

moved directly in the Standards Committee.

SportsML-G2

Alan Karben then went on to consider reasons behind the introduction of a SportsML-G2.

For publishers it would offer additional publishing and rights management functionality along with improved versioning features. Sports news articles would benefit from the improved metadata. General use of the new standard might also offer the possibility of content-neutral platforms for content permissioning and distribution.

For users, the common metadata model could be used for routing and processing, while they could also benefit from content-neutral platforms.

Some initial experimental SportsML-G2 samples are available on www.sportsml.org/dev. The SportsML 1.8 package is also at this location.

EventsML-G2

A look at the status of EventsML-G2 was provided by Johan Lindgren (TT) who explained that a wide range of information would be associated with an actual event. This ranged from calendars and day-books to the interchange of events information, along with assignments and other actions associated with coverage of the event. In some cases historical references would be involved, while one end result would be the production of news stories.

Handling events

A first consideration was to establish how an 'event' should be fitted into the NAR structure, with three options being to use the existing Concept Item, or NewsItem, or to create a special EventItem (based on the NAR AnyItem). Development of a structure to handle information specific to an event has resulted in an eventDefinition that can be used in a number of ways, such as in a NewsItem or a ConceptItem.

Date and time information is an essential part of EventsML-G2 with a series of use cases having been developed. Start (and End) Date and Time can be dealt with by using a Partial Date-Time Type, while the possibility of using the iCalendar model to deal with recurring dates is being investigated.

Much of the information to de-

scribe the 'What' of an event is already available in the NAR, such as details about participants, the subject of the event, and the organisers.

NewsML-G2

Intended for general journalistic content, NewsML-G2 will handle, for example: text articles; photos; graphics; audio and video; illustrated articles; and multimedia presentations. Working Group Chair Laurent Le Meur (AFP) explained that, in effect, the common functions from NewsML 1 have been taken into the NAR, while the general news markup is taken care of by NewsML-G2.

NewsItem

The key element of NewsML-G2 is the NewsItem (which consists of metadata and content), with all processors having to be able to support standardised NewsItem profiles. The NAR Concept, Knowledge and Package Items are also included in NewsML-G2 but are not an essential part of the standard and may be ignored by NewsML-G2 processors.

At the moment efforts are being directed to the implementation of specific sections of the new standard.

Photo content

For photos this includes: defining a list of physical characteristics; and



Agencia EFE

Agencia EFE is a major international news agency with services in Spain and Europe, South America and the United States.

Starting - in 1939 - as a press agency offering Spanish text news services, EFE began its international expansion in 1965, opening an office in Buenos Aires, Argentina.

With its headquarters in Madrid, EFE now has 25 offices in Spain and 35 abroad, with small branches and correspondents in 150 cities from 100 countries. Daily output is more than 3000 text items (in Spanish, Catalan, English, Brazilian, and Arabic), 1200 pictures, and a range of video, audio and multimedia services.

The communication system is based on a combination of a VSAT (Very Small Aperture Terminal) satellite network with world-wide broadcast coverage and simultaneous push-pull services on a IP (Internet Protocol) network with the main node in Madrid.

Databases

Main commercial databases are EfeData and Fototeca, which has two million pictures dating from 1998.



Manuel Fuentes described development of the EfeData database.

EfeData is the EFE historical text archive covering the past 20 years and provides news items and journalistic information, biographies and documents in Spanish as well as historical news items in Catalan, English Portuguese and Arabic. Spanish content includes all EFE output since 1988 with 15 million news items and over 15 thousand biographies

Over time EFEData has undergone a series of technical changes to meet new demands. When launched in 1986 it used BASIS-K, a full text database hosted on a MicroVAX. The current system uses TaWeb, a full text information system. IPTC standards are used - NITF 3.2 and the Subject NewsCodes for classification.

A major development was the arrival of the Internet with EfeData being the first product that EFE offered on the new media.



An overview of Agencia EFE was provided at the Autumn meeting by Jose Luis del Rey.

producing a bi-directional mapping between the IPTC Core and NewsML-G2, with a particular problem being to find a way of handling 'traditional' keywords. It is hoped that information to help with this will come from the new Photometadata Working Group.

Video content

The EBU have provided a series of specific requirements for video: a 'shot list' giving a structured description of segments of the content; handling of remote content in a stream or on removable media; and provision of separate physical metadata sets of media channels (audio and video).

However these requirements are complex and carry a risk of overloading the NewsItem. Also needed for video (and audio) are sets of physical properties and detailed time references.

Rights

The EBU have also made some proposals for handling rights, based on the practice in other EBU applications.

Specification

An important step is to draft a specification document, which will be referenced to the NAR documentation. The specification

will be processing oriented with typical implementation examples including:

- A photograph in multiple resolution forms, with metadata that is compatible with the IPTC Core;
- A vector graphic with a thumbnail of the image;
- An audio clip with an accompanying logo;
- A video clip with a shot list;
- An illustrated article.

NewsCodes

A major effort is under way to develop a new version of the Subject NewsCodes with a core group of delegates - Amy Sweigert (Associated Press), Kirsty Crawford (BBC Scotland) and Honor Craig-Bennet (Press Association) - working on the project.

Some initial conclusions were presented to delegates at the Autumn Meeting.

An underlying problem with the current Subject NewsCodes is that the structure is very limiting, for example it is not possible to create relationships or link terms. In addition it will be necessary to change the current data model for use with NAR based standards.

New version

The new version will not be limited

Amy Sweigert summarised work that has been carried out on the new system of Subject NewsCodes.



to a three level hierarchy and will support polyhierarchy, allowing a concept to have multiple parents. It will be possible for the same term string to refer to different concepts, which will be differentiated by unique IDs and definitions, while codes will not be tied to the structure. Use of relationships may also allow the creation of different views of the codes to suit the needs of different users.

So far as possible concepts will be defined in a way that is culturally neutral, and it is intended to complete the coverage in areas that have only been partly dealt with. Updated guidelines - based on the US NISO (National Information Standards Organization - <http://www.niso.org>) Standards - will be developed for the production of new terms and definitions.

Existing version

The existing set of Subject NewsCodes would be maintained to support current users, but will remain subject to the limitations of the current structure. So far as possible it is envisaged that changes to terms will be applied to both sets, while new terms will be also be added to the old set, in the nearest appropriate location.

However, the structure will be different, resulting in an increasing divergence between the existing Subject NewsCodes and the new version, so it will be in users' interests to adopt the new version when possible.

Taxonomy management

Consideration of the requirements of this project resulted in investigation of the advantages of using an appropriate taxonomy management system. A suitable system had been identified (from SchemaLogic) and a presentation of the system arranged for the Autumn Meeting - see panel on page 5.

NewsCode groups

A proposal to split the IPTC NewsCodes into groups that reflect their application was raised by Michael

MESH Update

Details of the MESH (Multimedia Semantic Syndication for Enhanced News Services) Project were given in IPTC Mirror Issue 135 (September/October 2006) and a further presentation was given to delegates at the Autumn Meeting by Paulo Villegas of Telefónica, who are a member of the Mesh consortium.

Telefónica are a global telecommunications operator, who use news content in a number of services such as TV news, 3G services and Internet news. They want their news content to be rich (multimedia) and adaptable, in a structured form that can be processed automatically, be personalised, and be amenable to rights management.

Knowledge structure

An extensive knowledge structure is under development by MESH, with IPTC standards being adopted in some areas. NewsML is being used for news document structures and metadata, while the Subject NewsCodes were the starting point for the news domain ontologies, with a first test version being produced in OWL (Web Ontology Language).

Initial target domains are 'Natural Disasters'; and 'Transport Accidents' which have been extended by the addition of further concepts. For example under 'Natural Disasters'. 'Earthquake' has the additional concepts of compensation, focus, houses, similar incidents, and time.



Paulo Villegas

Steidl. He explained that as IPTC Managing Director he receives many queries about the application of the NewsCodes, with the most common being "which of the 28 NewsCodes should I use to describe the content?". Grouping the NewsCodes by application should make their use clearer, but does not

have any technical effect.

The proposal was well received, and - after some discussion on which group was most appropriate for some specific codes - the following grouping was adopted:

Descriptive NewsCodes: Subject; SubjectQualifier; Genre; and Scene.

Administrative NewsCodes: OfInterestTo, Provider, Status, Urgency, Colourspace, AudioCodec and VideoCodec.

Transmission NewsCodes: Priority.

Exchange NewsCodes: CharacteristicsProperty; Confidence; Encoding; Format; HowPresent;

Taxonomy Management

Development of the Subject NewsCodes will involve the interchange of significant amounts of material between delegates (both those directly concerned, and others with specific interests). Interchange of amendments and comments will be an important feature of the work, while it will also be important to keep track of progress.

Discussions between those involved lead to consideration of the advantages of using a professional system for taxonomy management, resulting in a presentation to the NewsCodes Working Party by a team from SchemaLogic (www.schemalogic.com).

Presentation

Lowell Anderson (VP Marketing) started with a brief introduction to SchemaLogic as an organisation which provides business semantics management systems with a corporate governance approach. Business semantics is a consensus of terms and definitions and SchemaLogic aim to provide a structured way of managing the process - model the structure (taxonomy) and information relationships (ontology) - govern and manage changes - publish to subscribing systems - and collaborate to expand and maintain.

Principles

A review of the technology used was provided by Breanna Anderson, Chief Technology Officer who started by outlining the underlying principles. These included:

- Completeness - the system is designed to address both structural definitions and the encoding of information semantics;
- Neutral - to handle systems with different physical implementations and theoretical foundations;
- Flexible - to deal with systems that have divergent management philosophies;
- Humanistic - giving consideration to human usability factors as well as machine-orientated perspectives;
- Collaborative - be directly usable by all stakeholders;
- Evolutionary - facilitates fine-grain and incremental change and evolution;
- Granular - to help reuse, sharing and governance of model components.

Vocabulary Management.

A vocabulary can be thought of as a finite list of terms and relationships. With the SchemaLogic system terms can be shared by multiple vocabularies, and relationships can be set between terms in a vocabulary.

Relationship types include Hierarchical (broader-narrower), Entry (synonym) and Related (conceptual/associative). Different views of the terms and relationships can be produced.

The system allows collaborative management of taxonomies, with a web based governance system based on the SchemaLogic Server. Users have to be registered (with user name and password) and may have different rights - with specific restrictions on the addition, updating and deletion of objects. Typical change scenarios for the NewsCodes application would include browsing and searching entries, making additions, and proposing changes.

Other tools

Detailed consideration of how well the system would meet IPTC's needs to create and maintain the new Subject NewsCodes took place after the SchemaLogic team had left.

It was pointed out that this was an important decision for IPTC, as it was something that would have to be lived with for some time.

During discussions the availability of other tools was raised, with specific mention of the Factiva Synaptica Knowledge Management System. Fortunately Stuart Myles (Dow Jones) was able to arrange a presentation of this system (Factiva is a Dow Jones company) at short notice.

Recommendation

The NewsCodes Working Party reconvened the next morning, and after due consideration it was agreed that the Factiva system did not meet specific IPTC needs, and it was decided to ask the Management Committee to consider making the necessary arrangements for IPTC to use the SchemaLogic system.

This was later done and it is hoped that the system will be in operation before the end of the year.



Breanna Anderson, SchemaLogic Chief Technology Officer provided a technical overview of the SchemaLogic taxonomy management system. Other members of the Schema Logic team were Lowell Anderson (VP Marketing) and Gary Carlson (Chief Taxonomist).

Importance; LabelType; Location-Type; MediaType; MimeType; NewsItemType; Notation; Property; Relevance; Role; TopicType.

Correction

No requests had been received for additions to the NewsCodes, but there was a proposal to correct an error in the Sports section of the Subject NewsCodes.

On some previous occasion (possible some years ago) it was decided to deprecate an entry for the American Football League. However, in error the entry for Canadian Football was marked as deprecated, leaving the entry for the American League unchanged. To correct this it was agreed to reinstate the entry for Canadian Football, with the appropriate explanation, and mark the American Football League entry as deprecated.

Fast Track Jury

Working Party Chair John Minting informed delegates that the 'Fast Track' jury now had four members, in line with the revised Change Management Guidelines adopted at the 2006 AGM session.

Translations

Work on providing translations of the NewsCodes is continuing with a number of new translations being made available in early November. These included V17 of the Subject NewsCodes in Japanese (from NSK), Italian (ANSA) and German (dpa), along with German versions of the Subject Qualifier NewsCodes V 12, Genre NewsCodes V 8, and Scene NewsCodes V 1 - all provided by dpa.

All current NewsCodes (including the new translations) are available on www.newscodes.org.

NITF

Proposals for a NITF 3.4 were outlined by Working Party Chair Alan Karben (XML Team).

There had been a request to add a facility for including namespaced elements in the 'enriched text' areas of NITF instances, and it was intended to do this by the introduction of an 'any' construct. It was pointed out that this change would only be made to the XML Schema version of NITF 3.4, and not to the DTD.

It was also planned to make a change to allow the use of zero or more <abstract> elements, instead of the currently allowed zero or one,

The Challenge of New Media

A look at the challenge posed by new media and the way news syndication has changed was provided by Scott Calder (Mainstream Data) in a presentation during the Autumn Meeting.

He started by outlining the position of Mainstream Data, which is a major supplier of information distribution services to media sites in North America. It provides content aggregation, distribution applications for content syndication, processing and display.



Scott Calder

Market value

Looking at the market value of 'YouTube' in comparison to the value of established US newspaper publishers shows the relatively high value attached to a new market entry with an interesting concept. Also the value of Internet companies (such as Google and Yahoo) is much greater than that of newspapers. Overall it seems that a presence in the new media world has a high market premium.

One reason for this may be that newspaper readership is declining (in terms of population percentage) while younger readers are less common - but new media are attracting a younger audience. Another pointer is the way the number of weblogs is doubling every five to six months.

Audience interest

So newspapers, television and radio stations, web portals and mobile service providers are all battling for the same audience. The winners are integrating relevant multimedia content to keep their audience interested; providing compelling content on web sites; adding intelligence to their content by taking advantage of content metadata and relevance/indexing technology; continuing to be the best source of information for their local communities; and syndicating their content to generate additional revenue.

Mr Calder believed that actions that needed to be taken included reducing distribution costs, delivering multimedia content in web-savvy contents, and the development of new platforms for information exchange.

Content syndication

He went on to describe how the model for content syndication was changing. With the 20th century model information for content providers passed through distribution platforms to the media users (newspapers, broadcasters and web sites. Now there is a feedback loop with content from the media flowing to other media and back to the content provider. The next stage will introduce a body of "citizen journalists" providing content to the providers and direct to the media.

Success factors

Finally he outlined the success factors for the 21st Century:

- Proper strategy. Organisations have to recognise that in the multi-centric information world a single content supplier cannot expect to satisfy every customer. They will have to leverage their particular strengths, by partnering with complementary information companies, and there will be a focus on multimedia.
- Appropriate technology. Flexibility in distribution - for example with satellite of broadcast for large objects and web systems for text content. Real-time content will have to be seamlessly integrated with archive material.
- 'Ecumenical' Solution. Be prepared to work with unexpected technology and information providers; recognise that owning a smaller portion of a much larger business is better than owning a large part of a much smaller business.

making it possible to have multiple abstracts.

At the same time the <abstract> and <byttl> elements would be allowed to have 'commonAttributes' - rather than 'globalAttributes'. This would have the effect of adding class, style and xml:lang attributes and would make it possible to differentiate between multiple abstracts.

Following discussion a motion was passed to move NITF 3.4 to the Standards Committee for release.

NITF versions

Alan then turned his attention to the possibility of splitting the NITF to give two versions: a simplified core version and a fully-featured, stand-alone, power version (generally as with the NAR versions). At the moment the NITF has some 132 elements (and a lot of attributes) and a possible split would have around 55 elements in the 'core' version, with a further 80 elements for the 'power' version.

One effect of the split would be to remove much of the metadata from the 'core' version, making it more suitable for use as a possible text format for NewsML-G2.

A spreadsheet showing an initial split is available at www.nitf.org/dev.

During the Standards Committee it was announced that Stuart Myles (Dow Jones) would be taking over the position of Vice-Chair for the NITF Working Party.

NewsML 1

Working Party Vice-Chair Takahiro Fujiwara reminded delegates that a Beta version of the NewsML 1.2 XML Schema was released following the AGM.

He provided details of the comments that had been received, and explained that they has been taken in to the Schema to give a revised version (which was distributed with the pre-Autumn documents package).

Testing

Using this version Takahiro had arranged an extensive series of tests for the draft NewsML 1.2 XML Schema with the companies involved included Business Wire, Cybertech, East Co, IBM Japan, Microsoft Japan, Oracle Japan, and

PR Committee

Brief consideration was given to the names for the NewsCodes Groups (see the NewsCodes Working Party report for details). These names will be used on the NewsCodes web pages and in other marketing material.

Members were informed - by PR Committee Chair Walter Baranger (New York Times) - that steps are being taken to establish a corporate identity for the G2 family of standards. It is intended to engage a professional designer to do the work, which will probably result in a redesign of the IPTC Web site, the IPTC publications and IPTC marketing material

Following normal practice a Press Release was issued following the Autumn Meeting, focusing on the NewsCodes developments, detailing the NITF and SportsML releases, and providing a progress report on the G2 standards project. An archive of IPTC Press Releases is available on http://www.iptc.org/pages/prel_main.php.

Toshiba Solutions.

The tests were designed to check the following:

Compatibility of the NewsML 1.2 XML Schema with a range of other XML products including IBM BD2, Oracle database and Microsoft Biz-Talk and SQL Server;

Data transfer based on a simulation of Japanese use cases (between DTD and XML Schema based systems);

Validation of instances provided by AFP, Business Wire, Chunichi Newspapers and Kyodo News.

Namespace problems

Overall the tests were successful, but problems were encountered with the use of target namespaces (comments made during the Working Party session indicated that

such problems are not confined to the NewsML 1 XML Schema). The solution adopted for the tests was to use a XML Schema file that did not have namespace requirement.

However, Takahiro explained that the development group considered it important to deal with this issue properly as the use of namespaces was important for management and identification purposes. More importantly, it was important to ensure that systems using the NewsML 1.2 XML Schema would be able to receive input from older DTD based NewsML systems.

Because of this it had been decided to postpone release of the NewsML 1.2 XML Schema until the issue had been resolved - and it was hoped that this would be done before the next Meeting.

Validation problems with XMLSpy2006

Problems have been encountered with the use of Altova XMLSpy2006 with some versions of NewsML - though this problem was not present in earlier versions. It appears that the cause of the problem is that XMLSpy2006 has an internal conversion mechanism which takes a URN and links to an internally stored DTD. However, although XMLSpy2006 recognises the NewsML V1.0 URN, the only stored version of NewsML 1 is the V1.2 DTD, so the system returns an error.

A work around this problem has been found: Remove the "XML Catalog" data for NewsML V1.0 and 1.1 from the file "MainCatalog.xml" in the XMLSpy 2006 program files directory; For NewsML 1.0 or 1.1 copy the DTD files to the sub-directory Schemas/NewsML in the XMLSpy2006 program files.

Altova have recently released XMLSpy 2007, and the problem appears to have been resolved in this version.