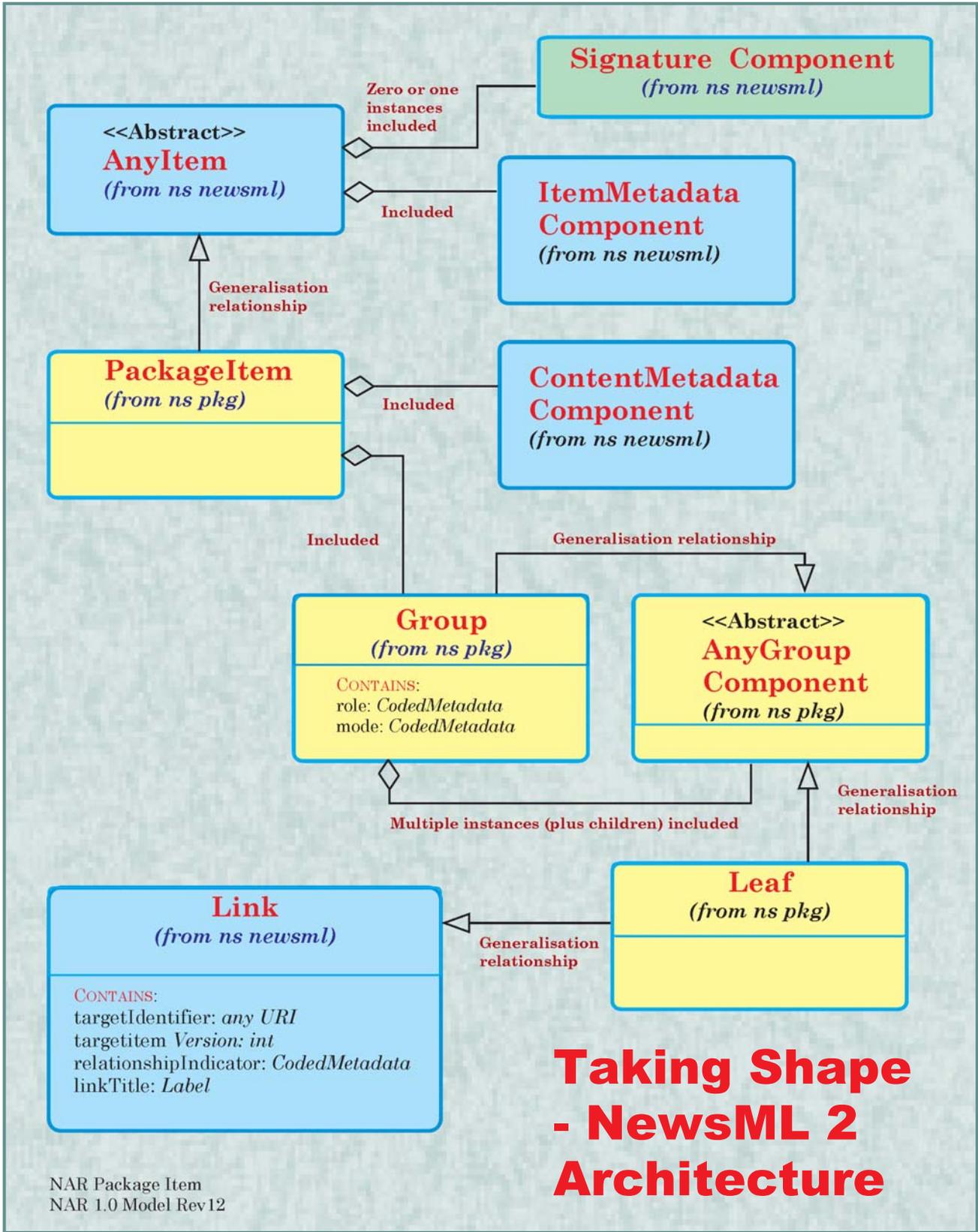




IPTC Spectrum

IPTC - INFORMATION TECHNOLOGY FOR NEWS



Taking Shape - NewsML 2 Architecture

IPTC Nominating Members

Agence France Presse (afp) - France - www.afp.com
ANSA - Italy - www.ansa.it
Associated Mediabase Limited - UK - www.mediabase.co.uk
Austria Presse Agentur (APA) - Austria - www.apa.at
BBC Monitoring - United Kingdom - www.monitor.bbc.co.uk
Business Wire - USA - www.businesswire.com
CCNMatthews - Canada - www.ccnmatthews.com
CNW Group Ltd - Canada - www.newswire.ca
Deutsche Presse-Agentur (dpa) - Germany - www.dpa.de
Dow Jones & Company - USA - www.dowjones.com
European Alliance of News Agencies - Europe - www.pressalliance.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
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
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 Telegrambyrå (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 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
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/index.html
EBU - European Broadcasting Union - Europe - www.ebu.ch
Eidos Media Srl - Italy - www.eidosmedia.com
eRoket.com - USA - www.eroket.com
Fingerpost Ltd - UK - www.fingerpost.co.uk
Harris and Baseview - USA - www.harrisbaseview.com
HINA - Croatia - www.hina.hr
IFRA - Germany - www.ifra.com
ITAR-TASS - Russia - www.itar-tass.com
La Repubblica - Italy - www.repubblica.it
Magyar Távirati Iroda Rt (MTI) - Hungary - www.mti.hu
Mainstream Data Inc. - USA - www.mainstreamdata.com
News Engin, Inc. - USA - www.newsengin.com
NewsLink - UK - www.newslink.co.uk
RelaxNews - France - www.relaxnews.com
Ritzau Bureau I's - Denmark - www.ritzau.dk
RivCom - UK - www.rivcom.com
Suomen Tietotoimisto Oy - Finland - www.stt.fi
XML Team Solutions Inc. - USA - www.xmlteam.com

International Press Telecommunications Council

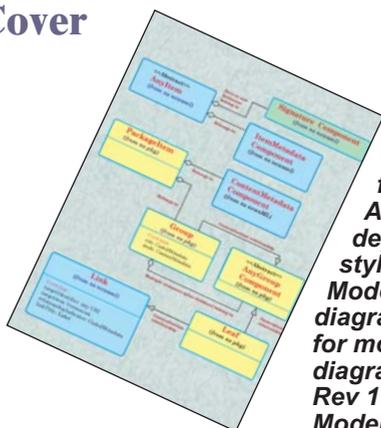
Chairman:
Stéphane Guérillot
Honorary Treasurer:
Henrik Stadler
Vice Chairmen:
**Walter Baranger; Geoffrey Haynes;
Rudi Horvath; John Iobst;
Peter Müller; Hitoshi Saito.**
Managing Director:
Michael Steidl
Editor:
Hugh Johnstone

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

Cover



*Package Item
for the NewsML 2
Architecture
depicted as a
stylised UML (Unified
Modelling Language)
diagram. See page 14
for more details. This
diagram is taken from
Rev 12 of the NAR V1.0
Model.*



**Information
Technology
for News**

Contents

Organisation

Development and Consolidation	4
IPTC Membership	5
IPTC Standards	6
Management Committee	7
Klaus Sprick	7
News Standards Summit 2005	8
Challenge and Response	10

PR Committee

Making News	9
-------------	---

Standards

Starting from the Basics	12
NewsML 2 Architecture Goals	13

NewsML 2 Architecture

Designing the Framework	14
NAR Metadata Handling	15
Power Conformance	16
Contributors	17
NAR Working Groups	17

NewsML 1 Maintenance

A Proven Success	18
NewsML 1 Applications	19

NewsCodes

Public Codes	20
Automated Categorisation	21

News Content

Ready to Process	22
Working Groups	23

News Industry Text Format

Well Proved	24
-------------	----

*This issue of the IPTC Spectrum has
hyperlinks for web addresses (in blue)
and for page references (in red).*

Development and consolidation

As an organisation IPTC is run by a Management Committee and Chair elected annually at a formal Annual General Meeting by the nominating members. The Committee act as Directors of the company and are responsible for development of the organisation. Day to day operations are overseen by the Managing Director.

Chair:
Stéphane Guérillot (AFP)

Honorary Treasurer:
Henrik Stadler (TT)

Vice-Chairs:
Walter Baranger (New York Times Inc)
Geoffrey Haynes (AP)
Rudi Horvath (APA)
John Iobst (NAA)
Peter Müller (SDA/ATS)
Hitoshi Saito (NSK)

Managing Director:
Michael Steidl

Web portal:
www.iptc.org

Recent activities have helped consolidate IPTC's position as the leading source of news exchange standards and laid the foundations for the development of a new family of standards that will offer a high level of interoperability.

Aims of IPTC can be summarised as being "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". In applying these aims the growing importance of the World Wide Web has become a major consideration so there has been increasing interest in systems for multimedia news and on-line publishing.

Fundamental; approach

The decision to start work on a successor to NewsML 1 (which is now over five years old) was taken early in 2004 with production of a business requirements document. At the same time work was underway on EventsML, while there was also an increasing appreciation of developments taking place in the underlying XML technology.

By the time of the Autumn 2004 Meeting it had become apparent that

a more fundamental approach would be appropriate. Further consideration resulted in a set of proposals that were considered by a special teleconference meeting of the Standards Committee, held in early January 2005. This meeting resulted in a decision to change the Working Party structure and re-focus development efforts on a new architecture for news standards.

NewsML 2 Architecture

Now known as the NewsML 2 Architecture (NAR) this is intended to provide a single generic model for exchanging all kinds of news information, and provide a common base for all new IPTC standards. This model is based on a view of the relationships between news (and topics) and the real world - further details on the concepts underlying the NewsML 2 Architecture are included in the Standards pages.

Established standards - such as the NITF and SportsML are not directly involved in the NAR, so the established user bases will not be affected, and appropriate development work will continue. However it is envisaged that future major re-



Retiring Chairman John Iobst (right) congratulates the incoming Chairman Stéphane Guérillot (left) at the Annual General Meeting.

leases of these standards will be designed to integrate into the NewsML2 family of standards.

Common structure

Use of the NAR will give IPTC standards a common structure and style which will make them easier to implement and understand. The design is modular with common components, management rules and processes that can be reused as appropriate. Steps have also been taken to establish a common XML Schema Style for all IPTC standards, again helping to reinforce the fact that they are part of a family of standards.

Metadata handling

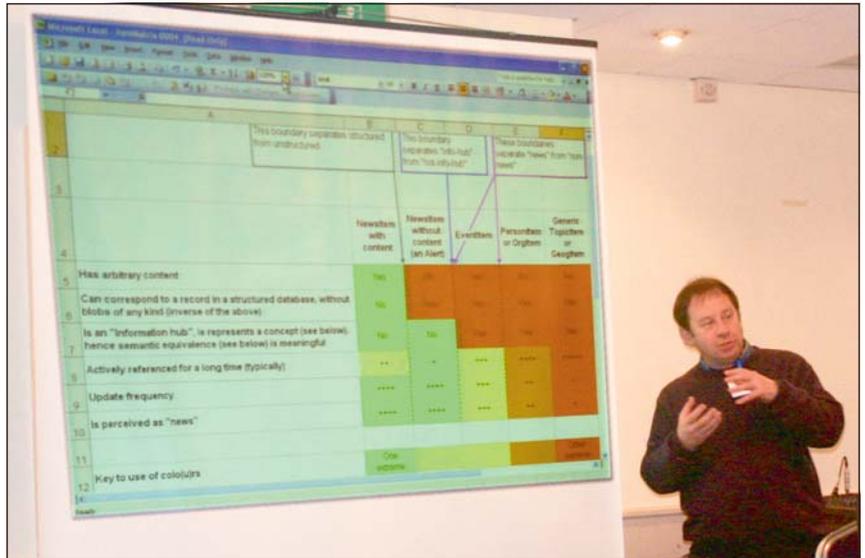
The news industry is a major source of information, and probably one of the largest users - and creators - of metadata, and an important feature of the NAR is a new way of expressing metadata. The NAR has been designed to be compatible with the Semantic Web of the W3C and steps have been taken to ensure it can be used with the underlying Resource Description Framework (RDF). However it is important to note that the implementation of metadata in the NAR does not require any knowledge of RDF.

Timescale

The decision to redirect efforts into the NewsML 2 Architecture has extended the original timescale for development of the first members of the new IPTC standards family - but it will also ensure that future standards work will be easier and quicker.

Consultants have been appointed to help reduce the time needed for NAR development, and to provide specialist expertise. As a first step Jay Cousins (RivCom) and Ulf Wingstedt (Cnet) investigated ways of expressing IPTC Standards using W3C XML Schemas and provided some Architecture Implementation Guidelines.

The same consultants are working on XML Schemas to implement the NewsML 2 Architecture. In addition Mark Birkbeck (x-port.net Ltd) has looked at ways in which the NAR metadata requirements can be reconciled with the W3C RDF abstract model.



Progress

Concentrating the available resources has resulted in rapid progress, and by the end of 2005 the NewsML 2 Architecture was sufficiently advanced for a first Experimental Phase to start. This is intended to help ensure that the NAR will meet users practical requirements, and help refine the technical specification.

It is hoped that results from the first Experimental Phase will be available for consideration at the Spring 2006 Meeting (to be held at the end of March 2006) so the NAR Technical Specification can be fi-

Demonstration and discussion are an important feature of Meetings, helping to ensure that delegates have a detailed understanding of how the work is proceeding. Here Mischa Wolf explains some aspects of the metadata approach that is being adopted in the NewsML 2 Architecture.

nalised. It will then be possible to move on to detailed development of the first of the new standards - a general news markup standard as a successor to NewsML 1 and EventsML - which should take place during 2006.

IPTC Membership

Membership of IPTC now includes a wide range of organisations representing different aspects of the news industry: news agencies, newspapers, photo agencies, broadcasting organisations, syndicators, public relation organisations, system suppliers, software producers, and other bodies involved in the news industry.

There are two main types of membership:

Nominating members have voting rights at General Meetings and in Committee and Working Parties. They can have up to three delegates at a meeting for each subscription. This type of membership is open to organisations and companies concerned with news collection, distribution and publishing.

Associate members are eligible to vote in Working Parties and can send one delegate to a meeting. Associate membership is available to organisations and companies concerned with news collection, distribution and publishing and to system vendors supporting the news industry. Special membership terms may be offered to new members from economically distressed areas and academic institutions.

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

Development

Most of the work on the NAR has been carried out by a small group of delegates (see page 17) using a

combination of emails - on dedicated development discussion groups that are restricted to IPTC members - and regular (two a

week) teleconferences.

For example in the period between the 2005 AGM and the 2005 Autumn Meeting work on the NAR involved nearly a thousand emails and around forty hours of teleconference time. As with other IPTC work the success of this project owes much to the individuals who contributed their time, and to their organisations for enabling them to make the commitment.

IPTC Standards

NewsML 1 - a media-type-independent XML-based standard for the representation of multimedia news throughout its lifecycle, including production, interchange and consumer use.
www.newsml.org

News Industry Text Format (NITF) - XML-based markup scheme for individual news items. It can be used as a standalone format for news distribution or as a content model for NewsML.
www.nitf.org

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

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

IPTC Core - photo metadata primarily for use with Adobe's Extensible Metadata Platform XMP. Successor to the widely used "IPTC Headers" and part of the IPTC4XMP effort.
www.iptc4xmp.org

Under Development

NewsML 2 Architecture - this is not itself considered to be a formal standard for public release but the architecture on which new standards will be based. Planned new standards include a general news markup (as a successor to NewsML 1) and

EventsML (intended for the exchange of information on event publishing, planning and coverage).

Legacy Standards

Information Interchange Model (IIM) - container for news information in any of the common news media (including text, photographs, graphics, audio and video). A subset of the IIM metadata sets was adapted to provide the "IPTC Headers".
www.iptc.org/IIM

Digital Newsphoto Parameter Record (DNPR) - container file format designed to carry digital news photograph data within the IIM.
www.iptc.org/IIM

IPTC7901 - original text message format developed by IPTC, and still in widespread use. The last revision of this standard was carried out in 1995.
www.iptc.org/IPTC7901

IPTC standards and supporting documents are all available for download from the links given above.

Established standards

Although efforts have been concentrated on the NewsML 2 Architecture, work has continued to ensure that the established standards continue to meet users' needs. There have been new releases of SportsML (V1.6) and the NITF (V3.3), while NewsCodes development is a continuing process.

A new Working Party was established to deal with any issues arising with NewsML 1, which has a substantial user base. The importance of NewsML 1 has also been recognised by its adoption and release as a Japanese Industrial Standard - see page 18.

Standards co-operation

Co-operation with other Standards bodies is an increasingly important aspect of IPTC's activities and discussions on metadata issues have been held between IPTC and W3C representatives. Following the success of the first News Standards Summit, IPTC were co-sponsors of the second News Standards Summit - see page 8.

A collaborative effort between IPTC, the IDEAlliance and Adobe Systems resulted in an XMP Schema and custom panels to allow the use of the IPTC Core in Adobe Photoshop.

In response to a request from Ifra a Colour Space Task force was set up to investigate ways of retaining the EXIF data associated with digital camera image files during agency processing.

Participation

Participation of non-members is encouraged with a series of public discussion groups, covering the NAR and individual standards - details of these groups are given in the sections dealing with individual standards. In addition relevant document are made available on

the IPTC Web site as work proceeds.

Meetings

Following the established arrangements there were three Meetings in 2005: The Spring Meeting was held in San Diego; the 40th Annual General Meeting in London; and the Autumn Meeting in Milan. All were well attended, and provided a wider forum for discussion of the work carried out by the development groups

In a new departure, as noted above, a special teleconference session of the Standards Committee was held in January. This approved changes to the Working Party structure so they were in place for the Spring Meeting.

Presentations

The Annual General Meeting was held in London at the invitation of Reuters, who arranged a stimulating series of speakers to complement the technical sessions. Delegates were welcomed by Geert Linnebank, Reuters Editor-in-Chief and Global Head of content with an address that encapsu-

lated many of the challenges faced by the news industry - see page 10.

Another aspect of the challenges faced by the news industry - this time the growth in volume and availability of financial information was looked at by Ken MacKenzie, Reuters Global Head of News Capabilities. A third Reuters speaker was Jeremy Lebrecht who provided an overview of the development and application of XBRL (eXtensible Business Reporting Language), along with a demonstration of how XBRL information can be handled in NewsML. The way that metadata is dealt with in XHTML 2.0 was explained by Steven Pemberton, W3C HTML Working Group and Forms Working Group Chair.

Two speakers from the Press Association (PA) addressed different aspects of news content. Phil O'Brien from EMPICS demonstrated the ShootLive system which provides real-time photography for new-media applications, and Jennifer Campbell, Managing Director Meteo Consult explained the Weather services offered by her organisation.

Klaus Sprick



The 2005 Spring Meeting saw the departure of Klaus Sprick, who stood down from the Management Committee on his retirement from dpa. After joining the German News Agency (dpa) in 1968 he quickly became involved in IPTC activities - which at that time were mainly concerned with safeguarding press telecommunication services.

As IPTC started development of standards for the news industry, Herr Sprick was appointed the first chair of the new Technical Committee, which produced the first IPTC technical standard - IPTC 7901 - in 1979. Election to the Management Committee followed in 1984, a position he held until his departure, including a term as IPTC Chairman from 1990 to 1993.

Klaus Sprick's considerable contribution to IPTC was recognised by a unanimous vote of the Committee of the Whole Council to make him the first Honorary Member of IPTC. In addition a dinner was held in his honour in San Diego (during the Spring Meeting) at which he - a keen sailor - was presented with a model of a J-class yacht as a memento.

Management Committee 2005-2006

There were a number of changes to the Management Committee at the 2005 Annual General Meeting. John lobst (NAA) had completed his three-year term as IPTC Chair and Stéphane Guérillot (AFP) was returned unopposed as the new Chair.

The departure of Klaus Sprick (see above right) left a vacancy on the Management Committee, and this remained open until the formal IPTC Annual General Meeting - which deals with the running of IPTC as a company - when a new Committee was elected. Members are:



Chair
Stéphane Guérillot
(AFP)



Vice-Chair
Walter Baranger
(New York Times Inc)



Vice-Chair
Geoffrey Haynes
(AP)

Vice-Chair
Rudi Horvath
(APA)



Vice-Chair
John lobst
(NAA)



Vice-Chair
Peter Müller
(SDA/ATS)



Vice-Chair
Hitoshi Saito
(NSK)



Honorary Treasurer:
Henrik Stadler
(TT)



IPTC Managing Director
Michael Steidl



A regular feature of Meetings has been presentations by members to bring various aspects of their activities to the notice of other members. During 2005 these presentations included the NewsEngin CatScan categorisation system, the services offered by Mainstream Data, and the Eidos Méthode editorial system.

Olympic briefing

The practice of arranging briefings for Olympic events continued with a visit to Torino - which is hosting the XX Winter Olympic Games - at the invitation of Mr Cristiano Carlutti, Torino Press Operations Managing Director.

Delegates were able to discuss

details of the press arrangements and meet the management team. This included Richard Palfreyman, Head of the Main Press Centre in Torino, who had previously met some of the delegates at the IPTC briefing for the Sydney Olympics 2000 (where he was head of Press Operations).

News Standards Summit 2005

Intended to help users and developers gain a better understanding of standards used by the news industry, and to show how the standards work in practice, the second News Standard Summit was held in Amsterdam in May 2005, in conjunction with the Xtech 2005 Conference.

The opening session provided an overview of the main standards in use (or under development), with a second session looking at ways the standards are being used in practical applications. Both sessions were chaired by Mischa Wolf (Reuters).

Brief details of the presentations are given here - the presentation slides, and audio files of the presentations are available from www.newssummit.org/2005, providing a continuing resource for interested users.

Standards

Some aspects of the work being carried out by the W3C were covered by Steven Pemberton, W3C HTML Working Group and Forms Working Group Chair who provided an introduction to Metadata in **XHTML 2.0**. Aims of the work on XHTML 2.0 are to make it more accessible and usable. Steps have been taken to integrate XHTML with RDF in a way that is accessible to the HTML community.

Ivan Herman, Head of Offices W3C, explained that the Semantic Web is a metadata based web infrastructure, with **RDF (Resource Description Framework)** being a general model for metadata statements. The **W3C Ontology Language (OWL)** can be used to define knowledge concepts and relationships, with a series of increasing levels of complexity.

The role of metadata in broadcasting was considered by Jean-Pierre Evain, Senior Engineer, EBU with the main types being administrative, technical and descriptive. There is a high level of complexity and efforts are being made to define common semantics, while an **EBU News Exchange Schema** has been developed. **TV-Anytime** provides metadata so consumers can search and select broadcast content.

Atom has a complex history but is now an IETF (Internet Engineering Task Force) standard,

known as a Request for Comments (RFC). Bob Wyman, CTO, PubSub provided a quick overview - with examples - of the standard and of **RSS** - another standard that has a complex history.

PubSub - Publish and Subscribe - provide subscribers with rapid notice of new content, and techniques are being developed to deal with the "Grey Web" - pages that search engines find difficult to index.

An update on the **PRISM (Publishing Requirements for Industry Standard Metadata)** and **ICE (Information and Content Exchange)** standards was provided by Dianne Kennedy, VP Publishing Technologies, IDEAlliance.

There was extensive coverage of current and planned IPTC standards:

NITF features and its application by dpa-infocom were considered by Hubertus Koehler, Chief Technical Officer, who also gave some information on the EU-funded **MINDS (Mobile Information and News Data Services)** project. **NewsML** and the **IPTC News Architecture** were explained by Laurent Le Meur, Technical Manager of Multimedia Development (AFP) IPTC News Architecture WP Chair.

Continued >

Future Attractions

Sponsors of the second News Standards Summit were IPTC, Ifra, and the IDEAlliance, and it was organised by Diane Kennedy (IDEAlliance), Harald Loeffler (ifra), Michael Steidl (IPTC) and Misha Wolf (Reuters).

The first News Standards Summit was held in the USA in December 2003. It is intended to hold further events but the date for the next News Standards Summit has not yet been fixed - it will probably take place after the release of the new IPTC NAR-based standards. Suggestions for the next News Standards Summit Program can be sent to nss_discuss@yahoogroups.com.

An outline of **EventsML** was provided by Arnaud Descamps, Chief Technical Officer, Relaxnews. Development of the **IPTC Core for XMP**, with an overview of **Adobe XMP**, was explained by IPTC Managing Director Michael Steidl.

Applications

Specific examples of the ways IPTC standards are being put to use were provided by a number of IPTC members:

Business Wire use of a NewsML 1 profile for their news distribution system - and the way it was introduced to their customers - was explained by Dean Large and Jayson Lorenzen.

Dave Compton outlined the development of Reuters News Systems - from IPTC7901, through the IIM to NewsML 1, and went on to consider the development of systems offering added-value con-

tent, such as the TopNews system.

AFP have moved towards the XML representation of multimedia news with the "Magazine Forum" being a pure NewsML platform, and the approach taken was described by Laurent Le Meur.

As a major news host the APA Database Host makes use of standards for the main operations of input categorisation and download. Rudolf Horvarth and Manfred Mitterholzer from APL-IT also demonstrated the APS-IT Power Search full-text retrieval system.

Some of the special features of mobile portals were explained by Kevin Smith, Architect, Vodafone Global Technology with reference to the Vodafone Live data service. Use of appropriate standards allows improved search and marketing facilities.

Making News

All aspects of IPTC's public relations are dealt with by the Public Relations Committee. This includes encouraging better understanding of the organisation and standards, and promoting the advantages of membership.



Chair: Walter Baranger (The New York Times)

As with other IPTC activities, maintaining the organisation's public image - and encouraging membership - relies heavily on the efforts of members, coupled with continuing input from the IPTC Managing Director.

Presentations at appropriate events have an important role in creating and developing interest in the standards - and in making contact with systems suppliers who work with the news industry. During 2005 such presentations were held at the NEXPO (USA) and Ifra (Europe) events, while the Managing Director gave a talk on "IPTC and Metadata" to the European Association of News Agencies.

Web site

The IPTC web site - and related sites for individual standards - provide the public face of the organisation, with the aims and activities clearly explained. Standards documentation and related information is available for free download. Web statistics show that the IPTC Web pages are well used with particu-

larly high download figures for items like the IPTC NewsCodes and other standards information.

Press releases

Press releases are issued on a regular basis to cover activities at the main meetings, and for major events such as the launch of the News Standards Summit and the start of the NewsML 2 Architecture Experimental Phase. Releases are distributed by IPTC members, who also issue press releases covering their IPTC-related activities.

Publications

Publications remain an important part of the publicity mix with the electronically distributed newsletter - IPTC Mirror - complemented by the annual IPTC Spectrum. This is distributed in both printed and electronic form and is intended to provide a less-technical overview of IPTC activities and achievements.

Both the IPTC Mirror and the IPTC Spectrum are available on the IPTC web site.

Challenge and Response

At the IPTC Annual General Meeting, Geert Linnebank, Editor in Chief of Reuters, looked at the challenges facing the news industry and suggested some of the ways in which the industry can respond.



The IPTC is an important standards organisation and one which Reuters is committed to, and very pleased to be, supporting. That is not least because of the challenges the IPTC, and the news industry it represents, face.

Why is this so, and what can we do about it? Well, as I see it, the issue is this, in a nutshell. The news industry - our industry - has been in the business of making B2B (business-to-business) standards for decades. And while there has been stiff competition within the industry, among ourselves, taken as a whole we didn't face any real competition from the outside.

So it did not matter much how difficult our standards were to understand - or implement - or use. And it did not matter much how slowly those standards of ours evolved. They were the only game in town. So we became a bit complacent, perhaps a bit lazy, even.

Change

And then, one day we all woke up and we found that the world had changed. So how did the world change? What happened?

What happened is that a comet hit the earth, so to speak. It sent, and continues to send, and will continue to send, shockwaves round the globe. This comet is, yes, you guessed it, the Internet.

We all know the Internet is much, much more than cheap connectivity. It dramatically changes the rules in the information space. It has brought immensely powerful natural selection into the world of standards and software. This in turn has resulted in the emergence of entirely new paradigms for information interchange.

Parallel environment

These new paradigms have been extremely successful, resulting in the emergence of a very different kind of "news environment", one which largely operates in parallel to ours, and one which is showing exponential growth where ours is static at best.

This new "news environment" is not based on the traditional one-way "broadcast" model that our industry grew up around, but rather on a fully networked model. Now some commentators

confidently predict that this new environment, of bloggers, alternative news co-operatives, aggregators, search engines and the likes is set to overwhelm the established industry of which we form part, and we can certainly argue about that, and at some length.

Standards

But what is much more difficult to argue with is the speed at which that environment is developing and increasing its reach. And that, I contend, is due because it is built entirely on standards. Not the kind of standards which we ourselves have traditionally developed and used. Not the kind that it takes our industry years to develop and to implement.

No, in this new news environment, this parallel universe, standards evolve like in warp time. Generation follows generation many, many times faster than in our own world. Some survive and evolve, while others simply get left behind. Take the growth of that blogosphere, which has been driven almost entirely by a standardised, low-cost publishing environment and also, to some extent, by RSS.

Simplicity

It is probably a truism to say that RSS owes its success to what it says it is - Really Simple - easy to implement and use. Now I recognise that some of our needs are more extensive, but I think there is still a lot we can learn from RSS about making standards accessible.

I talked about the two parallel universes, and hinted at the risks that the two remain separate. I firmly believe our challenge is to bring the two together in a way that is both beneficial to the thousands of millions of people who rely on the Internet to get and exchange information, and also to our industry with its professionalism, its creativity, its mission.

Adopting change

I do see some encouraging signs that this can be done, if we're prepared to adapt and adopt change. There is, for instance, the example of the BBC, a professional news organisation and

eminent representative of that industry of ours. The BBC increasingly succeeds in using its website, its online presence, and its brand, to solicit contributions from the public - both textual and visual. These it uses to enrich its news service. Safeguards are in place to protect the BBC from hoaxes or setups. And the contributors feel they're "part" of the BBC, which helps secure their loyalty.

What the BBC is doing here, if on a small scale still, is bridging the two parallel universes. In the process it is expanding its network of eyewitness reporters by millions. Compare that with Reuters 2,300 journalists worldwide - the very best and brightest, highly trained and very professional, but also potentially utterly outnumbered.

Vital role

I passionately believe that despite the emergence of the "parallel universe", the industry represented around this room has a vital role to play in the future. Our journalistic standards, our professionalism, our processes to validate are arguably more relevant today, not less, as the tide of raw, unverified, information, rumour, gossip and propaganda keeps rising like never before.

But unless we dare to venture outside our "walled gardens", challenge our "broadcast" approach and look for workable ways to become part of, and of harnessing, the Internet, there is a real risk that we lose our voice, and with it our relevance.

Co-operation

Back to standards. I think we all accept the need for co-operation - at least as a principle. But the reality is sometimes quite different. Our business is often fiercely competitive. We all really want that scoop, that exclusive. This means that we often work in isolation from each other. That's understandable and good, but it also makes it difficult to create industry-wide standards.

Our customers want us to standardise on the means by which they can link and bind with us. That doesn't mean they want us to standardise on everything - far from it, They want us to compete, and aggressively, on the quality of our content, on the coverage we provide, on its timeliness, its reliability, its angles, its voice. These are core strengths by which we will continue to differentiate ourselves.

Cost-effective

Now I think I speak for everyone here when I say cost does matter. We all need to be as cost-effective as we can - that's a simple fact of life. My belief is, standards really help us here. After the first-starter advantage

has worn off, using propriety technology usually equates to additional cost to the business. And if we can make savings by adopting standards then we can devote more to those areas where we do, and must, differentiate - primarily in the content itself.

We must ensure that the standards allow us to play to our strengths, and also deliver the best possible service to our customers and consumers. From that we will all gain.

Search

Let's not forget search. Recent developments in online search have revolutionised the way people consume information. How often do you land on a Web site via search rather than typing in a URL for the home page? We all use search to get roughly what we want.

Although competition is hotting up, Google remains the benchmark here. It has created ever more sophisticated search techniques bringing in video, mobile services and much more, and it continues to innovate. At the same time, it keeps users happy with a simple look-and-feel interface to all its services.

Metadata

Metadata - a hot topic everywhere, and not least here at the IPTC - is at the heart of effective search. Google knows this: their search algorithm is based on the implicit metadata of "who links to what". But it is not always perfect, and it resists the top-down metadata view.

We can ask ourselves, against the might of a Google is there really a role for us in metadata? Well, consider this ... when a news item breaks, no one can have linked to it! I believe this is where the real experience we all bring comes in. Working together on how we define, code and then search for content is core to delivering the high quality, multi-media services many of us here are committed to providing.

This address has been edited for publication

Public participation

An indication of the success that the BBC is having in obtaining contributions from the public is given by the response to the London bombings - the first of which took place on the 7 July, just over a month after the IPTC AGM. According to the BBC, within an hour of the first bomb going off they had received fifty pictures from the public, with around twenty-two thousand emails and text messages being received in the day.

In their terms and conditions BBC News state that contributors agree to grant them "a royalty-free, non-exclusive licence to publish and otherwise use the material in any way that we want, and in any media worldwide". They also make the point that the copyright remains with the contributor and that they will endeavour to give contributors appropriate credit.

Starting from the Basics

Overall aim of the new family of IPTC standards is to establish a single generic model for exchanging all kinds of newsworthy information. This will be achieved by the development of the NewsML 2 Architecture (NAR) which will provide the framework for future IPTC exchange standards.

Although NewsML 1 has been widely adopted by the news industry - see page 18 - feedback from users - and non-users - indicated that there were a number of significant areas in which improvements could be made. Specific points that needed to be addressed included the level of complexity, a lack of consistency in applications, and limited metadata handling capabilities.

XML developments

To some extent these reflect the increasing experience in the practical application of XML standards and developments in XML technology. It has to be remembered that NewsML 1 was an early XML application - it is now more than five years old, which is a long time in such a rapidly developing technology like XML.

Work on a successor to NewsML 1 started in spring 2004 and initial consideration of the requirements was combined with efforts on other planned standards such as EventsML, and existing standards like SportsML. As work proceeded it was decided that a more fundamental approach would be better, so work was started on the NewsML 2 Architecture.

Working structure

A revised working structure was introduced to support this work with the creation of a new NewsML 2 Architecture Working Party (page 14) responsible for producing the framework, and a renamed and restructured News Content Working Party (page 22) dealing with the individual content standards based on the framework.

Existing Working Parties concerned with the NITF (page 24)

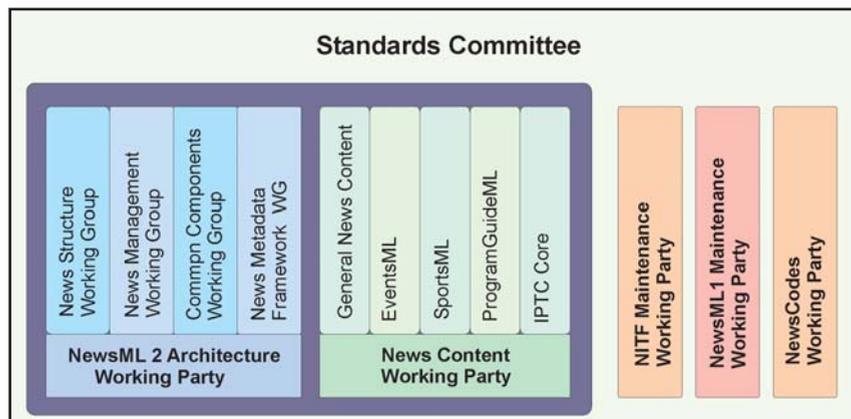
A new working structure was established to simplify development of the NewsML 2 Architecture. Activities related to the NAR are shown within the bold blue frame. There are two Working Parties, each of which has subsidiary Working Groups to deal with different aspects of the work.

Responsibility for the technical activities of IPTC belongs to the Standards Committee, which sets priorities, assesses available resources, establishes the Working Parties and oversees their work. Formal approval of new standards - and related material - has to be given by the Standards Committee before they are released.

Chair: Henrik Stadler (TT)



Web portal:
www.iptc.org



and the IPTC NewsCodes (page 20) continue unchanged, while a new Working Party was created for NewsML 1 Maintenance (page 18).

The decision to restructure the working arrangements and concentrate efforts on the NewsML 2 Architecture was taken at a special teleconference of the Standards Committee in January 2005.

Rapid progress

With the aims firmly established, work has proceeded at a fast pace, resulting in development of the NewsML 2 Architecture Model and Technical Specification drafts along with a draft XML Schema files by the end of the year. This made it possible to start an initial Experimental Phase (EP1) in December 2005. Results from EP1 (expected during March 2006) will help refine the concepts and the way they are implemented, to ensure that the aims are being met.

Abstract relationships

Underlying concept behind the NewsML 2 Architecture is the basic relationship between the abstractions of news and topics and the real world they report on and describe.

In this context news is considered as being “what has happened and is fit to be published” and occurs in the real world. The content of news - as written by a journalist, captured by a photographer or cameraman, or in the form of structured content - represents a more abstract level, though the actual technology used to report the news does not make any difference to the abstract “news content”. The NewsML 2 Architecture is designed to provide a wrapper for this “news content” which can be expressed in any media.

Dealing with the news needs more than a technical representation of the content, so a second level of abstraction is used to provide information about the content, with the metadata being applied to give a formal description, summary or categorisation.

Applying a management layer to the content and metadata results in a “News Item” which can be readily identified and processed within news systems.

In a similar way, Topics are taken

NewsML 2 Architecture Goals

- To simplify and unify the overall 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”.

as being “what is a matter of fact and is good to know”. In this case the real world is represented by factual knowledge of concepts in the real world. These may be real objects, such as people, organisations, locations and other material items, or abstract concepts represented by the codes used to identify the subjects of content.

Common framework

Basing new IPTC standards on a common framework - the NAR - means that operations will be carried out in a consistent way, making them faster to understand and easier to implement. The same function will always be handled in the same way, in all standards based on the architecture.

So far as possible the new structure will use generic building blocks, this means that development of new IPTC standards should be much simpler. In addition, implementers will be able to produce software components that can be used in different applications - such as for handling various types of news content.

A generic model has been adopted that will allow for future extensions. This means that it will be possible to extend standards to meet future - as yet unknown - requirements of the news industry.

Use of industry standards means that processing can be carried out

with standard tools. Underlying syntax is XML (Extensible Markup Language from the W3C) while the design makes use of W3C XML Schema and complies with the W3C RDF (Resource Description Framework). This should allow easy transfer of information to other XML-based standards and integration into the Semantic Web.

New standards

As the name suggests the NewsML 2 Architecture is intended as the underlying structure for new news-exchange standards. All new IPTC news exchange standards will be based on the NAR and it is anticipated that future versions of existing standards will be aligned with the NAR where possible and appropriate.

Present plans include:

A new standard for general news, to succeed NewsML 1. This will carry general news but will not provide packaging features;

Events ML, which will carry structured events information;

A major revision of SportsML which will allow use of the sports data structures as content of a NAR News Item;

Revision of the IPTC NewsCodes to use NAR technology to express a set of topics and their relationships;

Integration of ProgramGuideML as a NAR-based standard.

Designing the Framework

Task of the NewsML 2 Architecture (NAR) Working Party is to develop an architectural framework for the management and distribution of all types of news-related content. This framework will be the basis of all new IPTC news exchange standards.



Chair: Laurent Le Meur (AFP)

Vice-chair: Misha Wolf (Reuters)



Discussion groups:
<http://groups.yahoo.com/group/newsml-2>
<http://groups.yahoo.com/group/iptc-metadata>

Central element of the new IPTC standards family, the NewsML 2 Architecture (NAR) will provide the underlying structure for new content exchange standards.

Important features of the approach adopted are the development of a set of reusable Common Components along with an underlying structure - based on managed Items - that can be used for processing all types of news content, and a system for the representation of metadata.

Common Components

Common Components can be considered as building blocks that can be used in the construction of standards. At the lowest level, the data types represents a group of values and operations that can be carried out on the values. These include basic data types and more complex data types where the basic data type is extended by attributes to include additional information to more fully represent the value. Typical examples include dates, strings and numbers.

A basic component represents a single piece of business information and uses one of the data types to provide the model for the content. Basic components may be used on their own or as part of another component.

Aggregate components represent composite pieces of information with a specific business meaning, they can be made up from basic components and other aggregate components. For example *Administrative Metadata* is a Common Component which is made up of a series of other components - including such items as *Date Content Created*, *Creator*, *In-*

tended Audience and *Editorial Note*.

Examples of other Common Components include *Descriptive Metadata* (again made up from a set of other components) *Person*, *Organisation*, *Location*, *Label*, *Signature*, and *Rights Metadata*.

All Common Components have a consistent design and format, and are maintained for use in IPTC standards development. It is anticipated that some standards will also need dedicated components, and these will follow the same design principles.

Any Item

Content, both news and topics, is handled in the form of Items, which are manageable objects with a persistent, unique identity and a set of management properties that allow processing. An abstract *Any Item* acts as the model for all managed

NAR status

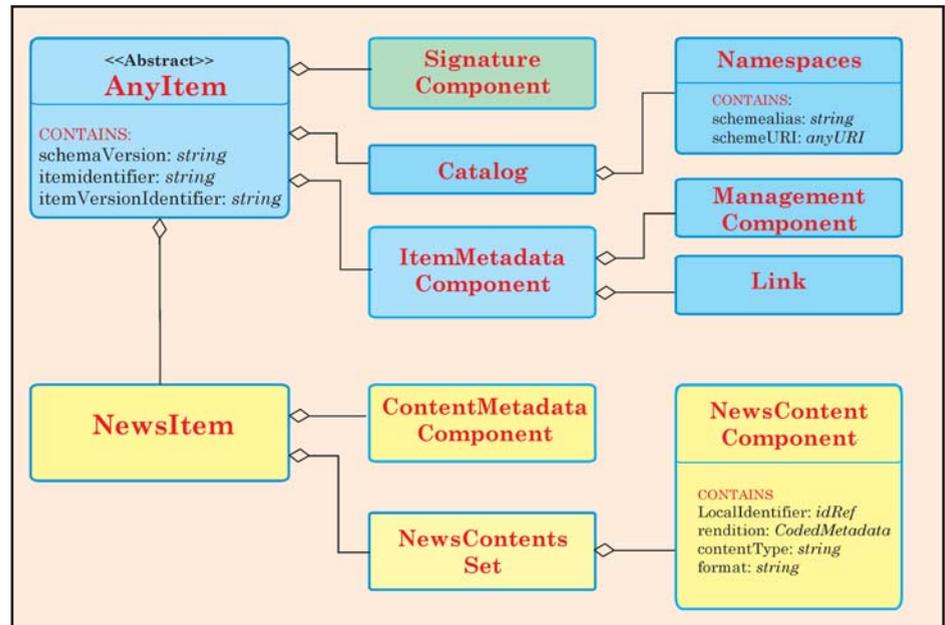
The overview of the NewsML 2 Architecture given on these pages is based on the version released in December 2005 (for use in the first Experimental Phase).

Development of the NAR is continuing and although the main features have been established it is important to remember that design changes may be made.

Representation of the News Item. Portions shown in blue are the Any Item, which is the model for the News Item, while specific News Item components are shown in yellow.

The Signature Component (in green) is part of the Any Item but can only be used in applications that conform to the “power” compliance level.

Connections indicate that components are included in the individual items. It is also possible to have a generalisation relationship where one part is a specialised version of the other. This type of relationship is shown in the Package Item diagram on the front cover.



items and three generic Items have been developed for use in the NewsML 2 Architecture. These are the *News Item*, *Topic Item*, and the *Package Item*. If any new content standard has specific requirements not met by the generic items a new Item will be generated - again this will be based on the *Any Item*.

Mandatory parts of the abstract *Any Item* are a *Schema Version* (for the XML Schema that specifies

the item), a unique *Item Identifier* (which has a form similar to that of a NewsML URN - RFC-3085 - but without version information) a *Catalog* (for scheme alias declarations, (see “NAR Metadata Handling” below) and an *Item Metadata Component*. It may also have an *Item Version*, a *Language indicator* (the default for the content of the item) and a *Signature Component* (see “Power Conformance” on page 16).

Management Component

In turn the *Item Metadata Component* contains a *Management Component* and a collection of *Links*.

Information directly related to the management of an Item is carried by the *Management Component*. This includes such information as indications of the nature of the item and of the content; individual dates for *Item Created/Modified/Released/Embargo Ends/Item Re-*

NAR Metadata Handling

Efficient, flexible and extensible mechanisms for handling metadata are an essential feature of the NewsML 2 Architecture.

Simple forms of metadata that have been identified for use in the NAR are String, Date, Integer and IRI (Internationalised Resource Identifier). This is a complement to the URI - Uniform Resource Indicator - that allows the use of a wider range of characters, such as non-Latin scripts).

In many cases metadata will take the form of coded values taken from a controlled vocabulary (or taxonomy), with a typical example being the IPTC Subject NewsCodes. To comply with the Semantic Web each term is identified by a URI (Uniform Resource Indicator). This URI is made up from a (separate) URI for the scheme and the code value.

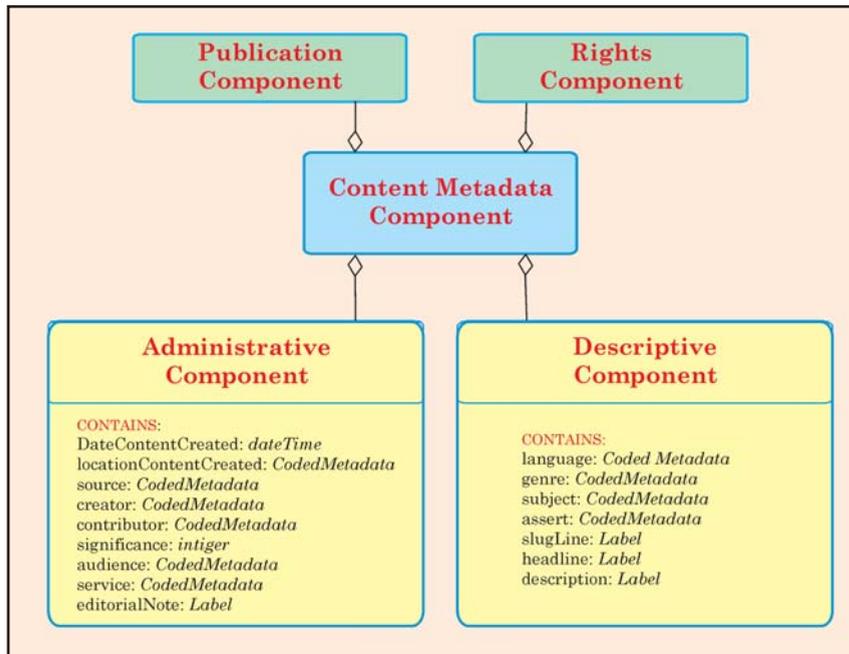
However, URIs can be fairly long - for example the URI for the IPTC Subject NewsCodes could be “urn:newsml:iptc.org:20001006:topicset.iptc-subjectcode:”. So to avoid having to include the complete URI each time the scheme is used a compact syntax is needed to handle them. In the NAR the intention is that scheme will be replaced

by an abbreviated form known as a scheme alias, and the code is then added after a colon. The resulting {scheme:code} pair is called a CURIE.

Any given item may use references to a large number of different taxonomies, while different users will want to use different taxonomies (including ones developed for their own use). To identify specific taxonomies the scheme alias parts of the CURIEs used in an item are included in the “Catalog” structure which is at the top of each of the NAR Items. This provides a mapping between the scheme alias and the URI of the scheme. Since this “Catalog” could be fairly large it can be stored in a central repository and simply referenced from the Item.

Provision has also been made for Labels, which are a specific forms of metadata that are intended to be human readable such as notes and instructions

The NAR is designed to be compatible with the Semantic Web of the W3C and the structure adopted is compatible with the Resource Description Framework (RDF). However, it is important to note that the implementation of metadata in the NAR does not require any knowledge of RDF.



Content Metadata diagram showing the Administrative and Descriptive metadata.

The Publication Component and Rights Component (shown in green) can only be used in applications that are compliant to the "power" level.

an article about a person to their biography, for example.

Derivation Links establish parent/child relationships. Typical applications include linking a translated article to the original one, or an edited picture to the original.

News Item

News content is carried in a *News Item*. The content may be in any media type or format, and may be structured. Typical examples are a news report, a picture or a video clip, and structured sports or events data.

Characteristics of a *News Item* include: short term interest (as news is volatile); it will usually be updated for only a short period, and may then be archived; it is expressed through a set of alternative renditions of some media content; it refers to an arbitrary set of concepts and entities; and it may be associated with other *News Items* or *Web resources*.

Since the *News Item* is derived from the *Any Item* it has all the associated management properties.

tired; the *Publish Status*, identification of the provider; and the conformance level the Item complies with.

Specific values for the *Publish Status* are; *Usable* - the Item may be published without restriction; *Withheld* - until further notice the Item must not be published or used, and must be withdrawn or retracted if already published; and *Cancelled* - must not be published or used and must be withdrawn on retracted if already published.

These status values can be qualified by: *Embargoed* - must not be published until the specified date and time; and *Retired* - may

be published and accessed but no new references should be created.

Taken together application of these values allows flexible management of items in an active news environment.

Links

Links provide a named relationship between the Item and another target item (or web resource) - links do not have to be included so the collection may contain zero to several links.

Navigation Links provide a connection between an Item and another related Item or a Web resource. This can be used to link

Power Conformance

Two conformance levels have been defined for the NewsML 2 Architecture - "core" and "power", with the "core" level designed to provide simplicity and interoperability between applications.

The "power" conformance level includes all the "core" functionality plus extra features. It offers a higher degree of flexibility, but at the cost of more complex programming, and some loss of interoperability (since not all applications will implement this level). However, "core" processors can be designed to deal with "power" level features by ignoring the information that cannot be processed.

So far only some explicit "power" features have been identified - these include a Signature Component, which makes it possible to ensure

integrity and message authentication for any type of data - this is provided by supporting the model and syntax defined by the W3C. Similarly there will be a Rights and Publication Components in the Contents Metadata. It will be possible to associate both Signature and Rights Components with individual content parts, as well as to a complete item.

Another area where "power" compliance levels will be an important feature is in metadata handling and as a first stage provision has been made to apply identification, details of the creator and the creation date to all metadata. With complex metadata types it will also be possible, for example, to indicate the equivalence of different codes in different schemes, or to give a degree of confidence in the accuracy of the metadata.

It also has a *Content Metadata Component* which carries the *Administrative and Descriptive Components* with the metadata associated with the Item. Descriptive metadata can generally be inferred when the content is used (such as language, genre, subjects, titles) while the Administrative metadata provides information that cannot be inferred, such as when and where it was created, and who by - see the diagram on [page 16](#) for further information.

The *News Content Set* wraps alternative renditions of the content, which can be included directly (as XML content or plain text), in encoded form (such as an image) or be remote and referenced by a hyperlink.

Topic Item

Intended to carry information about concepts (both named entities like organisations and abstract like the news subject), the *Topic Item* will generally only contain short, structures information about the topic and its relationships with other concepts.

Typically there will be long term interest for the content; the content will generally be updated infrequently, but over a long period; it will be focused on single concept or entity; and it will be referred to by a large number of other resources, particularly *News Items* and other *Topic Items*.

As with the *News Item* the *Topic Item* is derived from the *Any Item* so has the identification and management components, it also has a *Content Metadata Component* for descriptive and administrative metadata. A *Topic Content Component* supports properties specific to the concept being handled, along with links to associated topics. Provision is made for *Specialised Topic Content* which may be a XML structure. This is a flexible structure that allow inclusion of structures that represent, for example, information about a person or a location or an organisation.

Package Item

Third of the generic Items, the *Package Item* provides some structure to news related information - typically of medium-term interest. Structure is expressed through a hierarchy of groups and

items, while the items remain independent of the package which only keeps references to them. Typical items are the "top ten" list of news stories or the representation of a newspaper page section.

Again, since it is derived from the *Any Item*, the *Package Item* has the associated management properties, while it also has a *Content Metadata Component* and a *Group Set* which represents a tree of items. Root of the *Group Set* is a *Group Component*, while the tree may contain other *Group Components* along with links to *News Items* and *Topic Items* (and possible future items) in any order.

Each *Group Component* has a *Group Role* (to show the part it plays), a *Group Mode*, and a collection of *Group Components* and *Composition Links* to include external items or Web resources in the package (by reference only).

There are three *Group Modes*: Complementary and Unordered (this is the default); Complementary and Ordered (where items have to be use or displayed in a specific sequence), and Alternatives (where there are equivalent pieces of content - such as translations - with the user selecting one of them).

News message

Support for the exchange of items in the news workflow is provided by

a *News Message*, though use of this is optional as any other syndication protocol can be used for exchange if required. A simple structure is used with exchange properties located in a message header and the items being exchanged contained in an Item set.

Contributors

The NewsML 2 Architecture Model and Technical Specification documents incorporate work by the following :

Mark Birbeck (x-port.net), Dave Compton (Reuters), Jay Cousins (RIVCOM), Honor Craig-Bennett (Press Association), Arnaud Descamps (RelaxNews), Takahiro Fujiwara (East), Darko Gulija (HINA), Paul Harman (PA), Alan Karben (XML Team), Jeremy Lebrecht (Reuters), Laurent Le Meur (AFP), Johan Lindgren (TT), Jayson Lorenzen (BusinessWire), Stuart Myles (Dow Jones), Michael Steidl (IPTC), Miles Whitehead (Reuters), Ulf Wingstedt (CNET) and Misha Wolf (Reuters).

NAR Working Groups

Work on the NewsML 2 Architecture is carried out by a set of Working Groups. Their efforts are combined by the NewsML 2 Architecture Working Party to produce the NAR Model and Technical Specification, and to guide development of the XML Schemas.

Chair of the NAR Working Party is Laurent Le Meur (AFP) and Vice-Chair Mischa Wolf (Reuters) - [see page 14](#).

Common Components - production of components (in the NAR framework) that will be used in more than one content markup standard. Lead: Johan Lindgren (TT).

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

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

News Structure - developing an abstract news model for IPTC standards. Lead: Laurent Le Meur (AFP).



Johan Lindgren



Stuart Myles

A Proven Success

The NewsML 1 Maintenance Working Party deals with the functional specifications and requirements for NewsML 1, and specifies implementation guidelines with proposals for best practice, with the aim of promoting the use of NewsML 1 for packaging and syndication of multimedia news.

Chair: Dean Large (Business Wire)



Vice-Chair: Takahiro Fujiwara (EAST Co. Ltd)

Web site: www.newsml.org

Discussion group: <http://groups.yahoo.com/group/newsml>

Released in 2000, NewsML 1 is designed as a multimedia structural framework for news, that can be applied at all stages of the news life cycle.

When released it was envisaged that typical uses would be for the transfer of news between news agencies and their customers, publishers and news aggregators and between news service providers and end users, as well as within editorial systems.

An indication of the success of NewsML 1 in finding such uses is given in the table on [page 19](#). This gives brief details of some of the applications that have been developed - most of this information has been supplied to IPTC by the organisations concerned, and it seems likely that there are many more applications.

New standard

The NewsML 2 Architecture draws on the experience with NewsML 1 and makes use of recent XML developments and it is intended that there will be a new "general news content" standard to take over from NewsML 1.

Because of this it is not envisaged that any major modifications, or additions, to NewsML 1 will be undertaken. An important aspect of current work is establishing a clear and easy to implement transition path to the new standard.

Update

The last major revision of the standard was V1.2, which was released in October 2003. However, work undertaken to preparing NewsML 1 for submission as a Japanese Industrial Standard (see below) identified a number of minor errors in the NewsML 1.2 Functional Specification and these were corrected during 2005. At the same time the opportunity was taken to include a reference to the NewsML

URN specification (RFC 3085), which had not been established when the original documentation was produced.

Application efforts

Specific efforts have been made to promote the use of NewsML 1 in Japan with NSK (Nihon Shinbun Kyokai - The Japan Newspaper Publishers & Editors Association) setting up a series of task forces, initially to introduce NewsML 1 to the Japanese news industry, and then to provide support and further encourage its adoption.

These efforts mean that the standard is now widely used - for example Kyodo News have been delivering NewsML to newspapers and Web sites since 2002. As part of their application they developed an extensive series of DTDs by analysing legacy text formats so the information can be carried in NewsML ContentItems.

Kyodo News have adopted the NITF and the IPTC Subject NewsCodes for their services. Now that work on the new IPTC standards family is well under way, Kyodo are also working on the migration of their applications from NewsML 1 to the NewsML 2 Architecture.

JIS X7201:2005

An indication of the importance of NewsML 1 to the Japanese newspaper industry is given by its release as a Japanese Industrial Standard - JIS X7201:2005 - in July 2005.

Work to achieve this was initially carried out by NSK teams, followed by a JIS Working Party which included users, system providers, outside experts and academics. A Japanese version of the NewsML Technical Specification was produced, while the Implementation Guidelines are also available in Japanese, along with other supporting documents.

NewsML 1 Applications

Agence France Presse (afp) (www.afp.com) - NewsML compliant multimedia news feeds. The "Magazine Forum" is a pure NewsML platform.

ANA (www.ana.gr) - news feeds in NewsML.

ANSA (www.ansa.it) - "ipernews" real-time news collections are produced using NewsML.

APA (www.apa.at) - text feeds and multimedia feeds formatted as NewsML available.

Asia Corporate News Network (ACN) (www.asiacorpn.com) Financial and corporate press releases.

Belga (www.belga.be) - Multimedia feed (text, picture, audio and video) to media clients and telecom operators.

Business Wire (www.businesswire.com) - NewsML compliant news feed.

Chunichi Shimbun Group (www.chunichi.co.jp) - NewsML-based photo database for a group of six daily newspapers.

CTK (www.ctlk.cz) - multimedia feed in NewsML format. CTK's new multimedia editorial system is based on a XML format derived from NewsML.

Japan Newspaper Publishers & Editors Association NSK (www.pressnet.or.jp) Many members offer NewsML compliant news feeds and editorial systems for newspapers based on NewsML.

JCN Newswire (www.japancorp.net) Regulatory News Service for publication to financial terminals and the Internet.

Mainichi Newspapers (www.mainichi.co.jp) -

Palace system for content management is based on NewsML.

MarketWire (www.marketwire.com) - press release distribution of corporate news and media advisories.

NewsPress (www.newspress.de) releases available in NewsML format.

PA News (www.pa.press.net) "Top 10" story collections in NewsML, with associated pictures, to a wide range of traditional media and corporate organisations, covering a variety of topics such as news, sport, and entertainment.

PR Newswire UK (www.prnewswire.co.uk) A NewsML compliant editorial system.

Reuters (www.reuters.com) NewsML compliant multimedia news feeds.

SA/ATS (www.sda-ats.ch) - NewsML compliant multimedia news feed.

The Irish Times (www.ireland.com) NewsML compliant text news feed of the daily newspaper.

The Tokushima (www.topics.or.jp) Newspaper production system with full multimedia capability based on NewsML.

Thomas Publishing Company (www.thomasnet.com) ThomasNet Industrial Newsroom provides industrial news articles. Most news distribution is now done in the NewsML format. A daily snapshot of the news is available as NewsML - <http://news.thomasnet.com/NewsML/inr-newsml-daily.xml>.

United Press International (UPI) (www.upi.com) - NewsML compliant text news feed.

Wall Street Journal Online (www.wsj.com) - news feeds packages delivered in NewsML.

There is also a substantial group of suppliers offering NewsML 1 compliant systems. These include:

Athens Technology Centre (www.atc.gr) - News Asset Agency Edition is NewsML capable.

CCI Europe (www.ccieurope.com) - Editorial Systems use IPTC NewsML with IPTC NITF for two-way integration of third party systems for archiving, web content management, e-paper, wires, events, syndication etc.

Delta Solutions (www.deltasolutions.nl) - Software for the production and distribution of (audio)-news. NewsML is used for the communication with the editorial intranet and the distribution to end users.

Digital Collections (www.digicol.de) - Asset Management solution uses IPTC NewsML for integration of wires, editorial, web CMS, e-paper, syndication etc.

East Co (www.est.co.jp) - range of NewsML

based products for the news industry including a NewsML format checker.

Eurocortex (www.eurocortex.fr) - pure web solutions to manage (edit, search, export) newswire information and multimedia content for the Press and Media industry.

Netwyse (www.netwyse.fr) - Koorou content management system supports up and down NewsML content streams and allows updates to NewsML content.

Promind Systems Inc (www.xslmaker.com) - XSL Maker is a visual design tool for Web pages that can apply CSS styling to NewsML content.

Transtel (www.transtel.com) - "nm_fusion" suite for software applications for news content. NewsML feed up and down stream can be handled.

Xtenit (www.xtenit.com) - communications platform for managing content and delivering e-mails. Content management integration can handle NewsML feeds.

Public Codes

Responsibility for maintenance and development of the IPTC sets of metadata - specifically designed for news applications - lies with the NewsCodes Working Party, which also undertakes the creation of new code sets when appropriate.

Chair: John Minting (UPI)



Vice-Chair: Honor Craig-Bennett (PA News)

Web site:
www.iptc.org/newscodes

Discussion group:
<http://groups.yahoo.com/group/newscodes>

Metadata is the key to efficient use of information, and its value is much greater when common values are used within an industry. IPTC have developed, and maintain, an extensive set of metadata for use by the news industry, known as the IPTC NewsCodes. Although the NewsCodes represent considerable intellectual capital they are freely available, and can be downloaded from the IPTC web site.

Most widely used of the NewsCodes are the Subject NewsCodes, which also appear to be finding applications beyond the news industry. There is a three level structure with seventeen main Subjects, a second level series of SubjectMatters under each Subject, and a third level of SubjectDetails under the SubjectMatters. Over 1300 terms have been defined, and where appropriate several codes can be assigned to give a detailed description of the content.

Language independent

Each of the Subject NewsCodes has a "formal name" (a numeric code), a "name" and an "explanation". The names and explanations are developed and maintained in English (British), but can be translated into other languages for use. Although the words change the (numeric) code remains the same and is language independent. For some languages translations supplied by IPTC members have been made available.

A comprehensive overhaul of the Subject NewsCodes was carried out in 2004, with the addition of a substantial set of new entries, along with the provision of explanations for all entries. Since then

Updates to the NewsCodes are available as NewsML 1 and RSS feeds

there has been more gradual growth with additions in selected areas as users develop their classification systems. There is a formal mechanism for the submission and approval of new entries, which have to be proposed by a IPTC member - details are available on the NewsCodes Web site.

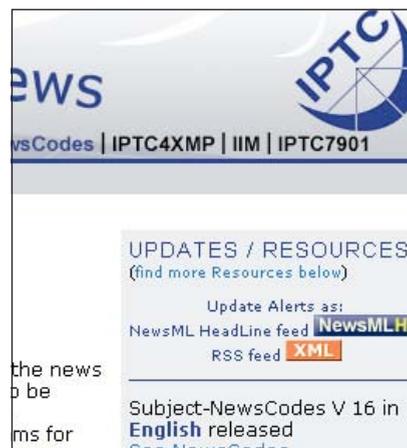
Replacement entries

One of the changes made to the Subject NewsCodes involved the replacement of an existing SubjectMatter heading by two new ones. Consideration of the implications of this decision resulted in the introduction of a formal policy for deprecating entries, which also deals with cases where it is considered appropriate to move or replace an entry.

Under this policy the deprecated heading is left in the Subject NewsCodes but marked up as deprecated with a recommendation that there should be no further use. The numeric code remains associated with the original entry so existing uses will not be affected and will continue to return the original meaning.

Versatile

Initially developed for use with IPTC standards such as NewsML and the IIM (Information Inter-



change Model - see page 6), there are twenty-six sets of NewsCodes in addition to the Subject NewsCodes (and the associated Subject Qualifiers). These sets cover a wide range and examples of the way they can be used include: Describing content (including Of Interest To, Genre, Role, Relevance, and Scene - along with the Subject NewsCodes); Handling news items (including NewsItem type, Status, Priority); Detailing technical parameters (including Encoding, Colorspace, Characteristics property).

Review and update

The NewsCodes have been developed over an extended period, drawing on IPTC members experience of the news industry. Following the overhaul of the Subject NewsCodes, an extensive review of the other NewsCodes sets is now under way, with the aim of ensuring that the entries meet current needs, and that all terms have proper explanations.

New sets of NewsCodes are developed where there is a clear requirement, and a recent submission for addition to the Subject NewsCodes was for a set of headings to cover International Organisations. Although it was recognised that there were practical advantages in having identifiers for such organisations it was not considered appropriate to include

them because they were entities rather than subjects.

However, the value of a set of Entity NewsCodes was appreciated and the possibility of generating one is under consideration - though it is appreciated that putting it together, and maintaining it, would be a major undertaking.

Updates of the IPTC NewsCodes are issued when changes or additions have been formally approved, and users can register to receive immediate notification of any new versions by NewsML or RSS feeds. Latest versions of all the NewsCodes are available for download in the form of TopicSets and there is a dedicated viewer for browsing and using them.

The NewsCodes Viewer provides an easy way to browse the codes and compare translations where available. Shown above are the Subject NewsCodes (the viewer works with all NewsCode sets) with the code tree to the left and the selected term at the top to the right. The Spanish equivalent of the chosen term, and its explanation are shown at the bottom.

New approach

Metadata handling has been an important feature of work on the NewsML 2 Architecture, with development of a new model for expressing the metadata. This has been designed to allow easier integration of the metadata with the W3C's Semantic Web and is compatible with RDF (Resource Description Framework). Introduction of NAR-based standards offers new opportunities - and raises fresh challenges - for development and application of news-specific metadata sets.

At the same time there have been suggestions that the three-level hierarchical structure of the Subject NewsCodes is starting to be a limiting factor in their development and there is a requirement for both internal and external links. Taking these factors into account a major initiative has been launched to look at the requirements for, and application, of new taxonomy standards for news applications. Although this will be a IPTC development effort, steps are being taken to seek input from other interested parties.

Automated categorisation

A major driving force behind the increasing use of the IPTC Subject NewsCodes is the increasing availability of automatic systems for applying the codes.

The CatScan system developed by NewsEngin (www.NewsEngin.com) is based on the IPTC Subject NewsCodes and also scans for organisations, people and places of particular interest to the user. It is said to be capable of scanning twenty typical news stories, or press releases, a second - or more than a million documents a day.

The system uses an array of marker terms related to each category and looks for matches between the marker terms and terms in the story. Each marker term has an associated score, and when there is a match the term's score is added to the total for the relevant category. If the overall score for the category exceeds a threshold, the category is applied to the story.

Specific advantages in using the IPTC Subject NewsCodes include the way that they allow a "bottom-up" approach with each category contributing to its parents, and that they are a familiar standard.

Ready to Process

The News Content Working Party oversees the development and support of standards for the mark up of all sorts of news-related material, with most of the work being carried out by Working Groups dedicated to specific interest areas.



Chair: Geoffrey Haynes (Associated Press - AP)

Web sites:
www.eventsML.org
www.sportsML.org
www.iptc4xmp.org
www.ProgramGuideML.org

Discussion groups:
<http://groups.yahoo.com/group/eventsml>
<http://groups.yahoo.com/group/sportsml>
<http://groups.yahoo.com/group/ProgramGuideML/>
<http://groups.yahoo.com/group/iptc4xmp>

When news content of a specific type is marked up in a standard way it can be much easier to deliver and process. The News Content Working Party investigates news areas that would benefit from the use of a common mark-up and works towards the introduction of appropriate mark-up systems (based on XML).

Where an area is considered of sufficient interest - in terms of the amount of news coverage involved - one of the first steps is to establish the business case. Assuming that this is favourable more detailed work on a standard may be undertaken, depending on the available resources. Where possible work can be carried out in collaboration with other standards bodies, though the particular requirements of the news industry means that this can be difficult.

SportsML

Longest established of the IPTC specialised content programs, SportsML was first released in March 2003. It has a modular structure with a core DTD to handle information that is common to many sports - including scores, schedules, standing and statistics (including wagering statistics).

This core is supplemented by plug-ins dealing with information that is specific to individual sports. For example, for baseball this includes the types of pitch and hit and ways a player can get out. Standard descriptions for the actions, and for teams, are held in resource files.

Current plug-ins cover American football, baseball, basketball, golf, ice hockey, motor racing, soccer and tennis. Most recent addition is the motor racing plug-in - at the moment this is mainly intended to

cover events such as Formula 1 and the Indy-car series. Extensions may be needed to cover other types of motor sport, but this will depend on user demand as SportsML developments are strongly user-driven. Similar considerations apply to plug-ins for additional sports.

Current release of SportsML is V1.6 which includes a number of enhancements to the core. The new motor plug-in relies on some of these enhancements made for V1.6, so is not backwards compatible with older versions of SportsML.

There is also a draft XML Schema version of SportsML. A download package containing the V1.6 DTD, draft XML Schema, resource files, documentation and examples is available from www.SportsML.org.

IPTC4XMP

For some time a series of datasets generally known as the "IPTC headers" have been used to attach metadata to images, using photo editing packages such as Adobe Photoshop. These datasets originally came from an early version of the IIM (Information Interchange Model - [see page 6](#)) and IPTC have since developed a much wider range of metadata.

Developed by Adobe Inc, XMP is a general way of embedding XML data in a file, and uses the W3C RDF (Resource Description Framework). Available as an open standard for use by developers, XMP has been widely integrated in the Adobe product range.

A collaborative effort between IPTC, Adobe Inc, and the IDEAlliance has produced an "IPTC Core" Schema for XMP that allows straightforward transfer of data from "IPTC Headers" to the XMP.

In addition custom panels have been developed to provide an interface to the IPTC Core within Adobe Photoshop.

A package including the IPTC Core Specification, technical documentation and a set of custom panels for use with Adobe Photoshop CS can be downloaded from <http://www.iptc4xmp.org>, where the RDF XML Schema for the IPTC Core is also available. In addition video tutorials on use of the panels - provided by David Riecks and the Stock Artists Alliance - can be found at <http://www.stockartistsalliance.org/tutorials>.

EventsML

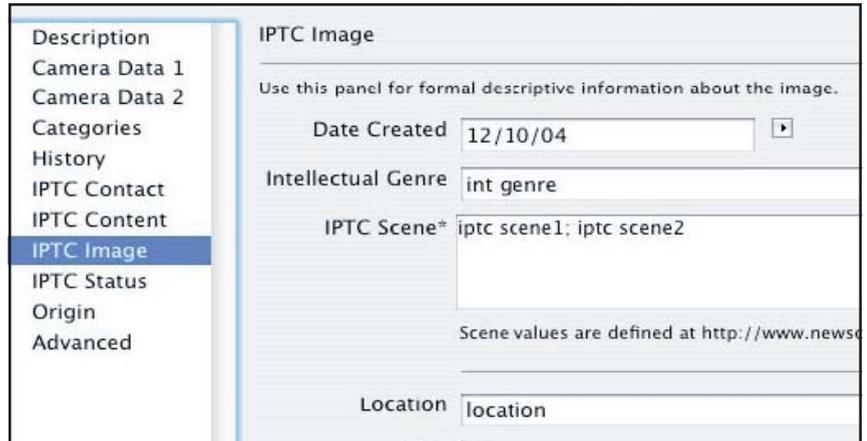
As the name suggests, EventsML is intended to be used for exchanging information about events. For this purpose events are defined as "something that happens", while for news applications the events also have to be subject to news coverage. Events may be planned or unplanned, with the special case of "breaking news" being capable of developing into a major area of activity.

Main interest areas are for:

Event publishing – communicating information about events, including associated news items,

Event planning – managing the coverage of breaking or upcoming newsworthy events, including support for gathering associated news items;

Event coverage – communicating



information about coverage of events by news organisations (often referred to as a "Daybook").

Typical uses of EventsML are seen as being: within news organisations, where the information might be used for a daybook or for the management and planning of news coverage; for event organisers to provide information to the public - either as individual events or as part of a series; and for aggregators (and on-line portals) to gather information about various events and publish the information in a consolidated format - or make it available to users from a central source. Initial work on the standard resulted in production of an EventsML Business Requirements document (this is available on www.eventsml.org).

However, it was realised that the work being carried out had many

There are four custom panels to allow the use of the IPTC Core with Adobe Photoshop CS. Shown above is the Image Panel which contains information about the visual content of the image. The other panels allow for contact information, description of the visual content of the image, and workflow and copyright details.

common features with the efforts being made to develop a replacement for NewsML 1, and this was a significant factor in the decision to concentrate efforts on the NewsML Architecture.

Following this decision the EventsML Business Requirements formed part of the input to the NewsML 2 Architecture (NAR) and direct development of EventsML was held back until the NAR was available. It is anticipated that EventsML will be one of the first members of the new IPTC standards family.

Working Groups

Development work on the news content standards is undertaken by individual Working Groups within the News Content Working Party. Current groups include:

SportsML

Lead: Alan Karben (XML Team Solutions)
Johan Lindgren (TT)

EventsML

Leads: Johan Lindgren (TT),
Dominic Chan (Canadian Newswire)

General news content

Lead: Laurent Le Meur (AFP)



Alan Karben



Johan Lindgren

Laurent Le Meur



Dominic Chan



General news content

Another standard that will be based on the NAR is the new general news content format that will follow on from NewsML 1.

In contrast to standards like SportsML and EventsML, the new standard will carry journalistic content that is typically unstructured - this would include text, images, audio and video. Where appropriate features of NewsML 1 will be carried forward to the new standard, but it will not be concerned with packaging and exchange functions (as was the case with NewsML 1).

One of the first steps was to establish the NewsItem as a central concept of the new standard, but it became apparent that this had

more general applications and so its development was moved to the NewsML 2 Architecture Working Party.

The NewsItem will probably include a management component, metadata constructs, a content metadata component, and a news content set that will contain alternate renditions of the content.

Programme data

ProgramGuideML (www.program-guideml.org) is an interchange standard for radio and television programme information, based on NewsML 1. It was adopted as a

candidate standard in 2004, but has not yet been brought into general use.

The possibility of reworking ProgramGuideML into a member of the NAR family is being considered - this might be as a way of handling data from TV-Anytime (a standard supported by the EBU that provides programme information for consumer use).

Other standards

The News Content Working Party also maintains a watching brief on relevant external developments. Where appropriate the group also

provides feedback to the developers of these external standards - in addition some members are actively involved in the work of other standards organisations.

One area of particular interest is financial data - this is a major source of information for the news industry - and it appears that XBRL (Extensible Business Reporting Language) is being more widely adopted. The approach being adopted by the NewsML 2 Architecture should make it much easier for future IPTC content standards to handle data produced using other MLs like XBRL.

NITF

NITF

NITF

NITF

NITF

NITF

Well Proved

Development and maintenance of the well-established News Industry Text Format (NITF) is carried out by the NITF Working Party

Chair: Alan Karben
(XML Team Solutions)



Vice-Chair:
Charles Tichenor
(The Associated Press - AP)

Web site:
www.nitf.org

Discussion group:
<http://groups.yahoo.com/group/nitf>

The NITF is a XML based standard that can be used to structure individual news articles, and provides information about the content and about the document itself.

Content information may include text formatting, specific identification of the subject of the content (using IPTC Subject NewsCodes), and enriched text elements that can be used to identify people, places, organisations, hyperlinks and so on. Content is not restricted to text but may also include tables, lists and images, for example.

Document information can include details such as when and where it was produced, copyright information and news management attributes.

Widely used

Longest established, of the IPTC XML-based standards (released in 1999), the NITF is probably the most widely used XML vocabulary in the news industry.

As a well documented and stable standard it is relatively straightforward to implement, so it is hard to assess just how popular it is. Some

indication is given by the fact that the NITF discussion group has some 440 members, and continues to grow.

Many of the points raised in the group appear to be from new users - with other members of the group generally being able to provide answers to questions raised and provide other guidance.

Changes

Similarly, the mature nature of the standard is reflected in the fact that there tend to be few requests for additions or modifications. These seem to arise as users develop more complex applications and generally tend to be minor.

Version 3.3 was released in 2005, and a typical change was modification of the <dd> element, which is used to hold the definition of a term - the change was simply to extend the content model for the element to include rich text.

XML Schema

Work on a XML Schema for the NITF is under way - this is a major development that does seem to be wanted by some users.