



IPTC Standards DRAFT

NewsML 2 Architecture

Version 1.0

Documentation

About Experimental Phase 2

Document Revision 2



Copyright © 2006 by IPTC, the International Press Telecommunications Council. All rights reserved.

IPTC address:

Postal mail: Royal Albert House, Sheet Street, Windsor, Berkshire, SL4 1BE, United Kingdom

Web: www.iptc.org

email: office@iptc.org

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.

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

This document is released under the terms and conditions of the IPTC Intellectual Property Policy:

http://www.iptc.org/download/public/IPTC-IntellectualPropertyPolicy_2006.2.pdf

Document file name (+ Word file extension “.doc”): **DRAFT-NAR_1.0-doc-AboutNARexpPhase2_1.doc**

Document URN: urn:iptc:std-draft:NAR:1.0:doc:AboutNARexpPhase2:1

(More information on IPTC URNs in RFC 3937)

Specification Versioning History

Version	Issue Date	Approved by	Remark
1	TBD	IPTC Standards Committee	Currently under review

Document Revision History

Revision	Issue Date	Author (revised by)	Remark
1	2006-05-08	Michael Steidl	Initial version, derived from the EP1,V5 document
2	2006-05-10	Michael Steidl	minor changes and updates for public release



Table of contents

- 1 About this document.....4**
- 1.1 Audience4
- 2 PART 1: Introduction to the NewsML 2 Architecture.....4**
- 2.1 Why the IPTC developed the NewsML 2 Architecture4
- 2.2 The goals4
- 2.3 The fundamental model5
- 2.4 Business advantages.....6
- 2.5 Future NAR-based IPTC standards.....7
- 2.6 The overall timeline7
- 3 PART 2: Experimental Phase #28**
- 3.1 About the Experimental Phase #28
- 3.1.1 Goals9
- 3.1.2 When – Who – How.....9
- 3.2 Available documents and files9
- 3.3 Experimental Phase #2 Guidelines10
- 3.3.1 Overview: what has been implemented (so far) – and what has been added or changed*) since to the Experimental Phase #1 specs:10
- 3.3.2 What is missing11
- 3.3.3 Notes12
- 3.4 Request for feedback.....12
- 3.4.1 General feedback.....12
- 3.4.2 Contributions.....12



1 About this document

This document provides a short introduction to the “NewsML 2 Architecture” developed by the IPTC and to its “Experimental Phase #2” from May 2006 to August 2006.

1.1 Audience

The IPTC NewsML 2 Architecture is primarily intended for, but not restricted to, use in the news industry.

[Part 1](#) of this document addresses persons at a news management and/or IT management level.

[Part 2](#) of this document addresses persons involved with the implementation of news technology.

2 PART 1: Introduction to the NewsML 2 Architecture

2.1 Why the IPTC developed the NewsML 2 Architecture

In 2003 the IPTC started to assess how its latest news exchange standard NewsML 1 succeeded in the news industry. A meeting of IPTC members was held that year in conjunction with the regular Autumn Meeting to receive a first round of feedback. Then the IPTC decided to conduct a detailed survey among members and non-members in late 2003 which resulted in an internal report in the Spring of 2004.

The main conclusions were:

- NewsML 1 merges the functionalities of wrapping content and packaging news items into a single concept – this was felt as overstretching a single model, making it too complex.
- NewsML 1 allows the expression of semantically the same information and structures in multiple ways – this is a burden to processors on the receiving side.
- The NewsML 1 syntax for expressing metadata values is too limiting.

In the Spring of 2004 the IPTC decided to tackle the conclusions from this assessment and to start developing a next generation of NewsML. After a first round of discussions and a first set of requirements were compiled it was felt a more basic approach than just a superficial NewsML 1 update was needed. This was the starting point for the development of the “NewsML 2 Architecture (NAR)”.

2.2 The goals

The basic goal of the NewsML 2 Architecture is to provide a single generic model for exchanging all kinds of newsworthy information, thus providing a framework for a future family of IPTC news exchange standards.

This new model should cover the features of existing standards like NewsML 1 and SportsML, and also standards which are under development like EventsML, ProgramGuideML or a future WeatherML.



The detailed goals of the **NewsML 2** Architecture are:

- To simplify and unify the overall design for representing newsworthy information.
- To be flexible, thus allowing lightweight “no bells and whistles” feeds and highly complex news feeds, based on the same model.
- To specify more details, leaving less space for interpretation.
- To streamline the processing model, providing only a single way to express specific structures and functionalities.
- To develop a new model for expressing metadata from the ground up.
- To provide an abstract model to be implemented by specific news exchange standards.
- To maintain, at the functional rather than syntactic level, a high level of backward compatibility with NewsML 1.
- To simplify the implementation of IPTC news exchange standards as a whole.
- To align IPTC news exchange standards with requirements from the “Information Highway”.

2.3 The fundamental model

The NewsML 2 Architecture is built on a basic view of the relationship between the real world and news or topics as its abstraction:

News covers “what has happened and is fit to be published”:

- The basic tier is the **real world**: events, persons, organisations, all kinds of objects and activities.
- The real world is reflected in the **content of news** as a first level of abstraction. For a first approach this abstraction is agnostic to any technology. Whether a reporter writes a story with a pencil on paper or types it on a typewriter or uses a keyboard in front of a PC does not change the nature of this content. It also does not matter whether news is reported by text, by photo, by audio or by video. But this is also the level where technology enters the news arena: The NewsML 2 Architecture defines a basic and generic technical wrapper for different kinds of content expressed by different types of media.
- But the technical representation of content is not enough; dealing with news in a professional way requires specific **information about the content**, in order to summarize or to categorise it, to represent or describe it in a more formal way. These metadata are required as add-ons to the basic content and form the second level of abstraction from the real world.
- Finally, a **management layer** is added to this model: the combination of content and metadata about the content makes a technical construct called a News Item in the scope of the NewsML 2 Architecture. Only this Item as a whole provides a global and unique identifier to keep track of it and a set of information about the Item as such, primarily administrative information like dates of creation and revision of items, embargo dates, who provided the item etc. The crucial feature of this management layer is allowing revisions of an Item in a straightforward way. This is done for an Item by maintaining its identifier over



all revisions and only incrementing a version number, thus allowing easy identification of an update to a previously released News Item.

Topics cover “what is a matter of fact and is good to know”:

- The basic tier is again the **real world**: events, persons, organisations, all kinds of objects and activities. Further to that this tier includes also abstract concepts which may be used to categorise the real world.
- In this case, the world is reflected by writing up **factual knowledge about real world concepts**: Information representing any kind of object – which may be a person, organisation, company, material object – or any kind of abstract concepts like terms used to categorize content. This kind of information can be understood as background or reference for news.

The NewsML 2 Architecture wraps a representation of the concept in a structure called a Topic Item, adding also the same **management layer** as for the News Item, thus providing the same set of management functionality to both types of Item.

Beyond items acting as a carrier for the payload of news or knowledge this construct can also be used for facilitating the exchange of packages of news and assigning workflow management for the creation of news and topics:

- **Packaging** is collating a number of News Items and/or Topic Items in a structured way. This is facilitated by a Package Item, providing a mechanism to include only references to existing News Items, Topic Items or other kinds of Items, even Web resources, and to structure them like a tree. This Package Item provides the same management layer as the News Item or the Topic Item and hence can be managed in exactly the same way.
- **Assignment** information is required to manage the workflow of a news company. It relates to the production of News Items and Topic Items as this requires resources. By using an Assignment Item, human and technical resources can be assigned to the creation of content. This Assignment Item provides the same management layer as all other items and hence can be managed in exactly the same way.

It is evident that all types of Item are siblings in a basic manner: they all inherit the same management layer, providing a consistent structure and processing model for all of them. And all Items are flexible containers for “information which has to be conveyed” in the exchange process.

All Items also provide means for plugging in Common Components. A Common Component is a well defined structure to be used in several types of the Item, to convey the same semantics. Examples are metadata to describe content, structures to represent a person, an organisation or a location, or a structure for digital rights management.

Finally, the NewsML 2 Architecture defines an **envelope for transmitting** one or more **Items**, this envelope is called News Message. You can throw in all kinds of Items mentioned above and it adds only a thin layer of transportation management, eg when the message was submitted, by whom and to whom.

2.4 Business advantages

- The same things are done in a consistent way: faster to understand, easier to implement. For example, all types of Item share the same management layer; hence the IPTC defines a



unique way of managing digital assets completely independent of the nature of the asset. Another example is the consistent, media-agnostic, mechanism for describing content.

- Reuse of building blocks from the NAR allows the reuse of software components, making implementations cheaper
The NAR relies on generic building blocks to a high extent. This allows implementers to write software components which can be used for different kinds of Item, for different types of content and for different news feeds.
- The generic model allows future extensions; hence investments in implementing NAR-based IPTC standards are future-proof. Nobody knows about all future requirements for the news industry. But as the NAR provides a flexible generic model, it can be extended for future standards. The basic outline of the NAR – a family of Items with common building blocks plugged in – will stay the same for the foreseeable future.
- Making use of industry standards: allows processing with standard tools. The NAR syntax is built on XML, the Extensible Markup Language of the W3C. Furthermore, the NAR makes use of W3C XML Schema and complies with the basic notion of the Semantic Web, the Resource Description Framework (RDF). This allows an easy transfer of NAR structures to other XML-based standards and the integration of news and topics into the Semantic Web.

2.5 Future NAR-based IPTC standards

The IPTC will build all new news exchange standards on the NewsML 2 Architecture and future version of existing news exchange standards will be considered for aligning with the NAR.

Hence expect:

- A successor to NewsML 1 – to convey general news with a rich set of content related metadata but not providing packaging features. This standard will be built on the NAR, making extensive use of the News Item.
- A next major version of SportML will allow the use of its sports data structures as content of a News Item.
- The new EventsML standards will be built on top of the NAR
- The draft ProgramGuideML standard may be revised to allow integration into the NAR.
- The IPTC defined controlled vocabularies – the IPTC NewsCodes – will make use of NAR technology to express a set of topics and their relationships and will collate all corresponding codes into a single NAR based “concept scheme”.

2.6 The overall timeline

At the time of the launch of the Experimental Phase #2 – May 2006 – the IPTC has agreed on this overall timeline for the development of the NewsML 2 Architecture and standards built on it:

- May 2006: the draft specifications for the NAR EP2 are released
- May 2006: the news exchange standard (see 2.5) specific Working Groups at the IPTC commence their work of specifying their standards by reusing and extending the NAR specifications.



- August 2006: the news exchange standard specific Working Groups provide drafts for their specs and propose and recommend changes to the NAR
- October 2006: at the IPTC Autumn meeting the membership will agree on the news exchange standard drafts and the amended NewsML 2 Architecture.
- November 2006: after applying changes to the NAR (and to the content standards – as required) an Experimental Phase of the news exchange standards is launched – until December 2006.
- January 2007: the news exchange standards specific Working Groups and the NAR group will draw their conclusions from this Experimental Phase.
- March 2007: final draft for the news exchange standards specific standards – and hence implicitly the NewsML 2 Architecture – will be proposed for approval by the IPTC membership at the Spring Meeting.
- Summer 2007: a comprehensive set of specifications and supplemental documentation will be released for public use.

3 PART 2: Experimental Phase #2

3.1 About the Experimental Phase #2

The IPTC has developed a Model of the NewsML 2 Architecture, its Technical Specification and a technical draft implementation using W3C XML Schemas, which allows experimental implementations of processors.

IMPORTANT:

- The current documents (Model, Tech Spec, Glossary ...) are **not** the final and comprehensive definition of the NewsML 2 Architecture. They provide all essential and basic design approaches and specifications, but some details and refinements are still missing. See details about this in section 3.3
- The NewsML 2 Architecture is the underlying framework for standards to exchange specific content or knowledge. As said in section 2.5 the details of these standards are still under development by the IPTC, hence only a more or less generic container is provided for the Experimental Phase 1. But any participant in this phase is invited to test conveying their kind of content inside the container.
- The focus of the current model is on a basic news management model only, a lot of further refinements are to be expected.
- The current specifications are not necessarily the final ones. Based on the results from this Experimental Phase modifications may be applied by the IPTC. Hence we appreciate any reports from hands-on experiences for being able to draw conclusions from real-world success or problems with the NewsML 2 Architecture.

More details on the intentions and on participation in the Experimental Phase #2 can be found below.



3.1.1 Goals

- To provide a common framework for the development of the next generation news exchange standards.
- Testing “real life” use cases against the NAR model and syntax, with a focus on the additions and changes to the specification since EP#1.
- Reporting back practical issues not spotted by the NAR developers.

3.1.2 When – Who – How

When in detail:

- 10 May 2006: a package of specification documents and a corresponding XML Schema implementation is available – this excludes the “Technical Specification” document for the moment. This document should be available in June 2006.
- All IPTC members and parties invited by IPTC members who are interested into the NewsML 2 Architecture development are encouraged to apply their use cases to the current draft.
- The IPTC Working Groups for General News Markup, EventsML and SportsML are invited to start building their standards upon the NewsML 2 Architecture.
- 29 August 2006: The Experimental Phase #2 closes. All users of the NAR are invited to send reports and requests for changes to office@iptc.org on this day by the latest.
- The NAR developer group will review the feedback.
- The NAR developers will discuss the proposed changes.
- The NAR developers will propose changes and additions to the specifications for EP2 based on their own development work since May 2006 and the recommendations from the IPTC Workgroups for General News Markup, EventsML and SportsML and other parties.
- 16 October 2006: Conclusions from the Experimental Phase #2 are drawn at the IPTC Autumn Meeting.

Who may participate: all IPTC members and parties invited by IPTC members.

How to participate: download the Experimental Phase 2 package, apply the draft implementation to your use cases and send your comments and replies (see section 3.4 for details) to the IPTC before 15 February.

3.2 Available documents and files

All documents and files are available from: www.iptc.org

Specifications: (in directory: <http://www.iptc.org/std-dev/NAR/1.0/specification/>)

- NewsML 2 Architecture Model introduction document: DRAFT-NAR_1.0-spec-Model-Introduction_1.doc
- NewsML 2 Architecture Core Model document: DRAFT-NAR_1.0-spec-CoreModel_2.pdf



- NewsML 2 Architecture Power Extensions Model document: DRAFT-NAR_1.0-spec-PowerExtModel_2.pdf
- NewsML 2 Architecture Technical Specifications: this document is not available at the start of the Experimental Phase #2, please expect it in June 2006.
- XML Schema files:
 - Framework-0.7.xsd
 - NewsItem-0.7.xsd
 - TopicItem-0.7.xsd
 - NewsMessage-0.7.xsd
 (XML Schema documentation files are available as HTML files in the sub directory XMLSchema-doc)
- File with IPTC metadata code scheme URIs and recommended aliases: IPTC-TempCatalog-inc_3.xml
- Draft IPTC NewsCodes required for NAR functionality NCD4NAR-Schemes_3.xls
(Excel file with a sheet for each scheme. For each code the full URI, a title and a definition is provided.)

Documentation: (in directory: <http://www.iptc.org/std-dev/NAR/1.0/documentation/>)

- NewsML 2 Architecture Glossary: DRAFT-NAR_1.0-doc-Glossary_1.pdf
- “NewsML 2 Architecture: About Experimental Phase 2” document: This document

3.3 Experimental Phase #2 Guidelines

3.3.1 Overview: what has been implemented (so far) – and what has been **added or changed*** since to the Experimental Phase #1 specs:

*) only the major changes and additions are indicated in this section, this is not an exhaustive list.

- The abstract **Any Item**.
 - A generic “core rights” component was added with label type values only.
 - Structures from GRDDL (see: <http://www.w3.org/2004/01/rdxh/spec>) may be used now.
 - In <itemMeta> the name of the element “instanceOf” was changed to “thread” and the element “altId” was moved to the “administrativeComponent”.
 - the <link> was constrained and opened up:
 - at he Core level only a “title” may be added
 - at the Power level are no predefined structures for metadata anymore but an extension point accepting metadata from any namespace.
- A set of general data types.
 - a Block Type for text to be rendered as text block was added.
 - ”administrativeComponent”: the element “altId” was moved in from “itemMeta”, the element “audience” was re-specified.
 - “descriptiveComponent”: was moved from the generic framework to the News Item specification as only news will provide a description of its content.



- A set of generic elements:
 - <childOf> was renamed to <broader> - to comply with other metadata languages.
- A set of **Common Components**: basic and sophisticated structures which are widely used.
 - **Aggregate Common Components** were added for persons, organisations (including companies), geopolitical areas, points of interest.
- Derived from Any Item: the **News Item** for general news content – almost completely implemented.
 - the model of the “contentSet” was changed, elements have been renamed
 - A catalog and a catalogRef element were added to allow the inclusion of scheme-URI/alias mapping information.
- Derived from Any Item: the **Topic Item** for providing information about a concept – almost completely implemented now.
 - The <contentMeta> wraps only administrative metadata now.
 - The <topicDefinition> component holds an extensible ConceptDefinition structure:
 - the basic ConceptDefinition structure is available for all types of concepts
 - for specific types of concepts, like persons, organisations (including companies), geopolitical areas and point-of-interest, the corresponding Common Components can be used to extend the basic ConceptDefinition.
- Derived from Any Item: the **Package Item** to facilitate packaging different kinds of content – almost completely implemented.
 - the sequence of elements in a “group” was changed to an unordered choice
- The **News Message** to wrap a set of Items.
- Major general changes:
 - for the whole NAR all namespaces defined in EP#1 were merged into a single one. The most obvious result from that move is that each item type has its own distinctive name for the wrapping element now: news:item → <newsItem>, topic:item → <topicItem>, pkg:item → <packageItem>.
 - namespace URI notation was changed from the urn scheme to the http scheme, hence also the namespace URI all other namespaces were merged into was changed!
 - Attributes to provide “internationalisation information” were added.
 - extension points were added (as necessary) to all major wrapper elements.

3.3.2 What is missing

- Most of a detailed Processing Model for the NAR.
- No details of a processing model of building “concept schemes” to represent controlled vocabularies, only the outline.
- No markup for text news defined:
 - As this will be covered by a special General News Markup working group currently no markup is defined. But as the NAR XML Schema allows the use of XML controlled by any schema in the directContent structure, experimental implementers are free to use the XML of their choice.



3.3.3 Notes

- IPTC NewsCodes URIs:
In the sense of the NewsML 2 Architecture each set of IPTC NewsCodes – a controlled vocabulary maintained by the IPTC – represents a scheme of codes and each code is represented by a URI.
As the NewsML 2 Architecture notates code URIs by Compact URIs (CURIEs), a formally defined URI for each scheme is required.
A list of all IPTC provided schemes and recommended aliases are available in this file:
IPTC-TempCatalog-inc_3.xml
- The NAR makes use of two kinds of IPTC provided controlled vocabularies = NewsCodes:
 - NewsCodes already specified by the IPTC
 - NewsCodes not specified by the IPTC yetDraft IPTC NewsCodes for the latter and required for NAR functionality are provided by the Excel file NCD4NAR-Schemes_3.xls with a sheet for each scheme

3.4 Request for feedback

Feel free to comment on anything.

3.4.1 General feedback

Please join and use the public “newsml-2” Yahoo group: <http://groups.yahoo.com/group/newsml-2>

3.4.2 Contributions

Any kind of contribution is welcome, e.g.:

- sample instances based on the NewsML 2 Architecture
- pieces of software to process XML instances based on the NewsML 2 Architecture
- XSLT stylesheets to transform XML instances based on the NewsML 2 Architecture

Before sending a contribution, consider this **DISCLAIMER**:

Any contribution is assumed to be provided under the IPTC Intellectual Property Policy (IPP) http://www.iptc.org/download/public/IPTC-IntellectualPropertyPolicy_2006.2.pdf

Clause 2 of the IPP says:

The Intellectual Property of the IPTC is developed largely from contributions from member companies and their staff. This is augmented by discussions in committees, conference calls, and electronic forums. Some of these activities may also provide opportunities for non-members to contribute information and opinions to the IPTC development process. The IPTC shall endeavour to ensure that to the extent that member and non-member contributions and related discussions are protected by intellectual property rights including copyright, patent, and trademark rights, those members and non-members will agree to waive the exercise of those rights against the IPTC if it uses any such protected material in its published standards and related documents (See also 8, Outside Standards and Related Materials). Referenced works, their titles, and extracts therefrom are also considered property of their source. Use of this material in Specifications and Materials will follow accepted IP practices.

Where to post contributions: into the “Files” section of the newsml-2 Yahoo list:

<http://groups.yahoo.com/group/newsml-2/files/Contributions/>