



IPTC Mirror

IPTC - INFORMATION TECHNOLOGY FOR NEWS

Making Steady Progress

Although the main focus of IPTC development work is now the integrated family of standards, the existing standards continue also require support and modification to meet current user needs. Efforts of the various Working Parties are being directed to meeting both of these needs, and activities during - and following - the Aarhus Meeting are considered here

NewsML Support

Work on the revised NewsML documentation is proceeding, with the Working Party Chairman Laurent Le Meur (AFP) having updated the outline, included revised requirements and added guideline chapters. Aim of the revision is to ensure simplicity, with references being given to show where more details can be found. Volunteers are needed to continue the development process and to help with this the draft documentation will be made available on the iptcmembers Yahoo list <http://groups.yahoo.com/group/iptcmembers/>.

The format used for the documentation was also considered. At the moment much of the material is in Microsoft Word format, but the Managing Director pointed out that an XML based format - such as DocBook would have practical advantages, making it easier to transform the content into different output formats.

The NewsML Namespace received further consideration in Aarhus. However,

since the AGM it has become apparent that further work is needed in this area and it is hoped it will be possible to agree a revised approach at the Autumn Meeting.

There appears to be requirement to include personal information in NewsML and the type of information needed is already dealt with by vCard. This is a non-XML format and although there have been a number of initiatives in this area none of them have yet been formally agreed. Accordingly it was decided to launch an initiative to produce a general XML-version of vCard as a standard that other users can agree on. As a first step efforts will be made to establish a technical committee under OASIS to look into the matter. If an agreed standard is produced it can then be taken into NewsML.

Reports from NewsML implementers indicate that there can be problems with the DOCTYPE declaration. Use of a declaration is recommended in NewsML but it appears abbreviated (local) versions do not work with some tools - only a full version of the declaration is acceptable. This means

that validation can put a significant load in the (IPTC) server, which is not intended for this purpose. It was agreed that to minimise this problem the DOCTYPE declaration should not be mandatory in NewsML instances, while information about possible parser problems with short form declarations will be included in the documentation.

Inclusion of general metadata in NewsML was raised at the Spring Meeting, with the main interest being in handling RDF metadata. Possible ways of doing this were considered during the Aarhus session but these would involve changes to the DTD and it was agreed that further investigation was required. In general it is considered important to use NewsML metadata structures wherever possible, to ensure compatibility is retained. However, it is recognised that there can be strong reasons for including other metadata structures and the probable approach will be for NewsML to carry external metadata but not to support it.

Following a request, consideration was given to the general question of how changes to NewsML should be proposed

2003 Autumn Meeting

8th to 10th October - Leipzig, Germany

Planned highlights include:

Expert paper on transferring specifications from DTD to XML Schema
 Presentation on Project Management at IPTC
 Discussion of new IPTC Website family
 Establishing a URN namespace for IPTC
 Proposal for a totally new TopicSet (ContentCharacteristics)
 NITF Version 3.2 ready for approval
 ProgramGuideML - 1.0 final draft

Following the success of the initiative at the AGM, steps have been taken to arrange Internet access for delegates in the conference room.

Contents

Working Party Reports

NewsML Support	P1
News Metadata	P2
EventsML	P2
Special Content	P3
NITF Support	P4
News Management	P4

2003 Autumn Meeting	P1
NewsML V2.0 requirements	P2
Rights Management	P3

This PDF version of the IPTC Mirror has active links to both the content and to external web sites

Clicking on the red links - Standards - will take you to the appropriate page, while the blue links - www.iptc.org - will attempt to open your browser and connect to the appropriate site.

(this also applies to changes to other IPTC standards). Formal proposals should be circulated three weeks prior to the meeting where they will be considered. This is done by submitting them to the Managing Director for distribution, and papers formally circulated this way should reach all member delegates. The Yahoo NewsML group is an alternative way of raising points for discussion but it should be remembered that points raised in this group will only be seen by those delegates who have specifically asked to join the list.

News Metadata

Opening the session Working Party Chairman John Minting (UPI) said that the Subject Reference System (SRS) continues to grow, and that various organisations had made contact with him about possible use. There had been a number of questions about improving the coverage, and his response had been to suggest that the organisations concerned should join IPTC so they could take an active part in the SRS development process.

He hoped that use of the "jury" system would reduce the time taken to consider proposed entries. In the event this has not been totally successful, partly because delegates appreciate the importance of getting entries right, and that it is difficult to make changes once an entry has been included in the system.

Additions

However, a series of additions were approved by the Working Party, including a group of sports entries (at both Subject Matter and Subject Detail level and as Qualifiers) which had been proposed by the Swiss news agency SDA/ATS; a proposal by dpa to change "violent demonstration" (under Subject Matter Unrests Conflicts and War) to "demonstration" - which can then be put in context by the use of additional terms; and a number of Subject Matter and Subject Detail additions originally proposed by HINA and revised by an ad-hoc group during, and following, the Nice Meeting. A proposal for a new Genre TopicSet entry of "Anniversary" (proposed by HINA) was also approved.

It was pointed out that a number of existing entries do not have explanations as they were not a requirement when the SRS was first developed. Following the Meeting efforts have been made to fill in as many of the missing explanations as possible and these will be considered for the Autumn Meeting.

A list of Geographic Regions had been produced following the Spring Meeting to see if it was possible to develop a common approach. Following discussion of the pro-

NewsML V2.0 - setting the requirements

Initial suggestions have been made for the planned new-generation version of NewsML (NewsML V2.0). The aim is for a simplified version that will be easier to implement but which will maintain compatibility with existing applications so far as possible. There will be simpler ways of handling the controlled vocabularies and some metadata sets should be common across all IPTC standards.

However, more work is needed to establish detailed requirements and approaches, and since time is limited a special "brainstorming" session has been arranged in Leipzig during the afternoon of Tuesday 7th October - the day before the Autumn Meeting formally starts.

posal it was decided that differing usages among members meant that there was no real application for the list - at least at the present time - so no further action was taken.

A number of TopicSets have been defined for use with NewsML. The structure is the same in most cases, but the TopicSets used for the SRS have an extended structure that uses a "Variant" attribute to identify different descriptions. The Managing Director suggested that including the "Variant" attribute in all TopicSets would make it possible to use the same Xpath expression to retrieve information from all TopicSets. This change was approved and allowed values for the "Variant" attribute will be defined in the TopicSet Guidelines (this approach will not involve changes to the NewsML DTD).

Language management

Another proposal by the Michael Steidl (IPTC Managing Director) was for a method of managing non-English versions of the SRS (and other TopicSets where appropriate). The intention is that the reference (English) version and the different language versions will have the same structure and be stored in a relational database available on the IPTC Internet server. File names will reflect the individual languages. Content management functions in the database will make it possible to identify: reference entries that are not present in specific language versions; entries in non-English versions that do not correspond to an entry in the reference version; and entries that

have been made subsequent to release of a specific version.

For local use, as for a NewsML TopicSet, both the English reference version and the different language versions will be kept as XML files (TopicSets). The "SubjectViewer" software will be available for viewing and editing SRS data in the different language versions.

Responsibility for translated versions of the SRS will be taken by individual member organisations, with an initial (and tentative) list being: Spanish - AP/UPI; German - dpa act as co-ordinators for the German language users; Arabic - UPI are preparing a translation; Swedish - TT; Danish - Ritzau; Japanese - NSK; and Chinese (traditional) - CINTEC. It was reported that an outside organisation was working on a Dutch translation, while a Georgian version had also been produced. Members were reminded that any outside translations would have to be sponsored by an IPTC member for adoption.

During the session an updated version of the Subject Reference System Guidelines was put forwards for consideration, with a number of changes being proposed following discussions. These changes have since been taken into a revised version of the Guidelines which will be considered at the Autumn Meeting.

EventsML

Two alternative mission statements for this new Working Party were presented to delegates by the Chairman Dominic Chan (Canada Newswire) and after some consideration and modification a final version was agreed - **To create an XML format to be used in notification of news worthy events such as press conferences for distributions to news media and other users who have an interest in the information for internal or external purposes..**

Three main objectives have been established for the planned standard:

An XML format that can stand alone or can be incorporated into other IPTC standards;

Contains relationship entity for relating to and from a news item;

Ability to extend the use of EventsML as a notification tool, but also incorporated into resource management such as an assignment desk.

A number of existing standards relevant to the EventsML initiative have been identified and there was a brief overview of them: **iCalendar** is an IETF (Internet Engineering Task Force) proposal - <http://www.ietf.org/rfc/rfc2445.txt> - and intended as an exchange format between applications and systems. Format is defined in

terms of MIME type and it was explained that this would let the objects be exchanged using several transport mechanisms such as SMTP, HTTP, file systems, desktop interactive protocols, point-to-point asynchronous communication and networks.

vCalendar was created by the IMC (Internet Mail Consortium). It is closely related to iCalendar and intended for the exchange of personal scheduling information.

xCalendar is an XML representation of the standard iCalendar format. An iCalendar DTD was published in 2002 as an Internet draft but appears to have expired.

skiCal is a product of the Structured Knowledge Initiative and is intended to improve the information infrastructure concerned with public events (such as concerts, sports competitions and conferences).

RDF Calendar is a representation of iCalendar using the Resource Description Framework (RDF).

Requirements

An initial set of events coverage requirements submitted by members had been collated and were broadly similar. However it was suggested that there might be advantages in working with other (non IPTC) interested parties to develop a standard with more general application.

Dominic Chan pointed out that there were a several possible approaches:

Adopt the RDF format which is already under development, but has the disadvantage that IPTC would have no control of future changes in the standard.

Adopt iCalendar or vCalendar - these are already in use but are not XML enabled (though xCal may be an alternative).

Create a new IPTC standard.

Collaborate with other organisations to create a new standard.

After some discussion it appeared that there might be advantages in trying to establish an OASIS initiative (IPTC is an OASIS member - see www.oasis-open.org for more information) to work on a new standard. But before taking this step it was decided that further efforts would be made to fully define IPTC requirements for events coverage.

Special Content

Television listings

The name ProgramGuideML has now been adopted for the television and radio listing system currently under development by NSK (the Japanese Newspaper Publishers and Editors Association). There is a dedicated web site www.programguideml.org with draft documentation and DTD (currently V0.7).

A brief update on the project was given by

Motayusa Mise (NSK) who explained that efforts had been concentrated on redefining the requirements. The system is designed to allow the exchange of data for programme guides. Information for a given programme is handled as a single unit which can be combined with others to give, for example, the daily programme for a channel. When used with NewsML the structure is contained within DataContent, but the system can also be used in stand-alone mode, and has been made media neutral.

It is planned to have a draft V1.0 ready for consideration and release at the Autumn 2003 Meeting, with formal approval and release of Version 1.0 planned for Spring 2004. There seems to be general interest from members that deliver programme listings, while it is likely that there will be more interest from outside parties once the standard has been released

Weather Markup

It seems that while weather organisations are interchanging data with well-formatted

systems, the formats being used are not really suitable for news applications. In addition individual commercial organisations tend to use their own systems.

One of the main applications for weather markup is likely to be on-line services but the business case for news agency applications has to be established, while it was suggested that there might also be a need to handle hydrographical data (such as tide information).

It was suggested that members with a specific interest in weather markup should make contact with one another to see if it is possible to establish the requirements, with the general aim of supplying customers with weather information in a standard format to make their processing easier. Any results will be considered at the Leipzig meeting with a view to starting more formal development if there is sufficient demand from members.

Public Relations

Martin Holyroyd (PR Newswire) said that work on XPRL (extensible Public Relations

Rights Management:

An overview of current Digital Rights Management (DRM) systems was provided by Geoff Haynes (AP), who explained that the main initiatives in this area are XrML (extensible Rights Markup Language - www.xrml.org) and ORDL (Open Digital Rights Language - <http://odrl.net>), and outlined the main features of each approach.

XrML. Originally developed at the Xerox Palo Alto Research Center (PARC), XrML is now controlled by ContentGuard, Inc, who state that they are committed to give governance and control of XrML to the international standards community. Because of this they have frozen their release of XrML at Version 2.0. This is the final release ContentGuard expects to post to XrML.org.

Microsoft are an investor in ContentGuard and is using XrML in its DRM solutions including the Media Rights Manager for Windows Media audio and video; Digital Asset Server for Microsoft Reader eBooks and Windows Rights Manager for Windows Server 2003. Other implementers include the Digital Magazine producers Zinio, and the software house DMDSecure (Digital Media Distribution Secure), while Sony have taken out a licence. In addition XrML is being used by MPEG-21, the OASIS Rights Language Technical Committee and the Open eBook Forum (OeBF).

ContentGuard claim to have patents which cover the use of rights expression languages and offer a number of license options, some of which are royalty free. Their main interest is now the development of XrML based solutions.

ODRL: The Open Digital Rights Language (ODRL) Initiative is aimed at developing an open standard for Digital Rights Management and promoting the language at standards bodies. Version 1.1 was released in August 2002, is freely available and has no licensing requirements. It has been officially accepted by the OpenMobile Alliance (OMA) - previously the WAP Forum - as the standards rights expression language for all mobile content. ODRL Version 1.1 has also been published by W3C as a Note for discussion, while there are also W3C Team comments on the ODRL V1.1 Submission.

In summary, it was explained that XrML can be considered as a broader, more general purpose language, while ODRL seems to be specifically focused on publishing and media applications, making it more compact.

Language) was proceeding with an XML Schema being produced. This has since been released and deals with three aspects - document release, clippings briefing and coverage reporting. Further information (and the XML Schema) is available from the XPRL website www.xprl.org.

Election Markup

In the USA the Voter News Service (VNS) experienced problems during the 2002 elections and has since been disbanded. AP has now taken over collection and delivery of election results and have developed their own system for this.

It was explained that the system had to maintain compatibility with legacy VNS systems (so is not optimal) and has primarily been designed for real-time messaging. AP also supply database dumps of results, along with the information needed to set up the database.

It was reported that the Election and Voter Services system being developed by an OASIS technical committee has started field testing.

SportsML

Formal release of SportsML V1.0 was in May 2003, following ratification at the Spring 2003 Meeting. Documentation was finalised before release and there were improvements to both the tutorial and the examples. The launch was described as "low key" since the final release was little changed from the Draft release in Autumn 2003.

Interest in SportsML continues to grow with the discussion group now having some 110 members. Notable new contacts include a supplier of statistics for Australian Rules Football; localsports.com (an internet sport service provider); the US "Fantasy Leagues" (this is a major business), and a programmer from the US Olympic Committee.

It is hoped that an XML Schema for SportsML will be available by the Autumn Meeting. Use of an XML Schema has the advantage that it is possible to rigidly define the formats - for example that a score must be a positive integer. There was a suggestion that the XML Schema version might become the reference form - say in 2004 - with the DTD being made available as a derived variant, though this will depend on the overall IPTC policy on the adoption of XML Schemas

Guidelines are being produced for the addition of new sports to the system - as plugins with the basic pattern containing: event-

metadata; team-metadata; team-stats; player-metadata; player-stats; and actions. The detailed content will depend on the characteristics of the sport concerned, and there will be a Style Guide for element and attribute names.

An immediate aim to include as many Olympic sports as possible and it is hoped to involve as many official and unofficial bodies (with understanding of Olympics data) as possible.

It seems likely that many of the sports can already be handled by the core DTD, but a first stage is to analyse the various sports and establish what additional information is needed. It was pointed out that this effort is not only relevant to the Olympics as coverage of the same sports is also needed at national level - and at college level in the USA.

Jean-François Ricard (AFP) informed members that the IOC news feed for Athens was being provided by the Sema Group. This feed was basically a text format and the software for it was virtually finished. It was suggested that efforts might be made to use the raw data feed to produce a SportsML feed. However, it was recognised that practical considerations might make it more realistic to only produce sample SportsML output from Athens.

A query was raised about the possibility of including betting data in SportsML. There seemed to be general approval of the concept of adding appropriate features though any decision will depend on the requirements of members and other users. A further point that needs to be taken into account is that some sports federations would have objections to the inclusion of betting data with their results.

NITF Support

Since the Chairman Alan Karben (XML Team Solutions) had been unable to get to Aarhus, a web meeting was set up to allow him to take part in the session from New York. Delegates were also able to log-in to the meeting - using the (much appreciated) Internet connection provided in the conference meeting room.

A number of, generally minor, adjustments had been made to give a draft NITF V3.2 and these were reviewed and modified where considered necessary. Agreed changes are fully backwards compatible (being additions and not changes) and are as follows:

Change content component of body.head

from headline? to headline* to allow for multiple headlines. These had been requested to make it possible to have alternative headlines for different media and it was pointed out that it also helped make the metadata handling in the NITF closer to that of NewsML.

Add xml:lang attribute to the %common-attributes, with the values similar to general XML practice as in RFC3066.

Changed dd's content model from (block)* to (block | %block.content;)*

Added common-attributes (class, style, and xml:lang) to: pronounce; classifier; distributor; pre; org; person; name.given; name.family; postaddr; and lang.

To improve consistency the remaining PCDATA references have been replaced with %plain-text entity.

References within the content models have changed to %general-text for a group of elements: rights.owner; rights.agent; distributor; copyrite; pre; location; sublocation; city; state; region; country; org; person; name.given; and name.family.

NITF V3.2

There had been an intention to produce a NITF V3.2 at the meeting, but after some discussion it was decided that a better approach would be to try and deal with other outstanding issues - such as "Ruby" handling with the intention of generating a final version for adoption and release in Leipzig. It was agreed that the handling of "Ruby" would generally follow the W3C guideline for an XHTML module. The appropriate changes have since been made to the NITF DTD and Documentation and a draft NITF 3.2 is available for comment at <http://www.nitf.org/3.2/>. A draft XML Schema for NITF V3.2 is also under preparation.

Comments were requested on the value of producing, or at least proposing, an XHTML version of NITF.

It was pointed out that there were two e-mail discussion lists dealing with the NITF and it was decided to simplify matters by only having a single list. Accordingly the little used nitf-standard list has been closed and all interested parties should make sure they are registered with nitf@yahoo-groups.com.

News Management

There was no News Management Working Party session at the 2003 Aarhus Meeting.