


News Architecture the last mile ...

Laurent Le Meur (AFP)
News Architecture W/P chairman, IPTC
Spring meeting, Vancouver / Mars 27, 2006


© IPTC – www.iptc.org



Agenda


- Minutes of the last meeting, matters arising
- Chairman's report:
 - NAR work status overview – 10mn
- Report on NAR EP#1 – 30mn
- Lunch
- 15h30 – 17h
- NSTR report (LLM) – 25mn
- CoCo report (MS) – 15mn
- NMDF & NMAN

© IPTC – www.iptc.org



- Minutes (NAR0510)
- Matters arising
 - One action: AP to send information about their processing of news. Done.


© IPTC – www.iptc.org



NAR - introduction

- NAR = **NewsML 2 Architecture**
- NAR is *not* an IPTC standard
- Great introduction in
 - IPTC introduction to EP1 (DRAFT-NAR_1.0-doc-AboutNARexpPhase1_5)
 - IPTC Spectrum, jan.06
- NAR provides a generic framework on which IPTC standards are built:
 - news: representation and management of textual stories, pictures, audio & video clips and illustrated stories (text associated with images).
 - events: representation and management of news events
 - sports: representation and management of sports results and statistics.


© IPTC – www.iptc.org



News Architecture

- Several goals:
 1. Simplify the processing of news objects
 2. Manage news, events, sports results and other news-related information the same way
 3. Use the latest XML technologies
 4. Make it compact
 5. Make it storage-friendly
 6. Make it semantically rich
 7. Keep compatible with the current model (not the current syntax)


© IPTC – www.iptc.org



NAR work status

- Since the Autumn 05 meeting
 - Public version of the model and tech spec drafts
 - Schemata v0.6
 - Experimental Phase 1 (dec 1 to feb 15)
 - Weekly calls (with [notes](#)), rapid evolution, currently: study of common components
- Spring meeting (March 06):
 - Report on EP1
 - Decisions taken
 - Some [open issues](#) still to be tackled
 - **Simplicity requires time ...**


© IPTC – www.iptc.org



Experimental phase 1

- Goals:
 - Allow IPTC members to get a grip on the NAR.
 - Test “real life” use cases vs the NAR
 - Bring back practical issues not spotted by the WP.
 - Help on the finalization of the NAR tech spec.


© IPTC – www.iptc.org 7



Three EP1 reports

- **AFP**: prototypes, Java, Open Source & XSLT
 - XML-Object binding (Java implementation of NAR objects)
 - NewsML 2 Persistence (RDBMS storage)
 - Navigation between concepts (GUI)
 - NewsML1->NewsML2 transform ([open-source](#))
- **Reuters**: prototypes, XSLT
 - NewsML1->NewsML2 transform
 - NewsML2->RDF N3 transform ([open-source](#))
- **Kyodo News**: review and tests
- Plus, from M.Steidl, a C# implementation of NAR objects


© IPTC – www.iptc.org 8



AFP EP1 report

- Study achieved by Ilm with 2 software developers
- Overall model clear easy to implement
- Still some issues
 - Multiplication of **namespaces**
 - Model of the Topic Item (metadata vs content)
 - Rigidity in metadata ordering
 - Complexity of NMDF structure
- Some minor points to (re)consider
 - Split of metadata between itemMeta and contentMeta
 - Presence of the subject/@about attribute
 - Validation of items in messages


© IPTC – www.iptc.org 9



Reuters EP1 report

- Study achieved by Dave Compton (Rtr's NewsML Technical Owner)
- NewsML1 vs NewsML2 comparison
 - Representation is much more compact, including topics and packages, content hints (eg caption) are more easily associated with content
 - Multiplication of **namespaces** is questioned
 - Usage & meaning of some properties need clarification
 - Some new properties (eg contentClass) imply a vendor-dependant mapping
 - Some Rtr properties are still treated as extensions (not agreed as *standard*)


© IPTC – www.iptc.org 10



Kyodo News EP1 report

- Study achieved by Hiroshi Shinotsuka
 - Encountered problem for properly handling multiple **namespaces** with older tools (XMLSpy 2004)
 - Remarks on some NAR properties
 - Answers to some specific questions exposed in the model & tech spec drafts.
 - Proposes to add a <NewsML> root element
- NewsML1 -> NewsML2 conversion
 - Issue with converting 100 content-related DTD to XML schemas for proper validation of the result.

© IPTC – www.iptc.org 11




Roadmap

[see STA0601.1-RoadmapDiagram](#)

- May 06
 - New version of the model and tech spec
 - Schemata v0.7 (expected from RivCom)
- April (end)-August (end) 06: EP2 (=> motion)
 - News Content WGs work of moving IPTC standards to the NAR
 - NAR refinements
- September-October 06:
 - EP2 conclusions and report
 - NAR final updates
- Autumn meeting - October 06
 - Formal approval of the NAR
 - Start of an IPTC standards Experimental Phase


© IPTC – www.iptc.org 12



Motion of the NAR WP

- [Request](#) to launch a second experimental phase.
- Using updated public drafts & xml schemata
- Actors: NewsContent and NewsCodes WGs
- Start: 26 April 2006
- Feed back expected : 29 August 2006 latest.
- => Final evolution of the NAR after EP2


© IPTC – www.iptc.org 13



NAR: Model

- “What it provides”
- Expressed as documentation + UML diagrams
- Introduction to:
 - A model for the creation of common components
 - A set of datatypes and common components
 - A conceptual model
 - A processing model
- two conformance levels (‘core’ and ‘power’)
 - **Therefore, soon, two documents**


© IPTC – www.iptc.org 14



NAR: technical specs

- “How it works”
- Expressed as documentation + XML samples
- Detailed definition of:
 - datatypes
 - common components
 - reusable items

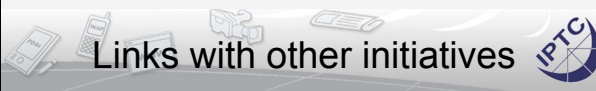
© IPTC – www.iptc.org 15



Completed by

- The XML schemata
- A comparison of NewsML 2 with other standards
 - IPTC7901
 - IIM
 - NewsML 1
 - Dublin Code
 - Atom (and RSS)
- A glossary of terms
- Discussion papers with design decisions
- An [XML Stylebook](#) for a consistent look


© IPTC – www.iptc.org 16



Links with other initiatives

- Exchanges with
 - RDF in XHTML
 - RDF interest group
 - W3C TAG
 - W3C international
- Maybe sometime, a more formal IPTC-W3C liaison

© IPTC – www.iptc.org 17



NAR Items

- A set of simple object called Items.
- All managed the same way, all with the same sets of metadata.
- **PackageItem**: package -> grouping of all kinds of Items.
 - Can include NewsItems, TopicItems or other PackageItems.
- **TopicItem**: knowledge -> representation of a concept, dedicated structure for each type.
 - Participates to the creation of an news related ontology
- **NewsItem**: news -> a news report , any media type, any format.

© IPTC – www.iptc.org 18

NewsItem

NewsItem

- Management properties
- Links to other Items
- Descriptive metadata
- Publication Metadata
- Signature metadata

thumbnail preview high definition

Courtesy ATC

© IPTC – www.iptc.org 19

TopicItem

TopicItem

- Management properties
- Links to other Items
- Descriptive metadata
- Rights Metadata
- Publication Metadata
- Signature metadata

Specialized content, structured

Courtesy ATC

© IPTC – www.iptc.org 20

PackagelItem

PackagelItem

- Management properties
- Links to other Items
- Descriptive metadata
- Publication Metadata
- Signature metadata

NewsItem TopicItem NewsItem

Courtesy ATC

© IPTC – www.iptc.org 21

NewsMessage

NewsMessage

Transmission metadata:

TopicItem PackagelItem NewsItem NewsItem

Courtesy ATC

© IPTC – www.iptc.org 22

Tension

- On one side, search for
 - Simplicity (the KISS principle)
 - Usability
 - Pragmatism
 - Interoperability by using a least common denominator
 - Web compatibility
- On the other side, search for:
 - Satisfaction of high profile news providers
 - Great rigour in the structure and processing
 - Control of content creation and usage

Courtesy ATC

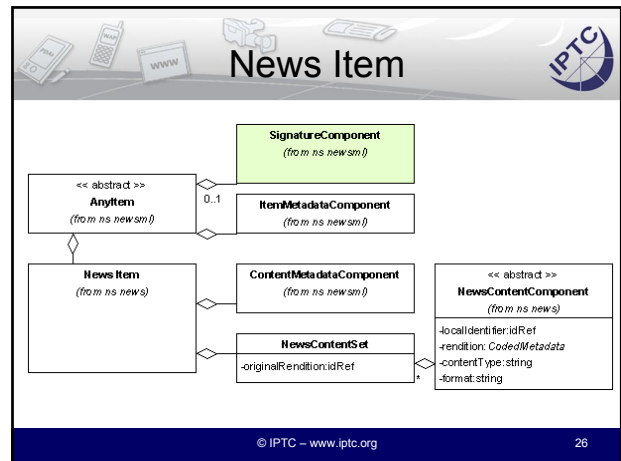
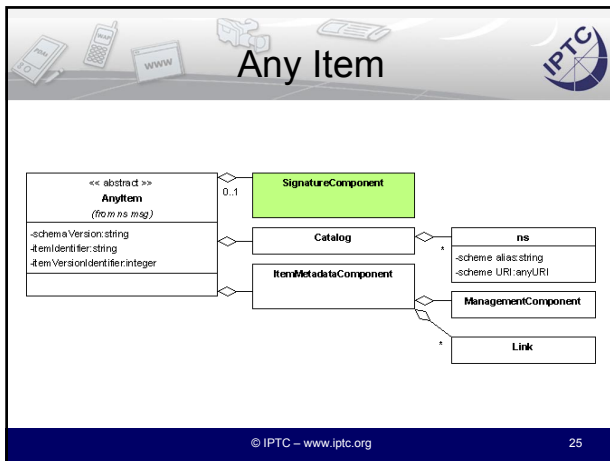
© IPTC – www.iptc.org 23

Conformance levels

- Two levels of complexity
 - “core” profile
 - “power” profile
- Provision for provider defined extensions
- Some modules of the framework belong to the “power” profile
 - Ex. rights, signature, metadata creator, multiple roles.
- The “core” profile is as easy to learn.
- The “power” profile offers top level features, and is a superset of the “core”.

Courtesy ATC

© IPTC – www.iptc.org 24



- ## Management properties
- Driven by the shared processing model
 - A set of properties in itemMeta, shared by all Items
 - Item Class (news, topic, package)
 - Content Class (text, photo ... person, location)
 - Instance Of a recurring report (x) (= *thread*)
 - Provider
 - Date item created, modified, released, embargo ends, retired
 - Status (usable, withheld, canceled)
 - Conformance level (core, power)
 - Editorial signal (x) (important corr., major add., prev. announced)
 - Generator Tool
 - Profile Name
 - File Name
- © IPTC – www.iptc.org 27

- ## Management (power)
- Power extensions
 - Alternative location (x)
 - Original Identifier
 - Alternative Identifier (x)
- © IPTC – www.iptc.org 28

- ## Links
- Links between items
 - From a News Item (e.g. text) to a News Item (e.g. picture)
 - From a Topic Item (e.g. person) to a News Item (e.g. an occurrence of a biography).
 - + Links from an Item to any other resource (on the Web)
 - Different variants:
 - “navigation” (see also)
 - “derivation” (eg translation, wrap-up)
 - “attachment” (illustrated text -> illustration)
- © IPTC – www.iptc.org 29

- ## Admin. metadata (1)
- Date Content Created
 - Location Content Created (x?)
 - **Accountable Person (new)**
 - Source of Information (x)
 - Creator (x)
 - Contributor (x)
 - Significance for Audience
 - Intended Audience (x)
 - Service (x)
 - Editorial Note (x)
- © IPTC – www.iptc.org 30

Descript. Metadata

- Language (x)
- Genre (x)
- Subject (x)
- Assert (x)
- Slug Line (x)
- Title (x)
- Headline (x)
- Description (x)

© IPTC – www.iptc.org 31

Sample (1)

```
<news:item schema="0.x"
  guid="urn:newsml:iptc.org:20051220:story1"
  version="1"
  xml:lang="en-GB" >
  <xxxxx />
</news:item>
```

Syntax still to be agreed-on (not RFC3085 compliant)

Language declarations (text only)

© IPTC – www.iptc.org 32

Sample (2)

```
<news:item schema="0.x"
  guid="urn:newsml:iptc.org:20051220:story1"
  version="1" xml:lang="en-GB"
  xmlns:news="urn:iptc:std:news:1.0:xmlns"
  xmlns="urn:iptc:std:newsml:2.0:xmlns" >
  <catalogRef href="..." />
  <xxxxx />
</news:item>
```

Namespace declarations

Catalog reference for scheme alias declaration

© IPTC – www.iptc.org 33

Sample (3)

```
<news:item >
  <catalogRef />
  <itemMeta />
  <contentMeta />
  <xxxxx />
</news:item>
```

Item related metadata (item management) And item links

Content related metadata (administrative, descriptive, rights and publication)

© IPTC – www.iptc.org 34

Sample (4)

```
<news:item >
  <catalogRef />
  <itemMeta />
  <contentMeta />
  <news:contentSet>
    <news:inlineXml />
    <news:inlineData />
    <news:remoteContent />
  </news:contentSet>
</news:item>
```

Wrapper for "alternative" representations of news content

Inline XML: an XML structure from an external namespace

Inline data: raw text (eg CDATA) of encoded binary data

Content referenced via a URL

© IPTC – www.iptc.org 35

Welcome to a sweet tasting sequel of:

Cracking the CoCo-Nut

NewsML2 Architecture WP
Common Components WG
Michael Steidl
(as stand-in for Johan Lindgren)
IPTC Spring Meeting 2006

Flash back

- Three levels of Common Components
 - **Datatypes**: most of them were defined in the scope of the NMDF work – find them in section 4 and 7 of the Tech Spec document, v 18
 - **Basic components**: structures primarily intended for reuse by aggregate components
 - **Aggregate components**: high level structures to represent a semantic set of properties.
- The work of the CoCo group focuses on Basic and Aggregate Components.

© IPTC – www.iptc.org 37

Work since the Autumn meeting

- These aggregate Common Components were discussed
 - On a **Basic Component** level:
 - Address CoCo
 - Location CoCo
 - Partial Date/Time CoCo
 - On a **Basic+ Component** level:
 - ContactInfo Coco, reusing Address
 - On an **Aggregate Component** level:
 - Person CoCo, reusing PartialDate and ContactInfo
 - Organisation CoCo, reusing PartialDate and ContactInfo
 - POI (Point of Interest) CoCo, extending Location and reusing PartialDate and ContactInfo.

© IPTC – www.iptc.org 38

What's exactly ...

Salzburg = a location

Mirabell = a company

The Mozartkugel™ = a product

Mozart = a person

© IPTC – www.iptc.org 39

What's exactly a Location/POI

Salzburg = a location = "geoArea"

with:

- ConceptId
- Concept(Sub)Type
- Name
- Description
- GPS coordinates
- Altitude

© IPTC – www.iptc.org 40

What's exactly a Location/POI

Role attribute + Address
Email addr
Phone number
Fax number
Web address
Instant Msg address

Salzburg = a location = a POI
extends geoArea
with:
ContactInfo
Directions
LocationDetails
Capacity
Facilities
OpeningHours
Commercial info

Address line
Postal Code
Postal Location
Area
Country (code)

© IPTC – www.iptc.org 41

What's exactly a Person

Partial Date
allows a full date with time:
2006-03-28T14:00:00Z
xOR: only parts of it:
-year = ???
-month = 03
-day = 28
-hour = 14

Mozart = a person
with:
ConceptId
Concept(Sub)Type
Name
Description
DateOfBirth (BTW: 250 years ago)
DateOfDeath
Gender
Expertise
ContactInfo

© IPTC – www.iptc.org 42

What's exactly a Company IPTC

Mirabell
= a company
= "organisation"
with:
ConceptId
Concept(Sub)Type
Name
Description
DateEstablished
DateDissolved
Sector
ContactInfo
Location

© IPTC – www.iptc.org 43

What's exactly ... IPTC

The Mozartkugel™
= a product
= a generic "concept"
...plus more
with:
ConceptId
Concept(Sub)Type
Name
Description
... more, like the recipe,
may be in a specialised
structure beyond IPTC
standards – using an
"extensibility point".

© IPTC – www.iptc.org 44

Summary IPTC

- These concept and entity components are "plug-ins" to the TopicItem
- They all share a basic set of properties:
ConceptId, Concept(Sub)Type, Name, Description
(I called it the "Supra CoCo" one time)
- Where ever sensible their structures are aligned (e.g. between person and organisation)
- There is a strict relationship between the "contentClass" property of the TopicItem and the plug-in CoCo.
- These CoCos were designed to cover a very wide range of applications: organisations, persons, locations of all kinds
- Specialisation is controlled by the (sub)type
- A set of code schemes for role, type and the like attributes to properties is required to maintain semantic equivalence !

© IPTC – www.iptc.org 45

Semantic equivalence: e.g. "name" IPTC

The examples below show how the very generic container <name> can be made more specific by using the role and part attributes. BUT: there should be an agreement on the codes used for these attributes, at least for one culture, to provide interoperability.

```

<name>Pablo Picasso</name>
<name part="pt:given">Pablo</name>
<name part="pt:family">Picasso</name>
<name part="pt:full" role="stl:formal">Pablo Diego José Santiago Francisco de Paula Juan Nepomuceno Crispín Crispiniano de los Remedios Cipriano de la Santísima Trinidad Ruiz Picasso</name>
<name>Marilyn Monroe</name>
<name role="scp:original">Norma Jeane Mortenson</name>
<name role="scp:baptismal">Norma Jeane Baker</name>
<name role="scp:married">Norma Jeane Dougherty</name>
<name role="scp:work scp:usual">Marilyn Monroe</name>
<name>Theodor Herzl</name>
<name role="scp:western">Theodor Herzl</name>
<name role="scp:hebrew">Binyamin Ze'ev Herzl</name>
<name>Zinedine Zidane</name>
<name role="stl:informal">Zizou</name>
<name part="pt:full" role="stl:formal">Zinédine Zidane</name>
<name part="pt:full" role="stl:formal pp:sort">Zidane, Zinedine </name>

```

© IPTC – www.iptc.org 46

Relationships btw concepts IPTC

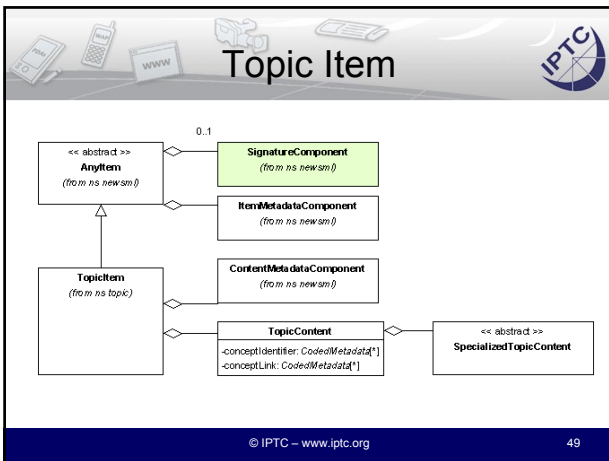
- Named relationships between concepts
 - A thesaurus of Themes can be created.
 - A Person may be related to an Organisation
 - An Organisation may be related to a Location
 - A Location may be related to an Event
 - An Event may be related to a Person
- These are semantic, directional relationships.
- RDF: these are "triples" (subject, predicate, object)
- A "weight" could be added on the relation

© IPTC – www.iptc.org 47

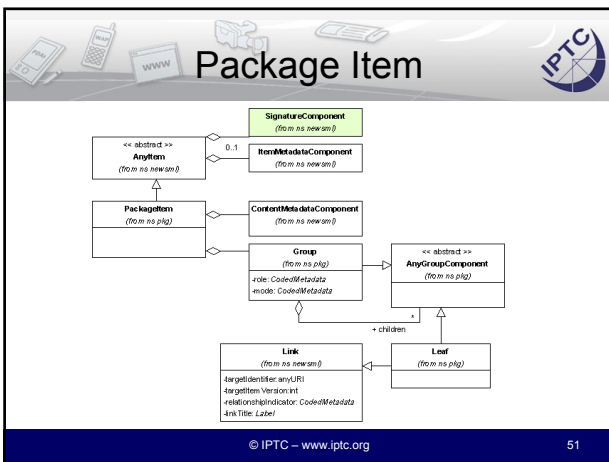
Semantic News IPTC

Courtesy AFP

© IPTC – www.iptc.org 48



- ## Topic Item
- This is the current focus of the group
 - Proposals for the enhancement of the model are under discussion
- © IPTC – www.iptc.org 50



- ## Metadata handling
- A mechanism for controlling values in schemes
 - Scheme alias/code pairs -> CURIE (foo:bar)
 - A Catalog (local or remote) maps scheme aliases to URIs
 - CURIE -> full URI = concept identifier
 - URI -> directory or database -> Topic Item
 - Topic Item -> description of the concept
- © IPTC – www.iptc.org 52

- ## How to use the model
- Specialized content WG will choose:
 - Is it news (a point of view)?
 - Check structure and processing of NewsItem
 - Create a class of NewsItem
 - Is it knowledge (a hub of information on a concept)?
 - Check structure and processing of TopicItem
 - Create a class of TopicItem
 - Is it something different?
 - Derive a new item class from AnyItem
 - Ex: AssignmentItem (EventsML)
- © IPTC – www.iptc.org 53