



IPTC Spectrum

IPTC - INFORMATION TECHNOLOGY FOR NEWS

NewsItem

- Management properties
- Links to other Items
- Content metadata
- Publication Metadata
- Signature metadata

thumbnail preview high definition

“The basic goal of the News Architecture is to provide a single generic model for exchanging all kinds of newsworthy information”

KnowledgeItem

- Management properties
- Links to other Items
- Content metadata
- Publication Metadata
- Signature metadata

```
graph TD; A[KnowledgeItem] --> B[KnowledgeItem]; A --> C[KnowledgeItem]; A --> D[KnowledgeItem]; B --> E[KnowledgeItem]; B --> F[KnowledgeItem]; C --> G[KnowledgeItem]; C --> H[KnowledgeItem];
```

IPTC Objectives

“To establish and maintain an open, apolitical international forum to promote and enable the exchange of news information in an efficient manner, while maintaining the highest technical quality. At the same time taking advantage of the advances in telecommunication and computing technology.”

Recognising the increasing use of computer based systems by news agencies, newspapers, and other news organisations, IPTC has concentrated on the development and application of standards for the high speed transfer of digital news information. Recently - in response to the influence of the World Wide Web - efforts have been directed towards systems for multimedia news and on line publishing. This includes the production of classification systems that make it possible to consistently identify news content, irrespective of the source of language of the service.

IPTC Standards

Current standards

NewsML 1: A media-independent standard for the packaging and management of multi-media news throughout its lifecycle. XML-based. www.newsml.org

NITF: Format for marking up textual news stories. XML based. www.nitf.org

SportsML: Standard for the interchange of sports data, including scores, schedules, standings, and statistics. XML-based. www.sportsml.org

IPTC Core: Metadata set primarily for photographs and used with Adobe's Extensible Metadata Platform (XMP). www.iptc4xmp.org

NewsCodes: Controlled vocabularies of terms widely used in the news industry. They include: an extensive Subject taxonomy; genres and scenes; and ratings for priority, urgency and relevance. www.newscodes.org

Standards under Development

The IPTC G2 Family of Standards will be based on the IPTC News Architecture which provides a framework and a set of common specifications and components. Use of a common style will make the G2 Standards easier to understand and implement.

NewsML-G2: Intended as a wrapper for general news in the form of text, photos, graphics, video or other media. It can be used for packaging any combination of these items.

EventsML-G2: An information interchange standard for newsworthy event information, including event publishing, planning and coverage. www.iptc.org/EventsML

SportsML-G2: Version of SportsML designed to integrate into the G2 Standards family.

ProgramGuideML: A specialised format for listings of television and radio program guides. www.programguideml.org

Legacy Standards

IIM: Container for news information in any of the common news media (including text, photographs, graphics, audio and video). Last revised in 1999. The **DNPR** is a container file format designed to carry digital news photograph data within the IIM. www.iptc.org/IIM

IPTC7901: Text message format, last revised in 1995. www.iptc.org/IPTC7901

Both the IIM and IPTC7901 are still in widespread use by the news industry around the world.

IPTC standards and supporting documents are all available for download and free use in accordance with the IPTC Intellectual Property Policy. www.iptc.org/goto/ipp

International Press Telecommunications Council

Chairman:
Stéphane Guérillot
Honorary Treasurer:
Henrik Stadler
Vice Chairmen:
Walter Baranger; Rudi Horvath;
John Iobst; John Minting
Peter Müller; Hitoshi Saito.
Managing Director:
Michael Steidl
Editor:
Hugh Johnstone

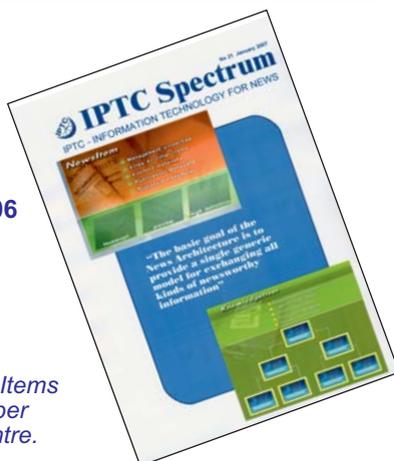
Published by the
**International Press
Telecommunications Council**
Royal Albert House
Sheet Street
Windsor
Berkshire SL4 1BE
England
Tel: +44(0)1753 705051
Fax: +44(0)1753 831541

E-mail: mdirector@iptc.org
editor@iptc.org

Web Portal: www.iptc.org

Development of the second generation - G2 Family of IPTC Standards has been a major undertaking, with a continued effort throughout 2006 giving a Release Candidate version of the underlying News Architecture.

Illustrations of the NAR Items provided by IPTC member Athens Technology Centre.



**Information
Technology
for News**

Contents

Organisation

IPTC Objectives	2
IPTC Standards	2
Members and Membership	4
A Major Investment	6
Management Committee	7
IPTC Discussion Groups	7
2006 AGM	8
Intellectual Property Policy	9

PR Committee	10
---------------------	----

Standards

Sustained Commitment	11
Working Parties and Groups	12

Photo Metadata	14
-----------------------	----

News Industry Text Format	15
----------------------------------	----

NewsML 1	16
-----------------	----

News Architecture

Common Components -	
Common Structure -	
Common Processing	17
Controlled Vocabularies	
and Qcodes	18
Person Concept	19
Any Item	20
Linking Features	21
News Message	22
Power Extensions	22

News Content

Information Exchange	23
NewsML-G2 Media	
Characteristics	23
Olympic Reports	24

NewsCodes

A Good Description	24
IPTC NewsCodes	25
NewsCodes and the NAR	26

IPTC Members

Nominating Members

Agence France Presse (afp) - France - www.afp.com
ANSA - Italy - www.ansa.it
Austria Presse Agentur (APA) - Austria - www.apa.at
BBC Monitoring - UK - www.monitor.bbc.co.uk
BBC Scotland - UK - www.bbc.co.uk/scotland
Business Wire - USA - www.businesswire.com
CNW Group Ltd. - Canada - www.newswire.ca
Deutsche Presse-Agentur (dpa) - Germany - www.dpa.de
Dow Jones & Company - USA - www.dowjones.com
Japan Newspaper Publishers & Editors Association (NSK) - Japan - www.pressnet.or.jp
Keystone - Switzerland - www.keystone.ch
KUNA Kuwait News Agency - Kuwait - www.kuna.net.kw
Kyodo News - Japan - www.kyodo.co.jp
Market Wire, a CCNMatthews Company - Canada - www.marketwire.com
NewsCom - USA - www.newscom.com
Newspaper Association of America (NAA) - USA - www.naa.org
ORF (Austrian Broadcasting Company) - Austria - www.orf.at
PA News Ltd - UK - www.pa.press.net
PLUS Coalition - USA - www.useplus.org
PR Newswire - UK - www.prnewswire.co.uk
Reuters Limited - UK - www.reuters.com
SDA/ATS - Switzerland - www.sda-ats.ch
The Associated Press (AP) - USA - www.ap.org
The New York Times Company - USA - www.nytimes.com
Tidningarnas Telegrambyra (TT) - Sweden - www.tt.se
TMNEWS-APCOM - Italy - www.apcom.it
United Press International (UPI) - USA - www.upi.com
World Association of Newspapers (WAN) - International - www.wan-press.org
Xinhua News Agency - China - www.xinhua.org

IPTC is an organization based on its members. The membership consists of news agencies, news agency alliances, newspaper publishers' organisations, individual newspapers and system vendors from around the world. Reasons for being an IPTC member include:

IPTC is the only organisation that addresses the news industry concerns for standardisation of information transfer formats.

IPTC is an organisation concerned with news agencies and their customers' information transfer problems.

IPTC fosters exposure to business ideas used around the world to distribute news.

IPTC encourages personal relationships among peers from around the world.

IPTC provides a world news lobby voice for standardisation of telecommunications services.

IPTC allows members to request research and development in areas of specific interest to their business activities.

Associate Members

AFX News Ltd - UK - www.afxnews.com
 Agence de Presse Belga - Belgium - www.belga.be
 Agencia EFE - Spain - www.efe.es
 Algemeen Nederlands Persbureau (ANP) - The Netherlands - www.anp.nl
 ANA, Athens News Agency - Greece - www.ana.gr
 AS Norsk Telegrambyrå - Norway - www.ntb.no
 Atex Media Command - Australia - www.atex.com
 Athens Techology Center - Greece - www.atc.gr
 Beijing Founder Electronics - China - www.founder.com.cn
 BVPA - Germany - www.bvpa.org
 Canadian Press - Canada - www.cp.org
 CCI Europe - Denmark - www.ccieurope.com
 Cepic - Coordination of European Picture Agencies Press Stock Heritage - Europe - www.cepic.org
 EAST Co., Ltd - Japan - www.est.co.jp/english
 EBU - European Broadcasting Union - Europe - www.ebu.ch
 Eidos Media Spa - Italy - www.eidosmedia.com
 Fingerpost Ltd - UK - www.fingerpost.co.uk
 HINA - Croatia - www.hina.hr
 IFRA - Germany - www.ifra.com
 ITAR-TASS - Russia - www.itar-tass.com
 Korea Press Foundation - Korea - www.kpf.or.kr
 La Repubblica - Italy - www.repubblica.it
 Magyar Távirati Iroda Rt (MTI) - Hungary - www.mti.hu
 Mainstream Data Inc. - USA - www.mainstreamdata.com
 Mecom - Germany - www.mecom.de
 Mediaspan - USA - www.mediaspan.com
 MENA - Egypt - www.mena.org.eg
 News Engin, Inc. - USA - www.newsengin.com
 Profium Oy - Finland - www.profium.com
 RelaxNews - France - www.relaxnews.com
 Ritzau Bureau l's - Denmark - www.ritzau.dk
 RivCom - UK - www.rivcom.com
 Suomen Tietotoimisto Oy - Finland - www.stt.fi
 Tera Digital Publishing - Italy - www.teradp.com
 XML Team Solutions, Inc. - USA - www.xmlteam.com

There are two types of IPTC membership:

Nominating membership is open to organizations and companies concerned with news collection, distribution and publishing.

Nominating members may send up to 3 persons per Contributory Unit (which is equal to a share) to a meeting, one or more Units can be subscribed. One nominating member representative (per Unit) may vote at General Meetings and Committees and all delegates can vote at Working Parties of the IPTC.

Associate membership is open to organizations and companies as for the Nominating membership and for system vendors supporting the news industry.

Associate members may send one person to the meetings, but they receive all papers and other material. Associate member representatives are only eligible to vote at Working Parties, not at General Meetings and Committees.

Further information on IPTC membership is available from www.iptc.org.

A Major Investment

Development of the new IPTC News Architecture (NAR) has proved to be a major undertaking and the process of refining the NAR model and specification to meet a demanding set of requirements was particularly complex and time consuming. It has also made heavy demands on the resources available to the organisation.

However the investment that has been made is now giving the desired results. A revised schedule for the development process has been produced, and it is anticipated that the first of the new-generation standards based on the NAR will be released towards the end of 2007.

Standards framework

The News Architecture is designed as a framework that will provide a common base for a new generation of IPTC standards. Since the new standards will have a consistent style, and make use of common components, rules and processes they will be easier to understand, and implement.

In addition, having an established structure means that it will be simpler and quicker to develop new standards, since the only new elements required will be ones that are specific to the subject of the new standard.

Consistent style

To reinforce the fact that the new standards belong to an integrated family they will be named in a consistent manner. Collectively the new news standards will be the "IPTC G2-standards" while individual standards will be known, for example, as EventsML-G2 and SportsML-G2.

The new general news exchange standard will be NewsML-G2, with this name benefiting from the success of NewsML 1 and maintaining the registered NewsML trademark. A corporate identity for the G2-standards will be developed to emphasise the close relationship between the standards.

Detailed work on the specification and model for the NAR is being carried out by a relatively small group of delegates, who have the necessary understanding of the technical requirements and how the news industry works. Their efforts are being complemented by the use of consultants to create and update a set of XML Schemas to provide an implementation, and to undertake a full quality assurance programme on the documentation.

Testing

As part of the development process the NAR has been subject to extensive testing with an initial Experimental Phase starting in December 2005. Results from this first test phase fed back into the continuing development process to give an updated model and specification for a second Experimental Phase that started in May 2006.

Feedback from this second phase, and from extensive discussions before, during, and after, the AGM and the Autumn Meeting, resulted in significant changes, with the resulting NAR Release Candidate released for public comment in mid January 2007.

The time-consuming nature of such tests - as a test package has to be put together and distributed, the actual tests carried out, and the results analysed before they can be acted on - has been a significant

Overview

A sustained effort has resulted in release of a draft News Architecture v1.0 and it is planned that the first G2-standards using this architecture will become available during 2007. Development of established standards continues with new versions of the NITF, SportsML and the IPTC NewsCodes, while these standards are finding further applications. Work has started in the area of photo metadata.

A formal Intellectual Protection Policy has been implemented. Efforts continue to improve public awareness of IPTC's standards and activities, and 2006 saw a significant increase in membership. Formal meetings were well attended with the working sessions complemented by a varied and informative series of presentations.

A naming policy has been established for the new standards family.

factor in the extended timescale experienced in this development. However, such tests are seen as essential to ensure that the News Architecture will provide a comprehensive and robust basis for the planned G2-standards family.

Other standards

Although the new News Architecture has absorbed a considerable amount of effort, the continuing contributions of members have allowed significant progress in other areas. There were new releases of both the NITF and SportsML - with DTD and XML Schema versions in each case - while a XML Schema for NewsML 1 is well under way.

These are established standards with substantial user bases complemented by a body of suppliers offering compliant systems. This makes them particularly attractive to users wishing to update older systems.

NewsML adoption

For example, in June 2006 the major Italian news agencies (Ansa, AGI, Apcom, and ADN Kronos) announced their intention to start delivering their news using the NewsML 1 format.

In another development the Korea Press Foundation have been investigating the possibility of NewsML adoption within its national news industry, and a NewsML Seminar was held in Soul in November 2006. As part of the seminar there were guest presentations from Michael Steidl (IPTC managing Director) and Takahiro Fujiwara (Vice-Chair of the NewsML 1 Maintenance Working Party).



SportsML

Similarly, interest in SportsML continues at a high level, and the standard was successfully used to provide coverage for the 2006 Torino Winter Olympics.

Comments from a Major League Baseball Club helped provide a revised and more detailed plug-in for this sport. A new plug-in has been developed for curling.

Suggestions for additional plug-ins to cover new sports are often received, but for these to be considered there also needs to be an offer to help undertake the necessary work. SportsML appears to have wide appeal outside the news industry - for example in fantasy sports applications - and this appears to be the source of some of the suggestions.

With all the standards the fact that they are freely available from the IPTC Web site means that it can be difficult to establish just how widely they are being adopted, and

Management Committee for 2006 to 2007. From left to right: John Minting (UPI), Honorary Treasurer Henrik Stadler (TT), Walter Baranger (New York Times), IPTC Managing Director Michael Steidl, IPTC Chairman Stéphane Guérillot (AFP), Peter Müller (SDA/ATS), John Iobst (NAA), Hitoshi Saito (NSK) and Rudi Horvath (APA). The Management Committees serves as the Board of Directors for IPTC, generates guidelines for the future development of the organisation, and is elected annually at the Annual General Meeting.

to find out exactly what they are being used for.

NewCodes

Further extensions to the IPTC NewsCodes have been agreed, and work is underway on a new generation of the IPTC Subject NewsCodes. To help with this IPTC have entered into a three-year agreement to use the SchemaLogic taxonomy management system. With this system the NewsCodes and related data will be held in a central repository allowing delegates around the world to make and access proposals, and interchange comments, during the development process.

Photo Metadata

Another area of new work is that of Photo Metadata with a Working Group having been established to deal with related issues within the IPTC. This group takes in

IPTC public discussion groups:

News Architecture G2 - <http://groups.yahoo.com/group/newsml-g2>

NewsML 1 and NewsML-G2 - <http://groups.yahoo.com/group/newsml>

News Industry Text Format - NITF - <http://groups.yahoo.com/group/nitf>

SportsML and SportsML-G2 - <http://groups.yahoo.com/group/sportsml>

EventsML-G2 - <http://groups.yahoo.com/group/eventsml>

ProgramGuideML - <http://groups.yahoo.com/group/programguideml>

Photo Metadata - <http://groups.yahoo.com/group/iptc-photometadata>

IPTC Core: metadata for XMP - <http://groups.yahoo.com/group/iptc4xmp>

NewsCodes - <http://groups.yahoo.com/group/newscodes>

ORGANISATION

activities associated with the IPTC Core (for XMP) and one of its first actions was to investigate the availability of photo software supporting this, and the older "IPTC Headers", making the information available on the IPTC Web site.

Results of this survey work has also helped to emphasise just how widespread the use of the "IPTC Headers" and the "IPTC Core" really is.

Photo Metadata Conference

As part of the process of improving awareness of IPTC's activities, and establishing user needs, the Photo Metadata Working Group intends to hold a Photo Metadata Conference "Working towards a seamless photo workflow" on the 7 June 2007.

In keeping with IPTC's practice of co-operation with other standards and industry bodies this Conference is being organised with Ifra and held in conjunction with the CEPIC (Coordination of European Picture Agencies Press Stock Heritage) Congress 2007 in Florence.

Development process

Standards development is undertaken by a series of Working Parties and Working Groups which carry out their work using the combination of a development discussion group (with membership restricted to IPTC delegates) and telephone conferences, along with meetings where possible.

As noted above, these efforts tend to be restricted to relatively small groups. In part this is because there is often a need for specialist knowledge, while delegates have to find time to make their contributions - while still carrying out their duties in the organisations that employ them. A lot of commitment is needed to take the work on - particularly for delegates that take lead responsibility in the Committees, Working Parties and Groups.

Considerable thanks are due to the individuals concerned, and to the organisations that let them have the time to take part.

Public awareness

Although participation in standards work is restricted to delegates from member organisations, care is taken to ensure that the general



There was a record attendance at the IPTC XXXXI Annual General Meeting, which was held in Vienna at the invitation of the Austrian Press Agency, who were very generous hosts. Welcoming delegates to the Meeting Wolfgang Vyslozil (Austria Press Agency CEO) - left - and Rudi Horvarth (APA-IT Managing Director) - right - explained that for them hosting the AGM was a way of saying thank you to the IPTC and its members, as APA had derived a lot of benefits from its long relationship with IPTC .



news industry - and other interested parties - are kept aware of plans and progress.

Information on activities and on the standards is posted on the IPTC Web site and a series of public discussion groups are maintained - see box on page 7. These groups also provide a forum for non-members to raise points of interest, and seek advice on specific aspects of standards use.

Intellectual Property Policy

The experience and industry knowledge provided by the members, coupled with the hard work of delegates in development, means that the IPTC standards represent a considerable amount of intellectual capital.

The standards, along with software, reports and other material, are made freely available for use by the news industry, and other parties, and an Intellectual Property Policy (IP Policy) has been introduced to set out the conditions under which the standards can be used. Details of these conditions are given in the panel opposite. The policy also includes two main licence agreements covering the use of specification documents

and software.

Meetings, conference calls and on-line discussions are carried out under the terms of the Policy. All IPTC members have explicitly accepted the terms of the IP Policy, and acceptance is one of the conditions of membership.

Membership

Increasing appreciation of the work IPTC is undertaking, and of the value of its established standards - along with sustained publicity efforts - have resulted in a significant increase in the membership.

As well as providing an additional presence in the Near East and Asia, the new members have provided further representation from the broadcast and photo areas. Delegates from the new members have also provided a welcome addition to the Working Parties and Working Groups.

Meetings

IPTC is an international organisation with world-wide membership and in recognition of this the regular working meetings - normally three a year - are held in different locations around the world to encourage as much participation as

possible. The 2006 meetings were held in Vancouver (Canada), Vienna (Austria) for the Annual General Meeting, and Madrid (Spain).

Locations and dates for the 2007 meetings are Cairo, Egypt (12 - 14 March), Tokyo, Japan (28 - 31 May) and Prague, Czech Republic (15 - 17 October).

Feedback

These formal meetings are where the output of the Working Parties and Working Groups are presented and discussed, providing valuable feedback for the development process. Once development is complete the standards are given a final review, with formal approval for release coming from the Standards Committee.

Informal discussions between delegates are another important feature of the meetings, allowing the interchange of techniques and ideas to the benefit of all concerned.

Meeting presentations

Presentations, both from members and from outside parties, are an important aspect of the Meetings, with the 2006 programme covering a wide range of interests. Presentations included:

- Syndication on the Web - Tim Bray, Director of Web Technologies at Sun Microsystems, and a major contributor to XML and Atom web standards.
- PLUS Picture Licensing - Jeff Sedlik, President and CEO of the Picture Licensing Universal System (PLUS).
- Metadata in Broadcasting - Jean-Pierre Evain from the EBU Technical department.
- Ars Electronica (creative use of the computer) - Wolfgang Bednarek.
- The newspaper has a future (plans for the launch of a new newspaper for the Austrian market) - Wolfgang Zekert.
- Advantages of taking NewsML into the Semantic Web - Raphaël Troncy, co-chair of the W3C Multimedia Semantics Incubator Group.
- MESH - Multimedia Semantic Syndication for Enhanced News Services. An overview of the project was provided by Nikos Saris from the Athens Technology Centre at the AGM with further as-

IPTC Intellectual Property Policy

The IPTC generally makes all of its Intellectual Property available to any interested parties. Such IP is made available under the following conditions:

a IPTC provides explicit licenses to use its Specifications and Materials. The licenses appear as "Non-Exclusive License Agreement for International Press Telecommunications Council Specifications and Related Documentation" and "International Press Telecommunications Council Software License Agreement".

b IPTC Specifications and Materials may be downloaded or copied provided that ALL copies retain the ownership, copyright and license notices.

c Specifications and Materials may not be edited, modified, or presented in a context that creates a misleading or false impression or statement as to the positions, actions, or statements of the IPTC.

d The name and trademarks of the IPTC may not be used in advertising, publicity, or products and their names without the specific, written prior permission of the IPTC. Any permitted use of the trademarks of the IPTC, whether registered or not, must be accompanied by an appropriate mark and attribution, as agreed with the IPTC.

e IPTC Specifications may be extended by both members and non-members to provide additional functionality (Extended Specifications) provided that the Extended Specifications and the related documentation make clear recognition of the existence and ownership of the IPTC IP and provided that the extensions are clearly identified and provided that a perpetual license is granted by the creator of the Extended Specifications for other members and non-members to use the Extended Specifications and to continue extensions of the Extended Specifications. The IPTC does not waive any of its rights in the Standards and Materials in this context. The Extended Specifications may be considered the intellectual property of their creator. The IPTC expressly disclaims any responsibility for damage caused by an extension to IPTC Specifications.

f IPTC Specifications and Materials may be included in derivative work of both members and non-members provided that there is a clear recognition in the derivative work and its related documentation of the IPTC IP and its ownership. The IPTC does not waive any of its rights in the Specifications and Materials in this context. Derivative work in its entirety may be considered the intellectual property of the creator of the work. The IPTC expressly disclaims any responsibility for damage caused when its IP is used in a derivative context.

The full IPTC Intellectual Property Policy statement and licences are available from www.iptc.org/goto/ipp.

ORGANISATION

pects outlined in Madrid by Paulo Villegas from Telefónica. Both of these organisations are participating in the Mesh project.

- Use of SportsML for a team database for ORF (the Austrian TV and Radio service) - Gerald Schinagl, ORF Systems Architect.
- Profium News Agency Solution - Essa Suurio, Profium Sales Manager.
- Agencia EFE - an overview of the history and activities of the Spanish News Agency EFE was provided by Jose Luis del Rey, while Manuel Fuentes described the evolution of the EFE news and photo databases.
- SchemaLogic Taxonomy Management - Breanna Anderson, Chief Technology Officer Sche-

maLogic.

This presentation was arranged to provide members of the NewsCodes Party with information about the system, and it was subsequently decided to adopt the Schemalogic system for NewsCodes management and development.

- The Challenge of New Media - Scott Calder, Mainstream Data.
- News domain research projects that have been undertaken at the Universidad Carlos II de Madrid - Professor Luis Sánchez Fernández.



One of the presentations at the 2006 Spring Meeting (in Vancouver) was from Jeff Sedlik of the photo-licensing PLUS Coalition. Appreciation of the benefits of working together has resulted in the PLUS coalition becoming members of IPTC, while IPTC are part of the PLUS Leadership Circle.

Public Relations

Ensuring that the news industry remains aware of what IPTC is doing, and encouraging organisations to join the organisation is the task of the Public Relations Committee, though much of the actual work is done by the members - as with other IPTC activities.

Press releases are issued to cover significant developments, generally following IPTC Meetings, which is when the main decisions are reached and standards released. These releases are freely distributed by member organisations, who also issue their own releases on IPTC related matters when appropriate.

An archive of previous press releases is maintained at www.iptc.org/pages/prel_main.php to provide a ready reference to the main announcements.

Similarly copies of the IPTC Mirror Newsletter and the annual IPTC Spectrum are available from http://www.iptc.org/pages/nlett_main.php with issues reaching back to the August 2001 issue of

the IPTC Mirror (Issue 100).

Standards naming

Establishing a name for the new set of standards based on the NAR proved time consuming, with input from many members and detailed consideration of several options before a conclusion was reached. This was to include the "G2" identifier as the last part of each standards name - as with NewsML-G2.

A particular advantage of this approach is that it maintains the identity of well established standards like NewsML (which is a registered IPTC trademark) and SportsML while making it clear that they are members of an integrated family of standards.

Steps are now under way to develop a unified corporate identity for the G2-standards, which will probably be extended to cover the IPTC Web site and the publications.

A similar task - but somewhat easier - was to establish the names to be used for identification of the various NewsCode groups.

Industry contact

Direct contact with the news industry has proved an effective way of getting the message across. For example immediately following the Spring Meeting a team consisting of PR Committee Chair Walter Barranger, IPTC Chair Stéphane Gué-

rillot and IPTC Managing Director Michael Steidl toured the exhibition area at NEXPO 2006 (in the USA) to contact relevant companies and issue invitations to a IPTC presentation held by the NAA Wire Committee.

WAN presentation

In his role as IPTC Managing Director Stéphane Guérillot was invited to give a presentation to the Digital Technology Round Table at the World Association of Newspapers (WAN) Congress in Moscow. His theme was the importance of IPTC standards to the success of print and digital technology.

Standards adoption

IPTC Managing Director Michael Steidl made two visits to Italy during 2006, giving presentations in conjunction with the decision of the main Italian news agencies to adopt NewsML and the IPTC Subject NewsCodes. Michael also participated in the CEPIC Congress 2006, where he outlined the features of IPTC photo metadata and took part in a panel discussion on the future of metadata.

Rounding off the presentations for 2006, Michael Steidl and Taka-hiro Fujiwara (Vice-Chair of the NewsML 1 Working Party) were guest speakers at the Korean Press Foundation NewsML Seminar held in Seoul in November.

Sustained Commitment

Overview

Hard work by the development team has formalised the relationship between real life and news content to give a Release Candidate for the News Architecture. A demanding set of objectives has been met with the specification subjected to extensive testing.

A new Photo Metadata Working Group has started work, establishing the availability of software for existing IPTC photo metadata standards and working towards the requirements for a seamless photo workflow.

The NITF has responded to developing user requirements with the latest release (v3.4) available in both DTD and XML Schema forms.

NewsML 1 continues to be an attractive proposition for news organisations wishing to adopt XML based systems, with an XML Schema version under development.

As a statement of intention “The basic goal of the News Architecture is to provide a single generic model for exchanging all kinds of newsworthy information, thus providing a framework for the new IPTC G2 Family of News Exchange Standards” appears relatively straightforward.

Converting this aim to a working standard - the NAR - has proved much less straightforward and has taken a lot of hard work, making heavy demands on the resources available to IPTC and requiring a lot of commitment from the development team. It has also proved particularly time consuming, but it is intended that the first standards based on the NAR will become available during 2007.

Established standards

Although success of NAR development is very important to the future of IPTC standards work as a whole, not all of the available resources can be devoted as the NAR. The established standards - NewsML 1, NITF, SportsML, NewsCodes and the IPTC Core - also need attention so they will continue to meet the needs of their large user bases.

Achievements in this area include new releases of both the NITF and SportsML in DTD and XML Schema versions, and development of a XML Schema for NewsML.

First steps have been taken towards a major revision of the NewsCodes - IPTC's family of controlled vocabularies. This is partly in response to the increased metadata demands of the G2-standards, but also reflects the importance of the NewsCodes as stand-alone standards which have applications beyond that of the news industry. Initial focus of this work is the production of a new-generation system of Subject NewsCodes.

Appreciation of the growing importance of metadata to photo-

graphic applications resulted in establishment of a Photo Metadata Working Group. This group will deal with all photo metadata related areas within the IPTC.

Concept

Underlying concept behind the NAR is the relationship between “real life” occurrences and “news” in the broadest sense. Journalistic input is needed to convert the occurrence into “news” which may be in any media - text, audio, photo or video - or a combination of media. However, this is not enough on its own, for the “news” to be effectively processed, published, stored, and reused requires the addition of detailed information about the content - or metadata.

Other information can complement the news, and help provide a context, and this takes the form of concepts. Concepts may represent real objects, such as people, places and organisations, or more abstract ideas like football and business.

In the NAR individual pieces of news and concepts are handled as specific items which have unique identifiers and share a common set of management features. Consistency within, and between, items is achieved by the use of generic components which have precise meanings and processing models, and by a standard mechanism for handling metadata.

Use of a generic approach also means that it will be easier to provide extensions to meet future requirements, while retaining the underlying structure.

Industry standards

So far as possible the NAR makes use of industry standards that will allow processing with standard tools. The syntax is based on XML (Extensible Markup Language) and the design takes account of Semantic Web requirements, simplifying the transfer of news and concepts to other XML standards.

Working Parties and Groups



Henrik Stadler



Laurent Le Meur



Mischa Wolf



Johan Lindgren



Stuart Myles



Alan Karben



John Minting



Honor Craig-Bennet



Dean Large



Takahiro Fujiwara



Harald Löffler



Michael Steidl

Direction of IPTC technical activities is the task of the Standards Committee, which allocates resources for new developments, and the maintenance of established standards. Work is undertaken by a set of Working Parties and Groups dealing with specific areas and reporting back to the Standards Committee which provides formal approval of standards, and other material, so it can be released for public use. *Chair: Henrik Stadler (TT).*

News Architecture (NAR) Working Party: Development of a generic architectural framework (NAR) suitable for the management and distribution of all types of news-related content. This includes guiding development of XML Schemas, which are being produced by outside consultants. The NAR will be the basis of the new G2-standards family. *Chair: Laurent Le Meur (Agence France Presse). Vice-Chair Misha Wolf (Reuters Ltd).*

During the initial development phase work on different aspects of the NAR was carried out by a series of Working Groups:

Common Components - components that will be used in more than one of the new content standards. *Lead Johan Lindgren (TT).*

News Structure - an abstract model for the NAR framework. *Lead: Laurent Le Meur (AFP).*

News Management - processing models for all types of news content covered by recent IPTC news exchange standards. *Lead: Stuart Myles (Dow Jones).*

News Metadata Framework - specification of the ways in which metadata will be expressed, referenced and managed in all new major versions of IPTC standards. *Lead: Misha Wolf (Reuters).*

These Working Groups have completed their tasks and have now been incorporated directly into the NAR Working Party.

News Content Working Party: Oversees the maintenance and development of standards for all types of news related content mark up, with work being the responsibility of a series of Working Groups. *Chair (interim): Henrik Stadler (TT).*

NewsML-G2 - a standard based on the NAR that can be used for the mark up of any kind of general news content. *Lead: Laurent Le Meur (AFP).*

EventsML-G2 - a NAR based format for the exchange of information on newsworthy events. *Lead: Johan Lindgren (TT).*

SportsML - SportsML is a XML based Sports Markup Language. *Lead: Alan Karben (XML Team Solutions). Vice-Lead: Johan Lindgren (TT).*

ProgramGuideML - a XML based standard for the exchanging TV and Radio program listings.

NewsCodes Working Party: Responsible for the maintenance of established IPTC metadata sets and the development of new sets as appropriate. *Chair: John Minting (UPI). Vice-Chair: Honor Craig-Bennett (PA News).*

NITF Maintenance Working Party: Maintenance and further development of the News Industry Text Format (NITF). *Chair: Alan Karben (XML Team Solutions). Vice-Chair: Stuart Myles (Dow Jones).*

NewsML 1 Maintenance Working Party: Promotion of NewsML as the standard packaging and syndication mechanism for multimedia news with maintenance of the functional specification and production of implementation guidelines. *Chair: Dean Large (Businesswire). Vice-Chair: Takahiro Fujiwara (EAST Co. Ltd).*

Photo Metadata Working Group: Acts as a special interest group regarding all photo metadata issues of the IPTC, providing support to all current IPTC standards in all photo related areas, including the development of specific photo metadata standards. *Lead: Harald Löffler (Ifra). Vice-Lead: Michael Steidl (IPTC).*

The NAR Model is independent of the way it is implemented. However, a XML Schema implementation will be provided, as it is believed that this will be the type of software most commonly adopted. Alternatively the model could be implemented using object-oriented software, in Java, or in C#.

It is important to remember that the NAR is not itself a news exchange standard. Different types of news content have very specific requirements and these are catered for by a set of individual standards, the IPTC G2-standards.

Initial efforts have been concentrated on a general news standard NewsML-G2 (as a successor to NewsML 1), EventsML-G2 and SportsML-G2.

News exchange standards

All of these news standards will have the NAR as their underlying framework, using the common items and components. Since the standards will have a consistent structure, and a consistent way of managing individual items and of dealing with the associated metadata, systems will be easier to understand and implement. The approach also provides a high level of compatibility for the information dealt with by each of the standards.

Another aspect of the “real life” approach is that information about a specific concept will probably be available from a number of different information providers. This information will have a consistent structure so it will be possible to use the separate items together, giving a whole that will be more than the sum of its parts. It is anticipated that extensive use of this feature may lead to a market of knowledge information.

Aims

In more detail the aims for NAR development are:

- To simplify and unify the overall

Underlying concept of the News Architecture is the relationship between things happening in the real world and the resulting news content. Some specific metadata is associated with the content, with common metadata components being used for description and to provide management features.

design for representing newsworthy information.

- To be flexible, thus allowing lightweight “no bells and whistles” feeds and highly complex news feeds, based on the same model.
- To specify more details, leaving less space for interpretation.
- To streamline the processing model, providing only a single way to express specific structures and functionalities.
- To develop a new model for expressing metadata from the ground up.
- To provide an abstract model to be implemented by specific news exchange standards.
- To maintain, at the functional rather than syntactic level, a high level of backward compatibility with NewsML 1.
- To simplify the implementation of IPTC news exchange standards as a whole.
- To align IPTC news exchange standards with requirements from the “Information Highway”.

Profiles

The all-inclusive nature of these objectives has inevitably resulted in a degree of tension within the development process as, for example, attempts are made to reconcile the requirement for simplicity and interoperability with the aim of handling highly complex news feeds. Provision of alternative “core” and “power” profiles has provided a solution to this problem. A similar approach is also being considered for the NITF.

In the NAR the “core” profile is kept as simple as possible, for interoperability and ease of implementation, while the “power” profile provides a much higher degree of flexibility, at the cost of reduced interoperability and a more complex implementation. The power level also makes extensive provision for user extensions.

Processing

Processing systems have to specify which level of functionality they

The Item (individual instances derived from "AnyItem")					
Item components		Packaging	General News report	Person definition	Event news and definition
Management Component	item class	Package	News	Concept	News - or - Concept
	content class	Composite Text Photo Still Graphic Animated graphic Audio Video	Text Photo Still Graphic Animated graphic Audio Video	Person	Event
	signature	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Common Metadata Components	administrative	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	descriptive	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	rights	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
	publication links	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Optional Content Component	generic concept definitions			<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
		Group Structure = XML hierarchical structure of references to other Items	News Content Components. = XML markup; plain text; binary photo, graphic, audio, video	Person Content Component = XML marked up person definition details	Event Content Component = XML marked up event definition details
equivalent to ... in the „Real World“		none			

support, with “power” level systems providing all of the “core” functionality as well as the “power” extensions. The “core” level systems should still be able to deal with “power” level items by ignoring the additional information.

Advanced technologies

Similarly the aim of aligning standards with the “Information Highway” is complicated by the fact the

underlying technologies are themselves under active development. To some extent the generic design approach adopted for the NAR will make it easier to cater for new developments, but some advanced features of the NAR have been based on technology that is itself at an experimental stage.

Use of advanced XML techniques can raise another problem in that the software tools used for

application development are not always fully compliant with the latest proposals, and may not be totally consistent in their interpretation of the standards (tool developers also have problems in keeping up with XML developments).

Testing

The need to ensure that the NAR objectives are being met, in a man-

Photo Metadata

Introduction of the IPTC Core (for XMP) in 2005 provided a new way of handling the “IPTC Headers” which can be used to hold a wide range of information about an image.

The IPTC Core has proved to be a popular initiative, and a series of suggestions for improvements and extensions had been raised. At the same time a related effort had been under way with the Colour Space Task Force - in association with IFRA - investigating ways of ensuring that the EXIF-JPEG camera data would be retained through the processing.

As work in these areas proceeded it became clear that these initiatives were only dealing with some aspects of what was a much wider interest area - that of photo metadata in general.

Discussions with other interested parties - including new IPTC members with a special interest in the photo area - reinforced the belief that this was an area that needed serious consideration.

Working Group

Accordingly a new Photo Metadata Working Group was established at the 2006 AGM. Objectives of this group are to act as a special interest group for all photo metadata related issues within the IPTC.

This work is not focussed on a single standard but will support all current IPTC standards in photo related areas.

Specifically, this group will:

- Support the development of generic IPTC photo metadata standards.
- Act as a standing group of experts to respond to issues raised by external parties.

Software support

As one of the first actions of the new Group, a survey was undertaken to establish how software providers support the existing photo metadata standards of the Information Interchange Model (IIM) - the well known “IPTC Headers” are a subset of the IIM data fields - and the IPTC Core.

Over fifty software packages were identified and the full list, including details of which standards are supported and whether there is synchronisation between the IIM and IPTC Core values, has been made publicly available at <http://www.iptc.org/photometadata/software-support-list1.php>.

Subsequent investigations have shown that a number of organisations appear to be providing suggestions for use of the “IPTC Headers” though their recommendations of how fields may be used do not always seem to be in agreement with one another, or with the IIM. This confusion is a further indication of the need for proper consideration of the requirements for photo metadata.

Metadata requirements

To help establish the requirements a White Paper is being produced.

This will explain the importance of metadata for the news and photo industries and provide a summary of the metadata needed to cover Descriptive, Administrative, Rights and Technical Properties.

Technical considerations related to metadata implementation will be looked at. Specific consideration of typical photo workflows will cover the following: Photographer, News Photo Agency, Stock Photo Agency, Newspaper, and Magazine.

Seamless workflow

This work will be complemented by a one day conference, with the title “Working towards a seamless photo workflow”.

Goal of the conference is to simplify the convergence and global applicability of photo metadata, and maximise the widespread and consistent implementation of standards.

Participants will include both photo creators and users, along with photo standards organisations, software providers, and camera manufacturers. The intention is to analyse current metadata practices and establish the best way of achieving a seamless workflow from camera to the end user.

The Conference is being organised in conjunction with Ifra and held in conjunction with the CEPIC (Coordination of European Picture Agencies Press Stock Heritage) Congress 2007 in Florence. See <http://www.phmdc.org/> for further details.

ner that will be as easy to use and as trouble free as possible, resulted in the decision to undertake a series of tests at different development stages.

In each case this involved preparation of extensive test packages that included an introductory document, model specifications, implementation of the specifications as XML Schema, and supporting material. Producing all this was an added workload for the development team.

The initial Experimental Phase (EP#1) was carried out to test real

life use cases against the NAR model and syntax, with particular attention to the handling of single and multimedia test feeds, object orientated applications, the persistence of news objects (as in a relational database), navigation between news objects, and the use of style sheets for conversion between NewsML1 and NewsML-G2.

Results showed a number of areas where improvements could be made, while further changes and additions were introduced as a result of continuing development work during the test period.

Trial applications

The updated specification was then subjected to a second Experimental Phase (EP#2) where the main aim was to investigate practical aspects of using the architecture to build content standards.

Working groups for the new IPTC content standards (NewsML-G2, EventsML-G2 and SportsML-G2) were specifically asked to participate, along with other parties who could apply the NAR to their own use cases.

This phase proved particularly valuable in identifying a number of

News Industry Text Format

The NITF was the first XML-based standard developed by the IPTC (being based on an earlier SGML - Standard Generalised Markup Language - version) and is in widespread use within the news industry.

Indeed, it is believed that worldwide, the NITF is the XML vocabulary most commonly used by news publishers.

The NITF is a stand-alone format for news interchange with a <head> section containing information about the document itself, and a <body> section that carries the content.

Information in the <head> can include the title, codes to describe the article, and the subjects covered by the article; along with document metadata such as the publication date, rights information, urgency, and news management features.

Content of the <body> consists of the article itself, possibly with tables, lists and images, along with components like the headline and byline. An important feature is that a series of "enriched text" elements make it possible to mark up the text

to indicate, for example, people, places, organisations, emphasised text and hyperlinks.

Since the standard has been in use for some years (it was launched in 1999) with regular updates it has been developed to the stage at which it meets most user requirements. Current requests for changes and enhancements are relatively minor. For example with release v3.4 a change was made to allow the use of multiple <abstract> elements so that users could include alternative summaries of the document.

XML Schema

However, a significant step was taken with the development and formal approval of a XML Schema for NITF v3.3. This work was undertaken in response to requests from users, who wished to take advantage of the greater flexibility and more precise data control offered by the use of Schema-based applications.

Similar thinking has resulted in the development of a XML Schema for NewsML v1.2, while the new generation exchange standards (such as NewsML-G2 and EventsML-G2) have been based on XML Schema from the outset.

Extra features

Subsequent approval of the NITF v3.4 XML Schema provided further evidence of the reasons for

Schema adoption as this version offers a specific feature that cannot be provided in the DTD version of the standard. This is to allow the inclusion of namespaced elements (material from another XML Schema with its own namespace) within the enriched text area of a NITF instance (this allows the import of external material, with its characteristics set by its own XML Schema).

NITF profiles

Much of the metadata carried in the <head> is similar to that which would be contained in the administrative and descriptive metadata of NewsML-G2, and this could create a possibility for confusion and conflict if the NITF is used as a text format for NewsML-G2.

At the same time, for some simple applications, such extensive metadata about the document itself may not be necessary, and could prove off putting for some potential users.

Because of this consideration is being given to reworking the NITF to give "core" and "power" versions.

The "power" profile would offer much the same functionality as the current version of the NITF, while the "core" version would have much of the document metadata removed, while retaining the enriched text and other content formatting features.

areas in which improvements could be made. Discussion of the results at the Autumn 2006 Meeting raised further concerns, which were also taken care of to give a NAR Release candidate. This was made public in January 2007 for comment and it is planned that formal approval of the NAR structure

specification will be possible at the IPTC 2007 AGM.

Implementation

Implementation of the NAR in XML Schema is being carried out by consultants, who are also undertaking a quality assurance assessment to ensure that the

specification documents are consistent and provide a high level of clarity. The XML Schema implementation takes the form of a "Master" file which includes both "core" and "power" features, with a (internal use) XML Schema generator being provided to create individual Schema for each profile.

NewsML 1

Although work on the next-generation NewsML-G2 is under way, interest in NewsML 1 continues to run at a high level. There is a well established, world-wide, user base with considerable application expertise, coupled with a substantial group of system suppliers offering NewsML compliant systems.

Providing a structured framework for multimedia news, NewsML 1 is a XML based standard that can be applied throughout the news lifecycle. Typical applications include; in and between editorial systems; between news agencies and their customers; between publishers and news aggregators; and between news service providers and end users as well as for the creation of news content.

Metadata provision

The standard has extensive metadata to cover Administrative, Rights, and Descriptive and News Management requirements, with provision for human-readable versions of appropriate metadata items such as headlines, rights, dates and keywords. However, there is no need to use all of the available features, so relatively

simple implementations can be developed.

The planned successor standard, NewsML-G2, draws on the concepts underlying NewsML 1 and experience gained in its application. A key aim with NewsML-G2 is to maintain a high level of backwards compatibility with NewsML 1, so there should be straightforward update paths for current and prospective users when this is considered necessary.

Since efforts have been concentrated on the new standard, development of NewsML 1 has been frozen at v1.2. However, work on a XML Schema has been undertaken in response to user requirements.

Draft XML Schema

A Beta draft XML Schema for NewsML v1.2 was released in Summer 2006 and has been subjected to extensive testing to ensure its compatibility with other XML products (such as databases), the ease of data transfer between DTD and XML Schema based systems, and the general validation of NewsML 1 examples.

Some issue were identified during the tests, and formal release of the NewsML v1.2 XML Schema will take place when these are

dealt with. It was considered particularly important to ensure trouble free operation as a number of users intend to implement XML Schema systems that will work alongside established DTD based systems.

Continued appeal

Since NewsML 1 is an established, well proven, standard it has considerable appeal for news organisations wishing to introduce XML based news systems with multimedia capability.

In June 2006 the Federation of Newspaper Publishers in Italy (FIEG) announced that the major national Italian news agencies (Ansa, AGI, ApCom and ADN Kronos) had agreed to start delivering their news using NewsML 1. At the same time these agencies will be adopting the IPTC Subject News-Codes for news categorisation.

Applications

Italian system vendors will also be involved with the implementation of systems and interfaces to allow integration of NewsML content into existing systems and content databases. It is anticipated that there will be a gradual transition to the new standard over some years to simplify the adoption process for the news agencies customers.

In another development the Korean Press Foundation (KPF) has been investigating the advantages of adopting NewsML within the Korean news industry. As part of this effort a "NewsML Seminar" was held in Seoul during November 2006 with guest presentations from IPTC Managing Director Michael Steidl and Takahiro Fujiwara, Vice-Chair of the NewsML 1 Working Party of the IPTC.



Panel at the Korean Seminar held to consider adoption of NewsML, with Takahiro Fujiwara (East Co. Ltd, Vice-Chair IPTC NewsML 1 Working Party) replying to a question from the audience. Other members of the panel were Tae-Sung Jung (Yonhap News Agency, Chair of the Korea Press Foundation NewsML Forum), and Michael Steidl (IPTC Managing Director).

Common Components - Common Structure - Common Processing

Overview

Meeting the requirements for the new News Architecture has been challenging and time consuming, but the end result will provide a secure basis for the new generation of IPTC standards.

There is a set of reusable components, with content being handled by a consistent family of managed Items. A new mechanism for handling metadata has been developed. Provision is made for both news and related concepts.

A core version of the News Architecture provides for straightforward applications, with more demanding requirements catered for by power extensions, while provision has been made for providers to add features required for their specific business needs.

Aim of the News Architecture development is to produce a framework for the new second generation IPTC exchange standards - the IPTC G2-standards.

This has been done by producing a set of standard components and instructions that can be used to handle both “news” and related “concepts” in a consistent way that is independent of the content being handled.

Building blocks

Each of the G2-standards will use these News Architecture (NAR) building blocks in the same way so speeding their development, and application. The NAR model is flexible and generic and can be readily extended to deal with as-yet unplanned standards. Standards currently under development cover General News (NewsML-G2), Events (EventsML-G2) and Sports (SportsML-G2).

For users, the applications will be easier to understand and be faster and less expensive to implement. The NAR makes use of XML with the initial implementation using XML Schema. Use of such in-

dustry standards will allow processing with standard software tools, while the design will make it possible to take advantage of developments in the underlying XML technology.

Throughout development a major aim has been to ensure that the NAR makes due provision for the information that is needed in typical news applications, while keeping the design as concise and simple as possible.

Conformance levels

However, it is appreciated that some users will need a high degree of flexibility in the way they handle information so two conformance levels have been defined. The core conformance level (CCL) offers simplicity and a high degree of interoperability, while the power conformance level (PCL) offers greater flexibility at the expense of greater complexity and a reduced level of interoperability.

Content handling

A news exchange standard has to handle the actual news content, along with associated concepts,



News content is carried by the News Item, one of a family of Items that have the same basic structure, and the same administrative and management metadata features.

Diagram by Athens Technology Centre.

and metadata for descriptive and administrative purposes. There has to be provision to manage and package the information, and then to deliver it.

To do this the New Architecture has a precisely defined set of components:

- Building blocks - that provide ways of representing and processing specific pieces of information.
- News Structure - to provide a standard way of managing individual pieces of news content.
- News Metadata - a mechanism for expressing and managing descriptive information related to the content.

In addition the News Architecture provides the News Message as a mechanism for the exchange of structured information defined using the News Architecture.

Each of these constituent parts is considered in more detail below, with particular reference to the core compliance level. Information on the added features offered by the power conformance level is included in a separate section "Power Extensions" on page 22. Throughout, defined NAR terms are indicated by the use of italics.

Common Components

Basic building blocks of the NAR are the common components. These represent pieces of information and are context free, but when used in business messages they take on the specific semantics of the business context.

Depending on requirements a common component may be used on its own, or combined with other common components to create larger context-free structures. The reuse of such components helps to ensure design consistency and gives a consistent structure for the content across the IPTC G2 Family of Standards.

Datatypes

Finest level building blocks are the primitive datatypes - such as *integer* and *string* - which are found in XML Schema or software languages, and so are not specifically defined for the NAR. There are also simple datatypes produced by applying restrictions to the primitive types, so in the NAR the *Int1to9Type* is defined as an *inte-*

ger from 1 to 9. A typical application for the *Int1to9Type* datatype is to denote the editorial significance of content (in the *Administrative Metadata*).

Natural language

Where information is provided in a natural language - as with a human-readable label - the language details have to be included to ensure proper display. This is achieved by use of the *i18n* Internationalisation attributes. For example *IntlStringType* is an internationalised string, while *LabelType* is a string with the *i18n* attributes, some with a role qualifier to identify the function of the label.

Date and Time

Dates and times are taken care of with a group of properties. In addition to the normal calendar date with an optional time part these include a *TruncatedDateTimeType* which consists of a calendar date (and optional time part) that can be progressively shortened by omitting one or more parts from the end. This means it is possible to give just the month and year, for example.

Other datatypes deal with approximate, recurring, and ranges of dates and times.

A consistent naming convention has been used, and all datatypes are identified by having *Type* as

Controlled Vocabularies and QCodes

The News Architecture makes extensive use of values taken from schemes, such as controlled vocabularies. These controlled values are identified by the combination of the scheme and the code from the scheme to give a {scheme:code} pair which can be identified by processing software.

Schemes will be identified by URIs (Uniform Resource Indicators) and combination of a scheme URI and a code will give a concept URI. However, URIs can be fairly long, and the use of a series of such references in a news item raises practical problems (for example with transmission capacity).

Compact syntax

Accordingly a compact syntax has been developed to provide an efficient way of using controlled values in the NAR. With this an alias (in the form of a short string) is defined for each of the schemes. A controlled value is then identified by the combination of the scheme alias, a colon ":", and the appropriate code identifier (from the scheme). The resulting pair (scheme alias:code) is known as a QCode (Qualified Code). This the recommended way of using controlled vocabularies with the News Architecture.

Some required properties can only take a QCode as their value and appropriate controlled vocabularies will be available for use with NAR-based standards. However, information providers will also be able to use their own vocabularies for this purpose.

Term recovery

To recover the actual controlled vocabulary term, the scheme alias part of the Qcode is replaced by the appropriate URI. For this, every Item in the NAR has a catalog which contains a mapping between each scheme alias used in the Item, and the corresponding URI.

In some applications a large number of schemes may be used in an Item and, if required, the catalog may be stored as a remote resource and referenced with a hyperlink. Normally it is anticipated that information providers will use a consistent set of schemes which will be referenced for all Items they provide. It will then be possible for them to supply the set to their customers so it will be available as a local resource for processing systems.

the last portion of their name.

Properties

Pieces of business information are represented by a basic component or property which takes a datatype as the model for its content - a datatype does not have a specific business meaning on its own.

A property can be used on its own or combined with other properties to form a group and is identified by having *PropType* as the last part of its name - as with *TruncatedDateTimePropType*.

Qualified codes

For consistent representation, many metadata properties are taken from controlled vocabularies, and these are handled by the use of *QCodes* (qualified codes). Further details on the development and application of *QCodes* are included in the panel on page 18.

In some cases there may not be an appropriate term available from a controlled vocabulary and to allow for this the *FlexPropType* (flexible property type) may take the form of a value from a controlled vocabulary (*QCode*) or consist of a text string.

Concepts

The handling of concepts is an important feature of the News Architecture, and an aggregate *Concept Component* has been developed for this purpose.

There are two main types of concept, named entities - real objects such as people and places - and generic (or abstract) concepts. Generic concepts cover a broad range from themes - such as music and football - to specific emotions.

Properties common to all concepts are dealt with by the *Concept Type* which makes provision for an unambiguous identifier, an indication of the type of concept (in the form of a *QCode*), a *Concept Information Group* and an *Entity Details Group*.

Relationships

Information handled by the *Concept Information Group* consists of a natural language name and definition together with a set of properties that may be used to provide alternative identifications for the concept - *sameAs* - and to establish relationships between con-



Second of the family of NAR Items the *Concept Item* is designed to allow the handling of concepts in the same way as news information.

Diagram by Athens Technology Centre.

Person Concept

An indication of how complex aggregate components are built up from the basic datatypes and properties is given by the person details concept (*PersonDetailsType*) which contains:

Date of Birth (*TruncatedDateTimePropType*)
 Date of Death (*TruncatedDateTimePropType*)
 Gender (*FlexPropType*)
 Contact Information (*ContactInfoType*)
 Affiliation (*FlexPropType*)
 Occupation (*FlexPropType*)
 Skill (*FlexPropType*)
 Extension Point

Here Contact Information is itself a composite datatype (*ContactInfoType*) with:

Email Address (*ElectronicAddressType*)
 Instant Messaging Address (*ElectronicAddressType*)
 Phone Number (*ElectronicAddressType*)
 Fax Number (*ElectronicAddressType*)
 Web Address (*WebAddressType*)
 Postal Address (*PostalAddressType*)
 Extension Point
 (Both *ElectronicAddressType* and *WebAddressType* are strings extended by the additional of Roles - which take the form of *QCodes*.)

Again, Postal Address is a composite datatype (*PostalAddressType*) made up from:

Role (*QCodeType*)
 Address Line (*IntlStringType*)
 Locality (*FlexPropType*)
 Country Area (*FlexPropType*)
 Country (*FlexPropType*)
 Postal Code (*IntlStringType*)

Basic datatypes and properties are shown in blue, with composite components shown in red.

The other entity components are constructed in the same way, again making use of the common datatypes and components. So Contact Info for Organisation and Point of Interest uses the common composite *ContactInfoType*.

Note that the composite components include extension points. These are provided to let information providers add other properties that they need to meet their specific business requirements.

cepts - *broader, narrower and related*. These relationships may be used to create taxonomies (hierarchies of concepts) and thesauri (sets of concepts associated via the relationships).

Any given concept may be identified by a controlled value, but the nature of concepts means that the same concept could be present in more than one controlled vocabulary (possibly from different providers) and so have several identifiers. An example of this is the way that a company can be identified by a number of different "ticker" symbols. This is why the concept identifier has to be unambiguous (but not unique).

Entities

More detailed provision has been made for entities with the following aggregate components:

- *Person Details* - dates of birth and death, gender, contact information, affiliation, occupation, and skill.
- *Organisation Details* - dates of foundation and dissolution, contact information, business sector, and business location.
- *Geopolitical Area* - geographic co-ordinates, and geopolitical type.
- *Point of Interest Details* - geographic co-ordinates, opening hours, capacity, contact information, point of interest type, facility, access and location details.

Provider extensions

Extension Points are provided in aggregate components so information providers can include their own defined properties to meet specific business requirements. Similarly provision is made for the addition of provider-defined qualifiers to any property of an instance document.

Items

An item is the smallest piece of information that can be managed within the NAR, and to maintain consistency all Items are derived from an abstract *Any Item*. This defines a structure and sets of metadata for administrative and management purposes, all of which are inherited by the other Items. Further information on the

Any Item is given in the panel.

A set of four Items (*News Item, Concept Item, Package Item and Knowledge Item*) have been defined to cater for all of the currently planned G2-standards, but future standards may need additional dedicated Items. These will also be based on the *Any Item* and inherit the same basic parts, which will be complemented by application specific extensions.

NewsItem

News content - in any media type or format - is handled by the *NewsItem* with typical examples

being a news report (text), a picture, or a video clip. Typically there will only be short term interest in the content which will be updated over a short period but may then be archived. The content may refer to a set of concepts and entities, and be associated with other *News Items* or Web resources.

Administrative and management properties inherited from the *Any Item* are complemented by *Content Metadata* and a *News Content Set* - which may consist of a set of alternative renditions - such as an image in thumbnail, preview and high resolution versions.

Any Item

The Item is the basic piece of managed information in the News Architecture and the abstract Any Item is a template for all NAR Items. It provides a common structure and metadata:

Schema Version *An indication of the XML Schema version specifying the item.*

Conformance Level *Conformance level of the item (either "core" or "power").*

Item Identifier *A globally unique identifier - a guid. This is needed to identify the item as it moves through the workflow and is transferred between systems.*

Item Version *To allow for updates of an Item.*

Catalog *Identification of schemes used for metadata values (See "Controlled Vocabularies and Qcodes on page 18).*

Rights Information *At the "core" level this is a container for a set of properties related to rights and licensing.*

Item Metadata *Metadata that relates to the Item as a whole, and not only to the content.*

Item Management Group *Management properties including provider, creation date and time and class (this helps indicate the structure of the item) which are mandatory. Optional information includes an embargo date, the publishing status (this is "usable" by default), state of evolution, a recommended file name, editorial service, title (in a natural language), editorial note (also natural language) and an editorial signal (for the processing system)*

Item Link *See panel on page 21 for details.*

Extension Point *For including additional provider-defined properties.*

Content Metadata *Metadata directly related to the content carried by the Item.*

Specialised Items developed from the Any Item will include their own sets of descriptive metadata here.

Administrative Metadata *Included as part of the Content Metadata with a set of optional properties to describe features that are not directly presented in the content. These include the date of creation and modification, place of creation, editorial significance, source of information, creator and contributor (a person or an organisation) and the intended audience.*

Content may be present as an inline XML component or an inline data component (plaintext or base64 encoded). Alternatively the content may be remote and identified by a hyperlink. A group of content attributes may be used to provide an indication of the rendition, type and format of the content.

The *News Content Metadata* has its own administrative metadata (to deal with the content, and so separate from the Item administrative properties) along with a set of metadata - the *NewsDescriptiveMetadataGroup* - to describe the content. This group covers the language, genre and subject of the content, and makes provision for *Slugline*, *Headline*, and *Description* (individually defined).

The News Item is the main element of NewsML-G2, which is an exchange standard for general news in all media types. In this standard the news content metadata is extended to include a set of typical media characteristics.

Concept Item

Generally similar to the *News Item*, the *Concept Item* is designed to convey information about concepts. It has the common Item components - with management properties and administrative metadata - along with a set of *Content Metadata* and a *Concept Component* (as previously described).

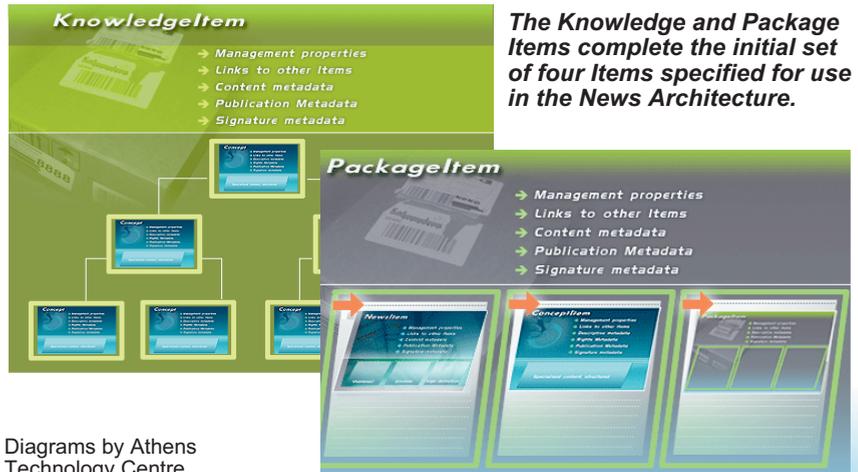
Main characteristics of a *Concept Item* is that it is focused on a single concept (which may be an entity), and that there is long term interest for the content, which will be updated infrequently but over a long period of time as the concept develops.

When a NAR compliant server is queried to obtain information on a concept it will return the appropriate *Concept Item* (or *Items*).

Knowledge Item

As the name suggests the *Knowledge Item* is intended to providing a way of presenting a specific set of information, such as the IPTC Subject NewsCodes or a provider's list of audience codes. Providers may use *Knowledge Items* to make sets of codes available to their customers.

The *Knowledge Item* has the normal Item components with the ad-



Diagrams by Athens Technology Centre.

dition of *Content Metadata* and a *Concept Set* - a set of concept definitions grouped together into a consistent structure, which is then managed and applied as a whole. Concepts in the set may be of different types and their identifiers come from separate schemes.

Normally the content of a *Knowledge Item* will be of long-term interest and updated infrequently but over an extended period (with evolution of the controlled vocabulary it holds).

Package Item

News content is often delivered as a related group of items, such as a collection of pictures, a "top ten" list

The *Knowledge and Package Items* complete the initial set of four Items specified for use in the News Architecture.

of news items, or a set of items relating to the same event. The *Package Item* provides a way of presenting a set of items in a structured manner, expressed as a hierarchy.

The *Package Item* does not carry any content directly, instead it provides a set of references to individual Items (or to Web resources), which are produced and managed independently.

In addition to the management properties and administrative metadata inherited from the *Any Item* the *Package Item* has its own *Content Metadata*, which includes a *Package Descriptive Metadata Group* (with the same features as

Linking Features

The use of links makes it possible to create a network of news resources and the link component provides a generic mechanism for linking NAR Items to one another and to Web resources.

Links are expressed as a hypertext reference (href), and to simplify processing the content type and size of the target can be included in the linking information. Typical functions are:

Navigation links between one Item and another Item or a Web resource. An article about a person might be linked to their biography, or sections of transcripts can be linked to one another in order.

Derivation links for the expression of parent/child relationships. Common application would be the link between the translation of an article and the original article, or between a processed image and the source picture.

Dependency links for use when external Items are needed to provide a full representation of the content of an Item. For example with an illustrated article the textual content of the News Item will contain a reference to the image (or images), which is represented by another News Item. A dependency link establishes the need to retrieve the picture News Item to produce a complete article.

Composition links are used to aggregate the Items in a Package Item.

the *News Descriptive Metadata Group*).

The content is a *Group Set* which represents a tree of sub-groups and references to items. Individual groups may have different roles and the elements in a group may be complementary or alternative, while their order may also be relevant.

Processing model

The News Architecture conceptual model is complemented by a Processing Model which provides guidance on the implementation of NAR compliant systems.

Specific aspects covered include:

- Accessing and checking catalogs, including remote catalogs.
- Obtaining human readable information about schemes and code, and retrieving all terms of a scheme.
- Processing the item status. Publishing status values are Usable, Withheld and Cancelled, along with Embargoed and Expired.
- Retrieval of linked resources and processing of the link property.
- Managing dates (and times) of items.

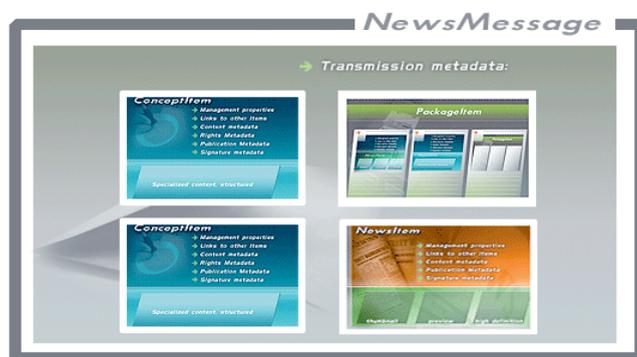
News Message

Designed as a mechanism for item exchange, the News Message is an optional part of the NAR.

Other exchange protocols, including SOAP (Simple Object Access Protocol), WebDAV (Web-based Distributed Authoring and Versioning), ICE (Information and Content Exchange) and the Atom Publication Protocol, which provide a wrapping message, are just as appropriate, with the choice depending on the information providers' normal practice.

The *News Message* has a simple structure with a message header and a set of items. Header information consists of: Date of transmission, sender, Transmission Identifier, Priority, Origin Destination, Channel and Extension Point, with the only mandatory element being the Date of Transmission.

Any defined NAR Item (initially the *NewsItem*, *Concept Item*, and *Package Item*) or combination of items can be carried, with the XML representation of each item being directly included in the *News Message*.



Athens Technology Centre.

Power Extensions

Simplicity and interoperability are key features of the News Architecture “core conformance level” (CCL), but some providers have particularly demanding applications, and their needs are catered for with the “power conformance level” (PCL).

Features available at the power conformance level are extensions to the core conformance level. This means that it will be possible for processing systems that are CCL compliant to deal with PCL instances by ignoring the additional information, while PCL systems will be able to handle CCL feeds.

Common components

Power extensions applied to the common components include;

- Addition of editing qualifiers to allow identification of the creator and details of modifications to metadata.
- Extended use of i18n attributes to allow fine grained control of language information.
- Extensive additions to flexible properties
- Development of composite concepts.

- Support for rich text and ruby mark-up in labels and blocks.
- Concept relationship properties have a qualifier to establish validity over time and additional qualifier for the relationship properties.
- Multiple sets of contact information are supported.

At Item level, there is provision for a digital signature to be applied to complete Items, or to parts of Items, while rights information can be applied to individual parts of the content.

RDF Compatibility

An underlying aim with NAR development was to achieve compatibility with the Semantic Web and to help achieve this the metadata model can be transformed to the Resource Description Framework (RDF).

One way this can be done is by using the GRDDL (Gleaning Resource Descriptions from Dialects of Languages) mechanism, and a GRDDL reference can be included at the root of a NAR item. However, it is important to note that use of the NAR does not require any knowledge of RDF.

Information Exchange

Overview

Development of the G2-standards for content exchange has had to wait for the IPTC News Architecture to become available. However, work on NewsML-G2, EventsML-G2 and SportsML-G2 is now under way. Popularity of SportsML continues to grow with regular updates and extensions.

Since the planned new generation content standards - NewsML-G2 and EventsML-G2 will be built using the News Architecture (NAR), detailed work on them had to wait until the NAR model was substantially complete. However, initial work on these standards was a major factor in the launch of the News Architecture project.

When the EventsML and NewsML business requirements were examined it was clear that there were many common features, and that it would be a poor use of resources to have two groups seeking solutions to the same problems. At the same time work was under way to investigate alternative ways of handling metadata (partly as the system used in NewsML 1 was seen as complex and difficult to implement).

New model

Initial moves were made to ensure closer co-operation in these areas, but further analysis of the relationships between news content and real life events resulted in the deci-

sion to concentrate efforts on production of a model that could be used for the exchange of all types of news content - the News Architecture. Separate standards for specific types of content would then be built on the framework provided by the NAR

With the NAR well advanced, the Working Groups responsible for the new content standards have been able to start detailed work.

General news

NewsML-G2 is the new standard for general news, handling text, photos, graphics, video or other media. It can be considered as offering similar functions to the news mark-up parts of NewsML 1, with improved metadata handling features and the consistent approach of the G2-standards family.

Main element of NewsML-G2 is the NAR News Item, which includes a set of "typical" media characteristics. This is only a small set and individual providers can add further characteristics that they consider necessary. The NAR Concept, Package and Knowledge Items are also included.

Events and Sports

EventsML-G2 is a way of allowing the exchange of information about events - or "things that happen", and for an event to be considered news it has to be covered in some way. Aspects to be handled include: publishing information about events; managing the coverage of an event; and providing information about how an event is being - or will be - covered.

Although details have still to be finalised, EventsML-G2 will define a set of event-specific properties which may be used in a NewsItem as content or in a ConceptItem to indicate persisting knowledge about an event.

SportsML-G2 will take advantage of NAR features to give enhanced publishing and rights handling features, and improved

Media Characteristics for NewsML-G2

Text: Word count.

Photo: Image Width, Image Height, Image Orientation, Image ColorSpace, Resolution.

Graphic: Resolution, Image Height, Width and Orientation (as for photos), Resolution Duration (as for audio and video).

Audio: AudioCodec; Duration, Audio Bit Rate, Audio Variable Bit Rate Flag, Audio Sample Size, Audio Sample Rate, Audio Channels.

Video: Image Height, and Width (as for photos), VideoCodec, Duration, Video Average Bit Rate, Video Variable Bit Rate Flag Width, Video Frame Rate, Average Bit Rate, Video Scan Technique, Video Aspect Ratio, Video Sampling Method.

The above set of media characteristics have been proposed for use with NewsML-G2. The set has been kept small, but users will be able to add additional characteristics that they consider appropriate for their content.

metadata rights management, as well as compatibility with other standards in the G2 family.

High interest

The established version of SportsML continues to attract a high level of interest - the public discussion group has over 500 members - and is regularly updated to provide enhancements and extensions with version 1.8 now available in both DTD and XML Schema forms.

The standard has a core module to handle information that is common to all sports, with a series of plug-in modules dealing with information that only applies to a specific sport.

Typical core information includes scores, standings, schedules and statistics, both for teams and for in-

dividuals. The scope of the information is such that many sports can be dealt with by the core alone. Coverage of the 2006 Torino Winter Olympics was provided this way - see panel. There is also provision for wagering statistics.

Plug-ins

Plug-in modules provide a high level of detail to cover actions that are specific to individual sports. For example during 2006 feedback from a Major League Baseball Club resulted in a number of enhancements to the baseball plug-in with a typical change being the addition of coverage for "umpire call".

Plug-ins are available for American Football, Baseball, Basketball, Ice Hockey, Soccer, Tennis, Golf, Motor Racing, with the most recent addition being for Curling.

Olympic Reports

Results for the 2006 Torino Winter Olympics were supplied to the websites of a large number of US newspapers using a system based on the SportsML core.

In the system (developed for the Associated Press by XML Team) the official Olympics WNPA feed was converted to SportsML and passed to a database. New SportsML output was then generated by a series of database queries, and formatted into HTML pages for supply to the newspaper web sites.

A Good Description

Overview

Maintenance updates of the IPTC NewsCodes have been complemented by a thorough review of all sets. Introduction of the G2-standards will make increasing demands, with additional NewsCode sets being required.

First steps have been taken towards the production of a new generation version of the Subject NewsCodes.

Introduction of the IPTC G2-Family of Standards family will place additional demands on the IPTC NewsCodes - which are a series of standard sets of meta-data that can be applied to news objects, and allow consistent coding across the industry and over time.

These NewsCodes are already in widespread use, with the established IPTC exchange standards and elsewhere, so there is a continuing programme of additions and updates to the existing sets in response to user needs. New NewsCode sets are also developed when specific requirements are identified.

Application sets

At the moment there are twenty-six NewsCode sets available for use, but their exact application is not always immediately apparent. To help make things clearer the codes have been sorted into four specific application sets - Descriptive NewsCodes, Administrative NewsCodes, Transmission NewsCodes and Exchange Format News-

Codes. Details of all of the Sets are given in the panel opposite.

Availability

All of the NewsCodes are available for free use, and can be downloaded from the NewsCodes section of the IPTC Web site (www.newsCodes.org) as XML files in the form of NewsML 1 Topic Sets.

In keeping with other IPTC standards the NewsCodes are developed and maintained in English, but several of the sets have been translated into other languages - including French, German, Italian, Japanese, and Spanish - by IPTC members - and these translated sets are also available.

Although the words for the terms, and their associated explanations, entries change with translation the "formal name" of each code is numeric and so is language independent.

There is also a NewsCodes viewer. This is a Windows application that provides an easy way to navigate through the codes and compare translations.

As with other IPTC standards use of the NewsCodes is subject to the provision of the IPTC Intellectual Property Policy - see <http://www.iptc.org/goto/ipp>.

Review
In preparation for the introduction of the G2-standards, and to meet the developing needs of established users, work has been di-

rected in two main directions, a comprehensive review of the existing NewsCode sets, and the development of new-generation sets. An extensive review of the Subject

IPTC NewsCodes

Descriptive NewsCodes

Taxonomies for description of the content of news items.

Genre Describes the nature, journalistic or intellectual characteristic of a news object, but not specifically its content.

Scene Describes the scene covered by the content.

Subject Code A three level system for describing content by. Some 1300 terms are available and several terms can be assigned to a single news object to give a very precise description of the content.

The top level has seventeen main Subjects, the second level a series of SubjectMatter under each Subject, and the third level provides SubjectDetails under the SubjectMatters.

Subject Qualifier Subject Qualifiers provide a narrower attribute-like context - typically for a sports-related subject code (such as the gender of participants, or indoor/outdoor sports).

Administrative NewsCodes

Taxonomies for the administrative properties of news items.

Audiocodexs Current audio-en/decoders, many of them controlled by international standards.

Audiocoder software based audio-en/decoders.

Colorspace Colour space definitions, such as RGB, YUV or CMY.

OfInterestTo Target audience for a NewsItem, based, for example, on demographics, geography or other groupings. (see also "Relevance" below)

Provider A unique ID assigned by the IPTC to a company, publication or service provider.

Status NewsML specific: The current usability of a NewsItem within NewsML.

Urgency Relative importance of a news object for editorial examination.

Videocodex Current video-en/decoders, many of them controlled by international standards.

Videocoder Software based video-en/decoders.

Transmission NewsCodes

A taxonomy with controlled values for the transmission process.

Priority Relative importance of a NewsItem for distribution.

Exchange Format NewsCodes

Taxonomies to support specific functionalities of the different IPTC news exchange format standards.

Characteristics Property Names (not values!) to describe physical characteristics of content - such as "width" and "height" for photos, or "sampling rate" for audio.

Confidence The degree of certainty that data assigned are correct.

Encoding Popular encoding schemes used to transform data.

Format Technical format of a content - JPG for a picture, MP3 for audio or NITF or PDF for a document.

How Present Describes the way in which a topic occurs in the content of a news object.

Importance Relative significance of the metadata applied to a news object.

Labeltype The type of a label attached to a news object. (Labels are portions of human readable text - unlike most other metadata which are considered to be primarily machine readable only.)

Location (Type) Identifiers for of regions of the world where events take place.

Media Type Description the type of media in a very general way, such as text or photo.

MIME Type More specific description of the type of media by use of IANA registered MIME types.

Newsitem Type General description of the type of content that a news item carries.

Notation Technical notation of a piece of content.

Property NewsML specific: The type of a NewsML Property element.

Relevance The extent in which a news object is relevant to the target audience specified by "OfInterestTo" (see above).

Role Role of an individual news object within a package of several news objects - for example "Main" (content), "Supporting", or "Caption".

Topic Type NewsML specific: The kind of thing that the individual thing represented by the topic can be characterised as.

NewsCodes marked in blue have been deprecated and should not be used for new applications. They are retained so previous applications will remain valid.

NewsCodes had been undertaken during 2005, so recent changes to this set have mainly consisted of additional terms, and some minor tidying up.

Consideration of the other NewsCode Sets resulted in a number of minor changes and additions, and a decision to deprecate the Notation NewsCodes and the LabelType NewsCodes as they no longer appeared to be used.

Two major sets for AudioCodec NewsCodes and VideoCodec NewsCodes were approved with the older Audiocoder and Videocoder NewsCodes being deprecated. Minor changes were made to a number of the other NewsCode sets.

When a NewsCode set is deprecated it remains available to ensure that existing users will not be affected, but is marked with a recommendation that there should be no further use.

Additions

Additions and changes to the NewsCodes have to be formally proposed by IPTC members in line with published Change Management Guidelines, but other organisations may submit informal proposals and seek support from IPTC members.

Proposals are then circulated to members before consideration by the NewsCodes Working Party when they may be approved, as submitted, modified or rejected (often when a proposal is turned down the Working Party will suggest ways that the proposer could submit a replacement proposal to meet their requirements).

Fast track

Since consideration of changes has to wait for a formal IPTC meeting there is also a mechanism for "fast track" approval of third level (Subject Detail) changes to the Subject NewsCodes. Approval is given by a standing jury and the requirement for this jury has been changed so it has to consist of at least four but not more than six delegates.

As with other proposals "fast track" applications are circulated to members for consideration and an other change to the procedure makes it possible for members to ask for consideration of a specific

term to be deferred for full consideration by the Working Party.

Updates

Updated NewsCode sets are issued when changes and additions have been formally approved - generally this is after a IPTC Meeting. Automatic notification of updates to the NewsCode sets is available as a RSS feed.

Subject NewsCodes

Initial development of the Subject NewsCodes took place over ten years ago and it has become apparent that the structure which was adopted at the time is no longer appropriate. In particular, the system has a fixed hierarchy while it is not possible to create relationships between terms.

Accordingly a Future NewsCodes Working Group has been set up within the NewsCodes Working Party. Although this group has not yet made any formal proposals, preliminary work suggests that a new data model will be needed to take advantages of the metadata features in the G2-standards.

Approach

The established seventeen main Subjects will probably be retained. However, the new structure will not be hierarchical but will support polyhierarchy so a concept can have multiple parents. Concepts will be differentiated by unique lds and definitions, while relationships

may be used to provide different user views of the codes.

Coverage will probably be extended to areas that have only been partly dealt with so far, with updated guidelines being introduced for the introduction of new terms and definitions.

It is envisaged that the maintenance of the existing Subject NewsCodes will continue, to support users. Where appropriate, changes and additions proposed for the new system will be also be incorporated in the Subject NewsCodes. This will not always be possible, though, and differences in the structure mean that there will be an increasing divergence between the two versions. Because of this it will be in users' interests to start using the new version.

Taxonomy management

Development of the new Subject NewsCodes involves the interchange of concepts, proposals and comments between Working Group members distributed around the world. To help with this it was decided to investigate the advantages of using a formal taxonomy management system.

Following presentations at the 2006 Autumn Meeting it was decided to adopt the SchemaLogic taxonomy management system (www.schemalogic.com). This will provide a central location for the semantic data with direct access for the delegates involved in the development process.

NewsCodes and the News Architecture

An important feature of the new News Architecture is the mechanism for handling metadata, and in many cases the metadata fields have to be populated with values from a controlled vocabulary.

Providers are free to use their own taxonomies, but IPTC will provide a recommended set of vocabularies for functional requirements. The existing NewsCodes will meet some of these requirements but a series of additional NewsCode sets will be needed for specific functionality in the G2-standards. These include concepttype, itemrelation, rendition, titlerole and whypresent, with the complete set being under development by the NAR Working Party.

Taxonomies used for any given News Item have to be identified by a catalog entry in the Item, and in most applications a set of appropriate vocabularies will be maintained by the users' processing system. The Package Item provides a straightforward way of handling providers' vocabulary sets for delivery to users.