



IPTC SPRING MEETING 2005

held at the
Holiday in on the Bay hotel
1355 Harbour Drive, San Diego, CA 92101, USA

3rd to 5th March 2005

Minutes of the News Architecture Working Party held on 3rd March 2005

Document history				[Document URN: urn:iptc:workdoc:nar:0502:1]
Revision	Issue Date	Pages	Author (revised by)	Remark
unrevised	2005-03-19	5+25	Michael Steidl/Laurent Le Meur	

Present:

Chairman:

Walter Baranger, New York Times Company
Scott Calder, Mainstream Data
Guthrie Collin, The Associated Press
Dave Compton, Reuters Limited
Jay Cousins, RivCom [as contracted consultant]
Honor Craig-Bennett, PA NewsLtd
Arnaud Descamps, Relaxnews
Takahiro Fujiwara, EAST Co. Ltd
Stéphane Guérillot, Agence France Presse
Darko Gulija, HINA
Klaus Herwig, Deutsche Presse-Agentur GmbH
Niels Hojer-Pedersen, Ritzau Bureau I's
Rudi Horvath, Austria Presse Agentur
John Iobst, Newspaper Association of America
Hugh Johnstone, IPTC
Alan Karben, XML Team Solutions, Inc.
Dean Large, Business Wire

Laurent Le Meur, Agence France Presse
Jaak Leenknecht, Agence de Presse Belga
Johan Lindgren, Tidningarnas Telegrambyrå
Harald Löffler, ifra
Jayson Lorenzen, Business Wire
Angelo Marrara, ANSA
John Minting, United Press International
Peter Müller, SDA/ATS
Stuart Myles, Dow Jones & Company
Karl Oanes, CCNMatthews
Jean-François Richard, Agence France Presse
Hitoshi Saito, Nihon Shinbun Kyokai
Hiroshi Shinotsuka, Kyodo News
Klaus Sprick, Deutsche Presse-Agentur GmbH
Henrik Stadler, Tidningarnas Telegrambyrå
Charles Tichenor, The Associated Press
Misha Wolf, Reuters Limited

Michael Steidl, IPTC Managing Director as Secretary

1 Chairman's introduction to the news working party

The WP Chairman said this Working Party was set up in the course of reorganising the IPTC working structure in January 2005. The scope of the work will be shown under item 2.

2 The IPTC News Architecture – introduction and overview

The WP Chairman made a presentation of the IPTC News Architecture – see slides 1 through 28 as Appendix 1 to these minutes.

Additional notes on this presentation:

Slide #5: The WP Chairman emphasised the IPTC has decided to move to W3C XML Schemas as primary tool to specify its XML based content markup standards.

Slide #11:

Misha Wolf said what exactly NewsML 2 is as a brand for some technology stuff has to be decided by the Management and Public Relations Committees.



Slide #12:

Misha Wolf said for the work on the News Metadata Framework more conference calls and more participants are required.

The WP Chairman responded we have to think about this carefully as we have to consider the (human) resources which are made available by the members.

DISCUSSION:

Johan Lindgren: the EventsML group does not have the requirement for two levels of conformance. Is this a must for all standards?

Misha Wolf said regarding the discussion on labels at the Yahoo group: it is required to have two levels of conformance and hence it has to be considered how to implement a specific level for a certain standard.

He added a requirement should be added to the NAR saying implementers should be enabled to do their own business within this framework. Currently it appears to him some of the NAR requirements are based solely on engineering considerations and not on business considerations. The IPTC has to have a way to meet the more complex requirements of Reuters' – maybe also other news agencies require them.

Darko Gulija demanded the provider should be able to extend a standard with specific requirements – while the basic features of a standard **MUST** be supported.

The WP Chairman said the rule must be: Providers are allowed to extend but not to change a standard.

Misha Wolf pointed at the current TopicSets as a good example for this kind of extension: Reuters use their own sets as they have created a very elaborate set of metadata which goes beyond what is provided by the IPTC TopicSets.

He added the IPTC should discuss if extended features as introduced by Reuters should be adopted for future versions of a standard.

Klaus Spring pointed at something he considers to be a key problem of making standards:

- either to include many options into a specification and finally having no standard anymore
- or to constrain having options and being able to model a solid standard.

He added he considers all IPTC members representatives participate in the standardisation work primarily for business reasons and not for engineering reasons only.

Stuart Myles said regarding conformance:

He thought the individual provider will decide which component or feature is covered by a “basic” and an “power” conformance level – in the presentation it was said this is to be decided by the IPTC.

The WP Chairman responded: the IPTC decides whether a component is in the “basic” or “power” level – the provider can decide to use the extended capabilities of the power level or not.

Stuart Myles proposed the IPTC members should be asked whether the current process of developing standards is appropriate.

Misha Wolf said he could not imagine how to get things done without a discussion between meetings. We have to balance available resources and the requirement to accomplish work.

The WP Chairman added currently about 10 persons are involved into the actual development work.



3 News Management Working Group (Stuart Myles)

The WG chair made a presentation of the work of the News Management working group – see slides 1 through 18 as Appendix 2 to these minutes.

DISCUSSION:

Misha Wolf asked if an item could opt out the management component?

The WG chair: yes and no, the component could not be opted out but everything except the identifier is optional.

Misha Wolf said if the status is part of a generic management component the allowed values have to be reviewed – for TopicItems a status like “retired” is required.

Karl Oanes asked if a mechanism for tracking distribution will be part of the management component.

The WG chair said this feature should only cover some kind of chaining of providers which were involved in the syndication process and this will not be part of the management component as this is a distribution issue.

He added IPTC standards can not solve the issue of checking business rules and the completeness of packages.

Darko Gulija said a “publishing” and a “rights” component could help for the case Karl raised.

The WG chair added currently no work is done on implementing something specific for a rights component.

Darko Gulija asked how to deal with indicating “outdated” news (e.g. a package of “today’s headlines” is not up to date tomorrow anymore)

The WG chair replied saying this would not be in the scope of management but rather in the scope of a “publishing” component.

Information about an embargo would be part of the management component, but would be expressed as a date, not as a status value.

Then it was discussed how to address the issue of allowing to store items in DBs easily without the requirement to update them for embargo state changes.

4 News Structure Working Group (Laurent Le Meur)

The WG chair made a presentation of the work of the News Structure working group – see slides 1 through 19 as Appendix 3 to these minutes.

Slide #10:

Misha Wolf said

DISCUSSION:

Misha Wolf asked how Reuters’ “Headlines” fit into this structure.

The WG chair replied this is a kind of General News and should be covered by NewsML.

5 Common Components Working Group (Johan Lindgren)

The WG chair made a presentation of the work of the News Structure working group – see slides 1 through 13 as Appendix 4 to these minutes.



Slide #9:

The WG chair emphasised the Common Components are for IPTC internal use only, they should not be available as discrete elements to the public outside.

DISCUSSION:

Stuart Myles asked how to reuse other non-IPTC standards.

The WG chair said we have to think about how to incorporate external standards, but currently this is left to the requirements expressed for such a component.

And we should be aware not to get involved with the development and maintenance of these standards – how would changes to these standards affect the IPTC component?

Stuart Myles pointed at Digital Rights Management standards and said he thinks this is a good candidate for such a common component.

Laurent Le Meur said we could think about an IPTC wrapper for external data structures.

Then the nature of metadata about a location or person were discussed:

- is this a common component

- or is it a kind of content component = a variant of a topicItem.

It was said if this is comprehensive information about a location or person it should be designed as a kind of topicItem.

But there is also a need for a construct inside e.g. a newsItem to reference this person or location related topicItem if the news content is about this person or content.

Finally the WG chair introduced the Common Components Model working document which can be downloaded from the Files section of the News Architecture Yahoo group.

6 News Metadata Framework Working Group (Misha Wolf)

The WG chair said the work of this group was in an idle state for some time but commenced to be at full power in February.

Then he made a walk through of version 8 of the News Metadata Framework Business Requirements document (DRAFT-NAR_1.0-spec-NMDF-BusinessRequ_8.doc) as available in the Files section of the News Architecture Yahoo group.

Finally he invited to join the Yahoo group on NMDF.

DISCUSSION:

Laurent Le Meur pointed at the extended approach to associate topicItems with content in this new framework. In the scope of NewsML1 only a topic inside a TopicSet could be referenced, now each topic should be a completely independent entity and it will be possible to assign associations also among topics creating some kind of network of topics this way.

Stéphane Guérillot: will the TopicSets survive?

The WG chair said the group is currently in the process of soliciting for requirements – so no decision made yet.



7 Presentation and discussion of the consultant's "Architecture discussion document"

Jay Cousins gave a presentation on the "Architecture discussion document" (available in the DRAFT-NAR_1.0-doc-ArchitectureDiscussionDocument_9.zip – package in the Files section of the News Architecture Yahoo group) – see slides 1 through 58 as Appendix 5 to these minutes.

After the presentation of the document Jay Cousins showed the decisions points arising from this document to the IPTC members. (Slides #53 through #57).

DISCUSSION:

- How to deal with these decisions as they require a lot of detailed considerations on very specific technology issues.
- It was proposed to extend the timeline for the project to allow the IPTC to discuss these decisions points for about two weeks after the meeting.

The WP Chairman said this discussion will be continued on the next day in an available time slot in the afternoon.

[continued discussion on 4 March 2005]

The WP Chairman showed a working document with current draft decisions for the points raised by the consultants.

This document was updated in the session to be presented to the Standards Committee on the final day of the Spring Meeting.


8 Any Other Business

There was no other business.

9 Date and Place of Next Meeting


6 – 9 June 2005, London, Bonnington Hotel in Bloomsbury

continued: Document Appendix 1,2,3,4 and 5 (+ 25 pages) >>>



News Architecture WP introduction & overview

Laurent Le Meur (AFP)
News Architecture WP chairman, IPTC
San Diego / March 03, 2005



NAR in the IPTC landscape

The IPTC (Annual General Meeting)

Management Committee					
Standards Committee					
News Architecture WP	News Structure WG	News Content WP	NITF Maintenance WP		
	News Management WG				
	Common Components WG				
	News Metadata Framework WG				
	NewsML WG	EventML WG	SportsML WG	ProgramGuideML	NewsML 1 Maintenance WP
Public Relations Committee					


© IPTC – www.iptc.org 2



A global framework

- Goal: achieve the "Roadmap 2005" ... in 2005, make the "Tower of standards" a real thing
- How: split the work, find the best frontiers, work for all standards
- Four WGs, four chairmen, good workforce
 - News management: Stuart Myles (Dow Jones)
 - News Structure: Laurent Le Meur (AFP)
 - Common Components: Johan Lindgren (TT)
 - News Metadata Framework: Misha Wolf (Reuters)


© IPTC – www.iptc.org 3



Charters

- NAR: Developing a generic architectural framework to manage and distribute all types of news related content.
- NSTR: Developing a structural/conceptual model for the representation of all types of news related content, to be adopted by all new major versions of IPTC standards.
- NMAN: Developing a management/processing model applicable to all types of news related content, to be adopted by all new major versions of IPTC standards.
- COCO: Developing components common to all new major versions of IPTC standards.
- NMDF: Specify how metadata will be expressed, referenced and managed in all new major versions of IPTC standards.


© IPTC – www.iptc.org 4



The help of consultants

- "Architecture Discussion Document"
- W3C XML Schema implementation
- Jay Cousins (Rivcom, UK/USA)
- Ulf Wingstedt (CNET, Sweden)
- Check the conceptual model against the capabilities of W3C XML Schemas
- Provide W3C XML Schemas templates


© IPTC – www.iptc.org 5



Today & tomorrow

- This overview (1 hour)
- Each WG has a slot (30 to 45 mn)
 - Present the work currently done
- Jay & Ulf have a large slot (2 hours)
 - Present the result of their work
- Tomorrow, Q&A session (1 hour)


© IPTC – www.iptc.org 6



NAR Goals

- Make it simple (for basic needs)
- Make it powerful (for high level needs)
- Make it easy to understand (clear concepts, ease of adoption)
- Make it modular (reuse btw IPTC standards)
- Make it **interoperable** (standard processing) (*)
- Make it **extensible** (for providers) (*)
- *Do not reinvent the wheel* (use other standards when available)
- Think forward (evolutivity)
- == Make it a worldwide standard ==


© IPTC – www.iptc.org 7



Extensibility

- From the consultant proposals:
- Extensibility is the ability to extend a content model to meet a specific requirement ->
 - Extensibility by design (common components)
 - Extensibility of content (controlled values with open schemes -> N MDF).
 - Extensibility of data model: extend the content model of a schema component in order to add further elements or attributes.
- Is the latter the mark of an 'extended' standard?
- Must be addressed at the COCO, NMAN, NSTR & N MDF levels plus at the level of each standard.


© IPTC – www.iptc.org 8



Interoperability

- Interoperability means the capability of two systems (of different type, model or manufacturer) to cooperate using exchanged information (whether connected to each other directly or through a communication system).
- The balance between interoperability and extensibility is still under investigation.


© IPTC – www.iptc.org 9



Warning on extensibility

- “The question of where extensibility is allowed and the scale of its use is a fundamental issue. Any extensibility will affect interoperability. For this reason, its use should be firmly controlled in order to preserve the interoperability and processing model of IPTC defined common components and of IPTC standards.” (NAR discussion document)


© IPTC – www.iptc.org 10



References

- NewsML requirements (2004)
- EventsML requirements (2004)
- Current generation of IPTC standards
 - NewsML 1.x, SportsML 1.x, NITF 3.x
- Work of other standards bodies
 - Atom, MPEG, PRISM, Topic Maps, RDF/OWL ...
- **It leads to a re-interpretation (re-shaping) of previous IPTC standards**
- **Don't be confused: NAR is not NewsML 2**

© IPTC – www.iptc.org 11



How do we proceed?

- From requirements (clearly stated needs)
- Via modeling
 - Conceptual model (extension from NewsML work)
 - Static view of the model
 - Processing model
 - Dynamic view of the model
 - News Management
 - Use of UML (for techies)
 - Object Oriented method <-> W3C XML Schemas
- To Specifications

© IPTC – www.iptc.org 12

Mails & Conference calls

- Discussions of dev forums (*)
- Conference call each Friday
 - Goal = reach agreement on subjects already discussed by mail
 - Each WG in turn
 - ~ 14:00 UTC
 - Advertised on the NAR-dev list with an agenda
- Notes sent to the NAR-dev after the confcall
- Please join at any moment to participate to the choices

© IPTC – www.iptc.org 13

Five dev forums

- For IPTC members only
- In <http://groups.yahoo.com/group/>
- News Architecture: [iptc-news-architecture-dev](#)
- News Management: [iptc-news-management-dev](#)
- News Structure: [iptc-news-structure-dev](#)
- Common Components: [iptc-common-components-dev](#)
- News Metadata Framework: [iptc-metadata-dev](#)
- Note: the newsml forum is not used for that purpose anymore.

© IPTC – www.iptc.org 14

Two users forums

- For all interested = external communication
- In <http://groups.yahoo.com/group/>
- News architecture: [iptc-architecture](#)
 - For structure, management & common components
- News Metadata Framework: [iptc-metadata](#)
- Will be active when information is ready to publish
- Note: it's a move in IPTC policy

© IPTC – www.iptc.org 15

A model, 5 documents

- News Management
 - DRAFT-NAR_1.0-spec-NewsManagement_2.pdf
- News Structure
 - DRAFT-NAR_1.0-spec-NewsStructure-Model_3.pdf
- Common Components
 - DRAFT-NAR_1.0-spec-CommonComponents-Model_4.pdf
- News Metadata Framework
 - DRAFT-NAR_1.0-spec-NMDF-BusinessRequ_5.pdf
- XML schema implementation
 - DRAFT-NAR_1.0-doc-ArchitectureDiscussionDocument_9.pdf

© IPTC – www.iptc.org 16

Main concepts

- News Structure
 - Abstract level
 - The Item: the unit of management -> News Management
 - The Component: a piece of content or a set of metadata -> Common Components & News Metadata Framework
 - Generic level
 - Package Item: a way to package Items
 - News Message: a way to exchange Items

© IPTC – www.iptc.org 17

News Structure (abstract)

The diagram illustrates the News Structure (abstract) as a vertical stack of components within a container labeled 'Item'. At the top is an orange box labeled 'Management Component'. Below it are three yellow boxes, each labeled 'Component'. To the right of the diagram, text explains the requirements for each level: 'Required, same for all standards' for the Management Component; 'Defined by an individual standard' for the generic Components; and 'Required or optional' and 'Single or repeatable' for the generic Components.

© IPTC – www.iptc.org 18

News Management

- A way to manage all types of news related content.
- An evolution of the NewsML 1.x management
- **Implementers of all IPTC standards will use the same news management**

© IPTC – www.iptc.org 19

Common Components

- Definition of 3 types of objects
 - Constructs
 - Properties
 - Data-types
- Constructs are the components that are imported in Items
- A construct contains properties and/or other constructs
- A property has a data-type
- **Implementers of all IPTC standards will use common components as much as possible**
- **Components created for one standard but useful for others will be added to the library**

© IPTC – www.iptc.org 20

Content vs metadata

Date: 2005/02/21
 Location: Brussels, Belgium
 Person: George W. Bush
 Person: Jacques Chirac
 Genre: reconciliation

BRUSSELS — Bush met with Jacques Chirac and hosted the French president at dinner. Chirac, who opposed the war, used conciliatory language. The dispute "in no way affects or in no way undermines the bedrock of our relations."

← about

Metadata

← about

Content (news or data)

Real life event

© IPTC – www.iptc.org 21

Metadata framework

- Properties (common components or specialized components) have values
- Property values may be controlled (only some values are allowed); but they are several types of rules for this.
- Property values may identify concepts
- Concepts have information associated to them (like names, descriptions)
- Concepts are associated to form ontologies
- The chunks of XML that describe concepts should be properly managed
- The N MDF WG looks for solutions for representing all this in a simple manner.
- **Creators of all new components will use the metadata framework**
- **Metadata will be processed the same way for all IPTC standards**

© IPTC – www.iptc.org 22

How to create a standard?

Item

Management Component

Component

Choose components in the Common library

Component	Component
Component	Component


Create specific components

© IPTC – www.iptc.org 23

Conformance levels (1)

- Make it simple! vs make it powerful!
- Levels should be layered
- All pieces of the puzzle exist in different flavors
 - Structure
 - Management
 - Common components
 - Metadata framework
- Addresses both structure and processing
- A global choice for a provider: no mix and match
- At least two levels: 'basic' (the core) and 'power'
- A choice for the processor:
 - "basic level compliant"
 - "power level compliant"


© IPTC – www.iptc.org 24



Conformance levels (2)

- 'basic' level: focus on *simplicity* and *interoperability*.
- 'power' level: focus on capability to fulfill *high level needs* of the news industry, and *extensibility*.
- The 'power' level is an *extension* of the 'basic' level; but it is not a 'do what you want' spec
- Requirement: to find a smart balance btw the features found at each level
- Target: a 80/20 ratio for implementers.


© IPTC – www.iptc.org 25



Next steps (1)

- Today: Presentation of the different pieces
 - News Management
 - News Structure
 - Common Components
 - News Metadata Framework
- Tomorrow
 - Q&A session
 - Request for confirmation: is the WP in the right direction?

© IPTC – www.iptc.org 26



Next steps (2)

- At the AGM in June (London)
 - Procedures for management of evolutions of requirements
 - Model (Structure/Management/Components) ready for adoption
 - NAR XML schema templates ready for adoption
- If possible, at the October meeting
 - A set of Common Components ready for adoption
 - A Metadata framework ready for adoption
 - NewsML and EventsML "specializations" ready for adoption

© IPTC – www.iptc.org 27



Thank you for your time



Check: <http://www.iptc.org>

© IPTC – www.iptc.org 28

IPTC SPRING MEETING

News Management Working Group
Thursday, March 3, 2005

Stuart Myles
Dow Jones & Company Inc
smyles@WSJ.com

1

News Management Working Group

- No minutes...
- Agenda
 - News Management
 - The current solutions in NewsML, NITF, etc.
 - Discussion of the proposed management component

2

News Management

- Often, news providers need to modify a news object which they have previously sent to a customer.
- For example, they may correct a headline, expand upon the body of a story or delete a piece of news altogether.
- This process of updating, deleting and modifying is known as "news management".
- IPTC needs to provide facilities for developing news over time

3

News Management

- Different news providers may have different news management policies
- The mechanisms that the IPTC provides must allow for providers and their customers to implement a variety of procedures
- News management isn't simple or obvious
- Defining sophisticated means of news management is important to the success of IPTC standards

4

IPTC News Management Today

- Currently, News Management is implemented within each IPTC standard
- For example, NewsML 1.X and NITF each provide their own news management facilities
- As other standards are created, such as EventsML and SportsML, they also need their own management mechanisms

5

IPTC News Management Today

- NewsML 1.X manages the evolution of News Items via several different mechanisms sprinkled throughout the DTD
 - Various types of identifier (newsidentifier, duid, euid, etc.)
 - The NewsManagement element with type, history, status and relationship information
 - The Update element
 - The Instruction element
- NITF has three "management" attributes within the docdata element, plus a "correction" element

6

IPTC News Management Today

- News management facilities within NewsML are uneven
- For example, how to manage things which are not NewsItems within NewsML, such as Topic Items? EventItems?
- NITF management does not have all the facilities of NewsML, although it has similar terminology

7

IPTC News Management Today

- News management needs to be “reinvented” (and therefore reimplemented) for each IPTC standard
- The exact syntax and mechanisms used for news management tend to differ between standards
- Combinations of IPTC standards can lead to duplicated and overlapping news management mechanisms
- Therefore, some items within a given news document can have more than one way to manage them, whereas others have no formal management mechanism at all

8

Time to change news management?

- As news providers and consumers use the existing standards, the strengths and weaknesses of the management mechanisms we created before have become clearer
- The recent set of “specialized content” standards in various stages of definition have made it clearer that management needs to be applied in more than two areas
- Broader rearchitecture of the IPTC standards presents an opportunity to redefine management

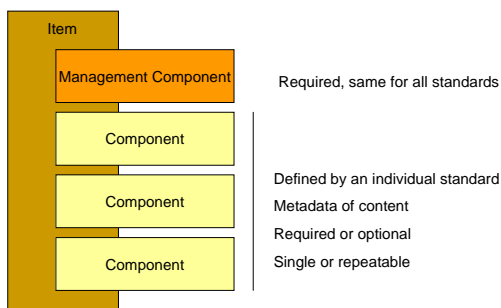
9

Proposal: Management Component

- As part of the rearchitecture of the IPTC standards, management is now a “component”
- The management component may be included within any “item” – news item, event item, topic item, etc.
- The management component is a specialized structure that governs the creation, evolution and destruction of items

10

Proposal: Management Component




11

Proposal: Management Component


- The management component brings the facilities from NewsML 1 to bear on all types of item, not just news
 - Several identification schemes (unique identifier, alternative identifiers, locators, filename)
 - Various kinds of “type” information (item class, media type, mime type, etc.)
 - Status
 - Importance
 - History and date information
 - Etc.

12



News Structure WG presentation


Laurent Le Meur (AFP)
News Structure WG chairman, IPTC
San Diego / March 03, 2005



Roots

- The work done in the scope of NewsML 2
 - First presented in Hong Kong (May 2004)
 - with three variants: Mini, Midi and Flexi-models
 - Evolution presented in Amsterdam (October 2004)
 - Based on the Mini model (the simplest)
- The choice to use the same structure for all IPTC standards -> News Architecture
 - Implication of other WGs
 - Creation of another work structure

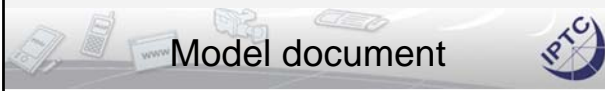
© IPTC – www.iptc.org 2



Charter & forum

- Charter: Developing a structural model for the representation of all types of news related content, to be adopted by all new major versions of IPTC standards.
- Conceptual model (most processing model -> NMAN)
- Forums at <http://groups.yahoo.com/group/>
- News Structure development:
iptc-news-structure-dev
- External communication:
iptc-architecture

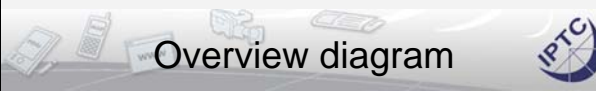
© IPTC – www.iptc.org 3



Model document

- News Structure
 - DRAFT-NAR_1.0-spec-NewsStructure-Model_3.pdf
- XML schema implementation
 - DRAFT-NAR_1.0-doc-ArchitectureDiscussionDocument_9.pdf


© IPTC – www.iptc.org 4



Overview diagram

- Please look at the News Structure model (in the current NAR-News structure model document)

© IPTC – www.iptc.org 5



About the UML diagrams

- Help for programmers
- Source code can be automatically created out of it.
- We can hope better implementations in C++, C#, Perl ... all OO languages
- If you don't understand it, read the text, there is more in it.
- But if you learn few 'glyphs', it is synthetic enough
- Other parties use it for clean design (e.g. OAIS:Open Archival Information System, MPEG21)

© IPTC – www.iptc.org 6

Two levels

- Abstract level: an object model.
 - Defines abstract classes (structure and processing)
 - “Specialized” classes inherit from these
 - Most processing model treated inside NMAN
- Generic level: a set of cross-standards classes, for packaging and exchange.
 - A specialized class inheriting from the abstract level
 - A generic class for broadcast oriented exchange

IPTC

© IPTC – www.iptc.org 7

Abstract classes

- Item (AnyItem): the unit of management
 - Simple or compound general news
 - Event description
 - Sports data
 - Package
- Component (AnyComponent): a piece of content or a set of metadata
 - Construct
 - Property

IPTC

© IPTC – www.iptc.org 8

An Item

The diagram shows a vertical stack of four boxes. The top box is labeled 'Item' and contains an orange box labeled 'Management Component' and three yellow boxes labeled 'Component'. To the right of the 'Management Component' box, the text reads 'Required, same for all standards'. To the right of the three 'Component' boxes, the text reads 'Defined by an individual standard', 'Required or optional', and 'Single or repeatable'.

IPTC

© IPTC – www.iptc.org 9

Metadata vs Content

- Currently no strict differentiation, both are components
- A construct (e.g. Person record) may be metadata in a certain context (extraction of a party in a NewsItem) but may be content in another context (part of an event description in an EventItem or maybe content of a PersonItem)

IPTC

© IPTC – www.iptc.org 10

A component

- Components have specific attributes
 - A local identifier, so they can be referred to or managed (updated/deleted)
 - *Maybe* assignment attributes (who creates it? is it reliable?)

IPTC


© IPTC – www.iptc.org 11

Updating Items

- Revise an item by sending modified pieces only
- Use cases (from simple to complex)
 1. Change a management property (e.g. status)
 2. Replace a metadata component
 3. Add a content component
 4. Delete/insert components
 5. Update xml information inside a component
- Original NewsML requirement: update metadata (use case 1 and 2)
- Could deprecate the U,A flags of the RFC30985

IPTC


© IPTC – www.iptc.org 12



Packaging & exchange?

- The NewsML WG (general news content WG) does not work on packaging & exchange anymore; packaging & exchanging news are 'cross-content' features defined by the NSTR WG
- It's up to the STA to define in which IPTC standard Package & Message classes should be described – it could be NewsML 2 or a specific standard.


© IPTC – www.iptc.org 13



PackageItem

- Goal: a news package with an internal structure
- Package:
 - Is managed
 - Has labels and description metadata
 - Has rights and publication metadata (power level)
 - May be extended with other components (power level)
 - Contains a hierarchy of references to items
- A package is an item, and so a package can be part of a package
- Example1: a main story, an analysis and a reaction (all news items).
- Example2: a story about a sports event and two pictures (news items), an event description (event item) and match results (sports item)


© IPTC – www.iptc.org 14



Group

- A Group contains ItemReferences and Groups
- Each Group and each Item has a specific *role* in its container Group
- Groups exist in three flavours (*mode*) : ordered, unordered, or alternative


© IPTC – www.iptc.org 15



ItemReference

- Item identifier + locator + hints
- Identifier: who is it?
- Locator: where is it?
 - Provides direct resolution of the identifier
- Hints: what is it?
 - No need to open the target item to know something about it
 - Name and description (basic level), maybe other descriptive metadata
 - Any kind of information given as components (power level), but no management information
 - May not be exact copy of the content of the item
 - Are included along with the item identifier (not "on the side")

© IPTC – www.iptc.org 16



getLatestVersion

- Power level
- Proposal for linking to the latest version of an item while giving the identifier of a given version
- No need to parse the item identifier to check the recommended process (not in RFC 3085)
- Needs a way to jump from an old version to the latest, on the processing side.
- Will be investigated further by the WG

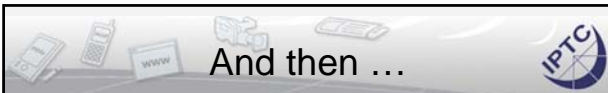
© IPTC – www.iptc.org 17



NewsMessage

- Used for exchange (broadcast)
- Exchange (transport) properties
 - dateExchanged
 - destination
 - channel
- Different namespace / item
- Contains a set of items of any kinds
- Alternatives : soap, syndication feed ...

© IPTC – www.iptc.org 18



And then ...

- Answer to consultant's questions
- Finalize the structure
- Include a Topicitem?
- Pass all processing given to NMAN?
- Present at the AGM:
 - The conceptual model completed
 - A draft syntax for generic classes (using agreed XML style)

© IPTC – www.iptc.org 19

IPTC

Common Components


What is that?

(DRAFT-NAR_1.0-spec-CommonComponents-Model_4)

© IPTC – www.iptc.org

IPTC

CommonComponents



© IPTC – www.iptc.org 2

IPTC

CommonComponents



Common Component "Corn Flakes"

Common Component "Milk"

© IPTC – www.iptc.org 3

IPTC

CommonComponents



Specialised Item "Coffee"


Using Common Component "Milk"

Container

© IPTC – www.iptc.org 4

IPTC

CommonComponents



Specialised Item "Breakfast"

Common Component "Milk"


Common Component "Corn Flakes"

Container

© IPTC – www.iptc.org 5

IPTC

CommonComponents



Common Component "Milk"

Specialised Item "Breakfast"

Common Component "Corn Flakes"

Other Container

© IPTC – www.iptc.org 6

CommonComponents

Common Component
"Corn Flakes"

Specialised Item
"Deli-dubious"

Other Component
"Wine"

Other Container

© IPTC – www.iptc.org 7

CommonComponents

Specialised Item
"Dessert"

Common Component
"Corn Flakes"
With added property – chocolate.

Other Component
"Ice cream"

© IPTC – www.iptc.org 8

CommonComponents

- Reuse of constructions
- Support standards work
- Ease implementation

© IPTC – www.iptc.org 9

CommonComponents

- Data Type (Example: String of certain length – "IPTC_string60_type")
- Property (Example: GivenName with a specific datatype.)
- Construct (Example: A <Person> with <GivenName>, <FamilyName>, <Title> etc)

© IPTC – www.iptc.org 10

CommonComponents

- DRAFT-NAR_1.0-spec-CommonComponents-Model_4:
- **anyComponent.** Core starting point for most of the actual Common Components.
- **Metadata components:** Signature, Rights, Publication, Label, Description.
- **Topic related components:** Person, Location, Organisation, Contact, Resource.

© IPTC – www.iptc.org 11

CommonComponents

- What will happen next?
- Continue work on Model document.
- Consultants will suggest Template.
- Produce the suggested components using the template.
- Maintain the CommonComponents.

© IPTC – www.iptc.org 12



CommonComponents

- <http://groups.yahoo.com/group/iptc-common-components-dev/>
- Email: johan.lindgren@tt.se

© IPTC – www.iptc.org 13

IPTC News Architecture

Jay Cousins
RivCom Ltd
jay.cousins@rivcom.com


Ulf Wingstedt
CNet Svenska AB
ulf.wingstedt@cnet.se

Expressing the IPTC News Architecture using W3C XML Schema

Background

The Consultants: Companies

- **RivCom** and **CNet** both provide independent, vendor-neutral consultancy to help organisations define, manage and communicate business information.
- Each has extensive experience in the development and adoption of XML standards to support publishing processes, having played leadership roles in the development of the NewsML, AdsML and IfraAdConnexion standards, among many others.

The Consultants: Jay Cousins

- Jay is a business process analyst and data modeller who works with XML and knowledge technologies for the creation, management, and distribution of information.
- Jay is Vice Chair of the AdsML Technical Working Group, and has co-edited the AdsML 1.0 and AdsML Structured Descriptions of Advertisement Objects specifications.
- Jay is a published technical author, having written on XML in the 'Professional Java XML' (2001) and 'Professional XML Meta Data' (2001) books published by WROX press.

The Consultants: Ulf Wingstedt

- Ulf Wingstedt, co-founder and senior technical manager at CNet responsible for XML products and services.
- Ulf has a long experience in working with media companies, including news agencies, newspapers, advertising media agencies and on-line media, applying XML standards to solve business needs.
- Ulf is currently working for Ifra as technical lead for the IfraAdConnexion and the AdsML Booking ad booking standards. He co-edited the AdsML 1.0 standard.

The Project

- An in depth study of the IPTC objectives and conceptual model of standards architecture
- Investigate alternative XML Schema representation of the architecture

Deliverables

- A decision support document with suggestions and samples
- A draft template for description of “common components”
- Participation at this meeting
- An implementation support package reflecting the decisions and choices made

The IPTC News Architecture Conceptual Model

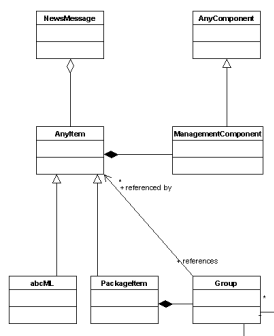
The IPTC News Architecture

- Important
 - Improve consistency, design and development of standards
 - Facilitate implementation and use
- Ambitious
 - The base for all future IPTC standards
 - Only very large organizations have done this before
 - E.g. UN/CEFACT (ebXML)
- The Challenge
 - To create a powerful and expressive architecture that is simple to maintain, understand and implement

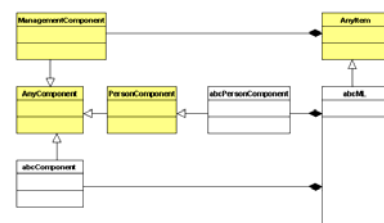
Base assumptions

- We have based our work on the versions of the News Architecture specifications available in the end of January 2005

The Abstract Conceptual Model



A Possible Conceptual Model for a abcML Standard



Overview of abcML standard as an implementation of a specific item type based on the conceptual model of the IPTC standards architecture. Common components (yellow) are being used as a base for specialized components for the abcML context.

Properties for an XML representation of the News Architecture

- Support for OO inheritance and object reuse
- Maintain the distinction between *semantic* layers of the news domain
 - Exchange, management, content metadata, content layers
- Maintain the distinction between *structural* layers
 - Abstract/generic, common components
- Provide a granularity of common components that simplifies reuse

Implementation of the Conceptual Model in XML

Reusable Common Components

- A reusable common component should be:
 - Well defined
 - Context free
 - Well designed with regards to granularity and semantic level

UN/CEFACT ebXML

- The most thorough work on component architectures today
- Has inspired many other groups including ANSI, OASIS and OAGIS
- Is likely too complex compared to the requirements from the IPTC, but could in a simplified version inspire the IPTC News Architecture's common component structure
 - The "80/20" rule

Proposed structure for the IPTC common components

- Data type
 - The lowest level, context independent without specific business semantics
 - Ex: String, Date, ShortString, CodeType, CountryCode
- Basic component
 - A single property or piece of business data with precise semantics. It is declared as a single data type.
 - Ex: Name, Age, SendingTime
- Aggregate component
 - A set of properties with precise business semantics. It can be composed of basic components, and other aggregate components.
 - Ex: Party (name, address), Location (lat, long), Administrative Metadata

XML Schema representation of the common component model

- Data types
 - Globally defined simple and complex types with simple content
 - Ex: ShortString: xs:string w. maxLength="50"
- Basic components
 - Globally defined elements
 - Ex: element: Name type="ShortString"
- Aggregate components
 - Globally defined (groups of) elements and complex types, with complex content
 - Ex: Person type:
 - Name ref="Name"
 - HomeAddress type="AddressType"

XML Namespaces

- Namespaces provide a generic mechanism for identifying “families” of XML structures (elements, types etc) and their associated processing rules (behaviour)
- When information objects are reused in new context, their semantics can be preserved by virtue of the namespace used to label its XML representation

Use of namespaces in the News Architecture

- Each item (“standard”) should have a separate namespace
- The common components should have a separate namespace - or namespace(s).
 - A single namespace for the complete set of common components, or,
 - Different namespaces for the three structural component types
 - Data types, basic and aggregate component
 - Or Different namespaces for semantically different components
 - RightsMetadata, Administrative Metadata etc.
- Simplicity – a single namespace for the common components

Validation of XML documents using XML Schema

- Two types of constraints:
 - Structures and content of XML documents
 - Possible to validate using XML Schema to a high degree, but not 100%
 - Processing and behaviour
 - Very limited support in XML Schema, has to be handled in application logic

Required extensibility

- Extensibility by design.
 - The ability to take existing data structures – i.e. common components – and assemble them in different ways in order to create new structures.
 - Useful for IPTC working groups creating new standards
- Extensibility of content.
 - The ability to extend the content carried and to extend the values of the content carried.
 - Useful for the IPTC and user organizations that need to adapt standards to local business practices
- Extensibility of data model.
 - The ability to extend the content model of a schema component in order to add further elements or attributes.
 - Useful for the IPTC and user organizations that have additional business requirements

Extensibility of Content

- Controlled vocabularies specified by the IPTC or local user organizations
- The generic approach:

```
<ProductCode codeList="company.se:ProdCodes" value="SPIK"/>
<ProductCode codeList="company.com:ProdCodes" value="NAIL"/>
```

- The explicit approach:

```
<ProductCode xsi:type="SwedishProductCodes">SPIK</ProductCode>
<ProductCode xsi:type="EnglishProductCodes">NAIL</ProductCode>
```

- The explicit approach allows XML Schema based validation, the generic approach can only be validated in application logic.

Guidelines for extensibility of Data Model

- Extensions should be clearly identified and designed in such a way that their use has minimal impact on other components and so their effect is isolated as far as possible.
 - For instance, use of namespaces in order to allow extended content to be unambiguously identified and ignored or processed by a processor as required by the implementation
- The architecture and modularity of (common) components should be such that they:
 - allow the extension of existing components with attribute or element content if required or, alternatively,
 - be such that the user can create new specific components to meet the needs of a specific domain.
- A document instance of an extended standard should be capable of validation to the same level as provided for in a non-extended standard.

Versioning and schema evolution

- Requirement
 - It must be possible to unambiguously identify the schema for an XML document instance
- How?
 - Namespace declarations
 - Schema file names
 - Schema name/version as document data
- Version data needs to be recorded in both schema and instances

Recording schema version in the schema

- A version attribute on the xs:schema element
 - "W3C good practice"
 - version="1.0.3"
- Make the namespace URI version sensitive
 - xmlns:car="urn:car-1.0"
- Make the file name version sensitive
 - CarML-1.0.xsd

Recording schema version in the instance

- Use the same version-sensitive namespace URI.
 - xmlns:car="urn:car-1.0"
- Specify a schemaVersion attribute on the root element of the instance document
 - schemaVersion="1.0.3"
- Identify the schema file name using the xsi:schemaLocation attribute.
 - CarML-1.0.xsd

Schema evolution: To be, or not to be compatible

- Examples of non-compatible changes
 - An optional and repeatable element is made optional with a maximum of 4 occurrences.
 - An optional attribute is made required
 - The data model is extended to add new attributes or elements that are not present in the previous version
 - The data model is changed to remove elements or attributes
 - ...

Backwards compatible changes

- Adding values to enumerations
- Optional extensions
 - Adding optional elements
 - Adding optional attributes

Major and Minor versions

- As often used on IT, a three level version model is proposed
 - Major versions are recorded using the first two digits
 - E.g: 1.0; 2.3
 - Minor versions are recorded in the third digit
 - E.g: 1.0.3; 2.3.2; 2.3.4

Proposed versioning policy:

- a major version change is defined as non backward compatible, and
- a minor version change is defined as backward compatible

Recording major and minor versions

- All three levels used in:
 - schemaVersion and version attributes in document instances and schema
 - E.g. 1.0.3; 2.3.4
- Only major level used in:
 - Schema file name
 - Namespace definitions
 - E.g: 1.0; 2.3

Versioning policy implications

- Major upgrades visible in instances, schema file name and namespace
 - Applications supporting the new version must be updated
 - Not possible for legacy applications to use new version data.
- Minor upgrades only visible inside instance documents and schema.
 - Legacy applications can use new version instance document with existing program logic (however not supporting new features) by using new schema
 - Applications supporting new functionality can query instance document to find exact minor version number.

Conformance levels

- Layers of functionality of a standard, grouped as “conformance levels”
- A conformance level structure should be:
 - Easy to understand
 - Include functionality that logically belongs together
 - Be layered, so that each higher level includes support for every lower level

A proposed model for conformance levels

- Conformance level 0:
 - Basic conformance, all base functions and connected content models must be supported.
- Conformance level 1:
 - Medium level feature, likely to be implemented by many applications
- Conformance level 2:
 - Advanced systems, using features important to less organizations.

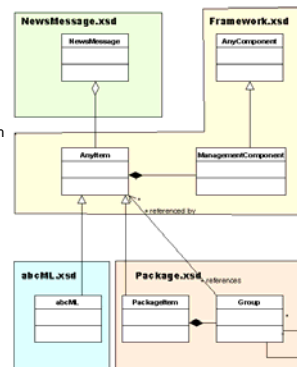
Expressing conformance in XML

- Functionality available in higher conformance levels may require elements or attributes that are defined as optional in the schema.
 - The data structures are thus mandatory for the higher conformance level, but these requirements are not expressed in the schema.
- Define separate schemas for each conformance level.
 - Likely confusing with several parallel version of the same standard in existence at the same time.
- A single schema can reflect requirements from different conformance levels by providing alternative content models
 - E.g: by choice, substitution groups or type substitution (xsi:type).
- XML document instances should include information about which conformance level the sender expects

Samples of XML Schema representation of the Conceptual Model

A schema file structure for the Conceptual Model

- Framework.xsd:
 - The collection of reusable structures including abstract base types (AnyItem and AnyComponent) and all common components.
- NewsMessage.xsd:
 - The generic messaging structure, i.e. the news envelope.
- Package.xsd:
 - The generic package structure for the assembly of news packages.
- abcML.xsd:
 - A specific item standard including reused common components and local structures.



Rationales for the proposed schema structure

- Every definition of an element, attribute, complex type or other construct should be made in only one place (i.e. a normalized set of components).
- The definitions and relationships expressed in the conceptual model should to a large extent be possible to express and enforce using XML Schema
- Each item level standard should be defined as separate schema since it should
 - be capable of stand-alone use (i.e. without any other item, or NewsMessage or Package)
 - be capable of validation outside the context of a news message.
 - have its own namespace.
- Every item must be able to re-use common components from the framework.
- As every item must be based on the AnyItem class, each item schema needs to import the framework, including the AnyItem type.
- AnyItem includes a ManagementComponent that is derived from AnyComponent defined in the framework. Thus, the common components and the AnyItem are closely connected and always used together. They are here in the same schema, the Framework.xsd.
- PackageItem is derived from AnyItem and should have its own namespace, as it is considered as yet another item. In the same way as other particular item level standards, PackageItem is then defined in a special schema in a separate file.

Use case walk through - XML Spy - Framework.xsd

■ The Framework Schema

- Namespace
- Abstract classes
 - AnyItem, AnyItemType, AnyAggregateComponent
- Common components
 - Data type – simple and complex
 - Basic component
 - Aggregate component

Use case walk through - XML Spy – Specific Item.xsd

- The CarML Schema
 - Namespace
 - Specific Item classes
 - Car extension of AnyItemType
 - Common components
 - Data types – creating new data types and controlled vocabularies by restriction
 - Restriction – controlled vocabulary creation and restricting to
 - Restriction – creating subsets of a controlled vocabulary
 - Aggregate component – restricting and extending existing components
 - Restricting – StrictPersonType
 - Extending – DriverType
 - CarML – instance
- The PackageItem Schema
 - The 'package' item, derived from AnyItemType

Use case walk through - XML Spy - NewsMessage.xsd

■ The NewsMessage Schema

- Namespace
- Carrying Items – including CarML in a NewsMessage
- Carrying Items – alternatives for validation of Item content:
 - Approach 1: The open envelope – any element, any other namespace, no Item validation
 - Approach 2: Only items endorsed by the IPTC – any element, specified namespace, no Item validation
 - Approach 3: Any item based on AnyItem – any element specified as a substitution for AnyItem element, Item validation
 - Approach 4: Only items of choice – pass-through schema, import the news message schema and the schema for allowed items

Applying the versioning policy to the schema architecture

- Subjects to versioning are:
 - The framework (i.e. the common components)
 - Individual standards (i.e. the items, including the PackageItem)
 - The news message

Versioning dependencies

- Only major version changes require updates by dependent schemas
 - File names and namespace definitions are maintained in minor version upgrades
- A change in the NewsMessage does not affect any other schema
- A change in a item schema may only affect the NewsMessage schema
 - depending on which approach for inclusion of items that is used
- A change in the Framework (common component) affects all importing schemas.

A Template for common components

A template for building common components

- A template for building common components in a consistent process and style
- Works by specifying specific characteristics of a common component
- Can be used to record entries for all 3 types of common component
 - Data type
 - Basic component
 - Aggregate Component

Template fields

- Each characteristic is recorded as a separate field
- Fields can be combined to create other fields
 - E.g. naming fields – can be qualified and combined to create dictionary entry name for a component
- Template contains a lot of data and is liable to evolve during the early stages of work
- Fields record characteristics concerning identification, component type, naming, constraints, cardinality, synonyms, declaration type, definition, synonyms

The template fields (1)

- Identifier
 - For the component
- Dictionary Entry Name
 - Unique name of the component in the component data dictionary. Made up by concatenating the object class term, property term, and representation term naming rule fields
- Common Component Type
 - The type of common component – data type, basic, aggregate, etc.
- Definition
 - The definition of the component

The template fields (2)

- Business Term(s)
 - Synonyms for the component
- Property Of
 - Identifies the component a component is part of
- The naming rule fields
 - Object class term – represents the logical data grouping to which a property belongs (if specified)
 - Property term – represents the characteristic or property the property represents
 - Representation term – represents the form in which the value of the property is represented
 - And 'Qualifier term' – used to qualify the other terms to specify semantics and ensure uniqueness of name
- Example – 'Publication Metadata' has a 'PublicationDate' property whose value is represented by a 'Date' data type.

The template fields (3)

- Declaration type
 - Identifies the type the component is declared as. Can be:
 - Data type
 - Reused basic or aggregate component – either a reference to an element or a type
- Cardinality
 - Specifies the occurrence of the component when it is part of another component
- Constraints
 - Specifies any constraints on the component. E.g. list of allowed values, maximum length
- Comments
 - Free form notes on the component.

Template use case walk through – CarML template example

- Excel-based spreadsheet
- Column for each field
 - Exemplified by 27 component entries for CarML. Entries for:
 - Data type – simple and complex
 - Basic component
 - Aggregate component

Conclusions Suggested decision points for the IPTC

Decision points (1)

- Extensibility
 - of Content
 - Use of generic or explicit schema type approach for validation of controlled vocabularies?
 - of Data Model
 - Location of structural extensions? Wherever a user organization prefer, or only in a few special contexts?

Decision points (2)

- Conformance
 - Is there a need for schema enforced conformance rules or not?
- Versioning policy
 - Does the proposed policy fulfill business requirements?

Decision points (3)

- Common Components
 - Should specialization of common components be allowed in items (standards) or not?
 - Is the proposed structure, definition and naming of the common components appropriate?
 - Data type, basic component and aggregate component

Decision points (4)

- Item construction
 - Can an item be made up of only aggregate components or a combination of aggregate and basic components?
- Namespaces
 - Should there be separate namespaces for the individual types of common components?
 - For common component layers – data type, property, construct?
 - For the different types of construct created?

Decision points (5)

- The Common Component Description Template
 - Is the provided template approach a good base for further work?
- Validation
 - Enforce validation of NewsMessage with payload, or only the NewsMessage level?

IPTC Spring Meeting, San Diego, 3 March 2005

Thank you


Jay Cousins
RivCom Ltd
jay.cousins@rivcom.com

Ulf Wingstedt
CNet Svenska AB
ulf.wingstedt@cnet.se

Expressing the IPTC News
Architecture using W3C XML
Schema