



IPTC Mirror

IPTC - INFORMATION TECHNOLOGY FOR NEWS

Autumn Meeting 2004-2

Both Specialised Content and News Metadata remain very active areas with the work carried out having a significant influence on the final form of the new IPTC standards family.

News Metadata

Updates of the various metadata NewsCodes remains a continuing process with SubjectQualifiers V11, Genre V7 and SubjectCodes V14 all having been released prior to the Amsterdam Meeting. Continuing with this an updated V15 of the Subject Codes was released after the meeting, taking in the latest approved additions

These additions included a series of submission that had been approved by the "Fast-track" jury. Working Party Chairman John Minting (UPI), thanked the members of the jury for their continuing efforts, which helped ensure that the users' requests were dealt with quickly.

Explanations

Delegates were informed that further work had resulted in almost all of the Subject Code entries now having proper explanations - around 500 entries had been dealt with, and specific recognition was given to the contributions made by the editors at AFP and UPI. However, some of the Subject Qualifiers still have to have appropriate explanations added.

A proposal to add a new SubjectMatter covering **medical conditions** was ap-

proved, as was a proposal to change the **metal goods and engineering** SubjectMatter to **manufacturing and engineering** as it was felt that the original heading was too restrictive.

Submission rules

Submissions for additions to the SubjectCode system are limited to a maximum of 20 terms at a time, but it was pointed out that there was no limitation on the number of submissions that could be made.

Accordingly it was decided to introduce a new rule giving a limit of one submission per week from any given party. The form for making submissions - available on the IPTC web site - has been modified to show the parent and/or grandparent for new entries, as this should help simplify consideration of the proposals.

NewsCodes Viewer

Development of the NewsCodes viewer (by Johan Lindgren of TT) has continued with the capabilities extended to cover both Subject Codes and SubjectQualifiers, and to give an editing programme that can be used for updating all of the IPTC NewsCodes.

The NewsCodes are handled as NewsML

TopicSets, with the data being held on an IPTC database. Users will be able to view and download the XML files using the Viewer, and provision has been made for file editing.

The intention is that the viewer will be made generally available when development is completed. However, the editing facilities will only be usable by users who have been allocated an editing key. Edited files will have to be returned to the IPTC Managing Director for addition to the database.

A demonstration of the programme by Johan Lindgren was well received by delegates, who believed that it would prove to be a very useful tool - with particular value for members undertaking translations of the TopicSets.

Translations

In this area members were informed that the Version 11 of the Subject Codes was available in both French and German versions, with a Spanish translation of V10, while updating of translations to V14 are underway. In addition an Arabic translation is being produced by UPI.

After some discussion it was decided that support for translations (on the IPTC web site) would be primarily for languages which are used in more than one country - such as German, Spanish, Arabic and French. However, some support would also be provided where members produce and maintain translations in other languages - for example Ritzau l's have expressed their in-

Summary

News Metadata: Additions to Subject Codes and review of policies regarding submissions, and translations - [Page 1](#) - and towards private extensions - [Page 2](#). Demonstration of the improved NewsCodes viewer - [Page 1](#). Working Party renamed as the NewsCodes Working Party with new support forum launched and an integrated approach for development of the NewsCodes - [Page 2](#)

NITF: Minor changes proposed for NITF 3.3 - [Page 2](#)

Specialised content: EventsML use cases and sample code -

[Page 2](#). Update of SportsML to V1.5 with improvements to both the core and specific sports - [Page 3](#). Draft XML Schema available for SportsML - [Page 3](#). Other markup language developments - [Page 3](#). AdsML - [Page 3](#).

2005 Spring Meeting - [Page 2](#)

The web edition of the IPTC Mirror has hyperlinks for web addresses (in blue) and for the page references (in red)

2005 Spring Meeting

3 - 5 March
San Diego, California, USA

Members intending to attend this meeting are reminded that they will need machine-readable passports for entry into the USA, and are recommended to check current USA visa requirements.

Registration will be required from 2 January to 31 January 2005, and can be carried out from the Members Only section of the IPTC web site.

tention to produce a Danish translation. Information on the provider of the translation will also be included - though it is not intended to offer translations produced by non-members.

Extensions

Further consideration was given to the question of private extensions to the Subject Codes. The Working Party Chairman, John Minting (UPI), expressed his concern that the development of too many private taxonomies (by IPTC members) could have an adverse effect on the SubjectCodes. Wherever possible he would ask members to submit their requirements for consideration as additions to the SubjectCodes. Against this some members pointed out that they had very specific requirements, involving the use of terms that would probably not be considered appropriate for general use.

Members were reminded that during the original development of the SRS a specific mechanism was included in the IIM (Information Interchange Model) to allow for private extensions. In NewsML private sets of codes can simply be handled as an additional TopicSet - and this approach could become even simpler with the planned standards developments.

NewsCodes Branding

Agreement on the principle of introducing a common NewsCodes brand covering all of the IPTC metadata taxonomies had been reached at the AGM and during the Amsterdam session consideration was given to some of the implications of this decision.

It was considered important that there should be a single interface for users, with the set of NewsCodes being presented as a single product. The appropriate internet domains have been registered and link directly to the appropriate section of the IPTC web site (www.iptc.org/NewsCodes/).

In addition a discussion forum has been established at <http://groups.yahoo.com/group/newscodes>.

There are already comprehensive guidelines for the SRS, and it was decided to see if they can be extended (and modified) to

provide a set of general NewsCodes Management Guidelines, with particular attention being given to management and extension of the codes.

Finally, the Metadata Working Party approved a change of its name to the NewsCodes Working Party, with the Chairman remaining John Minting (UPI) and the Vice Chair Honor Craig-Bennet (PA).

NITF

Unfortunately the Working Party Chairman - Alan Karben (XML Team) - was unable to attend the Amsterdam Meeting, but he was able to participate using a telephone link from Los Angeles. He informed members that the NITF discussion group now had more than 400 members, but that there had been relatively little demand for further modifications or additions to the standard.

Recent months had only seen a few requests for minor changes, along with some proposals for improvements to the documentation. There had been no proposals for new elements. It was anticipated that an updated NITF 3.3 - taking in the suggestions - would be put forward for approval at the Spring 2005 Meeting. It was also hoped that a XML Schema version would be ready for discussion at the same time.

Specialised Content

Following the AGM the agreed EventsML Business Requirements had been made available (through www.eventsml.org) and a series of conference calls undertaken to continue development.

Working Party Chairman Geoffrey Haynes (AP) informed members that one of the conference calls had involved a representative from NewsEngin Inc (<http://newsengin.com/>). This is a US company that provides systems for the management of editorial workflow and knowledge management, including "Event Tracker" which is intended for the co-ordination and organisation of events listings.

Feedback on the EventsML project in-

cluded a comment to the effect that the standard would be valuable, and that newspapers would probably be willing to pay for formatted events information.

Use Cases

An overview of EventsML was provided by Johan Lindgren (TT) who explained that the basic structure would provide for one or more "Event", and optional "sub-Events". Information on "Participants", "Affiliations", "Assignments" and "Resources" would be separate with appropriate linking.

Several use cases have been looked at by the Working Group and example were given with sample code. These examples included:

- 1) Coverage of a scheduled news event (such as a press conference);
- 2) A News Agency Daybook (providing information on the coverage of planned news events);
- 3) Sports Schedules (such as the planned fixtures for a given League);
- 4) Breaking News (essentially dealing with previously unplanned events); and
- 5) Leisure News listings (with provision of extensive editorial information).

Points raised during discussions on the presentation included the need to make use of appropriate standards for information such as date/time, telephone numbers and location (use of GPS data was suggested for this). The question of using Vcard for information about individuals was also considered, but members were reminded that this was considered inappropriate since it was not XML based.

A possible approach would be to produce a "superset" of information for IPTC use and provide mapping to other widely used standards. It was also pointed out that many standards had extension mechanisms to allow customisation.

Unambiguous identification of events is necessary and provision has been made for both a **GlobalId** - which might be assigned by the event organiser, and a **PubEventID** - which would be assigned by the information provider.

Since handling information on events would involve exchanges of data with the provider's customers some form of news management was considered necessary

Development considerations

Some aspects of EventsML development will be strongly influenced by other IPTC developments (such as the Global Metadata Framework and the standard XML Schema design). Guidance is required on the general approach to be adopted in these areas, and during discussions it was suggested that the approach adopted by AdsML might be worth investigating (see panel on next page).

Other Markup Languages

As part of the work of the Specialised Content Working Party a watching brief is kept on other markup language, and related developments, that may be relevant to the news industry.

In this area it seems that XBRL (eXtensible Business Reporting Language - <http://www.xbrl.org>) is now being widely adopted. Other languages dealing with financial information include MDDL (Market Data Definition Language - <http://www.mddl.org>), and RIXML (Research Information Exchange Markup Language - <http://www.rixml.org>).

In other areas the EML (Election Markup Language - http://www.oasis_open.org/committees/election/) (EML) has been recommended to member States by the Council of Europe for use in e-election and e-referendum applications, while steps have been taken to establish a Steering Group for XPRL (Extensible Public Relations Language - <http://www.xprl.org/>).

In general it was appreciated that efforts to develop a fully integrated approach might result in some development delays for standards like EventsML and SportsML, or could imply the need for the standards to be revised when common elements of the new IPTC standards family have been agreed and implemented.

Other Projects

In other areas delegates were told that there had been little further activity to report on the Weather Data project.

Associated Press (AP) reminded members that they had been using their own system for the past couple of years to cover elections in the USA, and that no significant problems had been encountered. The possibility of producing a general system IPTC for election results had been previously considered, but it had been decided that the business model was unsatisfactory, while significant problems were raised by the wider range of election systems in use.

Charles Tichenor from AP provided brief

details of an internal AP project to produce a ProfileML to handle information on people, organisations and place, typically with associated multimedia.

A very wide range of data was involved, so it was appreciated that development would be a major effort, but AP had a specific business need for this and were pushing ahead with development as fast as possible. As work continued the Specialised Content Working Party would be updated on the progress being made.

SportsML

The SportsML discussion group - <http://groups.yahoo.com/group/sportsml/> - continues to expand and Alan Karben (XML Team) told delegates that September had seen 20 more members joining, taking the total to over 330.

A series of proposed improvements had been discussed at the Spring 2004 Meeting and a formal SportsML 1.5 specification put forwards for approval at the Amsterdam session. This included changes and additions to both the core and the plug-in modules.

- Main changes to the core include:
- Provision for "Wagering Lines" to handle betting statistics used by newspapers and on-line services;
 - A group element to hold properties of, for example, conferences, leagues, and divisions;
 - An affiliation element to allow player and coach statistics to cover association with multiple teams;
 - Improved support for tournaments;
 - Better support for attendance statistics; and
 - Extensions to the associate element; and clarification of the documentation.

Changes to the individual sports modules mainly involve additional attributes to provide more comprehensive and detailed coverage of both team and individual performances. For example for soccer the **stats-soccer-defensive** has new attributes for *shots-on-goal-total* and *goals-against-total*.

Following discussion the SportsML 1.5 DTD was adopted and moved to the Standards Committee for formal approval.

Olympics coverage

During the 2004 Olympics, AP provided a SportsML news feed, which had been converted from an ANPA1312 feed by XML Team. Although the feed worked well, producing a "tight" Controlled Vocabulary had been a time consuming exercise that in-

involved development of an appropriate hierarchy. Handling the Winter Olympics events with SportsML appears to be feasible, but the detailed requirements need to be looked at.

XML Schema

A draft XML Schema for SportsML was also outlined by Alan Karben. At this stage the intention is that the XML Schema should validate existing SportsML instances but will have finer granularity with validation of values.

Design aims included: storing and validating Controlled Vocabularies such as teams, positions, players, and leagues with default data typing of attributes where there is not a maintained list; storing documentation for Sports ML elements and metadata in a standard way; and allowing for the inclusion (or exclusion) of Schemas for specific sports (baseball, basketball ...).

The proposal is for a "SuperSchema" which can be customised by users and calls in the *sportsml-core.xsd*; specific sport *xsd*'s as appropriate; filenames for all Controlled Vocabulary Resource Files used; and a *strict-type.xsd* file that makes references to the listed Vocabularies listed in the aforementioned Controlled Vocabulary Resource Files, or a *loose-type.xsd* file that just pulls in the default types (treating most values like character-data).

In addition use of the XML Schema comment mechanism - the `<annotation>` - will make it possible to incorporate detailed documentation for every element and attribute in the existing DTD, and to include properties of the Controlled Vocabulary values that might be useful for formatting engines and databases.

AdsML

AdsML is a standard for end-to-end advertising workflow, and the initiative is supported by the NAA and Ifra.

It is intended to cover the means by which data is exchanged for all kinds of advertising, in all media, and through all stages of the lifecycle of an advertisement.

First parts of the standards - the AdsML Bookings Specification and the AdsML Structured Descriptions Specification - have been released for public review.

See www.AdsML.org.