



IPTC Mirror

IPTC - INFORMATION TECHNOLOGY FOR NEWS

Progress in All Areas

Development of the new IPTC Standards Suite is a major undertaking, requiring sustained effort from all of the Working Parties along with the Standards Committee. Efforts made over the past months are starting to show results, as outlined during the Spring 2004 Meeting held in Athens.

NewsML Support

Reporting on activities since the last Meeting, Working Party Chairman Laurent Le Meur said that the NewsML discussion forum had seen a lot of activity, and from the comments it appears that the main participants are implementers. However, there had not been a lot of response to the requests for suggestions and comments for NewsML 2.

Both the News Standards Summit and the IPTC News Standards Survey had provided some interesting information on the public perception of NewsML, and this was being taken into account in the development process.

Documentation

Finalisation of the NewsML V1 documentation has been a major effort, with inputs from a group of around ten members, and Laurent thanked the members concerned for their assistance. This work has also identified a number of areas where the standard was unclear, leading to different interpretations by implementers. Following a final edit (to ensure consistency in such areas as presentation) the documentation

will be released through the NewsML Web site.

Although the documentation project identified a number of areas where improvements could be made, the intention is to freeze NewsML V1 in its present form, and concentrate on development of NewsML 2. It is hoped that users will be able to use the existing extension facilities to meet evolving needs without having to propose changes that would require a new release of Version 1.

It is intended to include sample NewsML feeds on the NewsML Web site as examples for new implementers. Members with suitable feeds, who are able to make samples available on a regular basis should contact the Working Party Chairman. Where possible supporting documentation would also be welcome.

Requirements

A set of draft requirements for NewsML 2 were presented for discussion - the draft was based on the original NewsML (V1) requirements with extensions identified by the NewsML brainstorming session held in Leipzig and by some requests made through the NewsML web site. It was

stressed that the draft requirements were intended as a starting point and subsequent discussion raised a series of points to be taken into account.

An ad-hoc group has been established to consider comments, and produce an updated set of requirements for consideration by the full Working Party. The group consists of the Working Party Chairmen Laurent Le Meur (AFP), and Stuart Myles (Dow Jones), with Darko Gulija (HINA), Christian Ratenburg (CCI Europe), Mischa Wolf (Reuters), and Michael Steidl (IPTC Managing Director).

Following the Meeting this group has held teleconferences, while comments on proposals has been sought from members - and the wider NewsML community - through the discussion group.

As work proceeds updated versions of the requirement document are being made available for comment on http://www.newsml.org/pages/intro_newsml2.php. It was also pointed out that experience gained with the NewsML 2 requirements would be applicable to the de-

NewsML in action

An indication of the practical application of NewsML was given by Christian Ratenberg of CCI Europe. He explained that a major aim of CCI systems is to add market value, and that they often have to do a lot of work integrating systems. This was the challenge that they addressed by using NewsML. In addition automation (with no manual labour) was particularly important for E-applications and this can also be dealt with using a NewsML interface.

General approach is to use a standard NewsML structure with a linked group of NewsML files, which may include text, images and other formats. If users can process the NewsML directly this is the best solution, otherwise an XSLT transfer is used to convert NewsML to the format required by the customer. CCI are also using a NewsML structure for representing advertisements.

Contents

NewsML Support	1
NewsML in Action	1
NewsML Solutions	2
News Management	2
NewsMetadata	2
UPI Categorisation	3
TopicSet updating	3
News Industry Text Format	4
Global Metadata Structure	4
Special Content	4
EventsML	4
ProgramGuideML	4
SportsML	5
Weather Data	5
Adobe XMP	5

velopment of the EventsML requirements, as there was a need to achieve a common style (which will form part of the Project Management Guidelines).

News Management

The main aim of the News Management Working Party is the development of a processing model for NewsML.

At the moment there are many options so it is possible to achieve the same results in several ways. This means that potential users may consider the standard to be complex (so limiting adoption) and also means that different implementations may not be compatible.

Creation of a Standard Model should simplify matters and it would be possible to define "profiles" for specific types of application and establish conformance levels for applications. Working Party Chairman Stuart Myles (Dow Jones) went on to identify a set of ten requirements for a processing model:

- 1 Provide implementers of NewsML software with all the information they need to ensure that their implementations are fully interoperable with other implementations.
- 2 Describe the data model, or abstract structure, of NewsML.
- 3 Describe how to build instances of the model from NewsML V1.x documents.
- 4 Describe how to build instances of the model from NewsML 2.x documents.
- 5 Be specific in exhaustive detail covering every model of NewsML that has model, or logical, significance.
- 6 Be written to be easily and fully comprehensible to implementers with a little scope for interpretation as possible.
- 7 Not constrain implementations beyond what is necessary to ensure that they do follow the NewsML standard.
- 8 Describe error situations and how to handle them.
- 9 Include as few optional parts of the standard as possible.
- 10 Include as few optional processing steps as possible.

It was recognised that some of these requirements might conflict with the general requirements for NewsML 2. In particular requirements 9 and 10 had been included to simplify implementation, but there was a NewsML 2 requirement that there should be a minimum number of mandatory features.

The possibility of using a formal language - such as UML (Unified Modelling Language) - was considered along with the creation of an API (Application Program Interface). While both of these have advantages in making implementation easier they would involve a lot of work and it was considered important to ensure that the objectives of the group remained within the resources of IPTC.

News Metadata

Recent months have seen an increasing number of requests for additions to the Subject Reference System and the Metadata Working Party Chairman John Minting

NewsML Solutions

EAST Co is a Tokyo based software developer which specialises in providing Microsoft Windows solutions. They are IPTC members and are represented at meetings by Takahiro Fujiwara, who is also a leading member of the NSK NewsML Team.

Applications developed by EAST Co include NewsML systems at Kyodo News for retrieving archived articles and for news distribution; and news editing systems for food industry newspapers (Syokuryo). They have also been consultants on the Mainichi Newspapers Palace System, and are working with Mainichi on a system to covert NewsML news data for display on mobile telephones (using a raster image to avoid the need for Japanese fonts).

Takahiro is now managing the development of professional NewsML tools in EAST Co, including a Format Checker for NewsML and Web Services for NewsML.

Format Checker for NewsML

Designed to be easy to use the checker has a Windows user interface for setup and operation. It is said to be a language-neutral (English) application which works in any area it is installed in. For example, it functions correctly on a PC where the regional and language settings are Chinese, Japanese or English.

The file to be checked is opened and the NewsML instance document is automatically validated by using an appropriate IPTC schema file (if required the user can specify which schema to use for validation). At the moment the Format Checker provides support up to NewsML V1.2. Three standard check profiles are provided - strict, normal and basis - and users can set their own profiles.

IPTC TopicSets - as a default Catalog and its TopicSets - are embedded in the Checker to assist compliance checking in cases where the TopicSet files that the NewsML instance

refers to are missing.

Any errors found are identified for investigation. In addition, when an error is selected a link to XMLSpy can be opened (provided that XMLSpy is available on the machine in use) and the element or attribute in question is highlighted.

See <http://www.topicset.com/checker/>.



Takahiro Fujiwara

Web Services for NewsML

When controlled vocabularies are used it is important that both the information provider and the user have the same information in their vocabularies - this means that both parties have to know how to identify the vocabulary, and how to update it. TopicSet.com is intended to provide a way for maintaining and using controlled vocabularies through a Web Service.

TopicSets are archived by TopicSet.com with direct access search facilities using SOAP. A NewsML Catalog is maintained to identify the revision number of the TopicSets being used, and users can customise the selection of TopicSets for their applications.

The system is also able to handle other formats - such as Topic Maps and Tab Separated Values - and it is hoped that this will help with future migration to NewsML 2.

As a first stage the service will be provided for the TopicSets managed by IPTC, items that have been added by NSK, and the TV-Anytime classification scheme. Consideration is also being given to external interfaces, such as to the ISO database. In addition facilities are available to create and modify TopicSets, allowing users to create their own private sets of data.

See <http://www.topicset.com>.

(UPI) said that he believed that this was an indication of how widely the system was being used.

A series of change requests were considered by the Working Party, with agreement being reached on the addition of various new subject matter entries, these included: a group of Japanese martial arts and Asian Games under *sports*; *illness* under *health*; and *religious education* under *education*.

Suggestions for name changes to minimise apparent duplication of subject details - for example changing *forecast* to *weather forecast* under *Weather* were rejected, mainly on the grounds that the Subject Codes form a structured system so any duplication is only apparent.

In addition a number of proposals for new subject details had been approved by the fast-track jury.

There had also been a request for a set of subject details to deal with a specific religious groups. Although the Jury had reviewed the proposals and agreed that a number of them were satisfactory it was considered that this was a particularly complex area that should be considered by the Working Party. After some discussion it was decided that the way of handling religions (and related matters) should be given more general consideration and an ad-hoc group was established for this purpose.

There had been a large number of requests for additions to the Subject Reference System in the period immediately before the Spring Meeting and it had been difficult for the fast-track jury to process them all before the meeting.

To prevent this becoming a continuing problem the update procedure was modified by the addition of the following paragraph: "A change request received by the Managing Director within 8 weeks of a formal IPTC meeting cannot be guaranteed to

UPI Categorisation engine

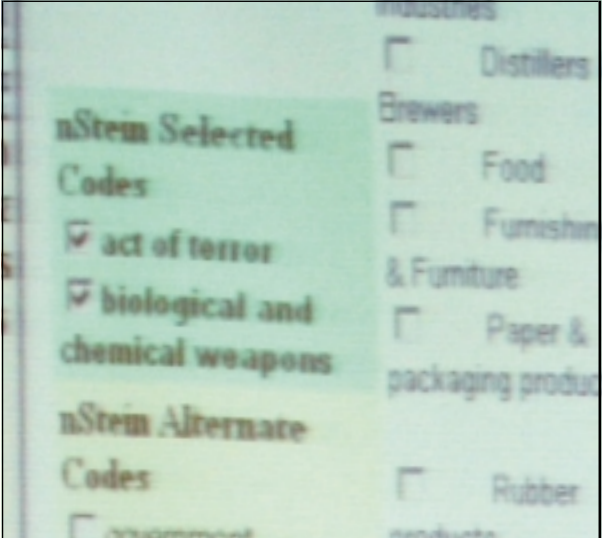
UPI are making extensive use of the IPTC Subject Reference System, with codes being generated automatically using a nStein categorisation system.

Once an editor has finished a story a "get codes" command directs to story to the system which automatically returns one or more recommended codes, along with a number of alternatives.

The editor can then accept the proposed codes and/or manually input other codes that they consider more appropriate - according to John Minting the editors say that the system saves them a lot of time.

There is a continuing learning process, and "unusual" code recommendations are referred back to nStein for analysis to help to refine the system, while the system also has to adapt as new codes are added to the SRS.

Recommended codes are offered to the editor for approval, but alternatives can be manually selected.



be dealt with by the ad-hoc Working Party before that meeting".

The Managing Director reported that all TopicSets had now been updated to the agreed common format, but further work was needed to ensure that all terms had appropriate explanations. The addition of a number of colourspaces used by the Japanese printing industry to the iptc-colorspace TopicSet was approved. It was also decided that future additions to the Provider TopicSet would be reviewed by the fast-

track jury.

An advisory service has been set up for notification of TopicSet updates with RSS and NewsML feeds being available on the IPTC web site.

TopicSet updates

Steps were taken to establish a formal update procedure for TopicSets that do not form part of the Subject Reference System. The approach adopted is essentially the same as that for SRS Subject Matter proposals:

Changes are sent to the Managing Director by (e-mail) and circulated to members for comment. Suggestions are then reviewed by an ad-hoc Working Party - taking account of any comments from members - to give an agreed final version that will be presented to the full Working Party for approval. Once approved the relevant TopicSet will be updated and posted on the IPTC Web site as soon as possible. Criteria for additions and updates are given in the box alongside.

Consideration was given to the possibility of developing some form of brand identity for all the IPTC TopicSets - since only five of them are included in the widely-known Subject Reference System. Although the general principle met with approval it appears practical difficulties - such as the need for professionally translated versions, and the need to make the best use of avail-

TopicSet additions and updates

- Criteria for inclusion.

1. An IPTC member – or an organisation sponsored by this member – must need to use the term/s and gain support from other members for the inclusion during the process below.
2. The term should relate to general news, not to a specific discipline and have a universal meaning. The exception is the term that has come into global usage although its origin may be specific to a local discipline.
3. The term is unique in its definition and not a synonym of an existing term.
4. Each new term must be accompanied by a precise explanation, in British English within the intended context of its use.
5. All requests from non-IPTC members must be sponsored by an IPTC member in good standing.
6. Each new term requested should be in lower case, in the singular unless it is a plural noun and in British English.
7. Requests shall be made using the form found at <http://www.iptc.org/IPTC/metadata/documentation/IPTC-TOSsubmission.xls>.

able resources within IPTC - meant that it was unlikely that any action would be taken in the near future.

News Industry Text Format

There were no major initiatives for consideration by the NITF Working Party, with the approval and release of V3.2 at the Autumn 2003 Meeting having dealt with most outstanding issues.



Christian Ratenburg (CCI Europe) has agreed to serve as Vice-Chair of the NITF Working Party.

A number of minor points have been identified as needing consideration, including: improving the content model for the "dd" element (data definition); re-instating the data-location attribute for the media reference element - this is for validation

purposes on older systems, and future use of the attribute would be deprecated; and clarification of the way email addresses should be handled.

All of these changes would be backwards compatible and is hoped that a formal proposal dealing with them (possibly for a NITF V3.3) will be ready for the AGM. In addition, both members and the NITF discussion group will be asked to suggest ways in which the NITF can be made more usable.

Christian Ratenburg (CCI Europe) agreed to take on the responsibility of Vice-Chair for the Working Party.

Special Content

Opening the session Working Party Chairman Geoffrey Haynes said that Special Content remained a very busy area, with a series of active working groups. EventsML is well under way, there are proposals for further development of SportsML, the launch version of ProgramGuideML was nearly ready, and formal commitment had been made to a new project dealing with Weather Data.

EventsML

After detailed consideration it had been decided that it would be impractical to try and develop EventsML in co-operation with OASIS. However there was clearly a high level of interest in this area (and from parties outside the news industry) and it had been agreed that the project would proceed within IPTC.

As a first stage a detailed Business Requirements document had been produced by a small working group (this was the first use of a Business Requirements document template that has been produced for use in the development of new IPTC standards). The document was presented to members by Chuck Tichenor (AP) and Johan Lindgren (TT) who had been responsible for it (along with Dominic Chan - Canada Newswire who was unable to be present) and generally well received, with delegates only making detailed suggestions for changes and improvements.

It was explained that there were some outstanding issues to be resolved, with a major consideration being the form of EventsML, Should it be an integral part of NewsML, or should it be a stand-alone program? The view of the working group was that EventsML should be usable on its own, but that it should be usable within NewsML and be in line with other IPTC standards developments.

Another question was the choice between using an XML Schema or a DTD. The recommendation from the EventsML working group was that the standard should be developed as an XML Schema, but that it should be possible to generate (automatically) a working DTD from the XML Schema.

Next stage is to produce the technical specification and create the XML Schema. This will involve continuing discussion on the Yahoo group to ensure the widest possible participation. It was agreed to change the name of this discussion group to reflect the name of the standard - the change was made shortly after the meeting (apparently without any problems) with the new name being eventsml@yahoogroups.com. In addition a new section has been added to the IPTC web site to deal with EventsML - <http://www.iptc.org/EventsML/>. Steps have also been taken to register the eventsml.com and .org domain names for use when appropriate.

ProgramGuideML

Reporting on progress, Manabu Miyaki (Yomiuri Shimbun) reminded members that following an initial meeting of the NSK team developing ProgramGuideML and the TV-Anytime Consortium it had become clear that although the systems had different applications there was a lot of common ground.

A presentation of ProgramGuideML and NewsML was given at the TV-Anytime forum held in November 2003 (in Kobe, Japan), and the general opinion of the forum was that "for broadcasters it would become a burden to create two types of

Global Metadata Structure

From the outset metadata was seen as an important part of the new IPTC Standards Suite, with the aim being to develop a common metadata structure for all the standards. The requirement was raised during several of the Working Party sessions in Athens, with a possible approach being identified during the Special Content session.

Some initial requirements for a common metadata structure were outlined by Alan Karben (XML Team Solutions) who also provided examples of ways in which metadata structures could be handled in an XML Schema. This led to considerable discussion and it became clear that the question of changing or replacing the TopicSet Structure was complex - and probably controversial.

For example, the East presentation included a video in which David Allen (the previous IPTC Managing Director) outlined the advantages and application of TopicSets, and pointed out that he considered that they made an important contribution to NewsML.

It appears that Japanese NewsML users consider that TopicSets are an integral part of NewsML1. Though they are often thought to be complex this is not the case for the computer systems that have to handle them, while they can be processed using standard XML tools. In considering changes the NSK group felt that it was important to ensure that bidirectional transformation was possible between the NewsML V1 and NewsML 2 metadata systems, with provision of a toolkit for this purpose. It was also thought important to minimise technical issues for users, and retain the functionality of the TopicSet system.

Similarly, during the general discussion it was made clear that a number of users have made a considerable investment in their TopicSets and would be reluctant to consider changes that might limit their usability.

It was decided that the best way to proceed with this difficult question was for it to be dealt with by the group of Working Party Chairman. Initial efforts will include consolidating the requirements, possibly with technical testing of possible approaches, ready for presentation to members (and other interested parties).

XML data - one for Personal Digital and Video recorders (an initial aim of TV-Anytime) and the other for ProgramGuideML”.

Subsequent discussions between the NSK group and members of the TV-Anytime Forum confirmed that the metadata for TV-Anytime was broadly compatible with that of ProgramGuideML. After some consideration it was decided that the best approach would be to include TV-Anytime metadata in ProgramGuideML and a liaison statement was agreed with the Forum.

The TV-Anytime metadata represents the programme information (such a title, broadcast time, and performers) and is taken into ProgramGuideML by the use of a namespace. Since TV-Anytime is based on an XML Schema the group have only produced ProgramGuideML in an XML Schema version. Appropriate steps have been taken to ensure that there is no clash of element names. Work on the revised version of ProgramGuideML is continuing with the aim of having a V1.0 ready for approval at the AGM, along with examples and documentation.

SportsML

Alan Karben (XML Team Solutions) informed members that interest in SportsML continues to grow with around 15 new members being added to the discussion list every month.

A series of improvements have been proposed for an updated SportsML V1.5, these include the addition of support for book-maker statistics; a new group element to hold properties of conferences, leagues, and divisions; an affiliation element - to allow the association of players (or officials) with multiple teams or for multiple years; improvements to the utility of coverage-attributes (for example to deal with games played in covered stadiums); improved support for attendance statistics; and clarification of some of the documentation. In addition there are a series of proposed enhancements for various specific sports DTDs.

Alan Karben also explained that work is under way to develop an XML Schema version of SportsML. Outline requirements include: storing and validating Controlled Vocabularies and their associated metadata; default datatyping of attributes that the publisher chooses not to maintain an updated list for; storing the documentation for SportsML elements and metadata in a standard way; and the inclusion / exclusion

of Specific Sport Schemas.

It is envisaged that this may be achieved by having a “SuperSchema” file, that contains references to (in effect calls-in as required) a series of other XML Schema files including: the sportsml-core.xsd; any/all specific sport xsd's; filenames for all Controlled Vocabulary Resource Files; and either a strict or a loose-type xsd file accessing the vocabularie or default values.

Weather Data

A first meeting of the Project Review Board was held to consider the proposal to start work on a Weather Data Definition (WDD), which is intended to provide a standard method of describing current, histori-

cal, and predicted weather information. This includes ground, water, atmospheric, and astronomic information that directly or indirectly influence, or are traditionally reported as part of, a weather package. The WDD would support data collection, archiving, and dissemination. The Review Board agreed that work should be started, using in-house IPTC resources.

First stage is production of the Business Requirements - this is being undertaken by Honor Craig-Bennet (PA) and any member wishing to contribute should contact her. It is also intended to establish a Yahoo discussion group to encourage wider discussion of the project (in line with the approach that has proved successful for the other standards).

Adobe XMP - Extensible Metadata Platform

An overview of XMP, with particular reference to its relevance to IPTC standards was given to delegates at the Spring Meeting by Gary Cosimini, Business Development Director, Adobe Systems International.

Mr Cosimini started by explaining that metadata can be considered like the labels on a can - it explains what is inside. In general metadata turns a file into an asset, which can be tracked and used by external assets, but at present many files do not have any metadata.



Gary Cosimini

A well-known exception is the “IPTC Header” which has been included in Adobe Photoshop for many years, and is used to hold specific information on the image - such as the caption and location details. The fields in this header came from an early version of the IIM (Interchange Information Model) but not all the fields are used as originally intended - partly because the IPTC fields were originally developed for transmission applications.

However there are a many other metadata needs, such as rights management, legal issues, assignment system links, content sharing and issues raised by the increasing use of digital cameras. The XMP concept has been developed to meet new demands with XMP standing for Extensible Metadata Platform. Adobe believe that metadata should be Open - build on standards; Extensible - support custom schemas or vocabularies; Universal - support diverse file formats, databases and containers; International - Unicode and multilingual; and Robust - keep metadata with asset.

XMP is an open standard (which uses RDF) that embeds XML data into data files. It is being implemented into Adobe products such as the Adobe Creative Suite (CS) and is available to developers via an open-source licence for use in other applications.

Because XMP is extensible it can handle existing metadata schemas. Since the IIM was introduced, IPTC metadata handling has also undergone considerable development and is now XML based. This means that the new IPTC metadata structures (both existing and planned) can be readily handled by XMP and IPTC and Adobe intend to co-operate on working out the best way to do this.

More information on XMP is available on: www.adobe.com/products/xmp.