



IPTC Mirror

Now for the Applications

With development of the new News Architecture well advanced and a second experimental phase under way attention at the AGM Working sessions turned to the application standards based on the new architecture and to the development of appropriate metadata structures.

At the start of the AGM working sessions, Standards Committee Chair Henrik Stadler (TT) explained that News Architecture (NAR) development was going ahead with a second Experimental Phase, while work on the NAR based EventsML was under way. Although a lot had been achieved it was still the middle of the work programme and more involvement (especially from new members) would be welcome.

Following the sessions of the individual Working Parties, the Standards Committee assessed the results and considered other aspects of IPTC standards work.

Consultancy

Delegates were informed that the Management Committee had agreed to make funds available for further consultancy work on the NAR. This will cover the development of a XML Schema generator to create individual XML Schema files for the News Architecture Core and Power conformance levels (from a Master file which contains all NAR features).

Note: the schema generator has now been delivered and is available for internal use by IPTC members at <http://groups.yahoo.com/group/iptc-news-architecture-dev/files/>

In addition consultants will be contracted to implement all NAR changes considered necessary as a result of the second Experimental Phase. This process will include Quality Assurance checks for the specification documents to ensure consistency and give a high level of clarity.

Namespace policy

Implementation of the NAR specification in a XML Schema involves the use of a XML namespace for the XML Schema. During development consideration was given to a policy for the conditions under which this namespace should be changed.

Since the NAR will be the underlying framework for new IPTC news exchange standards the NAR WP proposed that the policy should be adopted for general IPTC use. The policy is:

Autumn Meeting 2006

16 - 18 October
2006

Madrid, Spain
Hotel Melia Galgos

The NAR Model and Specification should be available for consideration as a candidate for public release.

- A XML namespace URI MUST be changed if XML instances which were valid under the previous version of the XML Schema will not be valid under the new version.
 - A XML namespace URI MAY also be changed at the discretion of its governing IPTC body.
- This proposal was approved by the Standards Committee.

Approvals

Formal approval was also given to a series of additions to the IPTC NewsCodes and to a change in the NewsCodes Change Management Guidelines (see the NewsCodes Working Party report - [page 5](#)), as well as release of the NewsML 1.2

Summary

Standards Committee

Further development consultancy - [Page 1](#). Namespace policy adopted - [Page 1](#). XML Schema Style and URL Task Forces wound up - [Page 2](#). New Photo Metadata Working Group - [Page 2](#).

NewsML Architecture Working Party

Experimental Phase 2 under way - [Page 2](#). NAR use cases considered - [Page 2](#).

News Content Working Party

Extension of the IPTC Core - [Page 4](#). EventsML-G2 development - [Page 4](#). NewsML-G2 progress - [Page 5](#). SportsML work for V1.8 - [Page 5](#).

NewsCodes Working Party

Subject NewsCodes additions - [Page 5](#). New AudioCodec and VideoCodec NewsCodes - [Page 5](#). Genre NewsCodes - [Page 5](#). NewsCodes updates

released - [Page 6](#). Change Management Guidelines revised - [Page 6](#). Future NewsCodes - [Page 6](#).

NITF Maintenance Working Party

NITF V3.3 now available - [Page 6](#).

NewsML 1 Maintenance Working Party

Beta release of NewsML 1.2 XML Schema - [Page 6](#).

Public Relations Committee

New standards family named - [Page 2](#). Presentations reports - [Page 2](#).

Semantic Web Approach - [Page 3](#).

MESH Enhanced News Services project - [Page 4](#).

SportsML at ORF - [Page 5](#).

Profium News Agency Solution - [Page 6](#).

The IPTC Mirror has hyperlinks for [web addresses](#) and for [page references](#)

XML Schema as a Beta version (see [page 6](#)).

Task Forces

Task Forces had been established for the consideration of XML Schema Style issues, and for URN issues, but Michael Steidl reported that there was now little activity in these groups. Accordingly it was decided to wind them up - when specific issues occurred they would be handled by the other Working Parties and Working Groups. However, the Yahoo discussion groups for these task forces have not been closed but are kept as archives.

Reporting on the Colour Space Task force Harald Löffler (Ifra) said that there had been little recent activity, but a special report prepared by Ifra on the subject had been posted on the discussion group (<http://groups.yahoo.com/group/iptc-colourspace-dev>). Ifra work in this area is continuing with a meeting with camera makers to discuss ways of making photo data more persistent from the outset.

During discussions it was pointed out that the new News Content standard (NewsML-G2) will include physical properties for the content, including some photo properties. Also, information handled by the IPTC Core (for XMP) does not include EXIF data (to avoid unnecessary duplication).

Photo metadata

Since photo metadata was clearly an important area the Standards Chair said he had decided to establish a new "Photo Metadata Working Group" which will be chaired by Harald Löffler and Michael Steidl. This group will merge the activities of the IPTC Core and Colourspace Working Groups and look at other issues regarding photo metadata.

A number of delegates - including some new members with specific interest in the photo area - expressed their willingness to contribute to the work. One of the first steps will be to establish the goals for this group, and both public (<http://groups.yahoo.com/group/iptc-photometadata>) and internal IPTC (<http://groups.yahoo.com/group/iptc-photometadata-dev>) discussion groups have been established.

Change management

Consideration was given to the possible introduction of a formal change management system for IPTC standards work. Requirements for such a system will be in-

Public Relations

Following a lot of deliberation, with input from many members, agreement has now been reached on the way the new family of exchange standards will be given a common "brand" image. This will be done by adding "G2" to all standards based on the new news Architecture (NAR). For example this means that the new general news exchange standard will be known as NewsML-G2, while a NAR compatible version of SportsML would be SportsML-G2. The whole family of standards will be the IPTC-G2 standards, and the intention is to develop a common logo to emphasise the brand image.

Presentations

A series of presentations had been made in the period between the Spring Meeting and the AGM as follows:

Nexpo 2006. Walter Baranger (PR Committee Chair), Stéphane Guérillot (IPTC Chair) and Michael Steidl (IPTC Managing Director) toured the exhibition area and visited relevant companies, inviting them to the NAA Wire Committee session where there was a presentation of IPTC developments. A further presentation is planned for Nexpo 2007.

World Association of Newspapers (WAN) Congress. Stéphane Guérillot was invited to the Congress as the IPTC Chair and gave a presentation to the Digital Technology Round Table, explaining why IPTC standards were important to ensure the success of both print and digital news technology. In addition he was able to make contact with many delegates from all over the world, some of whom expressed interest in IPTC Membership.

Cepic (Coordination of European Picture Agencies Press Stock Heritage) Conference. Michael Steidl was invited to give a presentation on IPTC photo metadata and join a panel discussion on the future of metadata (see report in the July 2006 Mirror).

Ifra Italia Michael Steidl reported that he was invited by FIEG (the Newspaper Publishers Association of Italy) to a meeting in Rome in April and then by Ifra Italia to join their annual meeting in Bologna in May to introduce Italy's news industry leaders to NewsML 1, the Subject NewsCodes and NITF. This spring the four major news agencies of Italy decided to offer news feeds which make use of these standards and to encourage users to move to them. (see report in the July 2006 Mirror).

vestigated, along with the availability of suitable (open source) software.

News Architecture

Outlining progress since the spring Meeting Laurent Le Meur (AFP) said that NAR development had continued using the established combination of e-mail discussion groups and conference calls - and that involvement from other delegates would be welcome. Experimental Phase 2 was under way, and it was anticipated that further NAR development would be carried out in parallel with the work on the applications - such as EventsML-G2 and NewsML-G2.

Use cases

Laurent went on to examine the business benefits of using NAR

based standards with a series of use cases: John (a stringer), Instant News (news agency) Global Photo (photo agency) and MyMedia ("newspaper").

Stringer

As "a stringer" John writes articles and takes photographs which he supplies to several publishers - he uses one standard for this, NewsML-G2. For his application the central feature is that the standard is simple with the same metadata model for all types of content.

Content in the form of a news item can be stored, sent, sold and exchanged (with an optional news message being available). Alternative renditions - such as a picture representing a video clip, or an audio extract from video - will have different formats, but will all come from the same resource.

Because the NAR is designed

with two conformance levels, it is easy to create applications for the core level. The NAR itself is built on a strict object model which should allow easy integration into editorial systems and ready development of low-cost tools.

(Laurent also mentioned that AFP have a prototype Java implementation and would be willing to consider co-operation with other members on development).

Agency application

The second case, "Instant News", is a news agency that uses the NAR as its internal model (for the content classes and the content management system) and uses NewsML-G2 and SportsML-G2. Since NAR applications are consistent they can be used to send Alerts, Breaking News, Events and Sports results.

Common sets of metadata are used throughout with an Administrative layer (repeatable, so it can cover several sources, for example) and a Descriptive layer which allows machine indexing of items in a precise manner, while codes can also be shared between organisations.

Items have a persistent identity with versioning, and extensive management properties driven by a common processing model. Content can be handled in multiple ways with different renditions, previews and cross-media publishing.

Another important feature is that NAR XML content can be made secure using a signature adopted from the W3C.

Photo agency

Photo agencies have specific requirements, and some of these were considered in the context of the "GlobalPhoto" agency. Here the descriptive metadata has to be language independent and there is a need for multi-language captions, while images will be offered in multiple resolutions. In addition there is a need for digital asset management.

Current practice relies on the use of keywords, though there is a move towards the use of controlled concepts. However, some issues relating to the compatibility of XMP metadata and the NAR have still to be addressed.

Complex rights management is dealt with by the NAR at the "power" conformance level - labels can be used at the "core" level - and users will have to select an appropriate rights package.

Cross-media publishing

Final case looked at by Laurent was "YourMedia". This would have been considered to be a newspaper some time ago but is now involved in cross-media publishing with multiple editions that include analysis, comment and a perspective on the news.

It uses a variety of sources (including input from readers). Here the main aims are to deal with the problems of information overload, and to create new applications. The

proposed approach involves reliable management of the news coupled with efficient management of knowledge - with the process being described as extracting the information DNA.

The NAR provides a framework for managing information about concepts, and four base concepts have been defined with specific properties:

- Organisation - founded, dissolved, sector, location, contact information.

Semantic Web Approach

The advantages and practicality of bringing NewsML into the Semantic Web were considered by Raphaël Troncy (with George Anadiotis) in a presentation at the AGM.

He started by looking at the importance of metadata as information that adds value to an asset and makes it machine processable, while semantic metadata allows easy exchange of news items, searches and delivery of personalised news content.

The IPTC metadata contains fields that include definitions and guidelines how to use the fields, while the metadata values are expressed as controlled vocabularies. However, the problem with definitions is that machines do not always understand them and fields with the same semantics may have different titles in different applications.

To overcome these problems needs formal semantics for both the content and the structure, and this is the role of the Semantic Web (SW). The Semantic Web is a framework that can make existing metadata standards and tools interoperable, and make them interoperable with the rest of the web.

Raphaël explained that there is already common ground between the NAR and the SW with the NAR having distributed resources (news items) that a globally and uniquely identified (with a URI), and making use of shared resources. He believed that moving to the SW would be a natural switch that would offer such benefits as giving better control of descriptions, enhanced searches for news topics, semantic interfaces, and unified news management with a semantic content management system.

An example showed how to create a News domain ontology with OWL Web Ontology Language and considered mapping of the IPTC NewsCodes into various SKOS thesauri - SKOS (Simple Knowledge Organisation System) supports the use of knowledge organisation systems (KOS) in the framework of the Semantic Web (<http://www.w3.org/2004/02/skos>). Transforms can be applied to allow conversion of IPTC metadata into its RDF (Resource Description Framework - www.w3.org/RDF) counterparts, with various tools being available to handle RDF data.

Finally, Raphaël introduced delegates to the work of the W3C Multimedia Semantics Incubator Group - he is the co-chair - which is looking at image and other multimedia on the Web, with a focus on interoperability with existing standards. He also encouraged input to the group and participation in its work - (www.w3.org/2005/Incubator/mmsem).

Raphaël Troncy and George Anadiotis are based at the Centrum voor Wiskunde en Informatica (CWI) in Amsterdam. The mission of CWI is to perform frontier research in mathematics and computer science and to transfer new knowledge in these fields to society in general and trade and industry in particular.



Raphaël Troncy

- Person - born, died, gender, affiliation, occupation, skill, contact information.
- Geopolitical area - gps coordinates, altitude, geopolitical type.
- Point of interest - extends geopolitical area with open hours, capacity, facility, access, details, contact information.

Named relationships between concepts can be established and the equivalence of concepts expressed. Use of concept identifiers makes it possible to obtain, and use, stored information about the concept - the development of knowledge bases containing the information may represent a new market.

Other features of the NAR make it possible to address communities with similar interests, for example news could be selected according to the target audience, the language, the subject, or the source.

News Content

IPTC Core

A look at options for the next version of the IPTC Core (for XMP) was provided by Michael Steidl, who explained that the general approach was to not add too much, with the primary use being for photographs. This could involve development of photo-specific NewsCodes, while it was also important to define a migration path between the IPTC Core and the NAR metadata structures.

Discussions have already been held with interested parties - such as stock photo agencies and photo libraries - with a number of requirements being established. These included:

- a taxonomy to describe the "real things" that form the subject of a photo;
- a way of handling named entities;
- increased coverage for "lifestyle" in the Subject NewsCodes;
- a rights information taxonomy;
- and improved coverage for stock photography to complement the IPTC news centric fields.

Future steps would include further investigation of the requirements and a look at the potential for collaboration with other bodies. There was also a need to encourage software producers to provide simple interfaces to taxonomies such as the IPTC Subject NewsCodes.

A subsequent decision by the Standards Committee set up a "Photo Metadata Working Group" which will take over the activities of the IPTC Core Working Group.

EventsML

Work on EventsML-G2 is moving ahead and Charles Tichenor (AP) said that a comparison had been made between the NAR EP#2 draft Schema and the original (pre-NAR) EventsML draft XML Schema. There appeared to be a good alignment in many areas though some issues still have to be dealt with.

Three ways of fitting events information into the NAR have been proposed; as a TopicItem (of type "event"); as a NewsItem (with event content); and as a specific EventItem derived from the AnyItem and using elements from both the TopicItem and the NewsItem.

The handling of Assignments and Resources also needs looking at.

Date and time

Date and time is a particular issue, with the following use cases identified:

- Known and certain date (and time);
- Partial date and time;
- Uncertain or approximate dates;
- Date ranges (also the specific case of historical eras);
- Recurring events (at regular or irregular intervals);
- Historical events;
- Holidays (on fixed or different dates).

Possible ways of handling this are to use a profile of ISO8601, to have a datatype constructed from a set of attributes; and to have both of these values on a single datatype with a defined precedence.

Finally Charles invited member representatives to join the

MESH Enhanced News Services

MESH (Multimedia Semantic Syndication for Enhanced News Services) is a project designed to "extract, compare and combine from multiple multimedia news sources, automatically create advanced personalised multimedia summaries, syndicate summaries and content based on the extracted semantic information and provide users with a 'multimedia mesh' news navigation system".

The project started in March 2006, will run for three years, and is co-funded by the European Union 6th Framework Programme. The Mesh consortium consists of 12

European organisations and an introduction to the project was provided at the AGM by Nikos Sarris of the Athens Technology Center (ATC), who are one of the participants.

Underlying motivation for the project is a belief that although there is a large amount of multimedia content available much of it is not really usable and users are unable to find the content of most value to them. In addition there are particular problems for smaller news producers because of the costs involved and difficulties in achieving high usage of their material.

Four main work areas are being addressed:

- Techniques for semantic analysis and annotation of multimedia content - a common knowledge base will be developed for text, image, audio and video analysis.
- Semantic comparison and combination based on content annotations to give a coherent and consistent set that can also be used for delivery according to user preferences.
- Generation of semantic-based media summaries, along with profiles of information sources (to help consumers assess their objectivity and credibility).
- End user solutions including access from mobile devices and personalisation systems.

Planned result of the work programme is a platform that will deliver content to users based on automatically extracted metadata and the user's preferences; create personalised multimedia summaries; merge content from multiple sources; and provide a dense web of content links for navigation across multimedia information. See www.mesh-ip.eu for more information.



Nikos Sarris

EventsML Working Group - as with other Working Groups work is being carried out using a discussion group (<http://groups.yahoo.com/group/iptc-eventsml-dev>) and telephone conferences.

NewsML-G2

Intended for the handling of general news NewsML-G2 provides similar news markup features to NewsML 1, but is based on the NAR. Laurent Le Meur said that the central object is the NewsItem which consists of metadata and content, and processors will have to support standard profiles.

There are also a TopicItem and a PackageItem but these do not have to be used.

There are three main media type domains to be handled - Photo, Video and Text.

In the Photo domain areas to be dealt with are the mapping between the IPTC Core and NewsML-G2, which involves finding a way to handle "traditional" keywords, and establishing physical characteristics for images.

The EBU have provided a list of video requirements. Details for a "shot list" (a structured description of content segments with time references) have been established but further work is needed on handling the time references.

At the moment it does not appear that any existing text markup fully covers the NAR requirements, but possible approaches have been considered within the NITF Working Party.

Current efforts are concentrated on testing the NAR (as the base of NewsML-G2) and drafting a processing-orientated specification for the new standard.

SportsML

Alan Karben (speaking over a Skype link) said there had been some delay in finalising SportsML 1.2 and posting it (on www.sportsml.org) but this had been done - a notification e-mail has also been sent to the SportsML discussion group (<http://groups.yahoo.com/group/sportsml>).

Work on SportsML 1.8 is under way with a plug-in for curling and some enhancements for the tennis XML Schema. Steps are being taken to make more of the XML Schema datatypes specific - for example the number of goals will be specified as an integer (rather than a string, as at present). Integration

of SportsML and the NAR is also under investigation.

Alan went on to inform members that a recent enquiry had been for Sports to handle Pro-fishing data. While such enquiries were welcome, what was really needed was detailed input from the proposers to actually produce the appropriate XML Schema.

NewsCodes

As normal a number of new terms had been proposed for addition to the Subject NewsCodes, some Subject Detail codes had been dealt with by the "Fast Track" jury while the new Subject Matter NewsCodes were considered by the full Working Party. This proved to be particularly time consuming and it was necessary to hold a second working session to complete the work.

During this second session it was pointed out by the Chair (John Minting - UPI) that documents covering the proposed additions had all been circulated for discussion well in advance of the meeting, so he was surprised at the level of comments being raised during the session.

Additions finally approved were, by Subject heading:

arts, culture and entertainment

history - as a SubjectMatter, *entertainment award* as SubjectDetail under *entertainment* SubjectMatter.

crime, law and justice tribunal - as a SubjectMatter.

human interest

plant - as a SubjectMatter,

fortune telling - as a SubjectDetail under *people* SubjectMatter.

lifestyle and leisure

wedding - as a SubjectMatter,

shogi as a SubjectDetail under *game* SubjectMatter.

sport

bodybuilding - as a SubjectMatter.

Audio and Video

Extensive sets of NewsCodes to cover audio and video coders had been submitted by the EBU and it was agreed to adopt the new AudioCodec NewsCodes to replace the old Audiocoders NewsCodes (which were deprecated) and to adopt the VideoCodec NewsCodes to replace old Videocoder NewsCodes (which were deprecated).

Genre

A request had been made to include the term "almanac" in the Subject NewsCodes, but the "Fast Track" jury considered that it would be more appropriate to include it in

SportsML at ORF

ORF are the main Austrian TV and Radio service (with a market share of up to 80%) and a major national internet platform. A description of how they are using SportsML was provided by ORF Systems Architect Gerald Schinagl.

The Sports database is the main information source of ORF editors and reports, and consists of an ASP NET Web application, SQL Server 2005 and a PocketPC application. It handles information on 40000 individuals, 3500 teams, 35000 sports events and over 40000 team sports events.

It was realised that implementing a database for team sports events - initially soccer - would be difficult and very expensive using a relational database, and it was decided to use SportsML with the SQL Server 2005 XML features for this purpose. The soccer data is stored in the SportsML format, so the database engine knows how the data is structured and how to access it. Stored information is accessed using XQuery.

Some information - such as results - has to be accessed frequently and this has been extracted and put into a relational database. However the information remains available in the XML database (typically with more detail).

Implementation started in December 2004 and the system has been in use since April 2005.

Sports information dealing with individuals - such as Alpine skiing, which is of particular interest in Austria - is currently handled using a relational database.



Gerald Schinagl described how ORF are using SportsML to delegates at the AGM.

the Genre NewsCodes, and a motion to do this was approved. Addition of the term "horoscope" to the Subject NewsCodes had already been agreed, but it was also decided that it would be better to include this term as a Genre.

NewsCodes Release

Following the AGM, there was a major release of the NewsCodes. This included the two new sets for AudioCodec and VideoCodec, along with new versions of the following sets: Subject (V17), Genre (V8), Format (V9), NewsItem Type (V5), and Topic Type (V6). Minor changes were also made to several of the other sets of IPTC NewsCodes. All of the NewsCodes are available for download from <http://www.iptc.org/NewsCodes>.

Guidelines

A proposal to amend the Change Management Guidelines for NewsCodes was approved after some discussion.

Agreed changes mainly related to the "Fast Track" procedure, introducing a clause that allows Members to make a binding request for consideration a single term (or a complete submission) to be discussed at the next Working Party Meeting instead of being decided by the "Fast Track" jury.

In addition the size of the "Fast Track" jury was changed to become "at least four but not more than six" while decisions require that at least half of the Fast Track Jury should be present.

The size of the ad-hoc Working Party (which considers Subject Matter request prior to submission at Meetings) was also changed to "at least four but no more than six". Another proposal to change the style in which terms are presented from lower case to Title case was not adopted.

Future NewsCodes

Delegates were informed that work on the development of a new generation of NewsCodes is under way and a "Future NewsCodes Working Group" has been formed (including some new members).

Members of the new Working Group managed to fit in a first session during the AGM, while work will continue with the normal combination of a discussion group

(<http://groups.yahoo.com/group/iptc-news-codes-dev>) and telephone conferences.

NITF

Delegates were informed that work finalising NITF 3.3 had been completed and that the standard had been released on the NITF web site (www.nitf.org) in both DTD and XML Schema versions.

Alan Karben (XML Team) informed delegates that tests had shown that the SportsML XML Schema would validate properly with a NITF document (XML Schema version) inside.

Some further consideration was given to improving acceptance of the NITF, and the possibility of developing alternative text formats for use with the new generation IPTC standards.

Proposals included: a "NITF Core" Schema, containing just the NITF tags needed for text markup; an XHTML Microformat; and a new ArticleML (www.articleml.org).

NewsML 1 Maintenance

Work on producing a XML Schema version of NewsML 1.2 started after the Spring Meet-

ing, with rapid progress having been made.

An overview of this work was provided by Jayson Lorenzen (Business Wire) who explained that although there were various experimental XML Schema for NewsML 1, they had not been extensively tested. Investigations revealed a number of issues, some of them functional, which had been dealt with, while the annotation and comments also had to be cleaned up.

The result was a draft experimental NewsML 1.2 XML Schema, which was being offered for release as Beta version. During discussions, the importance of having an efficient change process to deal with comments was stressed.

Motions were passed to agree the changes made, and to ask the Standards Committee to approve release of the Beta Version.

This experimental version of the NewsML 1.2 XML Schema is now available from www.iptc.org/std-dev/NewsML/1.2/specification and members are asked to test the schema.

Some issues have already been identified, and details are included in the directory containing the schema. Further updates to the experimental schema will be carried out as appropriate.

Profium News Agency Solution

Profium Oy (www.profium.com) are a Finnish systems vendor, who concentrate on the use of semantic technologies for content management and routing.

Their News Agency Solution has multimedia capability and is based on open standards including XML, NewsML and NITF. Content is stored in a central database, with the metadata in RDF format. Content metadata is matched to customer profiles, with delivery in the format required by the customer. A modular approach means that elements can be added to complement customers' existing systems.

Recent installations include systems for the Netherlands News Agency ANP and for AFP.

ANP wanted to introduce support for multimedia output, and to have NewsML content delivery capacity. To achieve this the Profium delivery system has been integrated with a Microsoft Exchange based editorial system.

With AFP the production system is designed to allow easy configuration to meet the requirements of new customers, and to package content in news ways. It has a content routing system integrated with an editorial module that supports real-time multimedia production.



Profium Sales Manager Essa Suurio described the main features of the Profium News Agency Solution.

Published by the **International Press Telecommunications Council**
Royal Albert House, Sheet Street, Windsor, Berkshire SL4 1BE, England.
Managing Director: Michael W Steidl (mdirector@iptc.org). Editor: Hugh Johnstone (editor@iptc.org)
Tel: +44(0)1753 705051 Fax: +44(0)1753 831541 Web Site: www.iptc.org