IPTC Standards DRAFT

NewsML 2 Architecture

Version 1.0

Documentation

Experimental Phase 1

Document Revision 5

Document file name (+ Word file extension ".doc")**: DRAFT-NAR_1.0-doc-AboutNARexpPhase1_4.doc**

Document URN: urn:iptc:std-draft:NAR:1.0:doc:AboutNARexpPhase1:5
(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	2005-12-30	Michael Steidl	Initial draft
2	2005-12-02	Michael Steidl	NewsML 2 renaming, facts for the Experimental Phase updated
3	2005-12-08	Misha Wolf	General clean-up
4	2005-12-11	Michael Steidl	Added more details
5	2005-12-16	Michael Steidl	minor updates to the wording and to the release information

Table of contents

1 About this document

This document provides a short introduction to the "NewsML 2 Architecture" developed by the IPTC and to the "Experimental Phase #1", lasting from December 2005 to February 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.

– In this case, the world is reflected by writing up **factual knowledge about real world concepts**: Information representing any kind of objects – which may be persons, organisations, companies, material objects – or any kind of abstract concepts like subjects used to categorize content. This kind of information can be understood as background or reference to 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 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 but not providing packaging features – 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 will 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".

3   PART 2: Experimental Phase #1

3.1   About the Experimental Phase #1

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 convey 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 #1 can be found below.

3.1.1 Goals

To allow IPTC members to get a grip on the NAR by:

– Testing "real life" use cases against the NAR model and syntax.

– Reporting back practical issues not spotted by the NAR developers.

– Helping to finalise the NAR Technical Specification.

3.1.2 When – Who – How

When in detail:

– 6 December 2005: Model and Technical Specification available.

– 19 December 2005: Full package available.

– 15 February 2006: End of test period. Test reports have to be submitted to office@iptc.org.

– The NAR developer group will review the test reports.

– 3 March 2006: A comprehensive report on the test period is filed for the IPTC Spring Meeting.

– 26 March 2006: Conclusions from the test period are drawn at the IPTC Spring Meeting.

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

How to participate: download the Experimental Phase 1 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/dev

Specifications:

– NewsML 2 Architecture Model document: DRAFT-NAR_1.0-spec-Model_12.pdf

– NewsML 2 Architecture Technical Specifications: DRAFT-NAR_1.0-spec-TechSpec_18.pdf

– XML Schema files:
    – Framework-0.6.xsd
    – NewsItem-0.6.xsd
    – TopicItem-0.6.xsd
    – NewsMessage-0.6.xsd

Documentation:

– NewsML 2 Architecture Glossary: DRAFT-NAR_1.0-doc-Glossary_1.pdf

– "About the NewsML 2 Architecture Experimental Phase 1" document: This document

3.3 Experimental Phase #1 Guidelines

3.3.1 Overview: what has been implemented (so far)

– The abstract Any Item.

– A set of general data types.

– A set of Common Components: basic and sophisticated structures which are widely used.

– Derived from Any Item: the News Item for general news content – almost completely implemented.

– Derived from Any Item: the Topic Item for providing information about a concept – basic structures only.

– Derived from Any Item: the Package Item to facilitate packaging different kinds of content – almost completely implemented.

– The News Message to wrap a set of Items.

3.3.2 What is missing

– No details for building "concept schemes" to represent controlled vocabularies, only the outline.

- Only stubs for named entity components, in particular for persons, organisations and locations, are provided. Expect further refinements on their structures.

- No markup for text news defined:
  As this will be covered by a special workgroup after agreement on the basics of the NewsML 2 Architecture, 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 Workarounds

- 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. As a temporary workaround the following URIs are provided:

Controlled vocabulary of:	temporary URI	temporary scheme alias
audiocoder	urn:iptc:concepts:audiocdr:	audiocdr
characteristicsproperty	urn:iptc:concepts:charprop:	charprop
colorspace	urn:iptc:concepts:colspace:	colspace
confidence	urn:iptc:concepts:conf:	conf
encoding	urn:iptc:concepts:enc:	enc
format	urn:iptc:concepts:fmt:	fmt
genre	urn:iptc:concepts:genre:	genre
howpresent	urn:iptc:concepts:howpres:	howpres
importance	urn:iptc:concepts:imp:	imp
location	urn:iptc:concepts:loc:	loc
mediatype	urn:iptc:concepts:medtype:	medtype
notation	urn:iptc:concepts:notat:	notat
ofinterestto	urn:iptc:concepts:ofintto:	ofintto
priority	urn:iptc:concepts:prio:	prio
property	urn:iptc:concepts:prop:	prop
provider	urn:iptc:concepts:prov:	prov
relevance	urn:iptc:concepts:rel:	rel
role	urn:iptc:concepts:role:	role
scene	urn:iptc:concepts:scene:	scene
status	urn:iptc:concepts:status:	status
subjectcode	urn:iptc:concepts:subj:	subj
subjectqualifier	urn:iptc:concepts:subjqu:	subjqu
urgency	urn:iptc:concepts:urg:	urg
videocoder	urn:iptc:concepts:videocdr:	videocdr

– The table below lists further temporary URIs for controlled vocabularies the IPTC has to create for NewsML 2:

Controlled vocabulary of:	temporary URI	temporary scheme alias
Conformance Level	urn:iptc:concepts:conflevel:	conformance
Content Class	urn:iptc:concepts:contclass:	contentClass
Item Class	urn:iptc:concepts:itemclass:	itemClass
Publish Status	urn:iptc:concepts:pubstat:	pubStatus
Item representation	urn:iptc:concepts:itemrep:	representation
Role of a description	urn:iptc:concepts:desc:	desc
Package group mode	urn:iptc:concepts:mode:	mode
Relationship with target	urn:iptc:concepts:rel:	rel
Kind of content rendition	urn:iptc:concepts:rendition:	rendition
Signal	urn:iptc:concepts:signal:	signal
Role of a title	urn:iptc:concepts:title:	title
Concept type	urn:iptc:concepts:type:	type
Why present	urn:iptc:concepts:why:	why

Be aware:

→ These URIs do not align with RFC 3739 for the "iptc" URN namespace - yet.

→ The scheme aliases are "recommended" only, not a normative specification.

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:
The IPTC considers all member and non-member contributions and related discussions to be provided free and clear of intellectual property protection encumbrances including copyright, patent, and trademark restrictions.

Where to post contributions: into the "Files" section of the newsml-2 Yahoo list:
http://groups.yahoo.com/group/newsml-2/files/Contributions/

3.4.3 "EP#1 Test Report"

The "EP#1 Test Report" should be sent by email to office@iptc.org on 15 February 2006 by the latest.

General Questions - A:
1. Did your XML tools validate the XML Schema without problems?
   If not: please provide vendor, product, product version of the tool.
2. Did your XML tools validate the example XML instances provided in the test package?
   If not: please provide which example(s) did not validate and which error message was shown.
3. Were you able to apply the values for metadata you intended to apply? Please consider the restrictions to values as explained in the experimental package documentation.
4. Were you able to apply structures (= XML elements and attributes) as intended?
5. Were you able to apply the kinds of content you intended?
6. Did you fully understand the semantics for XML elements and attributes from the specifications?
7. Did you have problems in understanding the model (= the hierarchy of elements and attributes)?
8. Did you have problems in understanding how the model and the implemented elements and attributes relate?

Specific **Questions, from the specification documents - B:**
(M = Model document, TS = Tech Spec document, G = Glossary document)
1. re M 5.3.1.1.1: Is this feature really needed in the standard, or should we leave it to provider defined extensions?
2. re M 5.3.2.3: Is it really a power feature?
3. re M 5.3.2.4: Should we restrict this feature to links pointing to Items.
4. re M 8.5: One of our requirements (4.8.2) is to be able to request "a complete collection of TopicItems belonging to a given scheme". How shall we implement it?
5. re M 8.5: Should we add a concept type (e.g. an person has 'person' as content class and 'artist' or 'musician' as concept type)?
6. re M 8.5: Is a CURIE the proper format for Concept Identifier, or should it be a full URI?
7. re TS 2.1: Should we place a punctuation character (eg "-") at the end of each namespace URN?
8. re TS 2.1: Should we change our namespace policy to not use URNs?
9. re TS 3.2.1.1: Please let us have your views on the question in orange.
10. re TS 3.2.1.2: Please let us have your views on the question in orange.
11. re TS 3.2.1.3: Please let us have your views on the question in orange.
12. re TS 3.2.1.4: Please let us have your views on the question in orange.
13. re TS 3.2.2.1: Can two different aliases be declared to be equivalent to the same URI?
14. re TS 3.2.2.2: Should we replace the alias declaration mechanism borrowed from Schematron with the xmlns attribute used for QNames?
15. re TS 3.2.2.3: Please let us have your views on the question in orange.
16. re TS 7.8: Is this a good selection of elements and attributes?
17. re TS 9.2.1: Should the values be hardwired into the XML Schema?
18. re TS 9.9: Should we make the time and time zone offset mandatory?
19. re TS 9.12.1: Should this be a direct attribute of the item?
20. re TS 9.12.1: Should the values be hardwired into the XML Schema?
21. re TS 9.17.2: Is representation a good name? Would type or form be better?

22. re TS 17.2.2: Is topic : link a good name, given that newsml : link connects objects (Items or Web resources) rather than concepts?
23. re TS 18.2.3.1: Should the role attribute be optional?
24. re TS 18.2.3.1: Should we provide an explicit sequence attribute for use with ordered groups?

Further comments and remarks not covered by the questions above will also be appreciated.

=== END of document ===