecommunications Council
NITF	News Industry Text Format
PDA	Personal digital assistant
XML	Extensible markup language

As we have seen, NewsML truly brings alive the cross-referencing power of the Internet, so it will also make it much more effective as a research medium, enabling consumers to find topics and threads more easily, and making cross-content database searching a simple reality rather than the time-consuming and frustrating task it can be today.

NewsML's ability to optimize bandwidth and storage costs will also have great benefits for consumers as well as publishers of information. The ability, for example, to screen out files above a certain size so that only their reference is included will make the Internet news experience much more rewarding for low-bandwidth users. At the same time, such a capability will enable publishers simultaneously to create rich-media content for users with high-speed Internet access.

Finally, NewsML's compatibility with other markup languages will enable news to be fully integrated into other aspects of the digital commercial world. For example, NewsML makes news interoperable with other environments and data models, which can connect topics, events, related financial instruments, and the required trading environment.

Implications for the publishing and broadcast industries

If NewsML is widely adopted, then – in combination with related standards such as XML – it does indeed have the potential to enable a range of positive changes in the business of publishing. However, there is still some way to go before some of the bandwidth and platform constraints that limit a true multimedia, access-anywhere news experience are lifted.

Importantly, though, the impact of NewsML is not limited to the online and wireless publishing media. As broadcast media switch to digital distribution, so they become positioned to take advantage of the benefits of a common publishing standard. In providing a proven, widely adopted and freely-published standard, NewsML

<http://www.iptc.org/NewsML>
<http://www.newsml.org>
<http://www.about.reuters.com/newsshowcase>

NewsML - A Revolution in News **REUTERS** 

Figure 4
NewsML — the “killer standard”?



Tony Allday is Head of Product Development and Operations for the Reuters Media Group.

In this role, he is responsible for the management and ongoing development of the various delivery systems which Reuters uses to deliver its services to Traditional and New Media customers globally. This includes all forms of text newsfeeds, pictures feeds, video and online multimedia offerings. He oversees the strategic direction of Reuters WNS video delivery network; its Multimedia Data Network for text, pictures and graphics; online pictures database and its new Internet Delivery System for XML-based news content.

Mr Allday has held various product, marketing and data management roles in his 15 years with Reuters, principally in the Media and Corporate Information Products area. He has worked in the information industry for over 20 years and, prior to joining Reuters, was

Managing Editor for Finsbury Data Services, producers of the Textline online business information service.

will enable broadcasters quickly to enter the interactive publishing space without having to divert vast resources to developing their own proprietary digital publishing strategies.

In the modern world of publishing and technology, it may be that NewsML is the “killer standard” rather than the “killer application” which enables rapid development – this time, for all concerned.

