



IPTC Mirror

IPTC - INFORMATION TECHNOLOGY FOR NEWS

Coming soon - Updated Standards

Work underway should result in the formal release of NewsML V1.01 and SportsML V1.0 at the Autumn 2002 Meeting, with the July San Francisco Meeting having dealt with many outstanding points and identified the areas that need further attention.

Standards

Main activity of the Standards Committee was the normal overview of the activities of the various working parties - considered in more detail on the following pages - along with formal approval of the recommendations for the format of Government Documents, and for release of NITF V3.1

Delegates were informed that London-based Members Hugin SA had been contacted by the London Stock Exchange who claimed copyright of the Tradable Instrument Display Mnemonics (ticker symbols) and were seeking to licence their use. It appears that this is mainly a UK problem as members from other countries had not, as yet, encountered any such problems. However it was agreed that IPTC were concerned about this and would continue to monitor the situation

This was the last session as Standards Committee Chairman for John Iobst (NAA)



who has passed on the responsibility to Stéphane Guérillot (AFP).

News Management

A draft Guideline on News Management with NewsML - produced by Working Party Chairman Stuart Myles (Dow Jones) - had been circulated to Members before the meeting and was extensively reviewed. The Guideline deals with three scenarios: *No Archive* - involving the complete replacement of a NewsItem; *Write Through* - with updating, deleting or replacing parts of a NewsItem; *Special Instructions*; and *Change of Revision Status*.

A series of comments had been received prior to the meeting and these, along with other points raised during the session, have since been incorporated into a revised release of the Guideline. This has been circulated to allow further consideration at the Autumn Meeting,

As presented the Guideline deals with the management of content, but there was some discussion about the possibility of extending news management to an earlier point in the production process. An example would be the release of an Advisory concerning the planned release of specific news content. Possible ways of achieving this were considered but there was some doubt whether the practice was strictly News Management, or if it should be dealt with within the News Agency Guidelines.

During the AGM, Stephané Guérillot (AFP) - top left - took on responsibility for the Standards Committee, with John Iobst (NAA) having stepped down to concentrate on his new duties as IPTC Chairman. This move left a vacancy for Chairman of the News Metadata Working Party, which was filled by John Minting (UPI) - bottom left.

AGM Technical Sessions

Standards Committee	P1
<i>Recommendations for the format of Government Documents and NITF V3.01 approved.</i>	
News Management	P1
<i>Guideline on News Management in NewsML</i>	
Specialised content	P2
Government Data	P2
<i>Recommendations reviewed and released.</i>	
Election reporting	P2
Financial Data	P3
SportsML	P3
RadioTVNewsML	P3
Weather reporting	P3
News Metadata	P3
<i>Subject Reference System revised, additional entries approved, procedure to be reviewed.</i>	
NewsML Support	P4
NITF Maintenance	P5
PR Committee	P5

Presentations

Future of News	P2
<i>Jerry Ceppos, Knight Ridder gives a personal view.</i>	
Chinese NewsML	P3
<i>Overview of recent developments and available tools.</i>	
Wireless News	P4
<i>ANSA offer a full news service on a PDA.</i>	
Automated Indexing	P5
<i>Randall Marcinko, Nstein</i>	
XML Adoption Realities	P5
<i>Bob Ainsbury, Xyleme</i>	
Autumn Meeting	P2
Latest Papers	P5

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 underlined blue links www.iptc.org will attempt to open your browser and connect to the appropriate site.*

Autumn Meeting 2002

9th to 11th October
Park Hotel
Amsterdam

Note: The venue for this meeting has been changed from Barcelona to Amsterdam to ensure adequate accommodation for delegates.

It seems that there might be some advantages in establishing a mechanism for news providers and users to agree on the management procedures being adopted for a given service. One proposal was for a series of specifically identified standard mechanisms so that Level 1, say, could refer to complete replacement (No Archive), while Level 2 would deal with the Write Through approach, and so on.

The possibility of extending the concept to allow auto-negotiation of news management was also considered, with general agreement that it might be of future interest.

Members attention was drawn to the management model originally developed in association with the CLIPPING project which includes a set of specific status definitions and might provide a way to help describe the NewsML processes. This was subse-

quently used as the basis for an appendix to the Guideline which explains use of the Status TopicSet for News Agencies. The document detailing the CLIPPING model has also been reissued for reference purposes.

Possible ways of distributing the final version of the Guidelines were also considered, with the main options being to publish them as a separate item, or to include them within the general News Agency Guidelines.

Provision of news management in the NITF has also been looked at with a proposal being submitted to the NITF Maintenance Working Party for consideration.

Specialised Content

Activity in this area continues to run at a high level with a number of different areas under active consideration.

Government data: An IPTC initiative has been launched to encourage Governments (along with other official bodies and large organisations) to provide their data in formats that are easy to handle. Walt Baranger (New York Times) explained that the initial impetus for this had come from US members following the President Clinton Impeachment Proceedings. These had resulted in thousands of documents being presented in a variety of formats (around 18 in total), and a considerable effort and expense was needed to get the information into a usable form.

Although there is no single right way, a set of recommendations have been produced, describing ways in which such information can be presented to make it easier to process. These recommendations were discussed in some detail and it was agreed that the preferred format would be Tagged Files using XML, with specific reference to both NewsML and the NITF. Other formats considered - in order of preference - are: Text only; Tagged files using HTML; and PDF. Specific information is also given on handling tables and photos, along with information on the availability and distribution of electronic documents.

The recommendation document is aimed at non-technical readers - such as elected officials - and it is anticipated that it will then be passed on to technical staff for implementation. The hope is that the technical people will have a better understanding of the technologies (such as NewsML) that should be applied.

The approved file has been posted on <http://www.iptc.org/site/govdocs.html> in all of the recommended formats and members are asked to bring it to the attention of appropriate parties.

Election reporting: Working Party Chairman Geoff Haynes (AP) said that progress was being made in this area with a number of inputs. Willis Fong gave a brief demonstration of the system that Canadian Press used for the 2000 elections, and details were provided of a Press Association appli-



Keynote presentation at the Annual General Meeting was given by Jerry Ceppos, (Vice President News, Knight Ridder) who looked at "The Future of News."

He believed that the actual delivery mechanism was not particularly critical as people will continue to need news, and journalists will provide the news in whatever way it is easiest for the user to access. What was more important was the worry that the news industry could give away its greatest advantage - its brands. In general the newspapers and news agencies have very valuable brands that can be leveraged because they can be trusted.

While news itself may be considered as a commodity, this is not the case for specific brands and what the industry needs to do it to continue to separate its brands from everybody else - whether distribution is in print or online.

Mr Ceppos went on to explain that he believed that one reason why this might not continue to be the case is the unwillingness of many journalists to change their habits and write for online applications. Newspaper news sites will need frequent updating to compete.

The Future of News

A second reason was "the lure of the certain". So much of the news industry's revenue is tied up with traditional sources that it can be difficult for the business side to focus on on-line services. The danger is for organisations to continue to think that they are in the newspaper business, rather than in the information business.

Finally, there was the danger of not maintaining credibility - one of the standards that newspapers need to apply is correcting mistakes in clean and obvious ways. This is something that the best papers have learned to do in print - and the same practice needs to be applied on-line.

In summary, Mr Ceppos said that he believed that there were three things that the news industry needed to do: get the journalists to care about online and file repeatedly on breaking stories (as many news agencies already do); get the financial people to look beyond the next quarter if possible; and to capitalise on their great brands - in both print and online services.

Mr Ceppos's speech was followed by a very lively question and answer session which explores some of the points raised in greater detail. A fuller version of the presentation, along with look at the subsequent discussion will be included in the IPTC Spectrum (to be published later in the year).

cation. OASIS (www.oasis-open.org) are continuing work on their Election Markup Language, with the Election Results Reporting system probably being of most interest to IPTC members.

AP are members of the Voter News Service (www.vnusa.org) which provides an election-night results service and is designed to handle a large number of short, abbreviated, messages. An XML based system is being developed for delivery of results in the USA November 2002 elections. The new version is planned to be more user friendly and it is believed that web sites may be major users.

In addition to the approaches described above, AFP have a DTD which has been used for French elections, and dpa are looking at a solution of their own.

Overall it was felt that there was sufficient interest to continue work in this area and see if it would be possible to produce an election reporting standard that would meet most members' business needs. As a first stage it was proposed to take a more detailed look at the efforts of individual members - and other organisations - with a view to the possibility of pulling them together into a common standard or making specific recommendations on the use of external standards. However, it was recognised that rapid progress would be needed.

Financial data: Irving Levine (Reuters) reported that there was significant activity in this area: RIXML have produced a revised

specification (www.rixml.org); a revised XBRL specification was produced towards the end of 2001 (www.xbrl.org), and MDDL (www.mddl.org) was being expanded to deal with bonds. There was also a new Reuters initiative MarketsML which was at the concept stage and designed to relate to other financial standards.

So far use of these standards appear to be limited, but it seems likely that both RIXML and XBRL applications will become more apparent during the next year.

SportsML: Activity in this area has been somewhat restricted over the past months, but the draft version has now been further developed to a Version 0.7, with the aim being to achieve a full V1.0 release later in the year. (www.sportsml.com).

RadioTVNewsML: Motoyasu Mise (NSK) reported that the Japanese group were holding regular weekly meetings to work on the draft specification. Listings for British, American and French programs have been looked at as part of this work. A web site covering the work has been established (<http://www.newsml.jp/radiotv/>) along with a mailing list (<http://groups.yahoo.com/group/radiotv-newsml/join>). It was hoped that more detailed progress report could be made at the Autumn Meeting.

Weather reporting: Initially there appeared to be little interest in continuing investigations into the handling of Weather data. However, during the Standards Committee meeting, members were informed

that a number of US newspapers appear to be active in this area so further investigations will be undertaken.

News Metadata

Following proposals made at the Spring Meeting a general cleanup of the Subject Reference System (SRS) has been undertaken. This involved application of consistent rules for spelling, capitalisation and use of plurals along with a look at the numbering system, and removal of duplicate entries.

Summarising the changes Metadata Working Party Chairman Stéphane Guérillot (AFP) said that although the SRS had been widely modified, the changes did not extend to any great depth. The SRS Guidelines have been updated to cover the new rules. It was also noted that the translated versions of the SRS will have to be updated.

The session went on to consider a series of proposals for new entries with agreed additions including:

Social Issues - abusive behaviour added as Subject Matter.

Politics - censorship as a subject matter. *Economics Business and Finance - economy (general), business (general) and finance (general)* added as Subject Matters - it appears that these had been omitted when the original set of entries had been produced.

Chinese NewsML Update

A brief update on the rapid progress being made by the Chinese NewsML initiative was provided by Jussie Chan from CINTEC. Aim of the initiative is to encourage the Chinese News and Media Industry to adopt NewsML and CINTEC is 90 per cent funded by Government sources with the remaining 10 per cent coming from other areas, including media and software organisations. Activities include general studies on NewsML; training and consultancy; running Seminars and Exhibitions; development of software tools; and participation in IPTC activities.

The development of tools is specifically intended to help lower the entry barrier for potential users, with a set of five now being available:

Chinese NewsML Text Editor - a text editor that lets users create, and edit news in a NewsML format.

Content Management System - allows news aggregation (with selection from a news archive), proof reading and publication in a categorised news package.

News Publishing System - integrates with the Content Management System to allow the generation of Web pages by the use of XSLT style sheets.

Taxonomy Authoring Tool - allows authoring and manipulation of taxonomies with support for the IPTC Subject Reference System.

XML Search Engine - demonstration using a native XML database with XPath queries. Designed to help developers who are more used to other storage technologies.

Further information on CINTEC activities, along with downloadable versions of the tools is available on <http://cnewsml.org>



Chinese NewsML Community
中文新聞標示語

About Us

Our Mission

The Chinese NewsML Community was set up by The Chinese University of Hong Kong by Innovation and Technology Commission, HKSAR. Its charter is to establish an NewsML standard and supporting tools in order to speed up the pace of e-business and enhance Hong Kong's commercial competitiveness in the world. The NewsML released by International Press Telecommunications Council, is designed for news operations which leads to better quality of services and higher efficiency. Although of NewsML have been recognized internationally, the news media and publishing of Hong Kong has only recently started to investigate it. Thus, there is an urgency for their potentials of NewsML and its associated technology.

Project Team

Principal Investigator: [Patrick F. WONG](#)

Senior Project Manager: [Ng Yung-Kit](#)

Project Engineer: [Ng Yung-Kit](#)

Trade dispute as a Subject Detail under *economy (general)*.

Wedding services as a Subject Detail under *financial and business services*.

Additional entries for *programme* and *exclusive* were added to the *Genre* Topic Set.

A number of other proposals were also looked at, but not approved as it was felt that they were already covered under another heading, or needed to be more closely defined or explained.

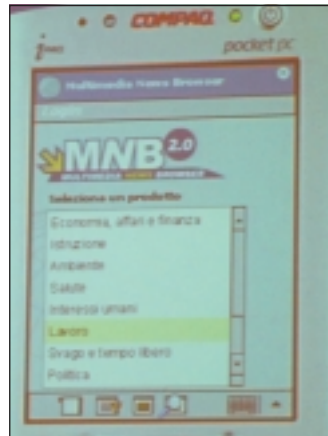
As with previous meeting, considerable time was spent considering the various proposed entries - a second session was convened for this purpose - but it was pointed out that the value of the standard depends on the rigour applied in setting it. It was decided that the Update and Fast Track change procedure needs to be looked at to ensure that it will both continue to meet the needs of the members and ensure that the SRS retains credibility as a major international standard.

Closing the meeting Stéphane Guérillot (who is taking on responsibility for the Standards Committee) thanked delegates for the support he had received over the years in what had been an interesting, and delicate, task of running the Working Party.

NewsML Support

Opening the session Chairman Irving Levine (Reuters) reminded members that the main aim was to produce a first update of NewsML (Version 1.01) -- two years after the standard was released. With this update the intention was to improve usability and address some minor shortcomings. There were a number of items that had remained open over the past few meetings and these needed to be dealt with.

Subsequent discussions looked at possible DTD changes in some detail, including: extension of the *DateLine* structure with lo-



Wireless News!

An application that lets users receive a professional news feed on a PDA using a wireless connection - and is said to be one of the first commercial integrations of these technologies - was demonstrated by Angelo Marrara (ANSA) during the AGM.

At the moment it can handle text and images, with audio to come. The system is Java based, so the application can be run on any device, while the underlying news format is NITF.

Users pay for the news content and the data transfer, with the target market including journalists and businessmen, though there may

also be interest for some consumer applications. Customers define the news profile that they want to receive - as opposed to existing commercial systems that tend to offer a limited subset of "top" stories - and the system can already handle eight news feeds.

The system also has the capability of using a PDA to send news - though this would also need facilities to enter the item into the PDA.

cation information; introduction of a *ByLineTitle*; and provision for *Event Location*. In the event this work raised a further series of points and arrangements were made for a follow-up teleconference to refine the various approaches and generate firm proposals. Recommendations resulting from this conference are given in the accompanying panel.

Examination of ways to deal with the "Original Content Creation Date" resulted in general appreciation that there was actually a series of related dates that needed identification. These included: date the image content was created; date captured; date modified (including creation of a digital version); and date published. There could also be individual copyright dates associated with content and *NewsComponents*. Further work is under way to establish the best way of handling all these dates.

Agreement was reached on extending *Genre* to allow zero or multiple references, and on including a *CommentType* attribute for the *Comment* element. However a proposal to change the handling of *Mime Type* was rejected as it was thought that this was best left individual publishers who could more readily identify those *Mime Types* that they use. Discussions during the meeting also covered a number of additions and modifications to the *News Agency Guidelines*.

There is a need to establish a *NewsML* namespace to allow for the introduction of *Schemas* and the possible inclusion of other XML content in *NewsML*. The approach adopted is to declare the namespace as an official (registered) URN that would have the form `urn:newsml:iptc.org;20001006:NewsML`. This is the same as the registered *NewsML* URN, but without the *RevisionId* and the *NewsItemId*. An approach has been made to the IETF to have this structure approved.

Practical considerations mean that some users would rather not include the document type declaration in their *NewsML* instances, but this means a way is needed to identify which version of the DTD is being used. To allow for this an optional version attribute has been included.

It is intended to publish the *NewsML* Schema at the same time as the *Version 1.01* DTD and it was agreed that this will now be achieved by publishing them - and future updates - using *NewsML* *Version 1.0*. This will help to ensure a unique ID for each updated version and make sure that the content can be handled properly. Investigations have shown that it will be possible to use a *Digital Signature* in *NewsML* applications. This can be done at various levels -

A teleconference held to consider some outstanding points from the San Francisco Meeting has proposed making the following changes to NewsML. These will be considered at the Autumn Meeting.

1. replace *Genre ?* with *Genre ** (to allow zero or multiple references)
2. add a *CommentType* attribute to *Comment* element
3. define a *NewsML* namespace with the format `urn:newsml:iptc.org;20001006:NewsML`
4. add an optional *Version* attribute to the *NewsML* element
5. clarify use of the optional *Scheme* attribute in the *Property* element or `%formalname; entity`.
6. add an optional *AllowedScheme* attribute to the *Property* element. *AllowedScheme* and *AllowedValues* would define scheme and topicset for *Value*.
7. clarify use of the *Encoding* element
8. Introduce the *ByLineTitle* element and restructure *NewsLines* to accommodate it.
9. Add the optional elements under *DescriptiveMetadata* to give a logical equivalent of the *DateLine*.

for example to sign content, complete NewsItems and parts of a NewsItem - and multiple signatures can be supported.

NITF Maintenance

Incremental development of the NITF continues, following the established pattern. A series of additions and changes were made following the Paris meeting, and a follow-up session held in Washington, and Chairman Alan Karben (Screaming Media) provided a brief overview of the planned NITF V3.1.

The name space has to be properly allocated, and contact has since been made with IETF (Internet Engineering Task Force) to do this. Date formats have been modified to conform to the way some members use ISO 8601 - as with NewsML - and comments added to the documentation to indicate the preferred approach.

Previous discussion had shown that there was a need for news management in the NITF and a method of doing this was reviewed. The aim was to achieve similar functionality to that provided in NewsML for applications where the NITF is being used in stand-alone mode (this is common practice in the US). This was achieved by the

addition of management-status and management-idref-status attributes to the docdata element. Allowed values are: usable; embargoed; withheld; and cancelled.

A request for the addition of a "Gender" attribute to the "Person" element resulted in some discussion. Although the need for the attribute was readily accepted, there was some doubt about how it could be applied when the gender was not defined. It was agreed to add a note to the DTD explaining that in such cases the attribute should not be used. The revised NITF V3.1 was then formally adopted and released. (www.nitf.org).

Work is continuing with the mapping of NITF to NewsML which is intended to help simplify use of the standards together with particular reference to metadata handling. The previously developed spreadsheet has been updated (to V0.4). There is also a draft XSLT stylesheet which can take a NITF document and return it as NewsML with the appropriate metadata, but with the body content retained as metadata. Both of these have been circulated for comment and it is hoped to complete both items by the end of the year. In addition a NITF to SQL stylesheet has been developed by Ray Snow to help loading NITF into a data-

PR Committee

Following the departure of Tony Allday, Walt Baranger (New York Times) was persuaded to take on the responsibility as Chairman pro-tem. Members were asked to provide input on any reactions they had received following previous releases.

A press release on the new government guidelines and on NITF V3.1 was published following the meeting - and is available on the IPTC web site. However, it was decided that specific promotion for NewsML and SportsML would take place following the Autumn Meeting, when it is anticipated that revised versions will be formally approved.

Agreement was also reached to undertake production of a standard format for IPTC press releases. In addition to the specific item and the release date the package will include background information on IPTC.

base.

A request had been made for the addition of an element to simplify the use of "Ruby" which can be used in Japanese applications to annotate the base text. After some discussion it was decided to see if it would be possible to adopt a similar mechanism to the plug in module used in HTML, or to use a namespace mechanism. Other features that would improve the internationalisation of the NITF will also be looked at.



Randall Marcinko

The AGM also benefited from a couple of lively presentations dealing with XML and the application of taxonomies such as the IPTC Subject Reference System.

Randall Marcinko, President and CEO of Nstein (www.nstein.com) discussed his organisation's approach to automated indexing, with particular reference to work they have undertaken to develop a system for UPI.

Their approach is to identify the underlying concepts in document, using a combination of statistical pattern analysis, lexical approaches using predefined lists, and linguistic analysis. These concepts are then used to derive the "linguistic DNA", or signature of the document which can be linked to an appropriate taxonomy to identify the appropriate categories.

In the UPI application the taxonomy is based on the IPTC Subject Reference System, and since the process is essentially language-independent it is able to handle material from variety of sources.

Bob Ainsbury of Xyleme (www.xyleme.com) took a look at "XML Adoption Realities". He explained that the industries that appear to be leading in the adoption of XML are News and Press, Financial Services and Insurance.

Bob Ainsbury



It seems that an important factor is the ease with which tagging can be automated, while one of the main reasons for adoption is to save money. There is only a limited tendency to try and make a total change, with an incremental approach being more popular. Attention has been focused on the Return on Investment (ROI), while many XML solutions have been produced internally. Adoption appears to be led by the major companies with smaller organisations being more selective in their approach.

Latest Papers

DIR 02/02	Amsterdam Schedule
CAT 02/08	Procedure for SRS changes
MAN 02/09	Amsterdam Agenda
NM 02/11	Amsterdam Agenda
NMA 02/17	Proposed Subject Code entries
NMA 02/18	Proposed Subject Qualifiers
NMA 02/19	Amsterdam Agenda
NMG 02/06	Amsterdam Agenda
NMLS 02/22	Definition of image quality
NMLS 02/23	Amsterdam Agenda
NMLS 02/24	Proposed changes to NewsML
PRC 02/05	Amsterdam Agenda
SPC 02/09	Amsterdam Agenda