



# **IPTC Standards**

---

## **NewsML 2**

### **Business requirements**

### **DRAFT 14 (July 2004)**

Copyright © 2004 by IPTC, the International Press Telecommunications Council.  
All rights reserved.

NewsML™ is a registered trademark of the IPTC.

**Document file name: NewsML\_2.0\_spec\_BusinessRequirements\_14**

**Document URN: urn:iptc:std-draft:newsml:2.0:spec:businessrequirements:14**

<b>Revision</b>	<b>Issue Date</b>	<b>Author (revised by)</b>	<b>Remark</b>
Draft 0.8	2004-04-05	Laurent Le Meur	Draft, for quick review by Christian Ratenburg, Misha Wolf, Darko Gulija, Michael Steidl.
Draft 1	2004-04-07	Laurent Le Meur	Draft, for open circulation after quick review.
Draft 2	2004-04-13	Laurent Le Meur	Internal draft before confcall.
Draft 3	2004-04-14	Laurent Le Meur	Internal draft after confcall.
Draft 4	2004-04-15	Misha Wolf	Preparing for publication on 2004-04-16.
Draft 5	2004-04-16	Misha Wolf, Laurent Le Meur	Draft, for open circulation.
Draft 6	2004-04-19	Michael Steidl	Formal revision, published publicly
Draft 7	2004-04-27	Laurent Le Meur	Internal draft after confcall.
Draft 8	2004-04-30	Laurent Le Meur Michael Steidl	Revision after receiving comments, published publicly.
Draft 9	2004-05-09	Laurent Le Meur	Internal draft after confcall.
Draft 10	2004-06-03	Laurent Le Meur	Internal draft after confcall.
Draft 11	2004-06-14	Laurent Le Meur	Internal draft after confcall
Draft 12	2004-06-15	Laurent Le Meur Michael Steidl	Formal revision, for public circulation
Draft 13	2004-06-24	Laurent Le Meur	Internal draft after confcall
Draft 14	2004-07-09	Laurent Le Meur	Formal revision, for public circulation

**Important notice: A change log is provided at the end of the document.**

# Table of Contents

---

<b>1</b>	<b>Management of This Document .....</b>	<b>6</b>
1.1	Approvals.....	6
1.2	Document Control Statement .....	6
<b>2</b>	<b>Terminology, definitions and references .....</b>	<b>7</b>
2.1	Terminology.....	7
2.2	Definitions, Acronyms and Abbreviations .....	7
2.2.1	NewsML entities and related terms .....	7
2.2.2	Generic terms .....	8
2.2.3	Standards Bodies.....	9
2.2.4	Standards.....	9
<b>3</b>	<b>Introduction .....</b>	<b>11</b>
<b>4</b>	<b>Business Objectives .....</b>	<b>12</b>
4.1	Primary Customers .....	12
4.2	Success metrics.....	12
4.3	Scope.....	12
<b>5</b>	<b>Use Cases .....</b>	<b>13</b>
5.1	Multimedia Package.....	13
5.2	Push distribution.....	15
5.3	Legacy format consolidation .....	16
<b>6</b>	<b>Requirements .....</b>	<b>19</b>
6.1	High level .....	19
6.1.1	News life cycle.....	19
6.1.2	Authoring .....	19
6.1.3	Storage .....	19
6.1.4	Exchange .....	19
6.1.5	Syndication .....	20
6.1.6	User agents.....	20
6.1.7	XML based .....	20
6.1.8	Presentation.....	20
6.2	NewsML model .....	20
6.2.1	Conceptual model .....	20
6.2.2	Processing Model.....	21
6.2.3	Construction .....	21
6.2.4	Details.....	21
6.2.5	Interoperability .....	21
6.2.6	No interpretation .....	21
6.2.7	One way.....	21

6.2.8	Patterns .....	21
6.2.9	Error handling.....	22
6.2.10	Application interfaces.....	22
6.2.11	Default behaviour .....	22
<b>6.3</b>	<b>Content representation.....</b>	<b>22</b>
6.3.1	Media-independence .....	22
6.3.2	Inclusion and reference .....	22
6.3.3	Encoding .....	23
<b>6.4</b>	<b>Metadata support .....</b>	<b>23</b>
6.4.1	News-related metadata .....	23
6.4.2	Minimal mandatory metadata.....	23
6.4.3	Metadata extensibility .....	23
6.4.4	Meta-metadata .....	24
6.4.5	Types of metadata values.....	24
<b>6.5</b>	<b>Metadata classes.....</b>	<b>24</b>
6.5.1	Characteristics .....	24
6.5.2	Administrative and descriptive metadata .....	25
6.5.3	Rights metadata .....	25
6.5.4	Management metadata.....	25
6.5.5	Links from news-items to other resources.....	26
6.5.6	IPTC legacy metadata .....	27
<b>6.6</b>	<b>Composition .....</b>	<b>27</b>
6.6.1	Alternative renditions .....	27
6.6.2	Groups of news-items.....	27
<b>6.7</b>	<b>Digital signature .....</b>	<b>28</b>
6.7.1	Signed news-item .....	28
6.7.2	Signed content .....	28
<b>6.8</b>	<b>Management .....</b>	<b>28</b>
6.8.1	Identification .....	28
6.8.2	Revision .....	29
6.8.3	News development.....	29
6.8.4	Metadata update.....	29
6.8.5	Management notice.....	29
<b>6.9</b>	<b>Labels .....</b>	<b>29</b>
6.9.1	Labels support.....	30
6.9.2	IPTC legacy labels.....	30
6.9.3	Topic references in labels .....	30
6.9.4	Hyperlink in labels.....	30
6.9.5	Line breaks in labels .....	30
6.9.6	Variants of Labels .....	30
<b>6.10</b>	<b>Exchange .....</b>	<b>31</b>
6.10.1	Exchange envelope.....	31
6.10.2	Exchange properties.....	31
<b>6.11</b>	<b>Forward and backward compatibility.....</b>	<b>31</b>
6.11.1	Conceptual model backward compatibility .....	31
6.11.2	Syntactic backward compatibility.....	31
6.11.3	Application forward compatibility .....	32
6.11.4	Bidirectional transformation .....	32
<b>6.12</b>	<b>Compatibility with other standards.....</b>	<b>32</b>
6.12.1	Support for and conformance with non-IPTC standards .....	32
6.12.2	Conformance with other IPTC standards.....	33

<b>6.13</b>	<b>Universality</b> .....	<b>33</b>
6.13.1	Internationalisation .....	33
6.13.2	Accessibility .....	33
<b>6.14</b>	<b>Usability</b> .....	<b>33</b>
6.14.1	Simplicity .....	33
6.14.2	Conciseness .....	34
6.14.3	Style & ease of use.....	34
6.14.4	Specifications .....	34
6.14.5	User Manuals and On-line References.....	34
<b>6.15</b>	<b>Standard maintenance</b> .....	<b>34</b>
6.15.1	Ongoing standard maintenance .....	34
6.15.2	Standard support .....	34
<b>7</b>	<b>Cross reference NewsML 1 vs NewsML 2</b> .....	<b>36</b>
<b>8</b>	<b>Licensing</b> .....	<b>37</b>
<b>9</b>	<b>Copyright and other Legal Notices</b> .....	<b>38</b>
<b>10</b>	<b>Appendices and Supporting Information</b> .....	<b>39</b>
10.1.1	IPTC documentation .....	39
10.1.2	Resources .....	39
<b>11</b>	<b>Change Log</b> .....	<b>40</b>
11.1	Transition to Draft 3 .....	40
11.2	Transition to Draft 4 .....	40
11.3	Transition to Draft 5 .....	40
11.4	Transition to Draft 6 .....	41
11.5	Transition to Draft 7 and 8.....	41
11.6	Transition to Draft 9 .....	41
1.2	Transition to Draft 10 .....	41
1.3	Transition to Draft 11 .....	42
1.4	Transition to Draft 12 .....	42
1.5	Transition to Draft 13 .....	42
1.6	Transition to Draft 14 .....	42

## **1 Management of This Document**

### **1.1 Approvals**

*<This subsection will be added at a later stage.>*

### **1.2 Document Control Statement**

This document's content and format are controlled by the IPTC Standards Committee. Any suggestions or inquiries should be directed to the Chairman of the IPTC Standards Committee or the IPTC Managing Director.

## 2 Terminology, definitions and references

### 2.1 Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY" and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC 2119].

### 2.2 Definitions, Acronyms and Abbreviations

#### 2.2.1 NewsML entities and related terms

<b>Digital asset</b>	A representation of some content. A digital asset may represent a piece of text, a picture, an audio or video clip or any other type of content used in the news industry. A digital asset may be contained in a news-item, or referenced from a news-item.
<b>Component</b>	A constituent of a news-item, defined as the association of metadata with content. A component is the result of the association of some content, physical characteristics, descriptive metadata and rights metadata relative to the content. A component bears a specific named relationship (i.e. a role) to its news-item container. The structure of a component (composition or association via references) is let to the implementation stage.
<b>Controlled vocabulary</b>	A controlled set of values associated with a given metadata element. May be a list of topic references or any other kind of controlled value.  “organized lists of words and phrases, or notation systems, that are used to initially tag content, and then to find it through navigation or search.” (Amy Warner). “A subset of natural language”(Wellisch). [BOX&ARROWS]
<b>Governed metadata</b>	Metadata-element which value is governed by a controlled vocabulary.
<b>Metadata element</b>	An xml element or attribute that supports information about some content.
<b>Metadata-set</b>	A consistent set of metadata (data about data) defined by a standardization body. A metadata-set is identified by a specific namespace. The term metadata vocabulary is not used as is could clash with the definition of controlled vocabulary.
<b>Metadata value</b>	The value of a metadata-element; it may be an arbitrary value, a rule based value or a value drawn from a controlled vocabulary.
<b>Named entity</b>	A named entity may be a person, place, event, organization, product name, object name or any other news related real life entity. This is a special kind of topic. An event is a named entity, a subject or genre isn't. Compare with proper name vs common name concepts.

<b>News-envelope</b>	An xml envelope used for the B2B exchange of news-items.
<b>News-item</b>	A unit of news as defined by the editorial or product construction practice of a news provider. A news-item is a compound document consisting of one or more components or news-items.
<b>Topic</b>	Any real-world thing or concept that can be referred to in a piece of news. Examples of a topic are Tennis, Censorship, the Iran-Iraq war, Tony Blair, Prime Minister of Pakistan, IBM, the United Nations, the Dyson vacuum cleaner, China, Kurdistan, Paris, the Kremlin, AIDS, aspirin, etc. The notion of topic covers generic concepts like subjects and genres, and named entities like people, places and organizations.
<b>Topic-item</b>	An xml structure that supports information about a given topic (especially a topic identifier, some description, relationships with other topics ...).
<b>Topic reference</b>	A reference to a topic-item, in the form of a topic identifier.

## 2.2.2 Generic terms

<b>API</b>	Application Programming Interface. “A boundary across which a software application uses facilities of programming languages to invoke software services. These facilities may include procedures of operations, shared data objects and resolution of identifiers.” (DAVIC). Set of functions defined in order to access, process or create news-items. Generally, an implementation in a given language (Java, .NET, Perl) is tightly associated with an API.
<b>DRM</b>	Digital Rights Management
<b>DTD</b>	Document Type Definition
<b>Globally unique identifier</b>	An identifier that is unique, unambiguous, and permanent. Being unique and unambiguous means that there is a 1:1 relationship between the identifier and the identified object. Being permanent means that its value never changes when time passes, and is never reused as an identifier for another object even if the original object disappears.
<b>News Provider</b>	A provider of news content. May be a news agency, a syndication company, a newspaper, a magazine ... or a blogger.
<b>NewsML API</b>	An API designed to process NewsML instances.
<b>NewsML enabled CMS</b>	Content management system (either editorial CMS or Web CMS or archiving system) able to handle news-items.
<b>NewsML instance</b>	A NewsML compliant XML document; it may be a news-envelope or a news-item.
<b>NewsML processor</b>	An application (user agent) that support the handling and processing of NewsML instances.



<b>NewsML workflow</b>	A workflow of NewsML documents between NewsML processors.
<b>Representation</b>	Two entities with corresponding included and excluded components have the same representation. Representation refers to the physical form something has.
<b>Revision</b>	Two entities that have the same identifier and the same revision level are at the same revision. Revision refers to how up to date something is and does not refer to the physical form of presentation.
<b>Resolution mechanism</b>	The processing rules for resolving a reference to a news-item or topic-item (must include a rule for detecting whether there is a resource at all, in order to deal with dangling references).
<b>Unambiguous identifier</b>	An identifier is unambiguous if it identifies one and only one object (but an object may have several different identifiers). i.e. there is a 1:n relationship between the object and the identifier; e.g. the name of a person is ambiguous (case of homonyms).

### 2.2.3 Standards Bodies

<b>IPTC</b>	International Press Telecommunications Council <a href="http://www.iptc.org/">http://www.iptc.org/</a>
<b>OASIS</b>	Organization for the Advancement of Structured Information Standards <a href="http://www.oasis-open.org/">http://www.oasis-open.org/</a>
<b>Unicode</b>	The Unicode Consortium <a href="http://www.unicode.org">http://www.unicode.org</a>
<b>W3C</b>	World Wide Web Consortium <a href="http://www.w3.org/">http://www.w3.org/</a>

### 2.2.4 Standards

<b>RFC 2119</b>	Key words for use in RFCs to Indicate Requirement Levels <a href="http://www.ietf.org/rfc/rfc2119.txt">http://www.ietf.org/rfc/rfc2119.txt</a> S. Bradner
<b>RFC 3066</b>	Tags for the Identification of Languages <a href="http://www.ietf.org/rfc/rfc3066.txt">http://www.ietf.org/rfc/rfc3066.txt</a> H. Alvestrand
<b>RFC 3066 bis</b>	Tags for Identifying Languages <a href="http://www.ietf.org/internet-drafts/draft-phillips-langtags-04.txt">http://www.ietf.org/internet-drafts/draft-phillips-langtags-04.txt</a> A. Phillips and M. Davis
<b>RFC 3085</b>	URN Namespace for NewsML Resources <a href="http://www.ietf.org/rfc/rfc3085.txt">http://www.ietf.org/rfc/rfc3085.txt</a> D. Allen, D. Rivers-Moore
<b>Unicode</b>	The Unicode Standard <a href="http://www.unicode.org">http://www.unicode.org</a>

<b>ISO 639</b>	Codes for the representation of names of languages <a href="http://lcweb.loc.gov/standards/iso639-2/iso639jac.html">http://lcweb.loc.gov/standards/iso639-2/iso639jac.html</a>
<b>ISO 3166</b>	Codes for the representation of names of countries <a href="http://www.iso.ch/iso/en/prods-services/iso3166ma/index.html">http://www.iso.ch/iso/en/prods-services/iso3166ma/index.html</a>
<b>ISO 15924</b>	Codes for the representation of names of scripts <a href="http://www.unicode.org/iso15924/">http://www.unicode.org/iso15924/</a>
<b>XML 1.0</b>	Extensible Markup Language (XML) 1.0 – Latest edition <a href="http://www.w3.org/TR/REC-xml/">http://www.w3.org/TR/REC-xml/</a> W3C Recommendation
<b>XML 1.1</b>	Extensible Markup Language (XML) 1.1 <a href="http://www.w3.org/TR/xml11">http://www.w3.org/TR/xml11</a>
<b>Namespaces</b>	Namespaces in XML 1.1 <a href="http://www.w3.org/TR/xml-names11">http://www.w3.org/TR/xml-names11</a>
<b>XML Schema</b>	XML Schema definition language 1.0 XML Schema Part 0: Primer -- <a href="http://www.w3.org/TR/xmlschema-0/">http://www.w3.org/TR/xmlschema-0/</a> XML Schema Part 1: Structures -- <a href="http://www.w3.org/TR/xmlschema-1/">http://www.w3.org/TR/xmlschema-1/</a> XML Schema Part 2: Datatypes -- <a href="http://www.w3.org/TR/xmlschema-2/">http://www.w3.org/TR/xmlschema-2/</a> W3C Recommendation
<b>IPTC 7901</b>	The IPTC recommended message format <a href="http://www.iptc.org/IPTC7901/">http://www.iptc.org/IPTC7901/</a> IPTC Standard
<b>IIM</b>	Information Interchange Model <a href="http://www.iptc.org/IIM/">http://www.iptc.org/IIM/</a> IPTC Standard
<b>NITF</b>	News Industry Text Format; an xml language created for the representation of textual news. <a href="http://www.iptc.org/NITF/">http://www.iptc.org/NITF/</a> IPTC Standard

### 3 Introduction

This document contains the business requirements for NewsML™, a media independent structural framework for the representation of news and specialized content, and reflects the work done to develop version 2 of NewsML.

News exchange is a business of moving around not only the core news content, but also data that describe the content in an abstract way (i.e. metadata), information about how to handle news in an appropriate way (i.e. news management data), and finally information about the news transportation or routing process itself (i.e. exchange data).

Having a strong background in developing and maintaining news exchange formats, the IPTC created NewsML as the most comprehensive and versatile way to move all types of data between media systems in order to make news exchange efficient and reliable.

The initial version of NewsML was approved in October 2000. Since then it went along with minor revisions: version 1.1 was approved in October 2002; version 1.2 was approved in October 2003.

This Business Requirements document covers only high level requirements and specifies all basic features for NewsML, its supporting environment and procedure for future development.

The Business Requirements will be supplemented by documents of an intermediate level, specifying a conceptual model and a processing model for NewsML. Finally one can expect all low level documentation to be delivered in the course of implementing the requirements and the models.

## 4 Business Objectives

NewsML™ aims to be the standard of choice for professional exchange of news.

Since its introduction in October 2000, many news providers, agencies, newspapers and other publishers have adopted NewsML. In order to drive further implementation, the IPTC has resolved to create a new version of the standard that builds on the strengths and tackles the weak points.

Publishers will continue to be encouraged to leverage content-specific standards (whether created by the IPTC or other organizations) alongside NewsML 2. The features of older news industry standards (e.g. ANPA 1312, IPTC 7901, and the IIM) will be supported, allowing publishers to migrate to modern technologies without losing the benefits of mature solutions. The IPTC will simplify and clarify the implementation of NewsML systems, balancing richness of functionality with ease of use.

### 4.1 Primary Customers

The primary intended users of the NewsML standard are:

- News agencies / syndicators
- Newspapers
- News aggregators
- Online portals
- Editorial system providers

### 4.2 Success metrics

Success will be measured by the level of adoption of any version of NewsML™ by international news providers and system vendors regardless of replacing older news exchange standards or being introduced for new news products like multimedia feeds.

### 4.3 Scope

The scope of the NewsML standard is limited to the definition of a media independent structural framework for the representation of news.

The development of version 2 of NewsML is part of the “IPTC Roadmap 2005”, an initiative aimed at widening the use of IPTC standards, making implementation easier and simpler, ensuring that the latest technologies are used when appropriate, and making the IPTC standards - currently having NewsML 1, NITF and SportsML approved and having ProgramGuideML and EventsML under development - a consistent family.

The definition of a controlled vocabulary mechanism is out of the scope of this effort.

## 5 Use Cases

We intend to describe the following use cases:

- **Multimedia package:** creation of a multimedia package for a web usage.
- **Consolidation of legacy formats:** how NewsML 2 can help a news provider consolidating several legacy formats (IPTC7901, ANPA1312, IIM, proprietary ones) and establishing a common set of metadata across all media.
- **Push distribution:** distribution of new via some push mechanism (satellite, workflow, exchange metadata, embedded content, signature of content)
- **News coverage:** coverage of a breaking news-item and its development over time (parallel to the EventsML breaking news use-case; see EventsML requirements; categorization, revision, alternative content, relationships between news-items, links to other resources)
- **International news:** handling of multilingual news.
- **Content management:** handling of news-items in a NewsML enabled editorial system (storage, management, indexing, searching)
- **Archiving:** how NewsML 2 can help maintain a news archive.
- **Semantic Web:** use of NewsML in a distributed environment (like the web) with navigation in a network of topics; compatibility with web standards.
- **Newspapers workflow:** how newspaper integrate news in their editorial systems (with B2B labels = newlines), how pages can be handled in NewsML format
- **Syndication:** distribution of news via some syndication mechanism (à la RSS/Atom, plus metadata and rights management and multiple representations)
- **Aggregation:** use for an aggregator receiving multiple feeds from multiple providers (controlled vocabularies)
- **Mobile consumption:** how a mobile java module consumes NewsML data

### 5.1 Multimedia Package

Use Case: Multimedia package	
<b>Context of Use</b>	Producer of 'ready to publish' news, targeted to web and 'mobile oriented' sites.
<b>Scope</b>	Editorial Goal
<b>Level</b>	Primary task
<b>Primary Actor</b>	News provider
<b>Preconditions</b>	The news provider has a NewsML enabled CMS, a formatting platform and a syndication system.
<b>Minimal Guarantees</b>	-
<b>Success Guarantees</b>	The news organization is able to manage collections of multimedia news stories and syndicate them to many client sites.
<b>Trigger</b>	A government set to be announced
Stakeholders	
<b>Stakeholder</b>	Arthur (multimedia editor) and Juliette (video correspondent)

Use Case: Multimedia package	
<b>Editorial</b>	Need to cover the news as a multimedia story, with text and associated photos; a video clip is added in a later revision; this story is inserted into a “Political news” web collection.
Description	
Step	Action
1	A new government set is announced. As the first leads go on the wires, Arthur, from the multimedia team, prepares a multimedia story on the event. He creates a news-item in the NewsML enabled CMS, names it “New government to be unveiled at 1700 GMT”, and edits the text in his XML-aware authoring tool.
2	Arthur picks a good picture in the NewsML enabled CMS. He creates a cropped thumbnail, optimized for small display and mobile usage (a close-up); the thumbnail is automatically added to the original picture document as a new rendition, so the newly created component can be reused later in other contexts. Arthur picks some other pictures from the CMS, for use as illustrations, files them in a personal electronic basket, creates cropped thumbnails if needed.
3	Arthur inserts the selected pictures in the news-item, and adds a contextual caption to each image. In a WYSIWYG mode, the pictures are positioned in the context of the story, each one visible as a preview picture alongside a given paragraph.
4	Now Arthur adds links to the story. The first link originates from the news-item itself, and points at a collection of background stories, written in advance for the event; a name (“background information”) is given to the link. A user selecting the link will be directed to a page where the abstracts of all the background stories will be shown, each story abstract being illustrated with the thumbnail of an associated picture.
5	Then Arthur selects a section of a sentence – the name of the future prime minister – and creates a link from it to a background story, in this case the ready to publish profile of the person. The story is now ready for publication.
6	Arthur inserts the story in a given set of ‘Political news’, at the top of the ordered collection, and validates the story update. The same story can be reused in other collections.
7	The updated collection (also represented as a news-item), the new story, the accompanying pictures and all associated stories are published on client web sites in an HTML format. The same collection is published on several i-mode site in CHTML, and basic WAP sites in WML. In the latter case, each story is truncated to the two first paragraphs, and no picture nor link is included (the distribution system does not even send the associated news-items to those sites).
8	After half an hour, Juliette sends a ready to publish video-clip to the provider’s NewsML enabled CMS. The associated NewsML instance contains a reference to the video-clip, a thumbnail and preview (automatically created), and a set of metadata (some automatically extracted from the video-clip itself, some entered by Juliette).
9	Arthur gets a notification of the insertion of a new item related to the event he focuses on; he checks out the story, and inserts the video-clip in the news-item, with a contextual caption; in a WYSIWYG mode, the video thumbnail is positioned in the context of the story.
10	Arthur validates (checks-in) the story update. A new revision of the news-item is created, and the collection in which the news-item appears is also updated and distributed to all client web sites.

Use Case: Multimedia package	
11	After some time and some revisions of the news-item, the story is of less importance; Arthur changes the order of articles in the 'Political news' collection, and places the "News government" article at the fifth position for some time (only the collection news-item is modified), then removes it from the collection. At this point, the article is automatically removed from all client web sites.

## 5.2 Push distribution

Use Case: Push distribution	
<b>Context of Use</b>	Producer of 'ready to publish' news, targeted to the news Industry
<b>Scope</b>	Technical Goal
<b>Level</b>	Primary task
<b>Primary Actor</b>	News provider
<b>Preconditions</b>	The news providers has a NewsML enabled CMS, a formatting platform and a multicast delivery network.
<b>Minimal Guarantees</b>	-
<b>Success Guarantees</b>	The news organization is able to deliver collections of multimedia news stories and syndicate them to all client sites in the format required by the customers. The delivery mechanism supports encryption and digital signature of the news.
<b>Trigger</b>	
Stakeholders	
Stakeholder	Interest
R&D	Needs to deliver the news as a multimedia story, with text and associated photos in a reliable (transparent backup support) and secure way (encryption and digital signature).
Description	
Step	Action
1	The first point is to organize the news content in terms of news feeds products. For example each news could be identified by the following quintuple: AGENCY:PRODUCT:SUBJECT:SUBJECT_MATTER:SUBJECT_DETAIL A customer profile should be a set of quintuple with the previous format.
2	The multicast delivery system (MDS) should support the definition of virtual channel on top of the physical multicast network. For convenience, we could define three channels: text, photo and audio/video channel. In the case of a text + photo news feed it could be useful to define also a fourth channel. Moreover, a special control channel should be defined to deliver administration information such as customer profile configuration, encryption key, ...
3	The MDS system should include a client and server application.

Use Case: Push distribution	
4	<p>The MDS server should provide the following features:</p> <ul style="list-style-type: none"> <li>- interface the news content management system in order to get the news content and the update</li> <li>- convert the content in NewsML if needed</li> <li>- encrypt and sign the content</li> <li>- fragment and delivery the news on the multicast network</li> <li>- implement some reliable mechanism (retransmission, ack/nack, etc)</li> <li>- implement the support for a transparent backup mechanism in case of fault on the multicast network</li> <li>- provide a user interface to define new channels and to administrate the customer profiles</li> </ul>
5	<p>The MDS client should provide the following features:</p> <ul style="list-style-type: none"> <li>- process the control message. Each control message is encrypted through the customer public key. So, only the right client could read the control message that contains its customer profile</li> <li>- control message contains the list of active channels for the customer and the customer profiles. The profiles should be used to check if the received news is compliant with the user needs</li> <li>- the client should be able to request the retransmission of missing packets and to switch to the backup channel in case of fault in the multicast networks</li> <li>- the client should apply on the news a stylesheet in order to convert NewsML news into the required format. This function could be used to maintain the compatibility with editorial system which do not support NewsML</li> </ul>
6	The MDS system should support the conversion of internal CMS links in public links.
7	The MDS could be used inside the conversion process for public links
8	The MDS should send the referenced items before to send the referring news
9	The MDS client should check for broken links in the case of NewsML news that point to a news-item that is not in the customer profiles. The client maintains a catalogue of received news and validates the news links by means of it. In case of news-items that are not present in the catalogue, the news-items will be deleted. A message should be provided in replacement of the deleted news-item.
10	The MDS client should be installed from internet. The client should be activated when it receives its control message out of the multicast network, with the list of channels and profiles. The message should be encrypted by means of the customer public key.
11	The MDS client should manage online any update to the user profiles and channels list.

### 5.3 Legacy format consolidation

Use Case: Legacy formats consolidation	
Context of Use	<p>Large news provider sending various feeds to its customers</p> <p>Small news provider using only parts of the overall functionality</p>
Scope	Technical/Editorial Goal



Use Case: Legacy formats consolidation	
<b>Level</b>	Primary task
<b>Primary Actor</b>	News provider
<b>Preconditions</b>	The news consumer has a NewsML enabled receiving software or CMS. The news provider has a NewsML enabled electronic DayBook/EventBook or multimedia editorial CMS or some other way to share metadata across various media desks.
<b>Minimal Guarantees</b>	
<b>Success Guarantees</b>	Ability to service all of the customers using the same software receiver. Ability to dynamically link related content from various media products
<b>Trigger</b>	A huge railroad accident near the capital
Stakeholders	
Stakeholder	Interest
R&D	A news provider FastNews.com uses ANPA1312 text feed for its US and IPTC7901 text feed for its European customers. For online market provider uses NITF. Photo customers are served using IIM or JPEG, while audio and video feeds are delivered using proprietary formats.  R&D was assigned the task to decrease the costs of software maintenance and to better integrate all of various news products.  To achieve that, FastNews decided to consolidate all its feeds to NewsML. For customers that are not able to receive NewsML feed, data would be converted on-the-fly back to the legacy format of their choice
Editorial	Needs to better integrate all of the content that is being produced (irrespective of the media) without needing to create explicit links between the particular items in various media products.
Description	
Step	Action
1	The main news desk sends out very short textual news (alert) with only the very basic metadata assigned to it.  For the customers that are not able to receive NewsML feed, receiving software automatically generates the legacy format of their choice.
2	The main news desk puts the information about a new event into the electronic EventBook/DayBook/CMS and assigns additional metadata to share it with other desks.  Information of a new event is sent out to EventML enabled customers.
3	The main news desk updates the alert with the data known up to this point to a short version of the news and assigns the additional metadata from the EventBook. (see the Breaking News usage case).  NewsML enabled customers see the update of the story and the metadata. For legacy format customers receiving software automatically generates a new story with reference in the text to the previous one (see also...)

Use Case: Legacy formats consolidation	
4	<p>When photos from the place of the accident get ready, photo desk assign the metadata on the event from the EventBook and add photo-specific metadata.</p> <p>For NewsML enabled customers, receiving software automatically relates the photos to the previously published textual versions. For the customers that are not able to receive NewsML feed, receiving software automatically extracts/generates the legacy format of their choice (JPG, IIM).</p>
5	<p>The main news desk sends out the full textual story with the data collected from various sources and with first comments from the officials. Story is assigned the general metadata from the EventBook plus the story-specific metadata like additional keywords or a comment to the customers what is being updated.</p> <p>Like in previous cases, NewsML enabled customers see the update of the story automatically related to all other relevant news-items and containing the full metadata, while legacy formats customers get the new story in the format of their choice with the metadata set suitable to that format.</p>
6	<p>The main news desk collects the background information about similar accidents from the corporate database, creates explicit link to the main news story and sends it out.</p> <p>NewsML receiving software links the background to the main news story and all related news-items and produces legacy format outputs for non-NewsML enabled customers.</p>
7	<p>Audio and video reports have been created and put to the corporate website and the download mirrors. A NewsML file containing relevant metadata and pointers to the content is sent out to the customers.</p> <p>NewsML receiving software, if configured to do so, automatically downloads the audio and video content to the customer's receiving station from the less stressed server and relates it to the other relevant information, or converts it to one of the legacy formats supported.</p> <p>If NewsML receiving software is configured not to download the content automatically, it just puts metadata about it to the receiver's CMS with the pointer to the actual content.</p>
8	<p>Multimedia desk creates a multimedia package from all the available sources and sends it out to the multimedia feed subscribers. Content that has already been pushed out need not to be sent again, but just linked. (See Multimedia package Use case).</p> <p>NewsML receiving software checks the content of the package and if some piece is missing, it automatically downloads it from the FastNews.com web server.</p>
9	<p>At the end of the day, link to the all related information is put to the top-stories-of-the-day collection. Again, only the pointers to the content are being sent out, relying on the receiving software to create the links to the actual content.</p> <p>NewsML receiving software allows recipients to extract only the information that is in format suitable to their needs, while still giving information about all the components of the package.</p>

## 6 Requirements

The sections below provide the requirements for the NewsML information exchange standard.

Each item is given a numeric identifier and a name.

Requirements are defined in short sentences. In many occasion a note (identified by the prefix "Note:") gives some highlights and background to the requirement.

Most of the items show a reference to the NewsML 1 version of the requirements (identified by "from NewsML 1 requirement xxx"); these cross-references are useful for checking that no original NewsML requirement is lost in the process by mistake. They will be moved to a bidirectional mapping table in a separate document at a later stage.

### 6.1 High level

#### 6.1.1 News life cycle

NewsML MUST be usable throughout the news life cycle.

(NewsML 1 requirement 200)

Note: This notion is intended to capture the characteristic processes through which news-items may pass between being created and being consumed. This life cycle includes but is not restricted to assignment, authoring & editing, storage & archiving, distribution, being searchable and so on. The next requirements of this group detail some specific aspects of this global requirement.

#### 6.1.2 Authoring

- 1) NewsML SHOULD be suitable for use with off-the-shelf XML authoring and editing software.
- 2) Undue configuration SHOULD NOT be required and extensive customization of editing applications SHOULD NOT be needed.

(NewsML 1 requirement 230)

#### 6.1.3 Storage

- 1) It MUST be possible for news-items to be stored and managed in a content management system and archived without transformation.
- 2) The conceptual model of NewsML SHOULD promote content reuse, i.e. reuse of news-items or components of news-items stored in a content management system.

(new)

#### 6.1.4 Exchange

- 1) It MUST be possible to exchange news-items using different kinds of push or pull distribution mechanism.
- 2) In particular, it MUST be straightforward for news-items to be carried over the Internet and on satellite networks.

(new)

Note: Some exchange protocols – eg SOAP - provide an XML envelope, some others are pure binary mechanisms. NewsML should support all kinds of mechanisms. NewsML provides natively an optional exchange XML envelope targeted to the multicast push mechanism often used by the news agencies (see the Exchange section).

### 6.1.5 Syndication

- 1) NewsML MUST be capable of representing the semantics required for news which pass through an arbitrarily complex chain consisting of the original provider and any number of other integrators, aggregators and distributors.
- 2) Value MAY be added at any point in this distribution chain.

(from NewsML 1 requirement 210)

### 6.1.6 User agents

The NewsML design SHOULD encourage implementers to deploy user agents that process NewsML.

(new)

### 6.1.7 XML based

The implementation of NewsML MUST be based on the family of XML standards.

(not new, but never said before)

### 6.1.8 Presentation

NewsML SHOULD NOT natively convey presentation semantics.

(NewsML 1 requirement 170)

Note: NewsML specifies the packaging of news components but does not specify the physical or temporal relationship between components when presented in a publishing medium.

## 6.2 NewsML model

Introduction: the NewsML model consists of a conceptual model and a processing model. The conceptual model captures the results of the requirements analysis from a structural angle and describes the most important NewsML objects. The processing model is a dynamic description of how NewsML objects shall be represented in-memory and handled by NewsML processors.

Note : Access rights management is not considered to be part of news processing. It will be applied through external applications. (from NewsML 1 note N310)

### 6.2.1 Conceptual model

NewsML MUST be based on a formal conceptual model (i.e. data model or abstract structure).

(new; maybe NewsML 1 requirement 730 is relevant)

### 6.2.2 Processing Model

NewsML MUST define a processing model, based on the NewsML conceptual model , which specifies how to handle instances of the objects defined by the conceptual model.

(new)

### 6.2.3 Construction

- 1) The NewsML processing model SHOULD describe how to build instances of the conceptual model from NewsML 1.x documents.
- 2) The NewsML processing model MUST describe how to build instances of the conceptual model from NewsML 2.x documents.

(new)

### 6.2.4 Details

The NewsML processing model SHOULD be specified in exhaustive detail, covering every aspect of NewsML that has model - or logical - significance.

(new)

### 6.2.5 Interoperability

The NewsML processing model SHOULD provide developers of NewsML processors with all the information they need to ensure that their implementations are fully interoperable with other implementations.

(new)

Note: Interoperability is the ability of software and hardware on different machines from different vendors to share data.

### 6.2.6 No interpretation

The NewsML processing model MUST be written to be easily and fully comprehensible to implementers with as little scope for interpretation as possible.

(new)

### 6.2.7 One way

The NewsML processing model SHOULD be written so that there is only one way to obtain one specific result.

(new)

### 6.2.8 Patterns

- 1) NewsML MUST define standard structural patterns for the representation and handling of the most usual types of news-items.
- 2) When dealing with news-items of a given type, all NewsML processors MUST handle these standard patterns.

(new)

### 6.2.9 Error handling

The NewsML processing model **MUST** describe error situations and how to handle them.

(new)

### 6.2.10 Application interfaces

NewsML **SHOULD** define a standard application interfaces, based on the conceptual model and applicable to the consumption of news.

(new)

### 6.2.11 Default behaviour

Default behaviour **SHOULD** be specified to allow sensible processing when optional elements (e.g. metadata) are not present in a specific news-item representation, or when unknown content is found and no application specific processor is available.

(from NewsML 1 requirement 720)

## 6.3 Content representation

### 6.3.1 Media-independence

- 1) NewsML **MUST** be able to handle content in the same way irrespective of its media type or format.
- 2) All media types **MUST** be treated equally.
- 3) NewsML **MUST** allow for the efficient representation of text, but NewsML **MUST NOT** assume that text is the primary vehicle for news.

(from NewsML 1 requirement 310)

### 6.3.2 Inclusion and reference

- 1) NewsML **MUST** support the inline inclusion of XML content, plain textual content and binary content.
- 2) NewsML **MUST** support, references to external content - i.e. content that is not included within the news-item instance but can be retrieved from an external location.
- 3) NewsML **MUST** support the explicit omission of content in news-items.

(from NewsML 1 requirement 520)

Note: Re item 1), there was a discussion about the need to include content that uses a different character set or character encoding than the main NewsML document (content with different character sets and encodings should be harmoniously integrated into a NewsML document.). But XML itself provides no mechanism for multiple charsets within a single document. Item 3) means that it is possible to include the description of some content, but give neither identification nor location of this content. The new XInclude standard provides a particular way of including XML content, and the WP will consider its impact on NewsML.

### 6.3.3 Encoding

NewsML MUST support the encoding of inline content.

(from NewsML 1 requirement 940)

Note: This is especially useful for the transport of binary content. NewsML 1 can represent a cascade of encodings (e.g. a Freehand graphics zipped then base64 encoded). Multiple-encoding has counter effects: it is not specified if a NewsML processor should or should not decode all encodings before content is stored as a separate resource, and the use of compression algorithms (e.g. zip) for transport purposes blurs the line between content file format and content encoding. Note that if zip is used for compression, the inner content should be atomic to fit with the NewsML packaging model. After discussion, the WP participants agree to keep only the basic encoding feature, and to rely on Characteristics to surface useful technical information about the content. What is considered as 'inline content' is what is encoded, and the media-type (mime-type or format) corresponds to this; if data is compressed by some application before being given to a NewsML processor, it is a matter of pre-processing or edition. Base64 is the most widely used encoding for binary content; in order to simplify NewsML processors, the encoding technique could be restricted to base64; but it could be wise to retain the ability to specify other encodings albeit for non-core implementations; it is a matter of specification. Remark: XSD schema support base64 (xs:base64Binary) and hexbinary (xs:hexBinary) as native encodings.

## 6.4 Metadata support

Introduction: The requirements for an IPTC Global Metadata Framework covering controlled vocabularies and topics will be defined by the IPTC GMF Working Group.

### 6.4.1 News-related metadata

NewsML MUST be capable of representing metadata necessary to support news as it passes through its lifecycle.

(from NewsML 1 requirement 150)

### 6.4.2 Minimal mandatory metadata

The number of mandatory metadata-elements SHOULD be as small as possible.

(new)

Note: This requirement tends to lower the minimum size of a news-item.

### 6.4.3 Metadata extensibility

- 1) NewsML MUST specify how a provider is able to add metadata-elements from other XML metadata structures or languages in their native format, using a namespace-based mechanism.
- 2) Such added elements MAY be alternative representations of metadata-elements defined by NewsML, or MAY form proprietary extensions.
- 3) A NewsML processor is only REQUIRED to process metadata defined by NewsML.

(from NewsML 1 requirement 151 & 152)

Note: This notion is intended to capture the importance of extensibility of metadata, and the need to provide an XML-standardized way to achieve this. A namespace based mechanism is preferred over the use of the current NewsML 1 Property element. NewsML should be able to transport any set of metadata elements represented in XML, be it Dublin Core, XMP or PRISM is its native form, in order to save undue processing for other applications (e.g. RDF applications).

#### 6.4.4 Meta-metadata

NewsML MUST support metadata information about any metadata-element, including the date of the last modification of the metadata-element and the importance of the metadata in the context of the news-item.

(new)

Note: This notion is intended to log in a standard way the evolution of metadata during news life cycle. As companies increasingly use automated systems to help editorial apply metadata, notions of confidence are crucial. In NewsML 1 this notion is supported by “assignment attributes”. These attributes represent: i) who has edited a metadata-element, ii) when the metadata-element has been modified, iii) a rating of the importance of the metadata ('High', 'Medium', 'Low'), iv) a rating of the confidence with which the current metadata was assigned ('Full', 'High', 'Medium', 'Low'), v) an indication of how strongly a particular piece of metadata applies (e.g. 'Prominent', 'Passing'). It was proposed (T.Fujiwara) to add information about who is assigned to the creation or modification of a given metadata-element.

#### 6.4.5 Types of metadata values

Each NewsML metadata-element MUST accept one or several of the following: i) arbitrary values, ii) rule-based values, iii) values drawn from a controlled vocabulary.

(new)

Note: A set of values may be closed (no term can be added by a provider) or open (a provider may extend the given set of values). The IPTC will define normative vocabularies, ie sets of mandatory values for some metadata elements.

### 6.5 Metadata classes

#### 6.5.1 Characteristics

- 1) NewsML MUST support a set of metadata-elements describing physical media-specific characteristics of content, including format and data size.
- 2) All characteristics SHOULD refer to content before encoding.

(new)

Note: NewsML seeks to let the content user know as much as possible about it before retrieving it. Characteristics specifically surface the metadata inherent to content that is of interest to content users. This information usually helps choosing between alternative content (e.g. a thumbnail vs a high definition picture). A provider may need to process the content in order to obtain some values, e.g. the number of words in a story, the format and resolution of a vector graphic stored as a zip file. This is actually the opposite approach to that detailed in the W3C document on 'Architecture of the World Wide Web', ([www.w3.org/TR/webarch](http://www.w3.org/TR/webarch)) where it says you don't (can't)



know the details till you get it; but the NewsML WP considers that our requirements go beyond what is described by the W3C.

### 6.5.2 Administrative and descriptive metadata

NewsML MUST support a set of media-independent metadata-elements that describe the content of a news-item, including items shown in the list below:

The date associated with the creation of the content. *Optional*

The location associated with the creation of the content. *Optional*

A creator of the content. *Optional and repeatable*

A contributor to the creation of the content. *Optional and repeatable*

A source of the content. *Optional and repeatable*

A subject (or category) associated with the content. *Optional and repeatable*

A style of expression associated with the content. *Optional and repeatable*

A language used in the content. *Optional and repeatable*

A topic (e.g. location, person, organization, event) associated with the content. *Optional and repeatable*

(from NewsML 1 requirement 116)

Note: Descriptive metadata tend to answer to the 5Ws of journalism (Who? When? Where? What? Why?) in a machine processable format.

### 6.5.3 Rights metadata

1) NewsML MUST support a set of metadata-elements that represent the copyrights and usage rights of a news-item considered as a package.

2) NewsML SHOULD support a set of metadata-elements that represent the copyrights and usage rights of each component of a news-item.

(new)

Note: This should preferably use an existing Rights Expression Language standard (e.g. MPEG REL).

### 6.5.4 Management metadata

NewsML MUST support a set of metadata-elements associated with the creation and management of a news-item, including items shown in the list below:

The globally unique identifier (i.e. unique, unambiguous and permanent) of the news-item. *Mandatory*

The revision number of the news-item previously identified (default 1 for the first revision). *Mandatory*

The news-item type (e.g. a simple text alert, a news agency picture, a multimedia document). *Mandatory*

The status of the news-item in the workflow (usable, embargoed etc.). *Optional*

The date of publication of the news-item. *Optional*

The date of last modification of the news-item. *Optional*

The date the embargo will end (or ended) on the news-item. *Optional*

The date the news-item will be (or was) cancelled. *Optional*

The creator of the news-item considered as a package. *Optional and repeatable*

The provider of the news-item. The entity responsible for making the news-item available. A list of providers (ordered or ranked) is possible in a syndication chain. *Optional and repeatable*

The editorial importance (urgency) associated with the news-item. *Optional*

(new)

Note: The relationships between the above elements and a chain of providers is to be studied and explained in the processing model. Should publishing information (issue number, page ...) be also defined?

### 6.5.5 Links from news-items to other resources

- 1) NewsML MUST support metadata-elements that represent qualified links from news-items to other news-items.
- 2) NewsML SHOULD support metadata-elements that represent qualified links from news-items to other resources.

(from NewsML 1 requirement 115)

Note: The relationship expressed between news-items is a permanent one. This is implemented as AssociatedWith in NewsML 1. It will be decided later if the qualifier (name of the association) is mandatory or optional in NewsML 2. The availability of the target resource is not required: in the case of a target news-item, the user of the link may have no right to access the target, and the target may even be created after its (future) identifier is used (case of previsions). If a URN resolution mechanism is provided, it is up to the resolution mechanism to cater with the possible unavailability of the target news-item.

Warning: There is a difference between news-item relationships used for navigation (e.g. “SeeAlso” or “IsTranslationOf” link) and composition of news-items, in the form of a collection of news-items in a compound new-item; the latter is described in the “Composition” section.

### 6.5.6 IPTC legacy metadata

NewsML metadata SHOULD be able to act as a replacement for existing IPTC sets of metadata (e.g. metadata supported by IPTC 7901, IIM 4 and NewsML 1).

(from NewsML 1 requirement 410)

Note: A metadata map should be provided between the different IPTC metadata implementations.

## 6.6 Composition

### 6.6.1 Alternative renditions

- 1) NewsML MUST support the alternative rendition of content in a news-item, i.e support alternative components with different content characteristics.
- 2) NewsML MUST provide a simple way to select one of the alternative components.
- 3) When several alternative components have a common root, it MUST be possible to determine the original rendition, i.e. from which the other renditions have been processed.

(from NewsML 1 requirement 320 and 325)

Note: Rendition here refers to the format, media-type, language or any other characteristic of content. Where more than one resolution of a picture is available it must be thus possible to identify the original high-definition alternative from which an image is derived. Alternative rendition is especially useful when content pieces are targeted to specific delivery channels. Alternative renditions of content have common descriptive metadata associated. This requirement fulfills the first item found in the W3C XML Accessibility Guidelines [XAG] (“Ensure that authors can associate multiple media objects as alternatives”).

### 6.6.2 Groups of news-items

- 1) NewsML MUST support a structure that allows news-items to be placed in named groups.
- 2) NewsML MUST support the definition of logically nested groups.
- 3) NewsML MAY support the representation of different kinds of groups of news-items:
  - ordered lists of complementary news-items
  - unordered lists of complementary news-items
  - lists of alternative news-items

(from NewsML 1 note N110)

Note: Metadata inheritance should not pass the barrier of a news-item, i.e. if a news-item is represented in another news-item (i.e. included by reference) it does not inherit from the parent news-item metadata.

Note: the original requirement 110 (“News-items have structure, i.e. are a composite of components of arbitrary media types with named relationships to each other”) was implemented in

NewsML as the ability to handle multiple complementary or alternative *physically nested* components (NewsComponents) in a news-item. It is agreed by the NewsML 2 WG that this structure leads to much confusion about what is a NewsComponent vs NewsItem, how embedded NewsItems should be managed etc... Such a structure does not seem necessary for the representation of a news-item, and is thus not described as a requirement.

## 6.7 Digital signature

Introduction: A digital signature provides integrity, message authentication, and/or signer authentication services for data of any type, whether located within the XML that includes the signature or elsewhere.

Note: The encryption of content is not supported by NewsML; encryption may be implemented at the transport level (e.g. SSL).

Note: The original requirement 820 (Signed Metadata) has not been implemented in NewsML 1 and has been dropped from the current requirements.

### 6.7.1 Signed news-item

NewsML MUST support the digital signature of news-items.

(from NewsML 1 requirement 810)

### 6.7.2 Signed content

NewsML SHOULD support the digital signature of each individual digital asset in a news-item, be it included by value or by reference.

(from NewsML 1 requirement 810)

## 6.8 Management

Introduction: A news-item is the unit of news as defined by the editorial or product construction practice of a news provider. A news-item is a package of components; this is a unit of management, exchange and storage for NewsML.

### 6.8.1 Identification

- 1) Each news-item MUST have a globally unique identifier, i.e. an identifier that is unique, unambiguous and permanent.
- 2) This news-item globally unique identifier MUST take the form of a URN.
- 3) The globally unique identifier of a news-item MUST be independent of its physical representation.

(from NewsML 1 requirement 510 & 530)

Note: The constraints of time and space as well as the demands of specific applications and delivery environments make it necessary to support different physical representations of the same news-item. Because the component parts of a news-item may be included and excluded in arbitrary combinations, there is no standard representation [or canonical form] for a news-item. Hence NewsML must provide an identification mechanism through which it can be determined

that two instances of news-items are the same or not, without referring to their content. A news provider is not obliged to provide a urn resolution mechanism.

### 6.8.2 Revision

- 1) NewsML MUST provide a deterministic mechanism to allow applications to distinguish earlier revisions of news-items from later revisions.
- 2) Any modification of content, metadata or labels associated with a news-item MUST imply an increment of the revision.

(from NewsML 1 requirement 640)

Note: A modification of management metadata also implies a new revision: this was not the case with NewsML 1.

### 6.8.3 News development

NewsML MUST support news updates over time, such as addition, overwrite and deletion of components of news-items, or a modification of the structure of a news-item (e.g. order of news-items in a collection of news-items).

(NewsML 1 requirement 620 & 630)

Note: Such an evolution may result from the different timeliness of availability of different media types (e.g. text may precede audio may precede video). Updates imply a revision of the supporting news-item.

### 6.8.4 Metadata update

NewsML SHOULD support the update of metadata elements only, without the need to exchange unmodified content.

(new)

Note: It should be possible to send out a new revision of a news-item that ONLY amends its metadata elements, something akin to the Amendment update in NewsML 1, a feature currently associated with management metadata only.

### 6.8.5 Management notice

NewsML MUST allow for providers to add non-publishable comments and statements to the news-items.

(new)

Note: NewsML should provide support for human-readable statements about content (eg giving details of Corrections), allowing basic markup (eg para breaks). Such feature will be covered by specialized labels.

## 6.9 Labels

Introduction: Labels expose aspects of news as natural language strings, and are typically displayed alongside the content of an object or in place of the object in a list of such objects,

providing a means of selection among them. Other kinds of labels are those notes, comments or instructions created by a news provider for use by recipient editorial teams. They are implemented as NewsLines and Comment in NewsML 1; examples are a headline or slugline.

### 6.9.1 Labels support

- 1) NewsML MUST support textual labels associated with news-items.
- 2) These labels SHOULD be represented in an expeditious manner.
- 3) NewsML SHOULD support different types of labels.
- 4) NewsML SHOULD let implementers create their own types of labels.

(from NewsML 1 requirement 120 + 1105)

### 6.9.2 IPTC legacy labels

NewsML labels MUST be able to act as a replacement for existing IPTC sets of labels (e.g. labels supported by IPTC7901, IIM 4 and NewsML 1).

(new)

### 6.9.3 Topic references in labels

NewsML SHOULD support the association of a part of a label with a topic (people, place, event, organization, subject etc.) via a topic reference.

(new)

Note: Topics references, especially references to named entities, can thus easily be translated to an hyperlink to any relevant resource during news processing.

### 6.9.4 Hyperlink in labels

NewsML SHOULD support hyperlinks in news-item labels.

(new)

Note: original request: “Within NewsLines, allow the use of the <a> element in addition to the currently allowed Origin element” (M.Wolf – Reuters)

### 6.9.5 Line breaks in labels

NewsML SHOULD support a lightweight facility for the representation of line breaks.

(new)

Note: structure in labels allows for better clarity, especially when they represent notes or instructions.

### 6.9.6 Variants of Labels

NewsML SHOULD support different labels in various languages in the same news-item.

(new)

Note: see also the Internationalization requirement that implies that some specific tags may also be present for internationalization purposes. The WG thinks about a generalization to other types of variants such as 'plain' and 'rich' or 'escaped-html' text.

## 6.10 Exchange

### 6.10.1 Exchange envelope

- 1) NewsML MUST define an exchange envelope, optimized for the multicast delivery of news.
- 2) The envelope MUST support the exchange of a set of news-items.
- 3) The envelope MUST support the exchange of a set of topic-items.
- 4) The exchange envelope MUST NOT be mandatory, so that other exchange mechanisms can be used instead.

(new)

### 6.10.2 Exchange properties

- 1) The NewsML exchange envelope SHOULD support properties relative to the exchange of news.
- 2) It SHOULD be possible to add workflow properties to the envelope, in order to extend the exchange mechanism with an external workflow control.

(from NewsML 1 requirement 260)

Note: NewsML allows for the capture and retention of some aspects of workflow process information as exchange properties. NewsML does not natively support workflow semantics; standard mechanisms may use the exchange properties to allow routing of news-items through the editorial and production processes.

## 6.11 Forward and backward compatibility

### 6.11.1 Conceptual model backward compatibility

- 1) Major versions of NewsML SHOULD be backward compatible at the conceptual model level.
- 2) Providers SHOULD be able to map NewsML instances from a given version to the next one without loss of features.
- 3) If some features are seen as being outside the scope of the NewsML schema, an alternative complementary standard SHOULD be proposed by the IPTC. Content representation, structure, metadata, news management are in the scope of the requirement.

(from NewsML 1 requirement 710 + extended)

### 6.11.2 Syntactic backward compatibility

- 1) Minor versions of NewsML MUST be backward compatible at the syntactic level.
- 2) New elements and attributes may be added as requirements become more explicit, but elements and attributes SHALL NOT be suppressed, so a NewsML X.x compliant processor can process all instances that are compliant with previous minor versions of NewsML X.

(from NewsML 1 requirement 710 + extended)

### 6.11.3 Application forward compatibility

- 1) NewsML processors SHOULD be forward compatible when dealing with different minor versions of NewsML. A forward compatible system needs to cope gracefully with unknown future features [FOLDOC].
- 2) NewsML processors which receive a NewsML instance of document type later than version X.x may not be able to handle any new elements and attributes with their version X.x compliant software. These systems SHOULD ignore any unrecognised elements and attributes, and elements contained by the unrecognised elements, without causing a system processing failure (*Must Ignore All* rule).
- 3) If systems are not built with this protection then they MUST be capable of recognising a document type other than version X.x and discard these documents.

(from NewsML 1 requirement 710 + extended)

### 6.11.4 Bidirectional transformation

The IPTC SHOULD provide a toolkit that allows for a bidirectional transformation between two consecutive major versions of NewsML.

(new)

Note: Such IPTC toolkit could for example be based on the XSLT technology.

## 6.12 Compatibility with other standards

### 6.12.1 Support for and conformance with non-IPTC standards

- 1) NewsML SHOULD be consistent with widely adopted techniques and published international standards and recommendations from other standards bodies.
- 2) The IPTC SHOULD liaise with other standards bodies as appropriate.
- 3) The IPTC SHOULD consider the incorporation of relevant new or updated standards into NewsML as they become available.
- 4) In addition NewsML SHOULD target maximum leverage from utilities, components and tools which are widely available for its processing.

(from NewsML 1 requirement 1001 + 1010)

Note: This aims at gaining maximum leverage from the intellectual effort that has gone into creation of such standards. Standards bodies include the W3C and OASIS.

As of 2004, a list of widely used standards and possible choices is:

- **XML namespaces** for xml modularity and metadata extensibility.
- **XML Schema** and/or **RelaxNG** for validation.
- **W3C Date-Time** [W3C.NOTE-datetime-19980827] as date format. A Date construct is an element whose child content is a W3C Date-Time string. The date format would allow for partial dates. Partial dates can have some details missing (only the year is known, only year/month etc...).
- **RFC 3066** as a specification for language tagging.
- **XLink** as a syntax for links.



- **RDF** for metadata representation.
- **Topic maps** for topic networking.

### 6.12.2 Conformance with other IPTC standards

1) NewsML MUST be able to act as a wrapper for content marked up with other XML based IPTC standards, like NITF and SportsML.

(new)

2) NewsML - used along with other XML standards like NITF - MUST be able to act as a functional replacement of older news industry standards like ANPA1312, IPTC7901, the IIM and NewsML 1.

(from NewsML 1 requirement 250)

Note: A NewsML instance can be transformed into and from a range of formats used in the news industry. Since NewsML provides for considerably greater richness of structure than other formats such back-transformations may involve a loss of information.

## 6.13 Universality

### 6.13.1 Internationalisation

NewsML MUST be fit for usage with any character set, language, and writing system. This particularly affects textual content and labels.

(from NewsML 1 requirement 340 + 10646)

### 6.13.2 Accessibility

NewsML SHOULD be designed in a way that lowers barriers to content accessibility for people with disabilities..

(new)

Note: NewsML design should follow the specifications found in the W3C XML Accessibility Guidelines [XAG], currently in draft status (last draft: 3 October 2002).

## 6.14 Usability

### 6.14.1 Simplicity

Simple things SHOULD be made simple. For example NewsML SHOULD:

- provide some short form of textual expression for news alerts;
- provide an efficient representation for captioned pictures.

(new)

### 6.14.2 Conciseness

The syntax of NewsML SHOULD be concise.

(NewsML 1 requirement 910 - part)

Note: NewsML is to be used in a wide variety of contexts with varying transmission capabilities. NewsML needs to provide the ability to express semantics in as concise a way as is possible while retaining the required flexibility.

### 6.14.3 Style & ease of use

NewsML SHOULD follow the styling conventions used by all IPTC standards, defined by the standards committee as part of the Roadmap 2005. For example, element names SHOULD be as specific and descriptive as possible.

(new)

### 6.14.4 Specifications

- 1) Core parts of NewsML, along with recommended extensions, SHOULD be available as XML schemas.
- 2) A DTD SHOULD be published for use by DTD compliant parsers.
- 3) Any rules that cannot be represented in an XML schema SHOULD be defined in the NewsML functional specifications.

(new)

### 6.14.5 User Manuals and On-line References

- 1) A set of documentation and guidelines with samples SHOULD be released at the same time as each new version of NewsML is released.
- 2) Those documents MUST be made available online, via the NewsML Web site.

(new)

## 6.15 Standard maintenance

### 6.15.1 Ongoing standard maintenance

- 1) A Working Party of the IPTC MUST provide ongoing maintenance for NewsML.
- 2) New versions of NewsML, which reflect modified specifications, SHOULD NOT be approved and published more frequently than annually.

(new)

### 6.15.2 Standard support

- 1) A web site providing all current reference and supporting material for NewsML MUST be maintained by the IPTC.
- 2) A public electronic discussion forum for asking for support and discussing NewsML issues SHOULD be provided by the IPTC.
- 3) The IPTC SHOULD assign appropriate persons to provide support to NewsML users by

electronic communication means as this is reasonable for a standardisation body driven by voluntary human resources.

(new)

## **7 Cross reference NewsML 1 vs NewsML 2**

*<This subsection will be added at a later stage.>*

## 8 Licensing

### LICENCE OF THE IPTC NewsML TRADEMARK

Use of the IPTC trademark shall be licensed by the IPTC ("the Licensor") to a Non-Member ("the Licensee") in consideration of the following obligations undertaken by the Licensee under the terms of this contract.

1. The Licensee recognises the Licensor as the sole owner of the intellectual property protected by the trademark.
2. The Licensee recognises that the Licensor has the right to grant licenses of the intellectual property protected by the trademark and has agreed to grant such a licence to the Licensee in the terms set out in this contract.
3. The Licensee shall not during the subsistence of this contract or at any future time register to use in its own name as proprietor any of the intellectual property protected by the trademark.
4. The Licensee shall not claim any right title or interest in the intellectual property or any part of it save as is granted by this contract.
5. The Licensee shall immediately call to the attention of the Licensor the use of any part of the intellectual property by any third party or any activity of any third party which might in the opinion of the Licensee amount to infringement of the rights protected by the trademark.
6. The Licensee shall not assign the benefit of this contract or grant any sub-licence without the prior written consent of the Licensor.
7. Use of the IPTC trademark is licensed only to those Licensees who comply with the requirements of the official published description of NewsML.
8. The Licensee promises to respect the integrity and quality standard of the trademark and shall refrain from all acts and omissions which threaten the integrity of the trademark as a mark of quality.
9. The Licensee shall communicate immediately to the IPTC any instances of actual or suspected misuse or non-compliance with the official published description of NewsML which come to the attention of the Licensee.
10. The Licensee shall, at the request of the IPTC Management Committee acting unanimously, accede to any reasonable request of the IPTC to inspect the address of the Licensee to verify compliance and each Licensee shall afford to the IPTC such assistance as is requested by the IPTC in response to the latter's reasonable enquiries in instances of suspected non-compliance with the official published description of NewsML requirements.

The Licensee shall from time to time provide the IPTC with the full address of its place of business and that place will be deemed the Licensee's address.

The IPTC reserves the right to terminate the use of the trademark by the Licensee at any time without notice or without the need to give reasons to the Licensee for such termination.

This contract shall be governed and construed in accordance with the laws of England and Wales whose courts shall be courts of competent jurisdiction.

## 9 Copyright and other Legal Notices

Copyright © 2004 by IPTC, the International Press Telecommunications Council.

All rights reserved.

NewsML™ is a registered trademark of the IPTC.

This project intends to use materials that are either in the public domain or are available by the permission for their respective copyright holders.

Permissions of copyright holder will be obtained prior to use of protected material.

In order of all interested parties to take advantage of NewsML™ only materials and information that their respective copyright holders will license to interested parties in a non-discriminatory manner will be used.

All materials of this IPTC standard covered by copyright shall be licensable at no charge.

## 10 Appendices and Supporting Information

### 10.1.1 IPTC documentation

**NewsML use cases**    NewsML use cases  
<http://www.newsml.org/xxx> *to be defined*

### 10.1.2 Resources

**FOLDOC**            Free On-line Dictionary of Computing  
<http://www.foldoc.org/>  
Ed Denis Howe

**XAG**                XML Accessibility Guidelines  
<http://www.w3.org/TR/xag>  
W3C Working Draft 3 October 2002

**BOX&ARROWS**    What is a controlled vocabulary?  
[http://www.bboxesandarrows.com/archives/what\\_is\\_a\\_controlled\\_vocabulary.php](http://www.bboxesandarrows.com/archives/what_is_a_controlled_vocabulary.php)  
by Karl Fast, Fred Leise and Mike Steckel

## 11 Change Log

### 11.1 Transition to Draft 3

- All instances of ‘must’, ‘should’, ‘shall’ replaced by capital letters, for better conformant with RFC2119 (‘shall’ not replaced in the licencing requirements).
- References to the original NewsML 1 requirements added inside parenthesis.
- Many changes in the structure, in order to delete unused parts; deletion of table captions
- Addition of some terms in the glossary
- Work on the introduction, on the business objectives
- Addition of the list of oncoming use cases
- Work on ‘scope of use’: It was too vague. Content reuse merged with content management as they fit together, and precision added in the item. 'Interchange' renamed 'exchange' (simpler), and item restated with more precisions; same for syndication (now a copy-paste from the original). Consultation is deleted; a NewsML instance can always be displayed, after proper transform. Transformation is moved to the ‘compatibility with other standards’ group. Now the list follows the main segments of NewsML usage (authoring, management, exchange and syndication of packaged news)
- Work on ‘content’: ‘media independence’ an ‘media equality’ are merged (same concept); ‘inclusion’ is renamed ‘inclusion and reference’ (more accurate); ‘Alert’ is added as a request of D.Gulija, agreed by all; ‘encoding’ is modified and base64 is not requested any more (implementation feature).

### 11.2 Transition to Draft 4

- A start has been made on the References section.
- The Definitions have been alphabetically ordered.
- More SHOULDs etc have been introduced into the Requirements chapter.
- Explanatory sentences sprinkled among the requirements have been made into distinct Notes.
- Some unclear sentences have been made clearer. Where the intent is unclear, questions have been added, enclosed in “[ ]” and terminating with “???”.
- The Internationalisation requirement has been changed from a SHOULD to a MUST.
- A separate Accessibility requirement has been introduced. Previously this topic was merged with Internationalisation.
- The use of styles has been improved, with Body Text used more widely.
- The Copyright notice has been moved to the start of the document, as is customary.
- The statement about the licensing of the NewsML trademark has been made into a separate chapter. It was previously included in the Requirements chapter, which doesn’t makes sense, as that chapter deals with requirements upon NewsML itself.
- The page numbering has been adjusted so that the last page doesn’t claim to be “N-1 of N”.

### 11.3 Transition to Draft 5

- The cross-references to NewsML 1 requirements have been made less visually intrusive.



- The separate Internationalisation and Accessibility requirements, introduced in the last iteration, have been brought together under the heading of “Universality”.

## 11.4 Transition to Draft 6

- Document URN reassigned (section 1.3)
- Section “Copyright and other Legal Notices” reintroduced as the statements there are required to be part of the formal requirements.
- minor format changes (blank lines in footer and header, case of TOC lines)

## 11.5 Transition to Draft 7 and 8

- Document file name, URN and revision history: all moved to page 2
- Addition of the use case “Push distribution” (ANSA – A. Marrara).
- Modification of “Business objectives”.
- Deletion of unused sections.
- Many updates in the requirements, especially ‘Content representation’, ‘Metadata’, ‘Management’, ‘Exchange’ and ‘Processing’ sections.

## 11.6 Transition to Draft 9

- Update of some definitions in the glossary.
- Addition of the Use Case “Consolidation of legacy formats” (Hina – D.Gulija).
- Content Representation: update of “Inclusion and reference”, addition of “Presentation” (moved from the Packages section)
- Metadata: extensive update
- Labels: extensive update
- Packages: update
- Addition of an empty “Cross reference NewsML 1 vs NewsML 2
- Sub-item numbering transformed (latin to numbers).
- Resources moved to Appendix, addition of a link to the use cases. Notes (RelaxNG and NUTS) removed from appendix.

## 1.2 Transition to Draft 10

- “Packages” section update; complementary components suppressed
- “Labels” section moved after management (news-item level)
- “Metadata” section divided in “Metadata support” and “Metadata classes” for better clarity. Items about metadata elements grouped in the latter. Addition of required metadata elements and explanation.
- “Presentation” requirement moved to the “Top level” section.
- “Conceptual model” requirement moved back to the “Processing” section

### 1.3 Transition to Draft 11

- “Management” section: update after conffall
- “Exchange” section: update after conffall
- “Global identifier” defined, “Unambiguous identifier” modified. “Unique identifier”, “Invariant identifier”: definitions suppressed.

### 1.4 Transition to Draft 12

- “Management” section: some updates during conffall
- “Usability” section: update during conffall
- “Forward and backward compatibility” section: update after conffall (no more reference to NewsML 1 and NewsML 2).
- Chapter 3: documentation roadmap added (mws)
- 2.2.4: RFC3085 added (mws)

### 1.5 Transition to Draft 13

- “Processing” section: updates during conffall; section moved up to 7.2 as “NewsML model” (in order to put more emphasis on the subject).
- “Universality” section: “Accessibility” item wording modified, as the Web content accessibility guidelines (WCAG) are out of the scope of NewsML and the XML accessibility guidelines (XAG) are still a draft; so it is problematic to include a reference to this document. A note added to “Packages”/ “Alternative rendition” about the relation between this requirement and XAG.

### 1.6 Transition to Draft 14

- “Metadata support” section: updates during conffall.
- “Metadata classes” section: updates during conffall.
- “Packages” section renamed to “Composition”
- “Primary customer” merged with “Business objectives”

--eof--