IPTC SPRING MEETING 2006

held at the
Vancouver Marriott Pinnacle Downtown
1128 West Hastings Street
Vancouver, British Columbia V6E 4R5 Canada

27 to 29 March 2006

Minutes of the News Architecture Working Party held on 27 March 2006

Document history				[Document URN: urn:iptc:workdoc:nar:0602:1]
Version	Issue Date	Pages	Author (revised by)	Remark
1	2006-04-21		Michael Steidl/	

Present:
Chairman: Laurent Le Meur, Agence France Presse

Najeh Al-Dulaimi, KUNA
Khalid Al-Haqhaq, KUNA
Walter Baranger, New York Times Company
Scott Calder, Mainstream Data Inc.
Dominic Chan, CNW Group Ltd.
Dave Compton, Reuters Limited
Honor Craig-Bennett, PA NewsLtd
Jean-Pierre Evain, EBU
Willis Fong, Canadian Press
Takahiro Fujiwara, EAST Co. Ltd
Andy Gleeson, BBC Monitoring
Walter Grolimund, Keystone
Stéphane Guérillot, Agence France Presse
Darko Gulija, HINA
Geoffrey Haynes, The Associated Press
Niels Hojer-Pedersen, Ritzau Bureau I's
John Iobst, Newspaper Association of America
Hugh Johnstone, IPTC Editor
Alan Karben, XML Team Solutions, Inc.
Paul Kelly, New York Times Company
Hiroshi Kashima, Nihon Shinbun Kyokai
Dean Large, Business Wire
John Minting, United Press International
Peter Müller, SDA/ATS
Stuart Myles, Dow Jones & Company
Fumiaki Nishino, Nihon Shinbun Kyokai
Hitoshi Saito, Nihon Shinbun Kyokai
Hiroshi Shinotsuka, Kyodo News
Henrik Stadler, Tidningarnas Telegrambyrå
Charles Tichenor, The Associated Press
Pauli Tölli, Suomen Tietotoimisto Oy
Rob Warner, Reuters Limited
Michael Wolf, Reuters Limited
Guowei Wu, Xinhua News Agency

Michael Steidl, IPTC Managing Director (MD) as Secretary

1 Minutes of the Meeting held on 24, 25 October 2005 (NCD0501.1)

It was proposed by Dean Large and seconded by Walter Baranger that the minutes be accepted as circulated. This was agreed unanimously.

2 Matters Arising

No matters were arising from the minutes not covered by this agenda.

3 Chairman's report

No special report was given, it was included into the following presentations.

4 Overview of the current state of the NAR development work: Report about the NAR Experimental Phase 1

(See presented slides for this item in the Appendix, slides 4 through 13)

Discussion on the Experimental Phase 1:

Why to use scheme-code-pairs in CURIE format?
The CURIE format saves a lot of space. It was decided to use URI-identifiers for concepts and their full syntax could get rather long. But URIs for concepts from the same scheme share a common part – the scheme URI – and this part is reflected by an abbreviation, the "alias".

Structure of PackageItems:
There is no specification how to design package structures except the general rules: a group may contain links to other items or to another group(s) of this PackageItem.
This structure highly depends on the use case and hence will be different for different providers.

Discussion on the Roadmap-Timeline: (see slide 12, Appendix)

Misha Wolf said the given timeline is very challenging, and there are issues with the sequence of work:
- an Experimental Phase 2 (EP2) should commence in May and its results should be implemented in October making a final version of the NAR.
- but in the time between May and October some more specification work will be done to the NAR as such and this has to be tested in some way.

The chairman underlined the main intention for the EP2 is to let content standards like SportsML or EventsML build their specifications on top of the NAR. For doing this a clearly defined state of the NAR specs are required. The feedback of the work of these standard groups has to be reflected in the NAR and hence this will require changes.

Misha Wolf underlined this May date does not indicate the NAR development will be stopped.

Alan Karben asked when to expect stable experimental XML Schemas to build a NAR version of SportsML on.
The chair replied saying this will be on the 8 May.

Darko Gulija emphasised no major changes will be done to the basic NAR design after 8 May, only elaboration on details and technical matters not covered by the version for the EP2.

The chairman showed a MOTION on launching an Experimental Phase 2:
This was moved by Laurent LeMeur and Dean large.

While discussing the motion the wording was amended.

Final MOTION: The IPTC Standards Committee requests all Working Groups from the IPTC News Content WP and News Codes WP to test a new draft of the News Architecture and provide written feedback during a period commencing on 8 May 2006 - after the release of new public drafts & XML schema - and ending on 29 August 2006 – so that the IPTC members can discuss the feedback at the autumn meeting 2006.
The goal is for the WGs of the News Content WP and the News Codes WP to develop a draft of a NAR compatible version of the IPTC standards they handle in the field of General News, Events, Sport or NewsCodes, and for the NAR WP to finalize the specification using comments from the other groups.
Processing of feedback:
- WG members have to send comments to the NAR-dev list.
- Each editor of a NAR document, or section of, should maintain a record of formal open issues.
- The NAR WP should prioritize issues.
- Issues should be discussed on the NAR-dev and newsml-2 lists.
- Each editor should make formal proposals to the group.
- After decision (during a conf call), each editor should update the document he maintains.

VOTE: This motion was agreed unanimously.

5 News Structure Working Group

(See presented slides for this item in the Appendix, slides 15 through 35 and 47 through 53)

6 News Metadata Framework

No special presentation was given.

7 Common Components Working Group

Michael Steidl gave a presentation as stand-in for the WG Lead Johan Lindgren.
(See presented slides for this item in the Appendix, slides 36 through 46)

8 News Management Working Group

No special presentation was given.

9 Any Other Business

There was no other business.

10 Date and Place of Next Meeting

Annual General Meeting, 3-6 July 2006, Vienna (Austria) at the Bristol Hotel.

=== END of document ===

News Architecture the last mile …

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

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 ….

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

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.

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)

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 …

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.

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

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

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)*

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.

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**

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

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

NAR: technical specs

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

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

Links with other initiatives

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

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.

NewsItem

Courtesy ATC

TopicItem

Courtesy ATC

PackageItem

Courtesy ATC

NewsMessage

Courtesy ATC

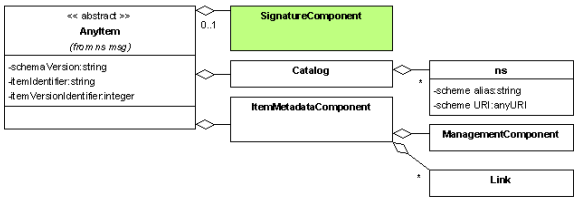
Tension

- 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

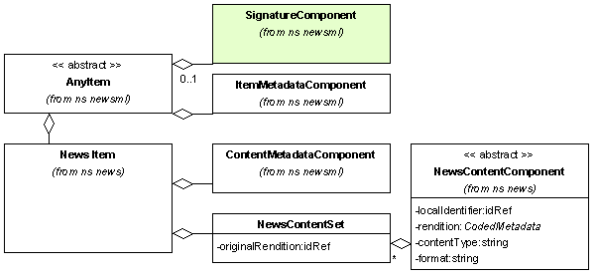
Conformance levels

- 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".

Any Item



News Item



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

Management (power)

- Power extensions
  - Alternative location (x)
  - Original Identifier
  - Alternative Identifier (x)

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)

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)

Descript. Metadata

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

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)

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

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)

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

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.

Work since the Autumn meeting

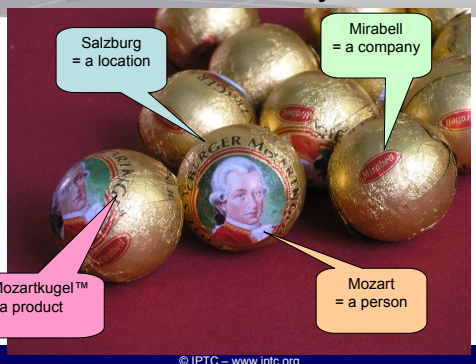
- These aggregate Common Components were discussed

On a **Basic Component** level:
- Address CoCo
- Location CoCo
- Partial Date/Time CoCo

On a **Basic+ Component** level:
- ContactInfo Coco, reusing Address

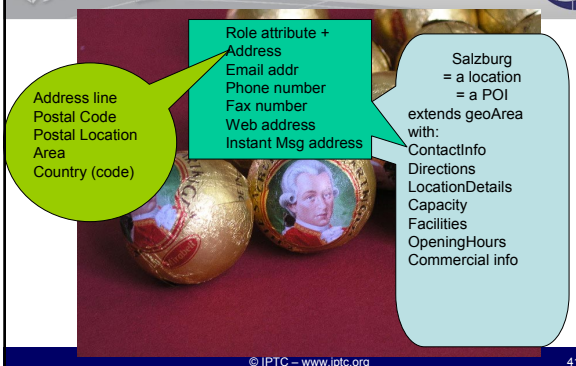
On an **Aggregate Component** level:
- Person CoCo, reusing PartialDate and ContactInfo
- Organisation CoCo, reusing PartialDate and ContactInfo
- POI (Point of Interest) CoCo, extending Location and reusing PartialDate and ContactInfo.

What's exactly …

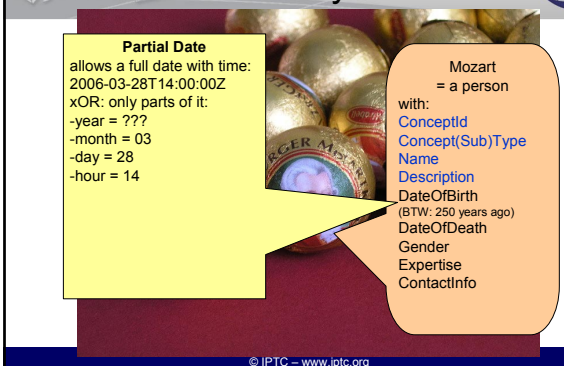


What's exactly a Location/POI

What's exactly a Location/POI

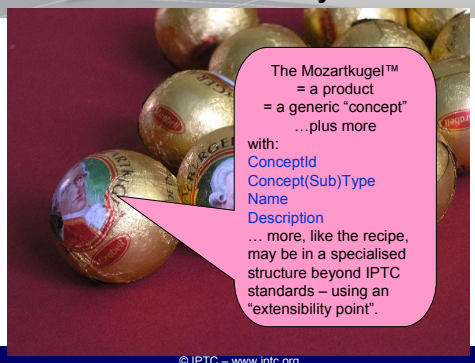


What's exactly a Person



What's exactly a Company

What's exactly …



Summary

- 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 !

Semantic equivalence: e.g. "name"

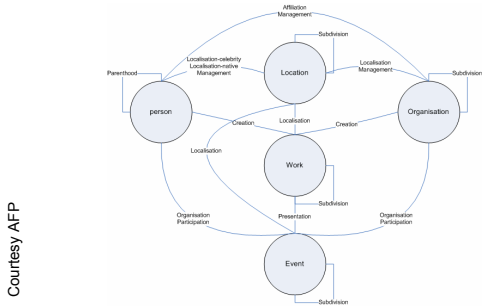
The examples below show how the very generic container <name> can be made more specific by using the role and part attributes. BUT: there should be an agreement on the codes used for these attributes, at least for one culture, to provide interoperability.

```
<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 Crispin Crispiniano de los Remedios Cipriano de la Santisima
    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>Zinédine 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 prp:sort">Zidane, Zinédine </name>
```


Relationships btw concepts

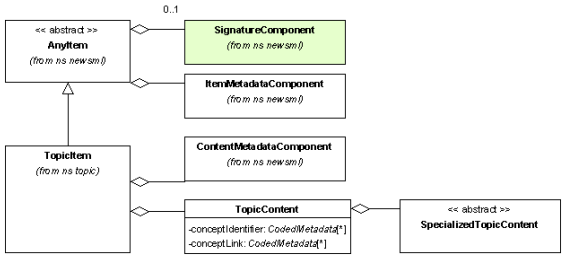
- Named relationships between concepts
  - A thesaurus of Themes can be created.
  - A Person may be related to an Organisation
  - An Organisation may be related to a Location
  - A Location may be related to an Event
  - An Event may be related to a Person
- These are semantic, directional relationships.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

Semantic News



Courtesy AFP

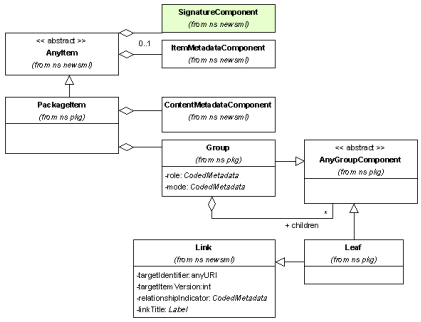
Topic Item



Topic Item

- This is the current focus of the group
- Proposals for the enhancement of the model are under discussion

Package Item



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

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)